

Insurance Claims Data & Analytics

Healthcare Data Analytics and Data Mining

Group 3

Jinman	Rong
Runxua	Cao
Ting	Xiao
Xinlan	Wu
Yunyi	Wang
Ziyu	Tang
(Audit) Peihan	Tian
(Audit) Fangzhou	Xiang



Introduction

This report analyzes the insurance claim data in 2016. Main topics are discussed as follows:

- The patient vignettes: What the patient journey in the US looks like?
- Service and cost profile of major insurances: Are they related to the patients' demographics?
- Analysis of the health crisis related to illicit drugs and prescription opioids use: Is it serious? What are the concerns?

Through this report, you will get to know the medical treatment scenario in the US, the patient personas of three major insurance organizations and their cost profile, and the insights of health crisis related to illicit drugs and prescription opioids use.



Patient vignettes

The seven short stories below demonstrate the disparate patient journeys of seven people. Some welcomed the birth of baby, some died in regret, and some were still struggling in the life. Through these stories, you will get to understand more about life.

“It is as natural to die as it is to be born.” – Francis Bacon.



Uniq ID: 507033

With the joy of becoming a mother, she has been waiting for the arrival of this day for a long time. She hoped that the process would not be too long nor too painful. Hopefully, everything was fine, and soon the day came after 40 weeks. As a young mother in her 25 living in VT in 2016, Northwestern Medical Center received her. The baby didn't seem eager to come out and see the world, permitting adequate time to schedule the availability of suitable accommodations.

After some necessary processes of diagnosing delivery, she was told that the fetus was in the correct position. Thus, delivery required minimal or no assistance, with or without episiotomy, without fetal manipulation or instrumentation. Just as the doctor said, she did not need cesarean but still some manually help from nurses. Then, there was a burst of a baby crying from the labor room. It was a new-born life!

She recovered quickly and only stayed in this hospital for 1 day and then transferred to another facility. Of course, the high expense of hospitalization was taken into account, in addition to other necessary costs of room, pharmacy, and surgery supplies. The total revenue charge was up to 6000 dollars. Fortunately, she had BLUE CROSS to cover part of the expense. Then she only needed to pay almost half of it. What a relief to a typical family like hers!



Uniq ID: 40436

This woman, age between 70-74, lived in Vermont. She was transferred from another hospital to the University of Vermont Medical Center in 2016 in an urgent situation.

At this age, people always had a lot of trouble. The same did to this woman, but she was even worse. She was conducted to two procedures and had 14 diagnosis. She had myocardial infarction, acute post hemorrhagic anemia, atherosclerotic heart disease, depressive disorder, hyperlipidemia, asthma, cough, transient ischemic attack (TIA) and cerebral infarction without residual deficits, nicotine dependence, presence of coronary angioplasty implant and graft, long term (current) use of insulin, long term (current) use of aspirin, ischemic heart disease and other diseases of the circulatory system and she even need a wheelchair.

Her heart was born to be not well, so she suffered from many kinds of heart diseases. The situation was even worse, when she was elder, so she had a heart operation and insert a coronary angioplasty implant. To prevent inflammation, she had to use a long term of aspirin. She also suffered from a long-term hyperlipidemia, so she had to use insulin for a long term. Even worse, the asthma caused her coughed frequently. Due to the dad health condition, she always felt pain and need to depend on nicotine, which was easily addicted to, but she had no choice.

This time, she transferred from hospital to a medical center may because of her implant inserted heart. The operation was finished and need further observation. The doctor gave her X-ray test, CT test and ultrasound test, all of them were common test to diagnose a heart disease. As she was in a bad condition and always complaint about it. The doctor gave her a stress test. Not surprise, she had a depressive disorder. However, what doctor can do was just sending her few dugs, suggesting her to be supported by a wheelchair and then sending her back to home. The total pay for her inpatient treatment is \$70275.41, all of it was paid by Medicare. With so much diagnosis, it was highly likely that this woman may went to the hospital soon. It seems that the tax she paid during most of her life really worked out.



Uniq ID: 859382

This day was the same as any other day of 2016. People worked, lived and loved as usual. But it was a day of spring, the trees were greener than yesterday.



The Emergency Department of Rutland Regional Medical Center was also the same as usual, watching people being treated, crying or dying. A young man in his early 30s was brought to the Emergency Department on this day. The onset of disease was unexpected. Before being brought to ED, he was not in any health care facility. His situation was very severe -- he suffered from acute respiratory failure with hypercapnia and cardiac arrest due to other underlying condition. Through Laboratory - Clinical Diagnostic, Electrocardiogram and Radiology - Diagnostic, doctors found Compression of brain and Anoxic brain damage nearly ruined his brain. He also had Ventricular fibrillation. What's more, he was diagnosed with poisoning by heroin and opioid dependence, which might be the source of everything on today. He spent a short time in ICU, with some respiratory services, such as IV solutions services, to help him breathe. In Emergency room, doctors used many methods and medicines. Unfortunately, nothing worked. He died in 48 hours.

The number of total charges was 13128.19. Surprisingly, it was a self-pay. However, he didn't have the opportunity to go to the toll booth himself.



Uniq ID: 1585831

The patient is a lady aged between 40-44 years old who was sent directly to emergency room in Rutland Regional Medical Center, but only stayed in the hospital for one day as her last day in 2016. Her charge was 17093.79 US dollars, paid by Medicaid. Her bill type was hospital based, inpatient final bill. The primary cost center is immunology.

She suffered from acute pulmonary edema, acute and subacute infective endocarditis, acidosis, cellulitis of right lower limb etc. She was dependent on opioid and had other psychoactive substance abuse.

She received treatment include intensive care, clinical and radiology diagnostic. She also scanned head using CT. Due to her Acute respiratory failure with hypoxia, doctors also provided respiratory services like EKG/ECG. As for pharmacy, she has been used IV solutions and medical/surgical supplies.

However, with the help of the doctors, she cannot make it and passed away in her middle age. Opioid abuse does serious damage to the health and we should enjoy every day instead of shorting our lives by drug abuse.



Uniq ID: 200760

In the first quarter of 2016, a young lady aged 18 to 24 was directly sent to University of Vermont Medical Center's emergency room due to a traffic accident. Her medial malleolus of left tibia was displaced, and shaft of left fibula has unspecified fracture. However, except for physical injuries, the woman was further diagnosed with gastro-esophageal reflux disease without esophagitis and the major depressive disorder. After the operation on the musculoskeletal system, she stayed in the hospital for four days as an in-patient.

She lived in a semi-private 2-bed room and accepted several services such as clinical diagnostic, radiology diagnostic, anesthesia, physical therapy and evaluation/re-evaluation after physical therapy. The surgical supplies including sterile supplies and other implants are used during the operation, and she was prescribed some medicine, too. Furthermore, she also used emergency room and operating room. All these services cost about 40,422 dollars.

Finally, she was discharged to go home and receive follow-up treatment by home health, such as a change of medication. The hospital charged 49,533 dollars eventually, which is hospital based, inpatient final bill. This bill will be paid by her insurance company, commercial insurance.



Uniq ID: 3692

This young man, aged between 18 and 24, was in terrible condition when he was taken to University of Vermont Medical Center in Milton in 2016.

It seems that he was assault by unarmed brawl or fight, got nondisplaced fracture of anterior process of left calcaneus. Apart from this, he also had serious mental problem. He was in bipolar disorder and current episode manic severe, and sometimes had suicidal ideations. Maybe this could be explained by his personal history of traumatic brain injury. Also, he had dependence on drugs like cannabis, nicotine and cigarettes. What's worse, Plantar fascial fibromatosis and Constipation was also bothering him.

The hospital provided drugs to treat his mood disorder and mental problems. To deal with his addiction to



alcohol, tobacco and drugs, the hospital offered him pharmacotherapy and detoxification services for substance abuse treatment for alcohol and drug rehabilitation. The treatment charged him 117,895 dollars, which is not a small number. But BLUE CROSS helped him pay part of the charge. After the treatment, the patient returned home. I hope he can follow the doctor's advice, take the medicine on time and stay away from tobacco and alcohol. After all, he is still young and has a chance to have a bright future!



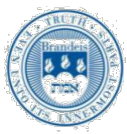
Uniq ID: 690326

She lived in Milton, MA. She was an elegant lady who was once pretty and charming, but in the end could not prevent herself from aging after 40. One day, she looked into the mirror and could not stand herself being out of shape anymore. She decided to make some changes finally.

However, she knew nothing about cosmetic surgery, so she first went to a non-health care facility to ask for some advice. Few days later, she showed to University of Vermont Medical Center half an hour before appointment time. After talking to her doctor, they finally confirm the surgical plan and set the detail of her hospitalization things. Finally, she took this surgery, including Alteration of Bilateral Breast (Open Approach) and Alteration of Abdominal Wall (Open Approach) operations. And she even chose to take gastric banding surgery to reduce the size of her stomach forever, preventing herself from going out of shape again. She stayed in the medical center for 3 days until recovered, during which she went through Pruritus, several times Tachycardia and muscle spasm. That was such a difficult time for her and cost her a lot of money, \$43,425.53 as total. She received 13 different services in the hospital and most of her money was paid for surgery related cost, including 3-days bed cost, surgical supplies and operating room services, respiratory services, recovery room, etc. Among all the services, operating room services cost her most, which is \$25,619. And she paid her bill all by herself!

Three days later, she left for home and hired someone to take care of her until she recovered completely. She became beautiful again as she expected. Although it costed her a large amount of money, she thought everything she did before was worthy, for the pleasure of back to beautiful is invaluable to her and since then she started loving her life again.

This story teaches us that if we do not have any insurance coverage, we must leave hospital as soon as possible. Otherwise, we will spend almost half-life savings even though the disease is cured. This lady will have no disease, and no money after getting out of the hospital.



Service and Cost Profile of Major Insurances

Charges of major diagnosis groups among three insurance payers

Data cleaning & manipulating process description:

Step 1: Filter the inpatient dataset by the column “PPAY” equals 1,2,6&7, which are the codes for Medicare, Medicaid, Blue Cross and Commercial Insurance. In this case, the latter two companies are combined to one large category as Commercial Payers. Those three companies are the final table column names.

Step 2: Aggregate charges by 25 types of Major Diagnosis Groups and sum them up. Then we round the values to \$Million to drop decimal points. Therefore, some Major Diagnosis Groups with fewer than 1 million patients will be omitted.

Our final result is the following table that shows among three insurance companies, the number of patients suffer from 25 types of major diagnosis groups.

Table 1 Charges of major diagnosis groups among three insurance payers

Major Diagnosis Groups	Commercial (million)	Medicaid (million)	Medicare (million)
BRAIN AND CNS	38	14	57
EYE	0	0	0
EAR, NOSE & THROAT	3	2	4
RESPIRATORY	24	22	78
HEART & CIRCULATORY	54	15	136
DIGESTIVE	32	14	63
LIVER & PANCREAS	12	11	18
MUSCULOSKELETAL	88	30	146
SKIN AND BREAST	6	4	11
ENDOCRINE	9	6	13
KIDNEY & URINARY	8	4	28
MALE REPRODUCTIVE	2	0	3
FEMALE REPRODUCTIVE	4	2	2
PREGNANCY, CHILDBIRTH AND THE PUERPERIUM	34	26	1
NEONATAL	28	31	0
SPLEEN & BLOOD	5	3	7
LYMPHATIC	6	3	8
INFECTION	18	14	61
MENTAL ILLNESS	10	15	25
SUBSTANCE ABUSE	1	4	3
INJURY, TOXIC EFFECTS	4	4	9
BURNS	0	1	1
ALL OTHER	6	3	15
TRAUMA	8	3	4
HIV	0	0	0

Inpatient services portfolio among three insurance payers

To better analyze the differences between three different inpatient services portfolios, we draw three pie charts for each insurance company. Furthermore, by analyzing three inpatient personas of each insurance company, we relate the differences in the portfolios according to the demographics (sex and age) of the patients.

Inpatient services portfolio and patient persona of Medicare

From the pie chart of Medicare, we observe that the top-5 MDCs are Musculoskeletal, Heart & Circulatory, Respiratory, Digestive, and Infection. Since Medicare mainly serves the elder people aged 65 and over, the “popular” MDCs are more related to the elderly¹.

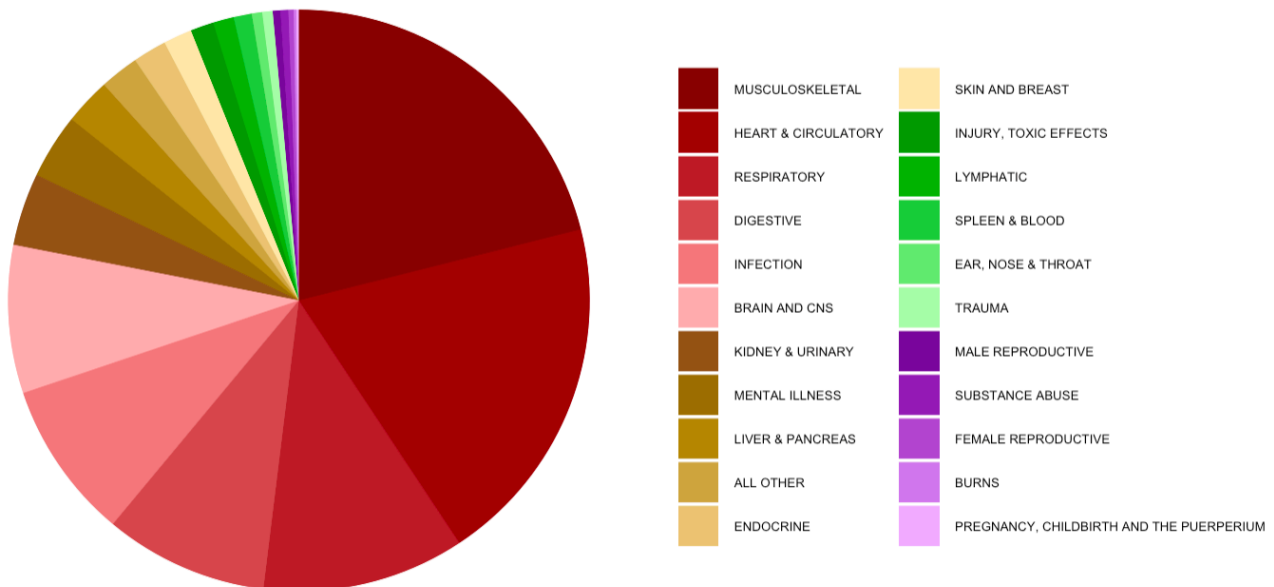


Figure 1 Inpatient services of Medicare

The Medicare’s patient persona analyzed from 2016 inpatient claim data shows as below.

Patient Persona of Medicare

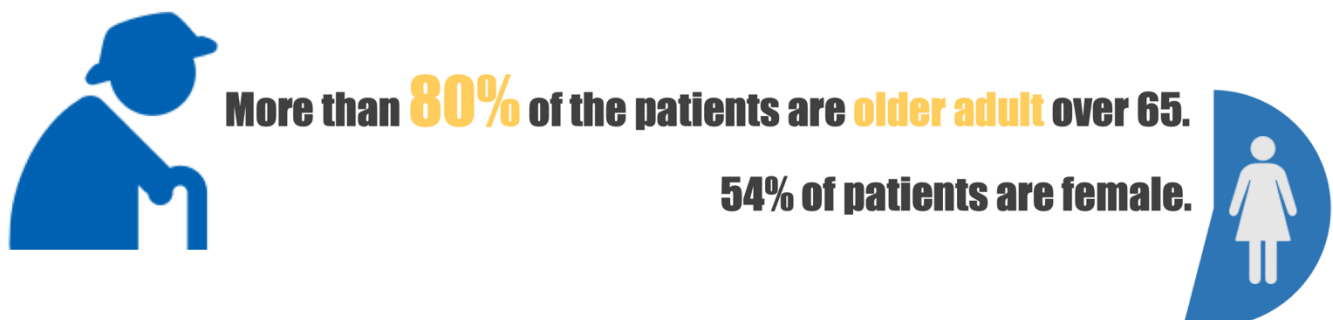
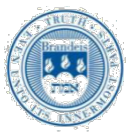


Figure 2 Patient persona of Medicare

¹ <https://www.kff.org/medicare/issue-brief/sources-of-supplemental-coverage-among-medicare-beneficiaries-in-2016/>



According to Efraim & Jeremy (2017), the elder people mainly suffer from osteoarticular diseases (OAD) and cardiovascular disease (CVD) caused by muscle weakness, infection due to the changes immune system, and chronic diseases that affect the digestive system². That is why Musculoskeletal, Heart & Circulatory, Respiratory, Digestive, and Infection are Top 5 charges of Medicare.

Inpatient services portfolio and patient persona of Medicaid

In Medicaid, the Neonatal, Musculoskeletal, Pregnancy, childbirth and the puerperium, Respiratory and Heart & Circulatory eat the largest portion of the money. The reason why Neonatal and Pregnancy are costing the Medicaid so much money is because Medicaid covers members who are women and infants³.

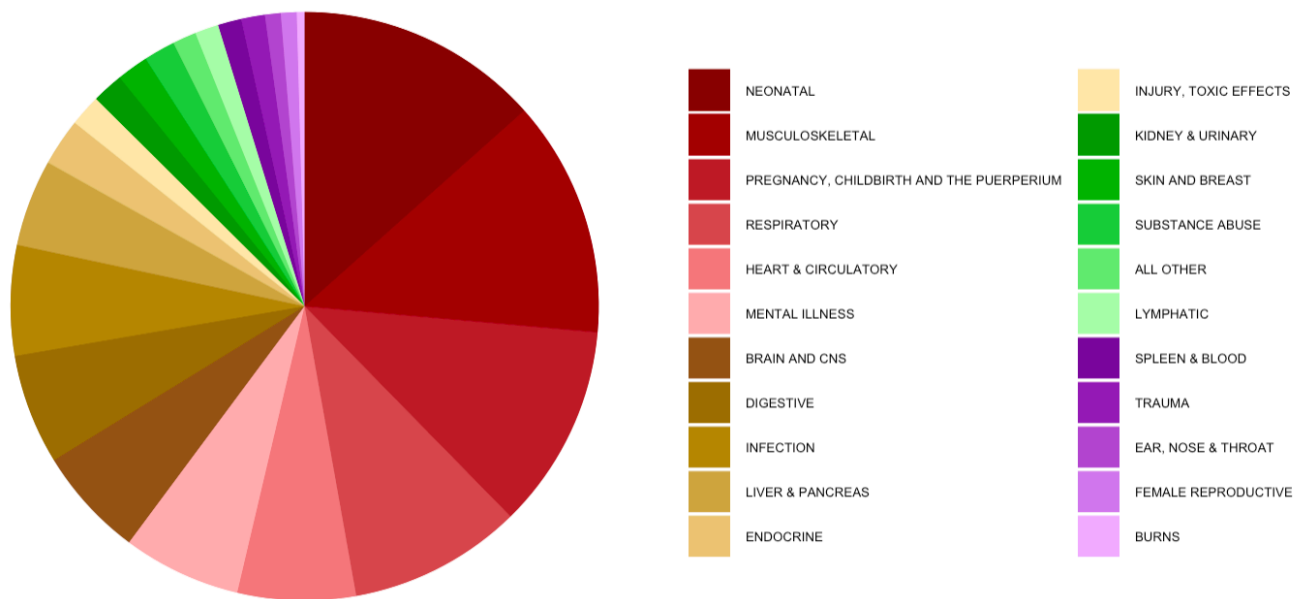


Figure 3 Inpatient services of Medicaid

The Medicaid's patient persona analyzed from 2016 inpatient claim data shows as below.

Patient Persona of Medicaid



Figure 4 Patient persona of Medicaid

² Jaul, E., & Barron, J. (2017). Age-related diseases and clinical and public health implications for the 85 years old and over population. *Frontiers in public health*, 5, 335.

³ <https://en.wikipedia.org/wiki/Medicaid>



Besides, the target customers of Medicaid include disabled and people with lower income. That is why musculoskeletal and mental illness rank among Top 5. Lower people suffer from daily life and are easy to feel stressed while disabled people may need bone-related treatment.

Inpatient services portfolio and patient persona of Commercial Payers

According to the pie chart, commercial payers mainly pay for Musculoskeletal, Heart & Circulatory, Brain and CNS, Pregnancy, Childbirth and the Puerperium and Digestive.

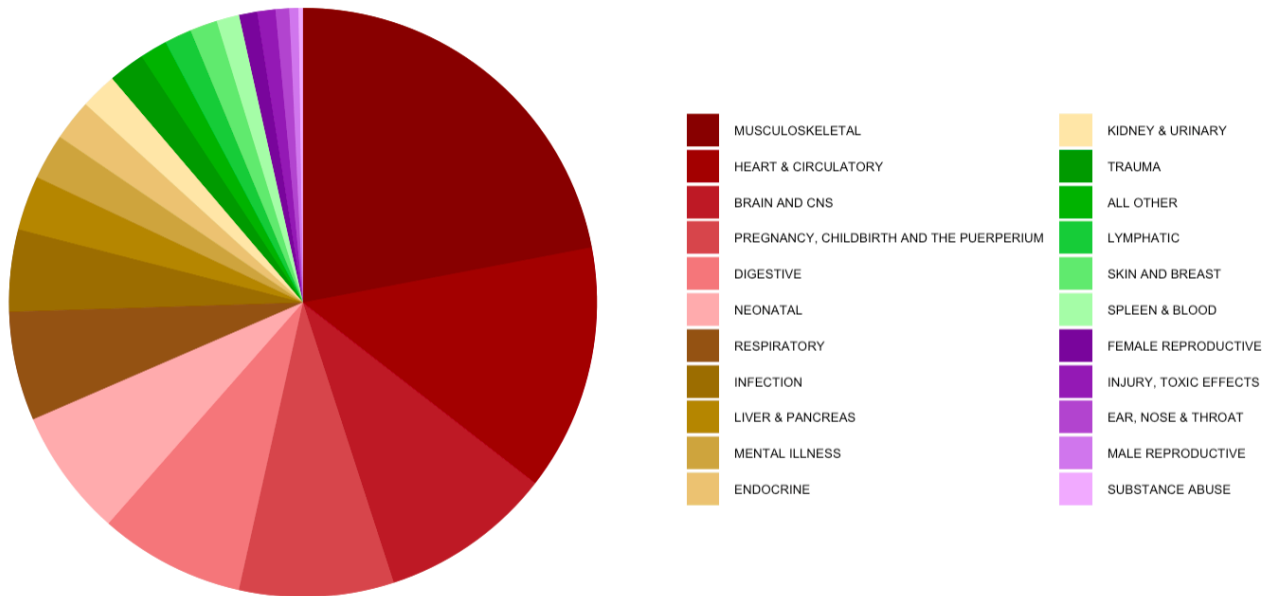


Figure 5 Inpatient services of Commercial Payers

Working age people are Commercial payers' mainly patients. According to national cancer institute, working age people occupy 50 % among patients who have brain and other nervous system cancer. They are more likely to get high incidence of brain system disease. Therefore, brain & CNS rank third among Top 5 charges and commercial payers spend more money it than Medicare and Medicaid⁴.

Patient Persona of Commercial Payers



67% of patients are at **working age** from 18 to 64.

Female patients becomes majority of the commercial insurance.

Infants and older adults aged 60 to 64 are the top 2, taking up

31%

Figure 6 Patient persona of Commercial Payers

⁴ <https://seer.cancer.gov/statfacts/html/brain.html>



Besides, Commercial insurance cover a wide variety of patients including women, child and elder people. Therefore, the costs of Musculoskeletal, Heart & Circulatory, Pregnancy, Childbirth and the Puerperium, and Neonatal are very high, resulting various types of disease charges that are age and sex oriented.

The Analysis of the Health Crisis

Every coin has two sides, so does medicine. In addition to enjoying their positive effects, people nowadays are more likely to face the problems drugs are bringing to them, like drug overdose, drug abuse. The governments and insurance companies are demanding the healthcare providers to pay more informational attention in submitting the drug abuse-related claim data.

Thus, in this section, the overall situation of drug use in the United States is to be examined. Six different dimensions of analysis will be applied; namely, numerical analysis, gender analysis, expense & coverage analysis, geographical & age distributions analysis, and diagnose type analysis, along with a typical case analysis based on synthetic narcotics.

1. Numerical Analysis

After selecting targeted ICD-10 classification, the table has 2151 rows in total, which means there are 2151 ED visits exactly have been diagnosed as drug user/abuser, accounting for only 0.8% of the total visits in the data set. Preliminary glimpse reveals that drug overdose seems to be not severe.

2. Gender Analysis

Then, we use Fisher test to check if the data support this gender bias myth. The elements of the matrix are numbers of male who use/don't use drug and female who use/don't use drug.

The 2x2 cross table for male/female for gender for Sole Proprietor is presented below:

Table 2 Gender differences in drug use/abuse

	Female	Male
Drug User	1141	1009
Non-Drug User	140553	123149

P-value = 0.8452

The result is not statistically significant. Thus, drug abuse is not related to gender. Consequently, the gender bias myth is solved here as either male or female could confront drug overdose/use issues.

3. Expense & Coverage Analysis

The total exact dollar amount for identified patients is 30,741,219.53. The absolute value sums up to thirty million. Additionally, the overall expense of these drug overdose patients has taken up 2.4% of the total spending in the dataset — a relatively high percentage in comparison to these people's number share.

Next, how do three insurance payers cover illicit drugs or prescription opioids use/abuse/overdose is analyzed.

The three major insurance payers are: Medicare, Medicaid and two major commercial insurances of "BLUE CROSS" + "COMMERCIAL INSURANCE", which is denoted as 'Commercial Payers'.

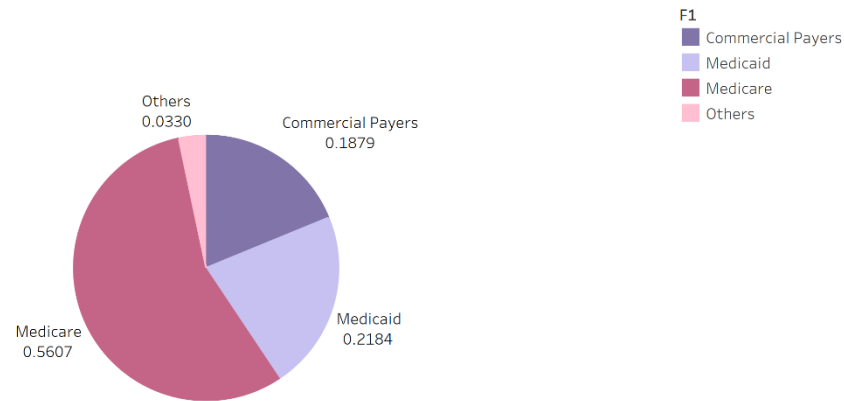
The shares of the targeted three insurance payers are shown below:



Table 3 Share of each of the targeted three insurance payers

Medicare	Medicaid	Commercial Payers
56.074%	21.840%	18.794%

Percentages of insurance payers



F1 and sum of F2. Color shows details about F1. Size shows sum of F2. The marks are labeled by F1 and sum of F2.

Figure 7 Share of each of the targeted insurance payers

Among the three major insurance payers, Medicare takes the largest percentage of the total payments (56.07%), Medicaid goes at second (21.84%) and Commercial Payers takes 18.79%.

The public financing seems to shoulder more than three fourths of the total spending on health crisis related to drugs. This is indeed a form of government financial waste.

Possible solutions could be that the government should strengthen the top-down control of drug overuse.

4. Typical Case Analysis—Synthetic Narcotics

Start with this part, we try to explore the reasons behind why drug use/abuse became so common these years. The first is the use of synthetic narcotics and its improper marketing, making many people unaware of the side effects.

We would like to count the patients who were brought to ED for diagnosis related to synthetic narcotics or amphetamines and select some typical case conditions to map the entire environment.

Table 4 Targeted diagnosis and corresponding patient numbers

ICD-10 codes	Number
T404xxx and T4362xx	156
T404xxx and T4362xx, ATYPE=1/2	155
T404xxx	114
T4362xx	43

156 patients have been brought to ED for diagnosis related to synthetic narcotics or amphetamines. When they were brought to emergency department, 155 of them are in emergent (ATYPE=1) or urgent (ATYPE=2) situation. Among these people, 114 of them have been diagnosed with drug use/abuse codes of T404xxx, which means, synthetic narcotics; and 43 use/abuse amphetamines(T4362XX).

Synthetic narcotics include Meperidine, Methadone and many other less potent drugs. The class is also known as “opiate analogs.” The abuse is very common.



Amphetamine is a kind of stimulants which can be used for enhancing performance. This is a very dangerous drug. The stereotype of person who drugs is, without formal job, mentally abnormal, and having poor interpersonal relationship. But this is wrong. There are some excellent people having very decent jobs and life who are drug addicts. They are called functional addicts. By using these kinds of drugs, these workaholics can do job very well. You can't find they are drug addicts until they are brought to ED.

5. Geographical & Age Distributions Analysis

In this part, the numbers of drug use/abuse cases related to zip code regions are to be analyzed to find which region has the relatively high amount of the drug crisis.

The three highest drug crisis regions are shown as follows:

Table 5 The highest three zips using drugs and their separate numbers

Txtzip	Totoal	Percentage
054	379	17.6%
057	249	11.6%
05701	226	10.5%

As the result shows, the highest use of the drug is in states 054, 057, and 05701, which have 379, 249, and 226 records, respectively. These areas are all in Vermont, which has one of the lowest population numbers in the United States, representing a high percentage of drug abuse among the total population. These severe situations could arise from flaws in the local health system or government regulations.

Next, the reasons behind the phenomenon perhaps are analyzed from the distribution of ages. The distributions are shown below:

Table 6 Percentage of age groups of the targeted three locations

054		057		05701	
Age group	Percentage	Age group	Percentage	Age group	Percentage
1	0.3%	1	0	1	0
2	6.75%	2	4.21%	2	1.1%
3	12%	3	11.2%	3	7.18%
4	12%	4	5.61%	4	9.39%
5	10.1%	5	7.48%	5	8.29%
6	6.13%	6	4.21%	6	8.84%
7	5.21%	7	7.01%	7	7.18%
8	6.44%	8	6.07%	8	4.42%
9	5.83%	9	12.6%	9	9.94%
10	4.91%	10	6.07%	10	6.63%
11	5.52%	11	7.94%	11	8.29%
12	8.9%	12	11.2%	12	9.94%
13	4.91%	13	5.61%	13	3.31%
14	11%	14	10.7%	14	15.5%

The table above shows the different age distributions in different locations (054,057,05701).

We can see that older people are likely to experience drug abuse among other groups, and in state 054, people aging between 18 and 34 are also likely to have drug abuse than other groups. Thus, state 054 should make some efforts to alert these kinds of people about the dangers of drug abuse, such as through TV and radio, to strengthen public awareness of drug abuse or put forward new laws to inhibit drug companies from selling risky drugs.



Besides, the government should think about reasons why older people or young people are likely to be troubled by drug abuse. For older people, perhaps they are in terrible health conditions and are prone to try all kinds of drugs to get rid of diseases, or they tend to believe the effect of some medicines and take them frequently. For younger people, they may suffer great pressure from their work or study. Thus, they may use drugs to relieve pains.

As a result, local government should care more about the two groups of people and set up specific policies to prevent them from drug abuse or put forward some suggestions which can alert the public about the dangers of drug abuse.

6. Diagnose Type Analysis

From this perspective, diagnose types for drug overdose are examined to give hints on which specific symptoms to lay emphasis for regulation and future improvement.

Summarizing the top 10 drug abuse cases can have the following table:

Table 7 10 most common diagnoses of drug use/abuse

Ranking	ICD-10	Num	Description
1	T401X1A	258	Poisoning by heroin, accidental (unintentional), initial encounter
2	T402X5A	256	Adverse effect of other opioids, initial encounter
3	T424X2A	123	Poisoning by benzodiazepines, intentional self-harm, initial encounter
4	T424X5A	114	Adverse effect of benzodiazepines, initial encounter
5	T40605A	112	Adverse effect of unspecified narcotics, initial encounter
6	T43222A	82	Poisoning by selective serotonin reuptake inhibitors, intentional self-harm, initial encounter
7	T402X1A	81	Poisoning by other opioids, accidental (unintentional), initial encounter
8	T424X1A	80	Poisoning by benzodiazepines, accidental (unintentional), initial encounter
9	T426X5A	75	Adverse effect of other antiepileptic and sedative-hypnotic drugs, initial encounter
10	T426X2A	74	Poisoning by other antiepileptic and sedative-hypnotic drugs, intentional self-harm, initial encounter

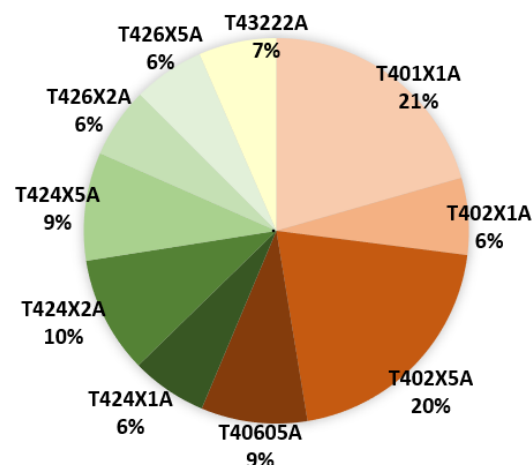
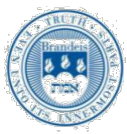


Figure 8 Percentages for 10 tops commonly drug use/abuse

The ICD-10 code of the T401X1A and T402X5A are the two most significant causes. They together account for 41% of all drug abuse cases in the emergency department. We can find that both T401X1A and T402X5A are from the T40 group, "Poisoning by, adverse effect of and underdosing of narcotics and psychodysleptics", which



may reflect social concerns that people are commonly under considerable pressure and try to use drugs to relieve pains.

Summarizing the top 10 abuse cases in different medicaments levels, we can get the following chart. The T40 and T42 are nearly the major problems for the emergency department. T40 is cause by narcotics and psychodysleptics drugs and T42 is cause by antiparkinsonism drugs. T43 only stands for a little proportion and T41 even not shown in the top 10 emergency drug abuse cases.

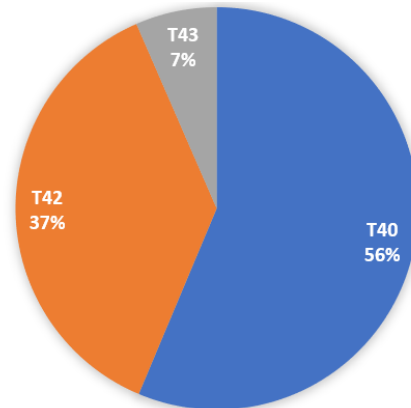


Figure 9 Top 10 abuse cases in different medicaments levels

In conclusion, the healthcare provider should keep more attention on the psychodysleptics drugs and antiparkinsonism drugs prescribed or sold to customers. Among these two kinds of drugs, heroin, other opioids used should be controlled more carefully.



Summary

The first section provides seven short stories of seven people, demonstrating the various patient journeys in the US.

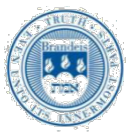
The second section analyze the cost profile of 25 major diagnosis groups among three major insurance organizations: Medicare, Medicaid and Commercial Payers. The cost profiles vary between these insurance organizations because the different patient personas they have. With the majority of elder people, Medicare expenses most on Musculoskeletal and Heart & Circulatory, while Medicaid expenses most on Neonatal, Musculoskeletal, and Pregnancy because it mainly covers members who are infants, women, disabled and people with low income. Commercial payers cover a wide range of people and has various types of disease charges that are age and sex oriented, from Pregnancy to Musculoskeletal.

The third section aims to explore the overall situation of drug use/abuse in the United States and give a preliminary analysis of how this situation comes into being and offer possible solutions.

The proportion of the people who experienced drug use/abuse only accounts for 0.8% of the total people who visited, while the expense of visits paid by these people took up more than 2% of the overall payment of all visits. The absolute amount equaled to thirty million. We can see that the United States is facing a severe drug overuse/abuse issue.

Most targeted people were from either a younger group or older group. Combining results from age distributions and diagnose type analysis, it can be seen that anti parkinsonism drugs may be the cause for high percentages in older-age groups and narcotics and psychodysleptic drugs in the younger-age group. This may because older people are more likely to be attacked by Parkinson's diseases and younger people to face significant pressure nowadays to use drugs to numb themselves or relieve stress. Thus, hospitals may look into different ways of curing Parkinson's diseases. In terms of younger people's health, more fundamental solutions like psychological counseling or more welfare may be taken into account. In addition, regulations are needed to curb misleading advertising synthetic narcotics.

In terms of geographic patterns, Vermont has a high percentage of drug abuse among the total population. We should not neglect the fact that the rate of people aged over 65 is higher than the average level in the United States. Consequently, these severe situations could arise from the demographic structure, along with flaws in the local health system or government regulations.



Appendix

```
library(data.table)
library(sandwich)
library(tidyverse)
library(lmtest)
library(ggplot2)
library(knitr)
library(psych)
library(dplyr)
library(tidyr)
library(scales)
library(RColorBrewer)
```

Question 1

```
ED=fread(file=~ /Downloads/VTED16.TXT")
inp_2016=fread(file=~ /Downloads/VTINP16_upd.TXT",
               col.names=c('age', 'sex', 'payer', 'charge', 'MDC'),
               select = c('intage', 'sex', 'PPAY', 'CHRGs', 'MDC'))
inp=fread(file=~ /Downloads/VTINP16_upd.TXT")
revcode=fread(file=~ /Downloads/VTREVCODE16.TXT")

patient_ed=ED[UNIQ==200760]
#patient_ed
patient_in=inp[UNIQ==200760]
#patient_in
patient_rev=revcode[Uniq==200760]
#patient_rev
#write.csv(patient_ed, file=~ /Downloads/patient_ed.csv")
#write.csv(patient_in, file=~ /Downloads/patient_in.csv")
#write.csv(patient_rev, file=~ /Downloads/patient_rev.csv")
```

Question 2

```
inp_2016=inp_2016[payer==1|payer==2|payer==6|payer==7]

MDC=fread(file=~ /Downloads/MDC.csv")
age=fread(file=~ /Downloads/age.csv", col.names=c('age', 'Age'), select = c('AGEGRP', 'AGE_GRP_DESC'))
gender=fread(file=~ /Downloads/gender.csv", col.names=c('sex', 'Gender'), select = c('SEX', 'SEX_DESC'))

inp_2016_medicare=inp_2016[payer==1]
inp_2016_medicare$name_payer=c('Medicare')
inp_2016_medicaid=inp_2016[payer==2]
inp_2016_medicaid$name_payer=c('Medicaid')
```




```
inp_2016_commercial=inp_2016[payer==6|payer==7]
inp_2016_commercial$name_payer=c('Commercial')
inp_2016=rbind(inp_2016_medicare,inp_2016_medicaid,inp_2016_commercial)
#inp_2016

charge_MDC=inp_2016[,.(charge=sum(charge,na.rm=TRUE)),keyby =.(MDC,name_payer)]
charge_MDC=charge_MDC[!is.na(MDC)&!is.na(charge)]
charge_MDC$charge=round(charge_MDC$charge/1000000)
charge_MDC=merge(charge_MDC,MDC,by=c("MDC"),all.x = TRUE)
#charge_MDC

table=charge_MDC%>%spread(name_payer, charge)
#table

table_commercial=table[,2:3]
table_commercial=table_commercial[!is.na(Commercial)&Commercial!=0]
table_commercial=table_commercial[order(-Commercial)]
#table_commercial

table_medicaid=table[,c(2,4)]
table_medicaid=table_medicaid[!is.na(Medicaid)&Medicaid!=0]
table_medicaid=table_medicaid[order(-Medicaid)]
#table_medicaid

table_medicare=table[,c(2,5)]
table_medicare=table_medicare[!is.na(Medicare)&Medicare!=0]
table_medicare=table_medicare[order(-Medicare)]
#table_medicare

#write.csv(table, file=~Downloads/tabulation.csv")
```

pie chart of medicare:

```
pie_medicare=ggplot(table_medicare, aes(x="", y=Medicare, fill=reorder(MDC_CAT_NAME,-Medicare)))
pie_medicare=pie_medicare+ geom_bar(width = 1, stat = "identity") +
  coord_polar(theta="y", direction = -1) +
  theme_void()
color=c("#800000", "#990F0F", "#B22D2D", "#CC5252", "#E67E7E", "#FFB2B2", "#8d5524", "#99700F", "#B28B2D", "#CC
A852", "#E6C77E", "#FFE8B2", "#1F990F", "#3CB22D", "#60CC52", "#8AE67E", "#BCFFB2", "#710F99", "#8B2DB2", "#A8
52CC", "#C77EE6", "#E9B2FF", "#990F20", "#B22D3C", "#CC5260", "#E67E8A", "#FFB2BC")
pie_medicare=pie_medicare+ geom_bar(width = 1, stat = "identity") +
  coord_polar(theta="y", direction = -1) +
  theme_void() +scale_fill_manual(values = color)+
  theme(legend.title=element_blank(),legend.text = element_text(size = 5))

## Coordinate system already present. Adding new coordinate system, which will replace the existing one.

#pie_medicare
#Length(unique(table_medicare$MDC_CAT_NAME))
```



pie chart of medicaid:

```
pie_medicaid=ggplot(table_medicaid, aes(x="", y=Medicaid, fill=reorder(MDC_CAT_NAME,-Medicaid)))
pie_medicaid=pie_medicaid+ geom_bar(width = 1, stat = "identity") +
  coord_polar(theta="y", direction = -1) +
  theme_void()+ theme(legend.title = element_blank())
color=c("#800000", "#990F0F", "#B22D2D", "#CC5252", "#E67E7E", "#FFB2B2", "#8d5524", "#99700F", "#B28B2D", "#CC
A852", "#E6C77E", "#FFE8B2", "#1F990F", "#3CB22D", "#60CC52", "#8AE67E", "#BCFFB2", "#710F99", "#8B2DB2", "#A8
52CC", "#C77EE6", "#E9B2FF", "#990F20", "#B22D3C", "#CC5260", "#E67E8A", "#FFB2BC")
pie_medicaid=pie_medicaid+ geom_bar(width = 1, stat = "identity") +
  coord_polar(theta="y", direction = -1) +
  theme_void()+scale_fill_manual(values = color)+
  theme(legend.title=element_blank(),legend.text = element_text(size = 5))

## Coordinate system already present. Adding new coordinate system, which will replace the existing one.

#pie_medicaid
#Length(unique(table_medicaid$MDC_CAT_NAME))
```

pie chart of commercial payers:

```
pie_commercial=ggplot(table_commercial, aes(x="", y=Commercial, fill=reorder(MDC_CAT_NAME,-Commercial)))
pie_commercial=pie_commercial+ geom_bar(width = 1, stat = "identity") +
  coord_polar(theta="y", direction = -1) +
  theme_void()
color=c("#800000", "#990F0F", "#B22D2D", "#CC5252", "#E67E7E", "#FFB2B2", "#8d5524", "#99700F", "#B28B2D", "#CC
A852", "#E6C77E", "#FFE8B2", "#1F990F", "#3CB22D", "#60CC52", "#8AE67E", "#BCFFB2", "#710F99", "#8B2DB2", "#A8
52CC", "#C77EE6", "#E9B2FF", "#990F20", "#B22D3C", "#CC5260", "#E67E8A")
pie_commercial=pie_commercial+ geom_bar(width = 1, stat = "identity") +
  coord_polar(theta="y", direction = -1) +
  theme_void()+scale_fill_manual(values = color)+
  theme(legend.title=element_blank(),legend.text = element_text(size = 5),legend.key.size = unit(0.
5, "cm"))

## Coordinate system already present. Adding new coordinate system, which will replace the existing one.

#pie_commercial
#Length(unique(table_commercial$MDC_CAT_NAME))
```

Personas of Medicare

```
personas_Medicare=inp_2016_medicare[,1:2]
personas_Medicare=merge(personas_Medicare,age,by=c("age"),all.x = TRUE)
personas_Medicare=merge(personas_Medicare,gender,by=c("sex"),all.x = TRUE)
personas_Medicare=personas_Medicare[order(Age,Gender)]

#personas_Medicare

table_age=table(personas_Medicare$Age)
#table_age
# 65+: (3781+3788+12233)/24029=0.8240876
```



```
table_gender=table(personas_Medicare$Gender)
#table_gender
```

Personas of Medicaid

```
personas_Medicaid=inp_2016_medicaid[,1:2]
personas_Medicaid=merge(personas_Medicaid,age,by=c("age"),all.x = TRUE)
personas_Medicaid=merge(personas_Medicaid,gender,by=c("sex"),all.x = TRUE)
#personas_Medicaid[order(Age,Gender)]

table_age=table(personas_Medicaid$Age)
#table_age
table_gender=table(personas_Medicaid$Gender)
#table_gender
```

Personas of commercial payers

```
personas_commercial=inp_2016_commercial[,1:2]
personas_commercial=merge(personas_commercial,age,by=c("age"),all.x = TRUE)
personas_commercial=merge(personas_commercial,gender,by=c("sex"),all.x = TRUE)
#personas_commercial[order(Age,Gender)]

table_age=table(personas_commercial$Age)
#table_age
table_gender=table(personas_commercial$Gender)
#table_gender
```

Question 3

```
remove(list = ls())
library(tidyverse)
library(data.table)
library(dplyr)
library(sqldf)
df=fread("~/Desktop/database/VTED16.TXT")
#3-1. How many ED visits exactly have been diagnosed as drug user/abuser?
df1=df%>%filter_at(vars(starts_with('DX')),any_vars(.%like%'^T4[0-3]'))
df1=unique(df1)
#3-2. Fisher Test
table(df1$sex)
FY=df1[which(df1$sex == 2), ]
MY=df1[which(df1$sex == 1), ]
F_all=df[which(df$sex == 2), ]
M_all=df[which(df$sex == 1), ]
```



```

gender=matrix(c(nrow(FY),nrow(MY),nrow(F_all)-nrow(FY),nrow(M_all)-
nrow(MY)),nrow=2,ncol=2,byrow=T,dimnames=list(c('Y','N'),c('F','M'))))
fisher.test(gender)

#3-3.
#the exact dollar amount for your identified patients in this question
sum(df1$CHRGs,na.rm = TRUE)

#Of the three insurances in Question 2, what was share of each of the total payments?
a=7
df1$PPAY[df1$PPAY==6]=a
#Medicare
sqldf('select sum(CHRGs)/30741220 from df1 where PPAY==1')
#Medicaid
sqldf('select sum(CHRGs)/30741220 from df1 where PPAY==2')
#Commercial_Payers
sqldf('select sum(CHRGs)/30741220 from df1 where PPAY==7')

#3-4.
ED_4=df1 %>% filter_at(vars(starts_with('DX')),any_vars(. %like% '^T404|^T4362'))
nrow(ED_4)
ED_4_2=ED_4 %>% filter(ATYPE==1)
nrow(ED_4_2)

#3-5.
myvars=paste('DX',1:20,sep = '')
myvars=c(myvars,'TXTZIP')
v=df1[myvars]
final=gather(v,'dx','code',-TXTZIP)
vv=final%>%filter(final$code%in%final$code[grep("T40|T41|T42|T43",final$code)])%>%group_by(TXTZIP)%>%summaris
e(total=n())
vv[order(-vv$total),][1:3,]

#3-6.
ED_=df1[,10:29]
data=as.matrix(ED_)
data2=matrix(data,nrow=43020,ncol=1)
data2=data.table(data2)
data2=data2[grep("^T40|^T41|^T42|^T43",data2$V1),]%>%group_by(V1)%>%summarise(n())
data2=data2[order(data2$n()),decreasing=T,]
colnames(data2)<-c("ICD-10","num")
data2[1:10,]

```