Data Analysis Report of Insurance Claims

Menghong Han, Xia Dai, Rui Cao, Mengchun Li, Xuhui Bai, Bo Chen
Team 2

Q1: Patient vignettes

According to this question, we basically merge tables together and select the necessary elements.

For Uniq 507033, 40436 and 690326, we are searching the inpatient table for the information.

For Uniq 1585831, 200760, 859382 and 3692, we firstly used inpatient but then found that the admission types are emergency, so we turn to the emergency for information we needed.

For the Uniq 507033, we derive the following table:

Column1	UNIQ	hnum2	ATYPE	asour	intage	sex	dstat	CHRGS	pdays	ccsdxgrp	REVCODE	REVCHRGS	REVCODE_DESC
1	507033	1	3	3	4	2	5	3233.29	1	11	300	341	Laboratory - Clinical Diagnostic
2	507033	1	3	3	4	2	5	3233.29	1	11	270	75.22	Medical/Surgical Supplies
3	507033	1	3	3	4	2	5	3233.29	1	11	720	1273.76	Labor Room
4	507033	1	3	3	4	2	5	3233.29	1	11	258	92.1	Pharmacy: IV solutions
5	507033	1	3	3	4	2	5	3233.29	1	11	272	334.33	Medical/Surgical Supplies: Sterile suppli
6	507033	1	3	3	4	2	5	3233.29	1	11	259	31.31	Pharmacy: Other
7	507033	1	3	3	4	2	5	3233.29	1	11	120	1002.13	Room & Board (Semi-Private 2 beds)
8	507033	1	3	3	4	2	5	3233.29	1	11	250	83.44	Pharmacy

This a 25-29 years pregnant lady, originated from a non-healthcare facility point with admission type elective stayed at Northwestern Medical Center for 1 day, and she was discharged home for family care afterward. Her insurance, Blue Cross, covered $3233.29. In the hospital, she lived in a semi-private room, which charged $1002.13 and used a labor room that costs $1273.76. She also received services from Laboratory-Clinical Diagnostic($341), Medical/Surgical Supplies($75.22), Pharmacy: IV solutions($92.1), Medical/Surgical Supplies: Sterile supplies($334.33), Pharmacy:Other($31.31), Room & Board($1002.13) and Pharmacy($83.44).

For the Uniq 40436, we derive the following table:

UNIQ	hnum2	ATYPE	asour	intage	sex	dstat	CHRGS	pdays	ccsdxgrp	REVCODE	REVCHRGS	REVCODE_DESC
40436	5	2	1	13	2	5	70275.41	1	7	272	741.33	Medical/Surgical Supplies: Sterile supplies
40436	5	2	1	13	2	5	70275.41	1	7	636	5703.7	Drugs Require Specific ID: Drugs requiring detail coding
40436	5	2	1	13	2	5	70275.41	1	7	324	348.82	Radiology - Diagnostic: Chest X-ray
40436	5	2	1	13	2	5	70275.41	1	7	343	784	Diagnostic Radiopharms
40436	5	2	1	13	2	5	70275.41	1	7	120	1692	Room & Board (Semi-Private 2 beds)
40436	5	2	1	13	2	5	70275.41	1	7	278	10146.48	Medical/Surgical Supplies: Other implants
40436	5	2	1	13	2	5	70275.41	1	7	352	4112.29	CT Scan: Body
40436	5	2	1	13	2	5	70275.41	1	7	730	104	EKG/ECG
40436	5	2	1	13	2	5	70275.41	1	7	250	293.61	Pharmacy
40436	5	2	1	13	2	5	70275.41	1	7	637	301.74	Drugs Require Specific ID: Self admin drugs (insulin adm
40436	5	2	1	13	2	5	70275.41	1	7	481	8387.29	Cardiology: Cardiac catheter lab
40436	5	2	1	13	2	5	70275.41	1	7	482	238	Cardiology: Stress test
40436	5	2	1	13	2	5	70275.41	1	7	483	2142.82	Cardiology: Echocardiology
40436	5	2	1	13	2	5	70275.41	1	7	402	764.38	Other Imaging Services: Ultrasound
40436	5	2	1	13	2	5	70275.41	1	7	762	3948	Treatment/Observation Room: Observation room
40436	5	2	1	13	2	5	70275.41	1	7	480	23275.36	Cardiology
40436	5	2	1	13	2	5	70275.41	1	7	341	5065	Nuclear Medicine: Diagnostic
40436	5	2	1	13	2	5	70275.41	1	7	300	2226.59	Laboratory - Clinical Diagnostic

From the table: A 70-74 years old lady with Diseases of the circulatory system who transferred from a hospital with admission type urgent stayed at the University of Vermont Medical Center (as of 2014) for 1 day and was discharged home for family care afterward. She received services from medical/ ASurgical Supplies: Sterile supplies($ 741.33), Drug Require Specific ID: Drugs requiring detail coding ($5703.7), Radiology- Diagnostic: Chest X-ray($348.82), Diagnostic Radiopharms ($784), Room & Board (Semi-Private 2 beds) ($1692), Medical/ Surgical Supplies: Other implants($10146.48), CT Scan: Body ($4112.29), EKG/ECG($104), Pharmacy ($293.61), Drugs Require Specific ID: Self admin drugs ($301.74), Cardiology: Stress test($238), Cardiology: Echocardiology ($2142.82), Other Imaging Services: Ultrasound ($746.38), Treatment/ Observation Room : Observation room ($3948), Cardiology($23275.36), Nuclear Medicine: Diagnostic($5065) and Laboratory-Clinical Diagnostic($2226.59)Her insurance coverage is $70275.41. Her specific costs and details of her services are shown in the last two columns.

For the Uniq 1585831, we derive the following table:

Column1	UNIQ	hnum2	ATYPE	asour	intage	sex	dstat	CHRGS	pdays	ccsdxgrp	REVCODE	REVCHRGS	REVCODE_DESC
1	1585831	8	1	3	7	2	8	17093.79	1	5	351	1648	CT Scan: Head
2	1585831	8	1	3	7	2	8	17093.79	1	5	320	656	Radiology - Diagnostic
3	1585831	8	1	3	7	2	8	17093.79	1	5	730	490	EKG/ECG
4	1585831	8	1	3	7	2	8	17093.79	1	5	410	3723.1	Respiratory Services
5	1585831	8	1	3	7	2	8	17093.79	1	5	250	2024.81	Pharmacy
6	1585831	8	1	3	7	2	8	17093.79	1	5	270	51.77	Medical/Surgical Supplies
7	1585831	8	1	3	7	2	8	17093.79	1	5	200	4450	Intensive care
8	1585831	8	1	3	7	2	8	17093.79	1	5	450	1227	Emergency Room
9	1585831	8	1	3	7	2	8	17093.79	1	5	300	2392.11	Laboratory - Clinical Diagnostic
10	1585831	8	1	3	7	2	8	17093.79	1	5	259	248.43	Pharmacy: Other
11	1585831	8	1	3	7	2	8	17093.79	1	5	258	182.57	Pharmacy: IV solutions

From the table: A 40-44 years old little girl with mental disorders who originated from a non-healthcare facility point with admission type emergency stayed at Rutland RegionaL Medical Center for 1 day and finally died. She received services from CT Scan: Head ($1648), Radiology- Diagnostic ($656), EKG/ECG ($490), Respiratory Services ($3723.1), Pharmacy ($2024.81), Medical/ Surgical Supplies ($51.77), intensive care ($4450), Emergency Room ($1227), Laboratory- Clinical Diagnostic ($2392.11), Pharmacy: Other($248.43), Pharmacy: IV

solutions ($182.57). So Her insurance coverage is $17093.79. Her specific costs and details of her services are shown in the last two columns.

For the Uniq 200760, we derive the following table:

UNIQ	hnum2	ATYPE	asour	intage	sex	dstat	CHRGS	pdays	ccsdxgrp	REVCODE	REVCHRGS	REVCODE_DESC
200760	5	1	3	3	2	6	49533.15	4	16	420	138.63	Physical Therapy
200760	5	1	3	3	2	6	49533.15	4	16	450	2582.43	Emergency Room
200760	5	1	3	3	2	6	49533.15	4	16	278	10696.77	Medical/Surgical Supplies: Other implants
200760	5	1	3	3	2	6	49533.15	4	16	320	833.28	Radiology - Diagnostic
200760	5	1	3	3	2	6	49533.15	4	16	272	409.96	Medical/Surgical Supplies: Sterile supplies
200760	5	1	3	3	2	6	49533.15	4	16	424	350.34	Physical Therapy: Evaluation/re-evaluation
200760	5	1	3	3	2	6	49533.15	4	16	370	2590.07	Anesthesia
200760	5	1	3	3	2	6	49533.15	4	16	120	6768	Room & Board (Semi-Private 2 beds)
200760	5	1	3	3	2	6	49533.15	4	16	360	14055.8	Operating Room Services
200760	5	1	3	3	2	6	49533.15	4	16	300	610.08	Laboratory - Clinical Diagnostic
200760	5	1	3	3	2	6	49533.15	4	16	250	1387.36	Pharmacy

From the table: A 18-24 years old little girl with injury and poisoning who originated from a non-healthcare facility point with admission type emergency stayed at the University of Vermont Medical Center (as of 2014) for 4 days and was discharged home for home health afterward. Her insurance coverage is $49533.15. She received services from Physical Therapy($138.63), Emergency Room($2582.43), Medical/Surgical Supplies: Other Implants($10696.77), Radiology-Diagnostic($833.28), Medical/Surgical Supplies: Sterile supplies($409.96), Physical Therapy: Evaluation/re-evaluation($350.34), Anesthesia($2590.07), Room&BOard($6768),Operating Room Services($14055.8),Laboratory-Clinical Diagnostic($610.08) and Pharmacy(41387.36).

For the Uniq 859382, we derive the following table:

UNIQ	hnum2	ATYPE	asour	intage	sex	dstat	CHRGS	pdays	ccsdxgrp	REVCODE	REVCHRGS	REVCODE_DESC
859382	8	1	3	5	1	8	13128.19	1	5	300	3339.63	Laboratory - Clinical Diagnostic
859382	8	1	3	5	1	8	13128.19	1	5	200	4450	Intensive care
859382	8	1	3	5	1	8	13128.19	1	5	250	923.99	Pharmacy
859382	8	1	3	5	1	8	13128.19	1	5	258	38.99	Pharmacy: IV solutions
859382	8	1	3	5	1	8	13128.19	1	5	410	2575.58	Respiratory Services
859382	8	1	3	5	1	8	13128.19	1	5	320	328	Radiology - Diagnostic
859382	8	1	3	5	1	8	13128.19	1	5	730	245	EKG/ECG
859382	8	1	3	5	1	8	13128.19	1	5	450	1227	Emergency Room

From the table: A male in the 30-34 age was sent to the Emergency Department of Rutland Regional Medical Center from a Non-Health Care Facility Point of Origin. He suffered from mental disorders and stayed in hospital for one day. He received the service from Laboratory-Clinical Diagnostic($3339.63), Intensive care($4450), Pharmacy($923.99), Pharmacy- IV solutions($38.99), Respiratory Services($2575.58), Radiology-Diagnostic($328), EKG/ECG($245) and Emergency Room($1227) so the total cost is $13128.19. But unfortunately, He died eventually.

For the Uniq 3692, we derive the following table:

UNIQ	hnum2	ATYPE	asour	intage	sex	dstat	CHRGS	pdays	ccsdxgrp	REVCODE	REVCHRGS	REVCODE_DESC
3692	5	2	3	3	1	5	117895.3	58	5	250	5608.26	Pharmacy
3692	5	2	3	3	1	5	117895.3	58	5	300	1716.92	Laboratory - Clinical Diagnostic
3692	5	2	3	3	1	5	117895.3	58	5	320	1873.08	Radiology - Diagnostic
3692	5	2	3	3	1	5	117895.3	58	5	730	52	EKG/ECG
3692	5	2	3	3	1	5	117895.3	58	5	124	106662	Psychiatric
3692	5	2	3	3	1	5	117895.3	58	5	450	1983.03	Emergency Room

A male aged 18 to 24 in urgent condition was sent to the Copley Hospital from the NON-HEALTH CARE FACILITY POINT OF ORIGIN. He suffered from Mental disorders and stayed in hospital for 58 days. He received services from Pharmacy($5608.26), Laboratory-Clinica Diagnostic($1716.92), Radiology-Diagnostic($1837.08), EKG/ECG($52), Psychiatric($106662)and Emergency Room($1983.03) so the total cost is $117895.03. Then he was sent home (own or family care).

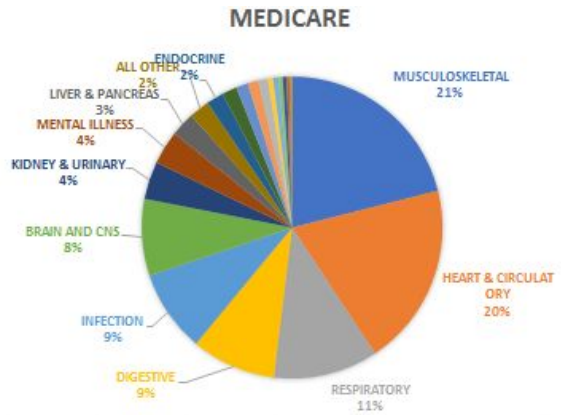
For the Uniq 690326, we derive the following table:

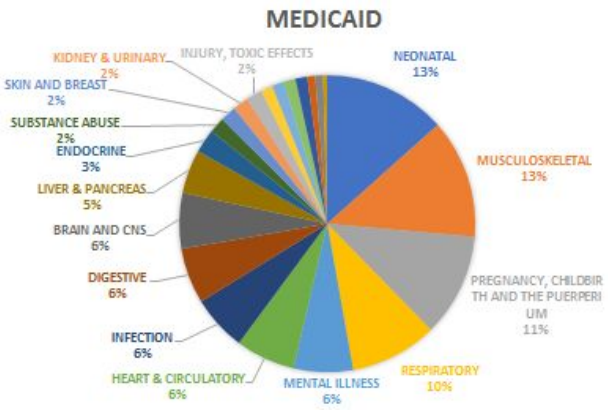
UNIQ	hnum2	ATYPE	asour	intage	sex	dstat	CHRGS	pdays	ccsdxgrp	REVCODE	REVCHRGS	REVCODE_DESC
690326	5	3	3	7	2	5	43425.53	3	18	730	52	EKG/ECG
690326	5	3	3	7	2	5	43425.53	3	18	921	1094.05	Other Diagnostic Services: Peripheral vascular lab
690326	5	3	3	7	2	5	43425.53	3	18	300	327.75	Laboratory - Clinical Diagnostic
690326	5	3	3	7	2	5	43425.53	3	18	250	1351.25	Pharmacy
690326	5	3	3	7	2	5	43425.53	3	18	271	146.18	Medical/Surgical Supplies: Nonsterile supplies
690326	5	3	3	7	2	5	43425.53	3	18	310	273.98	Laboratory - Pathology
690326	5	3	3	7	2	5	43425.53	3	18	370	4609.64	Anesthesia
690326	5	3	3	7	2	5	43425.53	3	18	360	25619.89	Operating Room Services
690326	5	3	3	7	2	5	43425.53	3	18	120	5076	Room & Board (Semi-Private 2 beds)
690326	5	3	3	7	2	5	43425.53	3	18	352	3273.01	CT Scan: Body
690326	5	3	3	7	2	5	43425.53	3	18	410	111.68	Respiratory Services
690326	5	3	3	7	2	5	43425.53	3	18	272	642.32	Medical/Surgical Supplies: Sterile supplies
690326	5	3	3	7	2	5	43425.53	3	18	710	847.78	Recovery Room

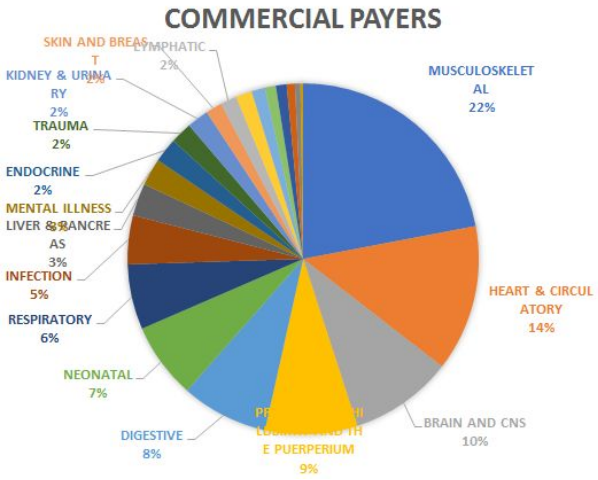
A female patient aged 40 to 44 went to Copley Hospital in elective admission type from NON-HEALTH CARE FACILITY POINT OF ORIGIN. Her diagnose was not classified in chart but the She received services from EKG/ECG($52), Other Diagnostic Services-Peripheral vascular lab ($1094.05), Laboratory-Clinical Diagnostic ($327.75), Pharmacy($1351.25), Medical/Surgical Supplies: Nonsterile supplies ($146.18), Laboratory- Pathology ($273.98), Anesthesia ($4609.64), Operating Room Service ($25619.89), Room& Board (Semi-Private 2 beds) ($5076), CT Scan: Body ($3273.01), Respiratory Services ($111.68), Medical/ Surgical Supplies: Sterile supplies ($642.32), Recovery Room ($847.78), so the total cost is $43425.53. After staying in the hospital for 3 days, she went home and may have own or family care.

Q2: Service and Cost Profile of Major Insurances

First of all, we summarized the fraction of each payment category for the 3 insurances, which is presented below.





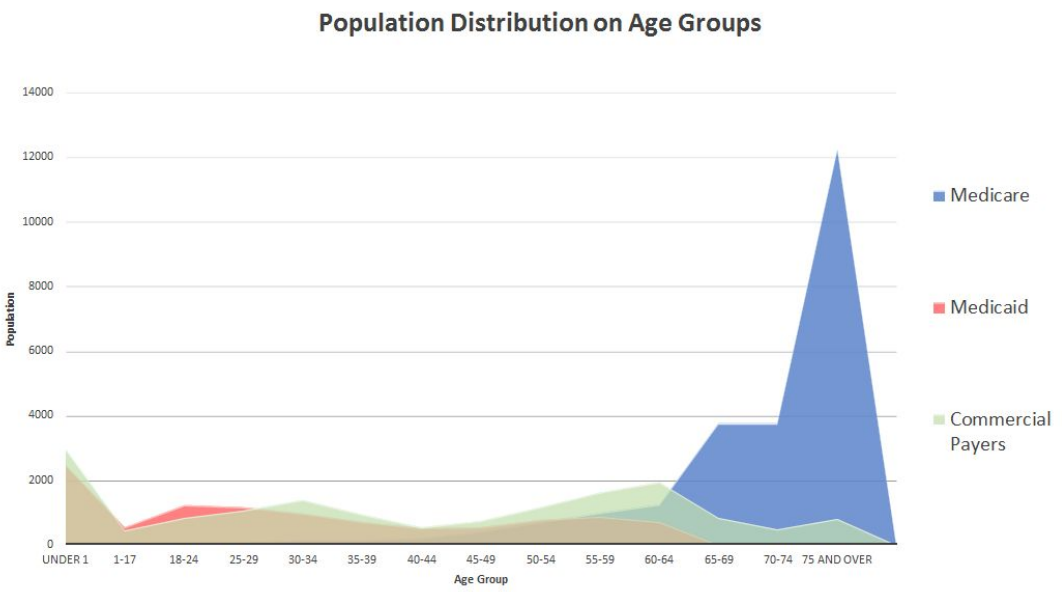


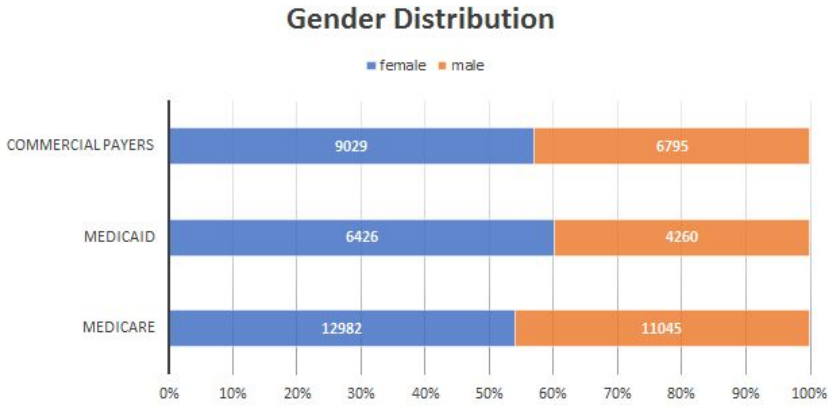
According to the graph, Medicare spent most funds on musculoskeletal services, while the Medicaid paid most for neonatal services. And the money from Commercial Payers went to spleen & blood category.

As the summary, the top-5 MDC of these 3 insurances are as below, with the biggest MDC group for each insurance highlighted:

Medicare	Medicaid	Commercial Payers
MUSCULOSKELETAL	NEONATAL	SPLEEN & BLOOD
HEART & CIRCULATORY	MUSCULOSKELETAL	INFECTION
RESPIRATORY	PREGNANCY, CHILDBIRTH AND THE PUERPERIUM	MUSCULOSKELETAL
DIGESTIVE	RESPIRATORY	EAR, NOSE & THROAT
INFECTION	HEART & CIRCULATORY	RESPIRATORY
	MENTAL ILLNESS	

By analysing by age and gender, we found as bellow:





The difference in the major MDC group for each insurance can be credited to the beneficiary's demographic divergence. As shown in our graph, Medicare mostly services people who are older than 60, while the other two insurances have relatively even distribution of users' age. This may explain why musculoskeletal illnesses and heart & circulatory illnesses cost most to Medicare: these illnesses are more common in older people than the middle-aged and young people.

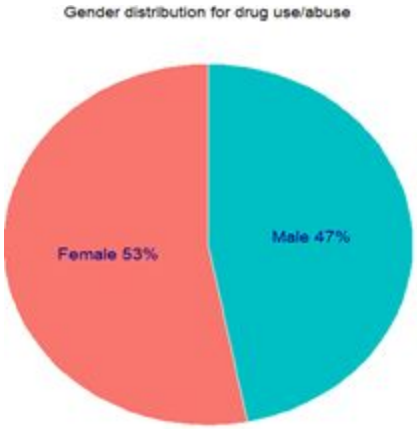
Notably, Medicaid services the middle-aged and young persons, and the proportion of female users in Medicaid (60%) are higher than that of the other two insurances. So, the neonatal, pregnancy, and childbirth appear in the top 5 MDC groups for Medicaid

Meanwhile, the demographics feature of Commercial Payers' users is relatively normal: cover most age groups evenly, and only slight differences in gender. So, the top 5 medical categories for Commercial Payers are those health issues that are common in most age groups: everyone is likely to be infected by a virus, or has some trouble with nose or throat.

Q3: Examining health crisis related to drug use/abuse/overdose

(1) Number of visits diagnosed as drug user/abuser in ED visits: 2151.

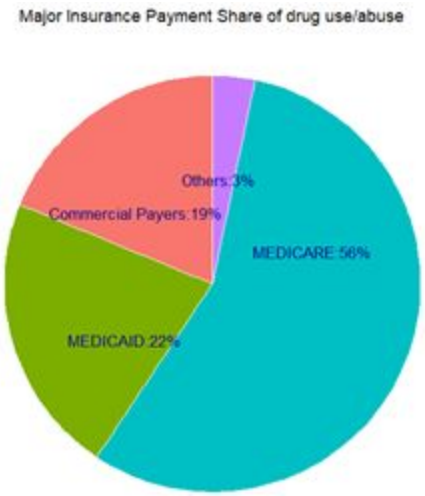
(2)



gender	Abuser number
Female	1,141
Male	1,009
Unknown	1

From the data we extracted and pie chart we plot, we can conclude the myth that the drug use/abuse has been a male problem and that women have much better protection staying away from drug use/abuse is false. Because the data shows there is almost no bias on gender in terms of drug use/ abuse, if there is, we might say women have worse protection because of the larger percentage of 53%. However, it is hard to say there is gender bias on drug use/abuse from the data because of the limited number of samples and almost similar proportions as 47% vs 53%.

(3)

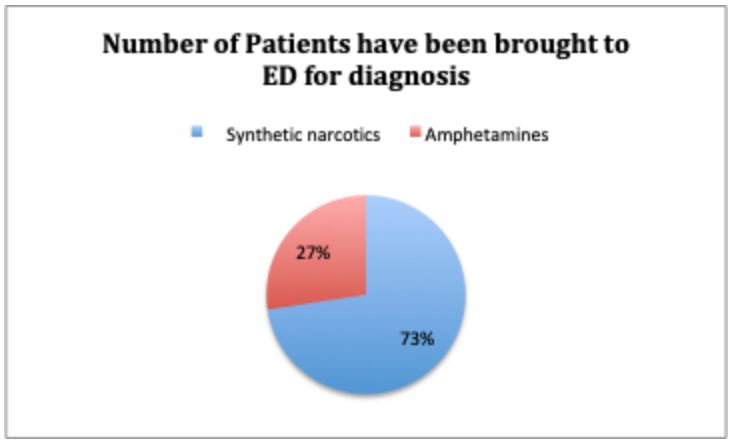


Principle Payment Source	Adjusted Charges	Share
MEDICARE	17,237,917.63	56.07%
MEDICAID	6,713,846.42	21.84%
Commercial Payers	5,777,497.88	18.79%
HMO	285,276.69	0.93%
SELF PAY	264,032.70	0.86%
MISSING OR INVALID	219,774.24	0.71%
WORKER'S COMPENSATION	137,479.26	0.45%
NO CHARGE	66,206.52	0.22%
CHAMPUS	24,591.95	0.08%
OTHER GOVERNMENT	14,596.24	0.05%

Total dollar amount of our identified patients spent on drug use: $ 30741219.53.

Above are the pie chart and data table for the three insurances payment share and others. We can see that the three insurance Medicare (56%), Medicaid (22%) and Commercial Payers (19%) took 97% of total drug use/abuse payments at a huge number, which is almost 30 million dollars. It shows us the three major insurance predominantly occupy or influence the drug use/abuser market.

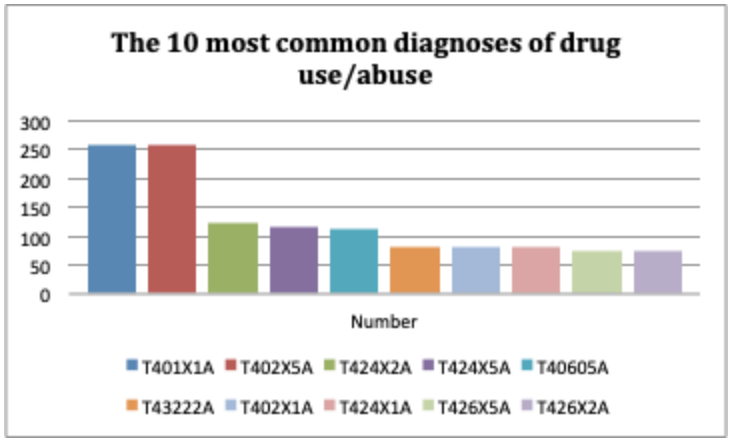
(4)



The number of patients has been brought to ED for diagnosis related to Synthetic narcotics is 114 while amphetamines are 43, a totally of 156.

(5) SOUTH BURLINGTON, COLCHESTER, ESSEX JUNCTION, MILTON (054),MIDDLEBURY (057), RUTLAND (05701) regions have the highest numbers of drug use/abuse cases.

(6) The 10 most common diagnoses of drug use/abuse are:



Code	Meanings
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	Diagnosis of 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, 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

Appendix

Q1

```
VTED16 <- read.csv("VTED16.TXT")
INP16<-read.csv("VTINP16_upd.TXT")
revcode<-read.csv("REVCODE.csv")
rev<-read.csv("VTREVCODE16.TXT")
colnames(revcode)<-c("REVCODE","REVCODE_DESC")
INP<-INP16%>%select(UNIQ,hnum2,ATYPE,asour,intage,sex,dstat,CHRGS,pdays,ccsdxgrp)
ED<-VTED16%>%select(UNIQ,hnum2,ATYPE,asour,intage,sex,dstat,CHRGS,pdays,ccsdxgrp)
rev<-merge(rev,revcode,by="REVCODE")
colnames(rev)[6]<-"UNIQ"
rev<-rev%>%select(REVCODE,UNIQ,REVCHRGS,REVCODE_DESC)
```

```
totalinp<-merge(INP,rev,by="UNIQ")
totaled<-merge(ED,rev,by="UNIQ")
First<-totalinp%>%filter(UNIQ==507033)
Second<-totalinp%>%filter(UNIQ==40436)
Third<-totalinp%>%filter(UNIQ==1585831)
Third<-totaled%>%filter(UNIQ==1585831)
Fourth<-totalinp%>%filter(UNIQ==200760)
Fourth<-totaled%>%filter(UNIQ==200760)
Fifth<- totalinp%>% filter(UNIQ==859382)
Sixth<- totalinp%>% filter(UNIQ==3692)
Seventh<- totalinp%>% filter(UNIQ==690326)
```

Q2

```
CREATE database db_hc_assignment_3;

USE db_hc_assignment_3;
alter table inpatient16 add primary key (uniq);

delete from inpatient16 where mdc = ' ';

drop table if exists tab1;
CREATE TABLE tab1
SELECT
        UNIQ,
   MDC,
   INTAGE,
   SEX,
   CHRGS,
   IF(PPAY = 1, 'MEDICARE', IF(PPAY = 2, 'MEDICAID', 'Commercial Payers')) AS insurance
FROM inpatient16 WHERE PPAY IN (1,2,6,7);

with aa as (
select b. MDC_CAT_NAME, a.insurance, ROUND(SUM(a.chrgs)/1000000, 0) AS
total_charge_$Million FROM tab1 AS a
LEFT JOIN mdc_desc AS b
ON a.MDC = b.MDC
WHERE insurance = 'MEDICARE'
GROUP BY a.MDC, insurance
ORDER BY a.MDC, insurance),
bb as (
```

```
select b. MDC_CAT_NAME, a.insurance, ROUND(SUM(a.chrgs)/1000000, 0) AS
total_charge_$Million FROM tab1 AS a
LEFT JOIN mdc_desc AS b
ON a.MDC = b.MDC
WHERE insurance = 'MEDICAID'
GROUP BY a.MDC, insurance
ORDER BY a.MDC, insurance),
cc as (
select b. MDC_CAT_NAME, a.insurance, ROUND(SUM(a.chrgs)/1000000, 0) AS
total_charge_$Million FROM tab1 AS a
LEFT JOIN mdc_desc AS b
ON a.MDC = b.MDC
WHERE insurance = 'Commercial Payers'
GROUP BY a.MDC, insurance
ORDER BY a.MDC, insurance)
SELECT aa.MDC_CAT_NAME AS MDC_category, aa.total_charge_$Million AS 'Medicare
($million)', bb.total_charge_$Million AS 'Medicaid ($million)', cc.total_charge_$Million AS
'Commercial Payers ($million)'
FROM aa
LEFT JOIN bb on aa.MDC_CAT_NAME = bb.MDC_CAT_NAME
LEFT JOIN cc on aa.MDC_CAT_NAME = cc.MDC_CAT_NAME;


# age analysis
with dd as(
SELECT b.AGE_GRP_DESC, COUNT(a.INTAGE) as total_per_age, a.insurance FROM tab1
as a
LEFT JOIN age as b
ON a.INTAGE = b.INTAGE
WHERE insurance = 'MEDICARE'
GROUP BY a.intage),
ee as(
SELECT b.AGE_GRP_DESC, COUNT(a.INTAGE) as total_per_age, a.insurance FROM tab1
as a
LEFT JOIN age as b
ON a.INTAGE = b.INTAGE
WHERE insurance = 'MEDICAID'
GROUP BY a.intage),
ff as (
SELECT b.AGE_GRP_DESC, COUNT(a.INTAGE) as total_per_age, a.insurance FROM tab1
as a
LEFT JOIN age as b
ON a.INTAGE = b.INTAGE
```

```
WHERE insurance = 'Commercial Payers'
GROUP BY a.intage)
SELECT dd.AGE_GRP_DESC, dd. total_per_age as Medicare, ee.total_per_age as Medicaid,
ff.total_per_age as 'Commercial Payers'
FROM dd
LEFT JOIN ee ON dd.AGE_GRP_DESC = ee.AGE_GRP_DESC
LEFT JOIN ff ON dd.AGE_GRP_DESC = ff.AGE_GRP_DESC
ORDER BY AGE_GRP_DESC;

# gender analysis
with gg as(
select sex, COUNT(sex) as total_per_sex, insurance FROM tab1
WHERE insurance = 'MEDICARE'
GROUP BY sex),
hh as (
select sex, COUNT(sex) as total_per_sex, insurance FROM tab1
WHERE insurance = 'MEDICAID'
GROUP BY sex),
jj as (
select sex, COUNT(sex) as total_per_sex, insurance FROM tab1
WHERE insurance = 'Commercial Payers'
GROUP BY sex)
SELECT if(gg.sex = 1, 'male', if(gg.sex = 2, 'female', 'unknown') )as gender, gg.total_per_sex
AS Medicare, hh.total_per_sex AS Medicaid, jj.total_per_sex AS 'Commercial Payers'
FROM gg
LEFT JOIN hh ON gg.sex = hh.sex
LEFT JOIN jj ON gg.sex = jj.sex;
```

Q3

1. diagnosed as drug user/abuser numbers

```
 create table vted16_abuse as
SELECT *
FROM vted16 t
WHERE substr(DX1, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX2, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
```

```
substr(DX3, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX4, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX5, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX6, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX7, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX8, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX9, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX10, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX11, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX12, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX13, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX14, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX15, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX16, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX17, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX18, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX19, 1, 3) IN ('T40', 'T41', 'T42', 'T43') OR
substr(DX20, 1, 3) IN ('T40', 'T41', 'T42', 'T43')

SELECT count(1)
FROM vted16_abuse
```

2. Gender Distribution

```
SELECT CASE WHEN sex = '1' THEN 'Male' WHEN sex = '2' THEN 'Female' ELSE 'Unknown'
END gender,
count(1) abuser_num
FROM vted16_abuse
GROUP BY CASE WHEN sex = '1' THEN 'Male' WHEN sex = '2' THEN 'Female' ELSE
'Unknown' END;
```

3. Expense and Insurance company share

```
SELECT t.PPAY,
d.ppay_desc,
round(sum(t.CHRGS),2) Adjusted_Charges
FROM vted16_abuse t
LEFT JOIN
dim_ppay d ON t.PPAY = d.PPAY
GROUP BY t.PPAY,
d.ppay_desc;
```

R

```
setwd("E:/MBA@Brandeis/Syllabus/193HS-256F Healthcare Data Analytics and Data Mining/HW3")
getwd()

library(dplyr)
library(ggplot2)

gen_data <- read.csv("HW3_Q32.csv", header = T)

summary(gen_data)

gen_data %>% filter(abuser_num>10) %>%
  mutate(prop = abuser_num/sum(abuser_num)) %>%
  ggplot(aes(x= 1, y = prop, fill = gender)) +
  geom_bar(stat="identity", width=1, color="white") +
  coord_polar(theta="y", start=0) +
  theme_void()+
  ggtitle("Gender distribution for drug use/abuse") +
  theme(legend.position = "none",
      plot.title = element_text(vjust = 0.5, hjust = 0.5)) +
  geom_text(aes(label = paste0(gender,round(prop*100), "%")),
            position = position_stack(vjust = 0.5), size = 5, color = 'darkblue')

pay <- read.csv("ppay_abuse_data.csv", header = T)

out_pay <- pay %>%  mutate(new_pay = case_when(PPAY =='6' ~ "Commercial Payers",
                                PPAY =='7' ~ "Commercial Payers",
                                TRUE ~ as.character(ppay_desc))) %>%
  group_by(new_pay) %>% summarise(adj_charges = sum(Adjusted_Charges)) %>%
  mutate(prop = adj_charges/sum(adj_charges))

write.csv(out_pay,"Insurance_payment_share_drug.csv",row.names = F)

out_pay%>%
 mutate(p_desc = case_when(prop < 0.1 ~ "Others",
                           TRUE ~ as.character(new_pay))) %>%
  group_by(p_desc) %>% summarise(adj_charges = sum(adj_charges), prop = sum(prop)) %>%
  ggplot(aes(x= 1, y = prop, fill = p_desc)) +
  geom_bar(stat="identity", width=1, color="white") +
  coord_polar(theta="y", start=0) +
  theme_void()+
```

```
  ggtitle("Major Insurance Payment Share of drug use/abuse") +
  theme(legend.position = "none",
        plot.title = element_text(vjust = 0.5, hjust = 0.5)) +
  theme(legend.position = "none") +
  geom_text(aes(label = paste0(p_desc,":",round(prop*100), "%")),
            position = position_stack(vjust = 0.3), size = 4.5, color = 'darkblue')


ed <- read.csv("~/Desktop/ed.csv")
library(raster)
library(tidyverse)
# 3(4)
select first three letters in DX1-DX29
ed3<-ed
ed4<-ed
ed5<-ed
ed4[,11:30] <- lapply(ed[,11:30], function(x)substr(x,1,4))
ed5[,11:30] <- lapply(ed[,11:30], function(x)substr(x,1,5))
# select T404X and T4362xx
edT404<-ed4 %>% filter(rowMeans(ed4[,11:30] == "T404")>0)%>% tally()
edT4362<-ed5 %>% filter(rowMeans(ed5[,11:30] == "T4362") > 0)%>% tally()

# 3(5)
# select zipcode
ed3[,11:30] <- lapply(ed[,11:30], function(x)substr(x,1,3))
EDzip<-ed3 %>% filter(rowMeans(ed3[,11:30] == "T40")>0| rowMeans(ed3[,11:30] == "T41") >
0|rowMeans(ed3[,11:30] == "T42") >0|rowMeans(ed3[,11:30] == "T43") > 0)%>%
group_by(TXTZIP)%>%tally(sort = TRUE)

# 3(6)
dx=ed[,11:30]
d=as.data.frame(table(unlist(dx)))
e=d[order(d[,2],decreasing=T),]
f=e %>% filter(substring(Var1,1,3) == "T40"|substring(Var1,1,3)  == "T41"|substring(Var1,1,3)
== "T42"|substring(Var1,1,3)  == "T43")
head(f,10)
```