**HS256F - Healthcare Data Analytics and Data Mining**

**Professor Moaven Razavi**

*February 16, 2022*

*Homework 4*

# INSURANCE CLAIMS DATA & ANALYTICS

**Group 4**

Christianah Adeoya

Cyndi Ng

Siddhartha Kumar

Zilin Luo

# INTRODUCTION

In the United States, the claims records are recognized as an “important source of service utilization data” (CDC, n.d.). Usually the data contains large portions of information across a geographically dispersed population. The data usually comes identifiable for different regions and captures a wide range of information detailing services (diagnoses and procedures) using ICD-10 codes, insurance coverage, and charges. The information captured in the claims data are useful in describing the patient’s pathway through the hospital system as well as the hospital’s service utilization per case (Razavi, 2022).

Some widely recognized sources of claims data include but are not limited to, Government organizations like Medicare and Medicaid, commercial insurances and HMOs, and private data vendors, amongst others (Razavi, 2022). However, while this is recognized as an important data source, some believe that a limitation to its use is the fact that it is put together with the sole aim of reimbursement and as such may not be an accurate representation of the patients’ true medical conditions (CDC, n.d.).

This paper seeks to analyze Vermont’s All Payer Public Use Files for 2016, detailing some patient vignettes, as well as analyzing the data set for service and cost profiles for some major insurance organizations, and examining the data set for the extent of the opioid and illicit substance use crises in Vermont.

The data set used in this report of the Vermont’s All Payer Public Use Files for 2016, containing sub-datasets for patients seen at the Emergency, In-patient, and Outpatient Departments, as well as a data set on the revenue codes and charges, and a codebook that explains the information in these data sets.

## PART 1 - Patient Vignettes

This section of the report details brief descriptions of specific patients' interaction with the health system from the point of origin, through the system and all the services they received during the episode. Also included in each patient's vignette are charge details for the services they received at the point of services, recorded using revenue codes.

**Method:** Using Microsoft Access, each patient's profile was searched for, and information pulled, from both the Emergency Department and Inpatient data files. The values for each code was pulled from the corresponding files and codes resource that came with the data set. The revenue codes data set was also used to detail the breakdown of the patient's charges per encounter.

**VIGNETTES**

---

**Patient 507033**

**Profile:** Patient 507033 is a 25 - 29 year old female with a Blue Cross insurance coverage, who was admitted to the inpatient department for no more than 1 day at Northwestern Medical Center for an elective procedure.

**Diagnosis:** Patient 507033 was admitted from a non-health care facility, possibly their home, principally for an uncomplicated delivery, after a 40-week gestational pregnancy. The patient's case fits into the Diagnosis Related Group (DRG) of a vaginal delivery without complicating diagnosis, and the Major Diagnostic Category (MDC) for inpatients of pregnancy, childbirth and the puerperium, as reported in the data set.

**Procedures**: While at Northwestern Medical Center, the following procedures were carried out on the patient:

- Delivery of products of conception, external approach - Principal procedure
- Drainage of amniotic fluid, therapeutic from products of conception, via natural or artificial opening.

Other obstetrical procedures to assist delivery were performed.

With no recorded arising complications, the patient was discharged after 1 day to their home with a final charge of $3,233.29. Using the revenue codes to verify, the patient was billed for the following services:

| Revenue Code | Description | Charge |
|---|---|---|
| 120 | Room & Board (Semi-Private 2 beds) | $1,002.13 |
| 250 | Pharmacy | $83.44 |
| 258 | Pharmacy: IV solutions | $92.10 |
| 259 | Pharmacy: Other | $31.31 |
| 270 | Medical/Surgical Supplies | $75.22 |
| 272 | Medical/Surgical Supplies: Nonsterile supplies | $334.33 |
| 300 | Laboratory - Clinical Diagnostic | $341.00 |
| 720 | Labor Room | $1,273.76 |
| **Total** | | **$3,233.29** |

**Patient 40436**

**Profile:** Patient 40436 is a 70-74 year old female admitted to the inpatient department of University of Vermont Medical Center on transfer from a hospital. Covered by Medicare, the patient was admitted principally for an urgent procedure related to cardiovascular events.

**Diagnosis:** Patient 40436 was transferred urgently from a hospital, and was diagnosed principally with a Non-ST elevation myocardial infarction. Coupled with this, the patient was noted to have had acute posthemorrhagic anemia, type 2 diabetes mellitus with diabetic neuropathy, atherosclerotic heart disease of native coronary artery without angina pectoris, and hyperlipidemia; which are all related disease

conditions. Alongside these, the patient was noted to have presented with unspecified asthma and a cough. Patient 40436 had a personal history of nicotine dependence, which could be attributed to smoking; they also had a coronary angioplasty implant and graft, a history of transient ischemic attack and cerebral infarction, long term use of insulin and aspirin (indicated for use in treating diabetes and some cardiovascular events respectively), and a family history of ischemic heart diseases and other circulatory system diseases.

**Procedure:** The patient underwent the principal procedure of dilation of the coronary artery to address the urgent myocardial infarction. A fluoroscopy of multiple coronary arteries using low osmolar contrast was also performed on the patient. In line with these, the patient fits into the Diagnosis Related Group (DRG) of a percutaneous cardiovascular procedure with drug-eluting stent without major clinical complications, and the Major Diagnostic Category (MDC) of heart and circulatory, as reported in the claims data.

With no recorded arising complications, the patient was discharged after 1 day to their home with a final charge of $70,275.41. Using the revenue codes to verify, the patient was billed for the following services:

| **Revenue Code** | **Description** | **Charge** |
|---|---|---|
| 120 | Room & Board (Semi-Private 2 beds) | $1,692.00 |
| 250 | Pharmacy | $293.61 |
| 272 | Medical/Surgical Supplies: Sterile supplies | $741.33 |
| 278 | Medical/Surgical Supplies: Other implants | $10,146.48 |
| 300 | Laboratory - Clinical Diagnostic | $2,226.59 |
| 324 | Radiology - Diagnostic: Chest X-ray | $348.82 |
| 341 | Nuclear Medicine: Diagnostic | $5,065.00 |
| 343 | Diagnostic Radiopharms | $784.00 |
| 352 | CT Scan: Body | $4,112.29 |
| 402 | Other Imaging Services: Ultrasound | $764.38 |
| 480 | Cardiology | $23,275.36 |
| 481 | Cardiology: Cardiac catheter lab | $8,387.29 |

| | | |
|---|---|---|
| 482 | Cardiology: Stress test | $238.00 |
| 483 | Cardiology: Echocardiology | $2,142.82 |
| 636 | Drugs Require Specific ID: Drugs requiring detail coding | $5,703.70 |
| 637 | Drugs Require Specific ID: Self admin drugs (insulin admin in emergency-diabetes coma) | $301.74 |
| 730 | EKG/ECG | $104.00 |
| 762 | Treatment/Observation Room: Observation room | $3,948.00 |
| **Total** | | $70,275.41 |

**Patient 859382**

**Profile:** Patient 859382 was a 30-34 year old adult male admitted to the Emergency Department of Rutland Regional Medical Center from a non-health care facility point of origin. Admitted as a self-paying patient, this patient was admitted for no more than one day.

**Diagnosis:** Patient 859382 was admitted to the ED with the principal diagnosis of poisoning by heroin, accidental (unintentional), initial encounter. Other diagnoses from further assessment through their stay include acute respiratory failure with hypercapnia, compression of the brain, uncomplicated opioid poisoning, anoxic brain damage, and cardiac arrest due to other underlying conditions.

**Procedure:** In managing patient 859382's case, primarily, the care providers administered respiratory ventilation (less than 24 consecutive hours), inserted endotracheal airway into the patient's trachea, via natural or artificial opening to address the patient's respiratory failure. A single cardiac output performance was also carried out on the patient.

The patient was transferred from the ED to the ICU during the care process, however, from the data provided, the patient passed on after having a cardiac arrest within 1 day of their admission. Their care incurred a charge of $13,128.19. Using the revenue codes to verify, the patient was billed for the following services:

| Revenue Code | Description | Charge |
|---|---|---|
| 200 | Intensive care | $4,450.00 |
| 250 | Pharmacy | $923.99 |
| 258 | Pharmacy: IV solutions | $38.99 |
| 300 | Laboratory - Clinical Diagnostic | $3,339.63 |
| 320 | Radiology - Diagnostic | $328.00 |
| 410 | Respiratory Services | $2,575.58 |
| 450 | Emergency Room | $1,227.00 |
| 730 | EKG/ECG | $245.00 |
| **Total** | | **$13,128.19** |

**Patient 1585831**

**Profile:** Patient 1585831 was a 40 - 44 year old adult female admitted to the Emergency Department of Rutland Regional Medical Center from a non-health care facility point of origin. Admitted with Medicaid coverage, this patient was at the hospital for no more than one day.

**Diagnosis:** Patient 1585831 was admitted to the ED with the principal diagnosis of poisoning by heroin, accidental (unintentional), initial encounter. Other diagnoses from further assessment through their stay include acute respiratory failure with hypoxia, and acute pulmonary edema. Some of the other diagnosis recorded for the patient during their stay include acute and subacute infective endocarditis, acidosis, cellulitis of right lower limb, poisoning by benzodiazepines, hypokalemia, hyperglycemia, viral hepatitis C, puncture wound in the right foot, intentional self-harm by a sharp object, diarrhea, and bradycardia. The patient's case progressed to cardiac arrest due to other underlying conditions.

**Procedure:** In the management of this patient, the principal procedure performed by the care providers included respiratory ventilation (less than 24 consecutive hours), and a single cardiac output performance.

The patient was transferred from the ED to the ICU during the care process, however, from the data provided, the patient passed on after having a cardiac arrest within 1 day of their admission. Their care incurred a charge of $17,093.79. Using the revenue codes to verify, the patient was billed for the following services:

| **Revenue Code** | **Description** | **Charge** |
|---|---|---|
| 200 | Intensive care | $4,450.00 |
| 250 | Pharmacy | $2,024.81 |
| 258 | Pharmacy: IV solutions | $182.57 |
| 259 | Pharmacy: Other | $248.43 |
| 270 | Medical/Surgical Supplies | $51.77 |
| 300 | Laboratory - Clinical Diagnostic | $2,392.11 |
| 320 | Radiology - Diagnostic | $656.00 |
| 351 | CT Scan: Head | $1,648.00 |
| 410 | Respiratory Services | $3,723.10 |
| 450 | Emergency Room | $1,227.00 |
| 730 | EKG/ECG | $490.00 |
| **Total** | | **$17,093.79** |

**Patient 200760**

**Profile:** Patient 200760 was an 18-24 year old commercially-insured female who was admitted to the Emergency Department of the University of Vermont Medical Center from a non-health care point of origin. All indications point to the fact that the patient had been in an accident prior to presenting at the medical center.

**Diagnosis:** The patient was diagnosed principally with having a displaced fracture of the medial malleolus of the left tibia (closed fracture) and an unspecified fracture of the shaft of the left fibula (closed fracture). During their time at the medical center, the  patient was also recorded to have had

Gastro-Esophageal Reflux Disease (GERD) without esophagitis, as well as a single episode of a major depressive disorder.

**Procedures:** In the management of patient 200760, an open procedure of repositioning the displaced left tibia using an internal fixation device and an intramedullary internal fixation of the fractured left fibula. The patient spent a total of 4 days at the medical center and was charged $49,533.15. They were discharged to their home under the care of a home health agency, ideally with a written plan for care. Using the revenue codes reported, below is a breakdown of the patient's charge:

| **Revenue Code** | **Description** | **Charge** |
|---|---|---|
| 120 | Room & Board (Semi-Private 2 beds) | $6,768.00 |
| 250 | Pharmacy | $1,387.36 |
| 272 | Medical/Surgical Supplies: Sterile supplies | $409.96 |
| 278 | Medical/Surgical Supplies: Other implants | $10,696.77 |
| 300 | Laboratory - Clinical Diagnostic | $610.08 |
| 320 | Radiology - Diagnostic | $833.28 |
| 360 | Operating Room Services | $14,055.80 |
| 370 | Anesthesia | $2,590.07 |
| 420 | Physical Therapy | $138.63 |
| 424 | Physical Therapy: Evaluation/re-evaluation | $350.34 |
| 450 | Emergency Room | $2,582.43 |
| **Total** | | **$40,422.72** |

Worth noting however is that looking through the entire data, there seemed to be a discrepancy in the amount the patient was charged ($49,533.15.) per the information in the ED and IP files, and the charges accounted for in the revenue codes ($40,422.72) to the tune of $9,110.43.

**Patient 3692**

**Profile:**
Patient 3692 is a 28-24 year old adult male admitted to the University of Vermont Medical Center from a non-health care facility point of origin. He is using Blue Cross Insurance. He stayed in the hospital for 68 days and went back home for own or family care.

**Diagnosis:**
Patient 3692's principal diagnosis was bipolar disorder. His other diagnosis included suicidal ideations, personal history of traumatic brain injury and Cannabis dependence.

**Procedure:**
Patient 3692's principal procedure was pharmacotherapy for substance abuse treatment. Another procedure was detoxification services for substance abuse treatment.

| **Revenue Code** | **Description** | **Charge** |
|---|---|---|
| 124 | Psychiatric | $106,662 |
| 250 | Pharmacy | $5,608.26 |
| 300 | Laboratory - Clinical Diagnostic | $1,716.92 |
| 320 | Radiology - Diagnostic | $1,873.08 |
| 450 | Emergency Room | $1,983.03 |
| 730 | EKG/ECG | $52 |
| **Total** | | **$117,895.29** |

**Patient 69032**
**Profile:** Patient 690326 is a 40-44 year old female admitted to the inpatient department of University of Vermont Medical Center, which is located in Burlington, between the month Apr-June in 2016. The discharge status of this patient was family care or own. The total charge was 43425.53 and paid by herself. The patient was admitted principally for cosmetic surgery.

**Diagnosis:** The patient was diagnosed principally with an encounter for cosmetic surgery. Coupled with this, the patient was noted to have pruritus, tachycardia problems, and other muscle spasms. Alongside these, the patient was noted to have presented with bariatric surgery status.

**Procedure:** The patient underwent the principal procedure of alteration of bilateral breast with open approach. Alteration of abdominal wall, an open approach, was also performed on this patient. This patient stayed in the hospital for 3 days. Both of these operations are on the integumentary system. In line with these, the patient fits into the Diagnosis Related Group (DRG) of other skin, subcut tiss & breast proc w/o CC/MCC, and the Major Diagnostic Category (MDC) of skin and breast, as reported in the claims data.

| **Revenue Code** | **Description** | **Charge** |
|---|---|---|
| 120 | Room & Board (Semi-Private 2 beds) | 5076 |
| 250 | Pharmacy | 1351.25 |
| 271 | Medical/Surgical Supplies: Nonsterile supplies | 146.18 |
| 272 | Medical/Surgical Supplies: Sterile supplies | 642.32 |
| 300 | Laboratory - Clinical Diagnostic | 327.75 |
| 310 | Laboratory - Pathology | 273.98 |
| 352 | CT Scan: Body | 3273.01 |
| 360 | Operating Room Services | 25619.89 |
| 370 | Anesthesia | 4609.64 |
| 410 | Respiratory Services | 111.68 |
| 710 | Recovery Room | 847.78 |
| 730 | EKG/ECG | 52 |

| | | |
|---|---|---|
| 921 | Other Diagnostic Services: Peripheral vascular lab | 1094.05 |
| **Total** | | **$ 43,425.53** |

## PART 2 - Service and cost profile of major insurance

This section sought to explore the service and cost provided by 3 major insurance payers, Medicare, Medicaid and Commercial Payers. In this report, BLUE CROSS and COMMERCIAL INSURANCE were combined to serve as Commercial Payers. Further analysis was carried out on the share of MDCs in the percentage of all costs for each payer.

**Method:** Using Python, the data for 3 major insurance payers was filtered and renamed. The non-classified, unknown, and missing rows were also excluded from the new dataset. Then this subdataset was grouped based on Payer and MDC so that the total value of charge for the MDC for each payer could be calculated. The result was in Millions of dollars and was rounded to drop any decimal points. Furthermore, 3 pie-charts were drawn to present the graphical view of the inpatient services portfolio for each insurance.

**Result and Analysis:** The result of cost portfolio of 3 major insurances is shown below:

| **MDC** | | **Medicare** | **Medicaid** | **Commercial** |
|---|---|---|---|---|
| 1 | BRAIN AND CNS | 57 | 14 | 38 |
| 2 | EYE | 0 | 0 | 0 |
| 3 | EAR, NOSE &THROAT | 4 | 2 | 3 |
| 4 | RESPIRATORY | 78 | 22 | 24 |
| 5 | HEART & CIRCULATORY | 136 | 15 | 54 |
| 6 | DIGESTIVE | 63 | 14 | 32 |

| | | | | |
|---|---|---|---|---|
| 7 | LIVER & PANCREAS | 18 | 11 | 12 |
| 8 | MUSCULOSKELETAL | 146 | 30 | 88 |
| 9 | SKIN AND BREAST | 11 | 4 | 6 |
| 10 | ENDOCRINE | 13 | 6 | 9 |
| 11 | KIDNEY & URINARY | 28 | 4 | 8 |
| 12 | MALE REPRODUCTIVE | 3 | 0 | 2 |
| 13 | FEMALE REPRODUCTIVE | 2 | 2 | 4 |
| 14 | PREGNANCY, CHILDBIRTH AND THE PUERPERIUM | 1 | 26 | 34 |
| 15 | NEONATAL | 0 | 31 | 28 |
| 16 | SPLEEN & BLOOD | 7 | 3 | 5 |
| 17 | LYMPHATIC | 8 | 3 | 6 |
| 18 | INFECTION | 61 | 14 | 18 |
| 19 | MENTAL ILLNESS | 25 | 15 | 10 |
| 20 | SUBSTANCE ABUSE | 3 | 4 | 1 |
| 21 | INJURY, TOXIC EFFECTS | 9 | 4 | 4 |
| 22 | BURNS | 1 | 1 | 0 |
| 23 | ALL OTHER | 15 | 3 | 6 |
| 24 | TRAUMA | 4 | 3 | 8 |
| 25 | HIV | 0 | 0 | 0 |

Table 1: Dollar value of charge for the MDCs for 3 insurance companies ($Million)

The table 1 indicates the total charge of each MDC paid by Medicare, Medicaid, and Commercial Insurances respectively. Cost ranges from under 1 million to around 146 million. These numbers were also calculated by percentage for every insurance in the following pie charts.

**Medicare:**

Figure 1: Chart showing the share of the charge for each MDC item in Medicare

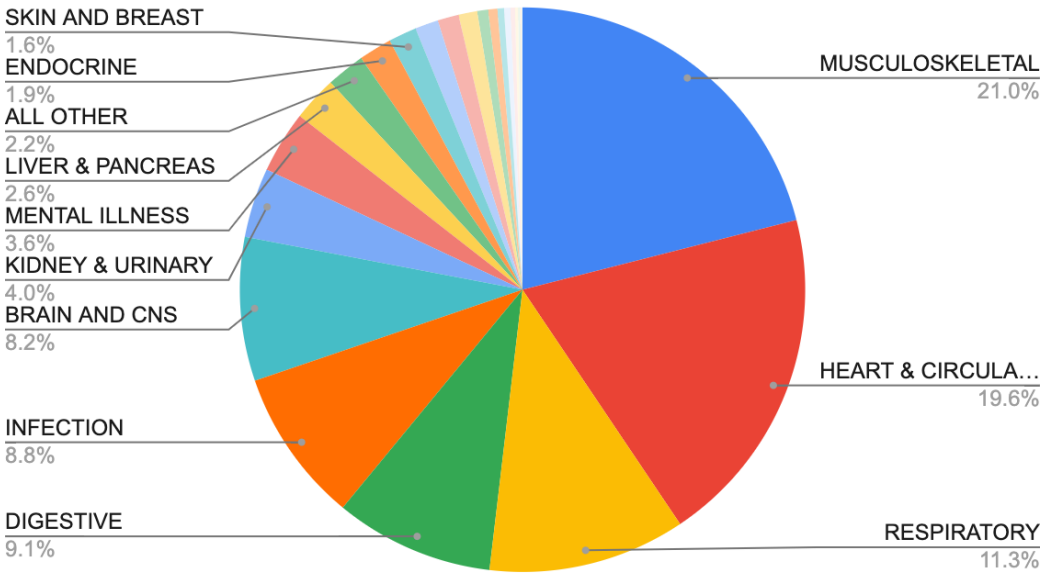



Chart showing gender distribution in Medicare beneficiaries

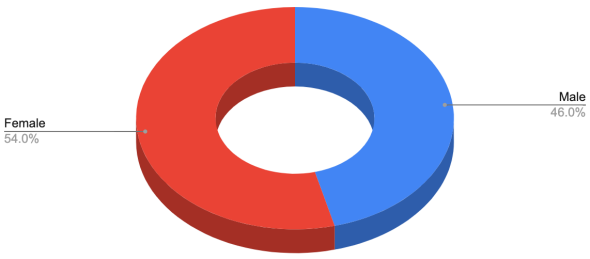
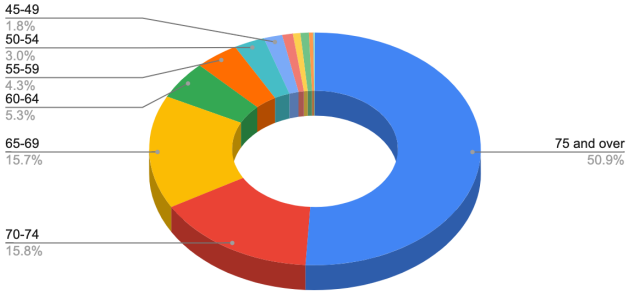



Chart showing the age distribution of Medicare beneficaliers




From the pie chart of Medicare we are observing that the top-5 MDCs are musculoskeletal, heart and circulatory, respiratory, digestive, and infection and the reason MDC musculoskeletal is costing the Medicare so much money is because the majority of members, over 80%, covered by Medicare covers is individuals over 65-year old, musculoskeletal disorders are widespread in elderly persons. Heart and

circulatory problems are also common in older people, so this item also eat large portion of charge in Medicare.

**Medicaid:**

Figure 2: Chart showing the share of the charge for each MDC item in Medicaid

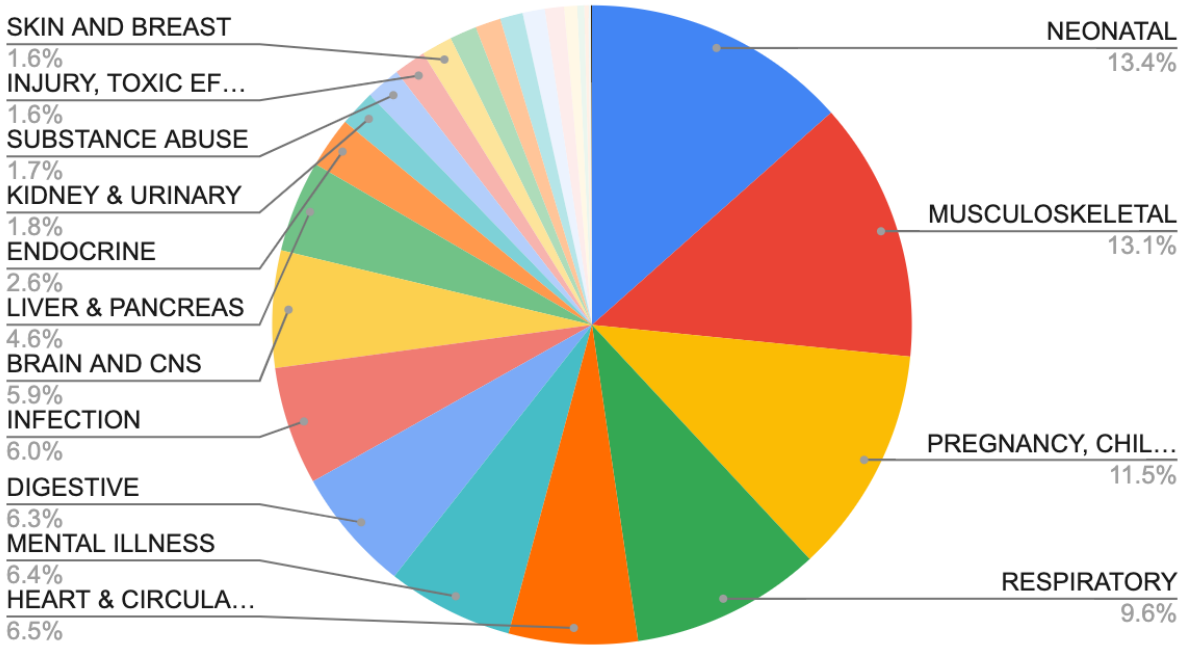



Chart showing the gender distribution of Medicaid beneficiaries

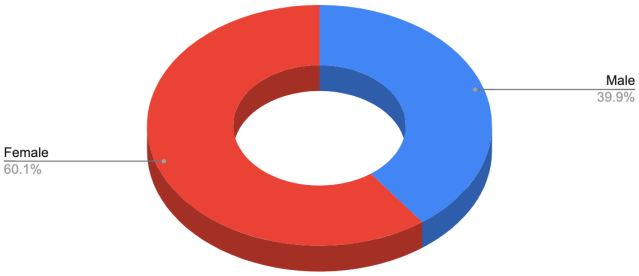
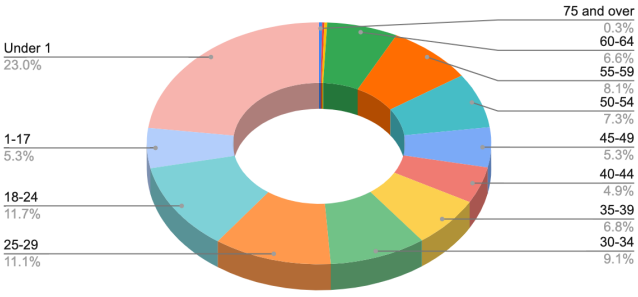



Chart showing the age distribution of Medicaid beneficiaries



The Top-5 MDCs shown in the chart are neonatal, musculoskeletal, pregnancy, childbirth and the puerperium, respiratory, heart and circulatory. Females occupied 60% of Medicaid beneficiaries and 23% of individuals in Medicaid are under 1. The reason why a large portion of the charges are related to women and children is that the beneficiaries of Medicaid are those residents in low income. It means that these people could not spend enough resources and money looking after the health of family members, even vulnerable groups such as pregnant women and children.

**Commercial Payers:**

Figure 3: Chart showing the share of the charge for each MDC item in Commercial Insurances

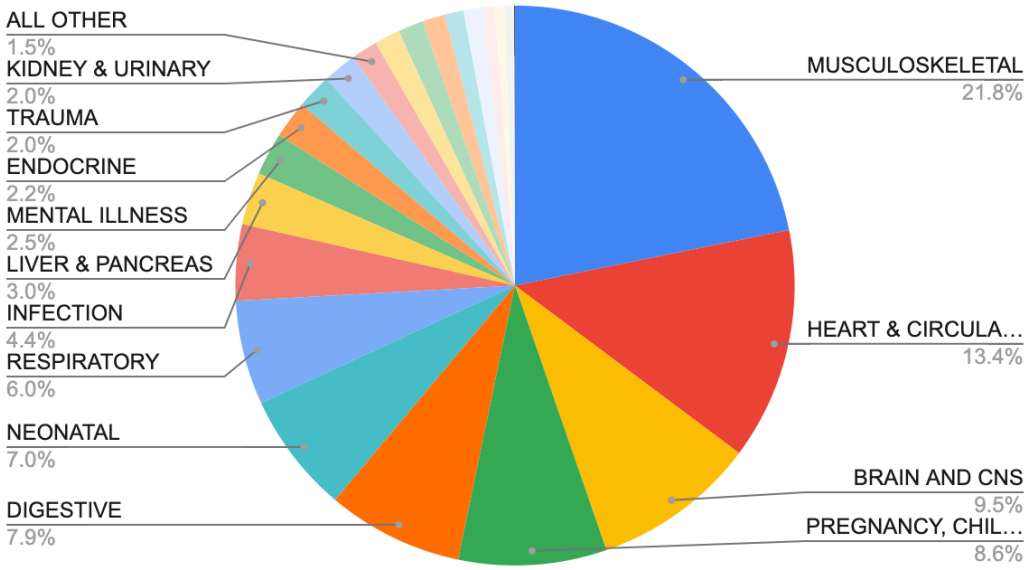



Chart showing the gender distribution of Commercial Insurances beneficiaries

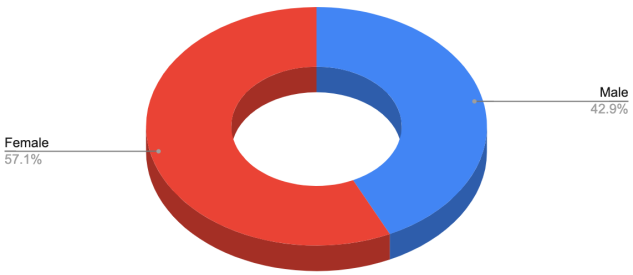
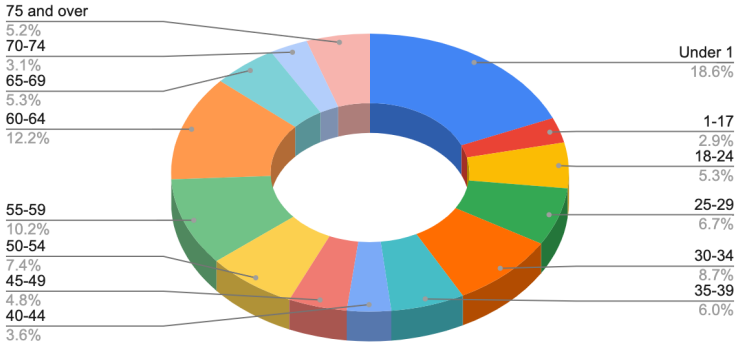



Chart showing the age distribution of Commercial Insurance beneficiaries



From the pie charts we can see that musculoskeletal, heart and circulatory, brain and cns, pregnancy, childbirth and the puerperium, and digestive are the top 5 spending categories. In the age distribution graph, the age group from 50 to 64 covered around 30%. These people are not eligible for Medicare Plan and they are also the prone age group that are affected by musculoskeletal, heart and circulatory, so much money spent on these two items. Similarly, females aged from 25-39 and children under 1 also take over a large part of beneficiaries, so pregnancy and childbirth and the puerperium are also the major MDC paid by commercial insurances.

## PART 3 - Enormity of the health crisis related to use/abuse/overdose of illicit drugs and prescribed opioids

This section details the analysis to quantify the enormity of the health crises in Vermont as it relates to the use or overuse - abuse,  or overdose of both prescribed opioids and illicit drug substances.

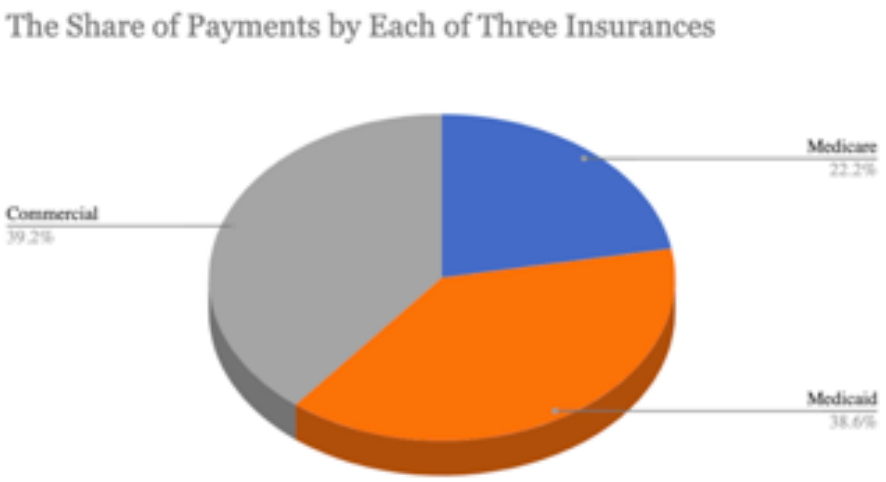
**Method**: Using STATA, a sub data set was created containing those who presented with a diagnosis code related to substance abuse. The resulting dataset was then analyzed for trends

**Result and analysis**:

1. 2,151 visits have exactly been diagnosed as drug users/abusers.

2. The gender bias myth is easily quashed by the aforementioned data. Of the 2,151 drug users/abusers, 1009 are males while 1,141 are females. Thus, since more females are reportedly drug users/abusers, therefore the gender bias does not exist.

3. A sum total of $2,244,410 was spent at the ED for the identified patients. The breakdown is as follows:




4. As per the analysis, 156 patients have been brought to the ED for diagnosis related to synthetic narcotics or amphetamines.

5. Zip codes within the range of 05400 to 05499 (excluding Burlington to Saint Albans), and within 05700 to 05799 (excluding 05701), are the regions with the highest cases of drug users/abusers.

6. According to the National Survey on Drug Use and Health (NSDUH), carried out by Substance Abuse and Mental Health Services administration (SAMHSA) (2021), the ten most common diagnoses related to drug use and abuse are marijuana, cocaine (including crack), heroin, hallucinogens, benzodiazepines, prescription pain relievers, opioids, inhalants, methamphetamine and Central nervous System Stimulants.

## REFERENCES


CDC. (n.d.). *Administrative Claims Databases | Information on Data Sources | Vision & Eye Health Surveillance System | Vision Health Initiative (VHI)*. CDC. Retrieved March 1, 2022, from https://www.cdc.gov/visionhealth/vehss/data/claims/index.html

CMS.gov. (2021, December 1). *MCBS Public Use File*. CMS. Retrieved February 18, 2022, from https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/MCBS-Public-Use-File

Folland, S., Goodman, A. C., & Stano, M. (2017). *The Economics of Health and Health Care: International Student Edition, 8th Edition* (8th ed.). Taylor & Francis Group.

Razavi, M. (2022). *Insurance Claims Data & Analytics*.

Substance Abuse and Mental Health Services Administration. (2021). *Section 1 PE Tables – Results from the 2020 National Survey on Drug Use and Health: Detailed Tables*. SAMHSA. Retrieved March 2, 2022, from https://www.samhsa.gov/data/sites/default/files/reports/rpt35323/NSDUHDetailedTabs2020/NSDUHDetailedTabs2020/NSDUHDetTabsSect1pe2020.htm

## APPENDIX 1

```
SELECT VTINP16_upd.UNIQ, VTINP16_upd.CHRGS, VTINP16_upd.DX1, VTINP16_upd.DX2,
VTINP16_upd.DX3, VTINP16_upd.PX1, VTINP16_upd.PX2, VTINP16_upd.PX3,
VTREVCODE16.REVCODE, VTREVCODE16.REVCHRGS, VTREVCODE16.REVUNITS,
VTREVCODE16.PCCR, VTREVCODE16.ccsproc
FROM VTINP16_upd INNER JOIN VTREVCODE16 ON VTINP16_upd.UNIQ =
VTREVCODE16.Uniq;
```

## APPENDIX 2

```
import pandas as pd
import numpy as np
INP = pd.read_csv("VTINP16_upd.csv",usecols=['PPAY','MDC','CHRGS','intage','sex'])
INP['sex'].replace(' ', np.nan, inplace=True)
INP["MDC"].replace(' ', np.nan, inplace=True)
INP["CHRGS"].replace(' ', np.nan, inplace=True)
INP.dropna(subset=['intage','sex','CHRGS','MDC'],inplace = True)
INP["CHRGS"]=pd.to_numeric(INP["CHRGS"])
INP["sex"] = INP["sex"].astype(int)
INP["MDC"] = INP["MDC"].astype(int)
PP = [1,2,6,7]
INP = INP.loc[INP["PPAY"].isin(PP)]
conditions_1 = [
   INP["PPAY"]== 1,
   INP["PPAY"]==2,
   ((INP["PPAY"]==6)|(INP["PPAY"]==7))
]
values_1 = ["Medicare","Medicaid","Commercial"]
INP["Payer"] = np.select(conditions_1,values_1)
df = INP.groupby(["Payer","MDC"])["CHRGS"].sum().reset_index(name ='total_charge')
df["total_charge(M)"]= (df["total_charge"]/1000000).round()
df["total_charge(M)"]=df["total_charge(M)"].astype(int)
df.to_excel("final_result.xlsx")
INP_Medicare = INP.loc[INP["Payer"]=="Medicare"]
INP_Medicaid = INP.loc[INP["Payer"]=="Medicaid"]
INP_Commerical = INP.loc[INP["Payer"]=="Commercial"]
INP_Medicare = INP_Medicare.groupby(["MDC"])["CHRGS"].sum().reset_index(name ='total_charge')
INP_Medicare["share"]= INP_Medicare["total_charge"].div(INP_Medicare["total_charge"].sum())
INP_Medicare.sort_values('share',ascending=False)
INP_Medicare_f = INP_Medicare.groupby(["sex","intage"])["intage"].count().reset_index(name
='intage_num')
INP_Medicaid = INP_Medicaid.groupby(["MDC"])["CHRGS"].sum().reset_index(name ='total_charge')
INP_Medicaid["share"]= INP_Medicaid["total_charge"].div(INP_Medicaid["total_charge"].sum())
```

```
INP_Medicaid_f = INP_Medicaid.groupby(["sex","intage"])["intage"].count().reset_index(name
='intage_num')
INP_Commerical = INP_Commerical.groupby(["MDC"])["CHRGS"].sum().reset_index(name
='total_charge')
INP_Commerical["share"]=
INP_Commerical["total_charge"].div(INP_Commerical["total_charge"].sum())
INP_Commerical.to_excel("commerical.xlsx")
INP_Commerical_f = INP_Commerical.groupby(["sex","intage"])["intage"].count().reset_index(name
='intage_num')
INP_Commerical_f.to_excel("Commercial_f.xlsx")
```

## APPENDIX 3

. import delimited "VTED16.TXT"

keep if substr(dx1, 1, 3) == "T40" |substr(dx1, 1, 3) == "T41" |substr(dx1, 1, 3) == "T42"|substr(dx1, 1, 3) == "T43"|substr(dx2, 1, 3) == "T40" |substr(dx2, 1, 3) == "T41" |substr(dx2, 1, 3) == "T42"|substr(dx2, 1, 3) == "T43"|substr(dx3, 1, 3) == "T40" |substr(dx3, 1, 3) == "T41" |substr(dx3, 1, 3) == "T42"|substr(dx3, 1, 3) == "T43"|substr(dx4, 1, 3) == "T40" |substr(dx4, 1, 3) == "T41" |substr(dx4, 1, 3) == "T42"|substr(dx4, 1, 3) == "T43"|substr(dx5, 1, 3) == "T40" |substr(dx5, 1, 3) == "T41" |substr(dx5, 1, 3) == "T42"|substr(dx5, 1, 3) == "T43"|substr(dx6, 1, 3) == "T40" |substr(dx6, 1, 3) == "T41" |substr(dx6, 1, 3) == "T42"|substr(dx6, 1, 3) == "T43"|substr(dx7, 1, 3) == "T40" |substr(dx7, 1, 3) == "T41" |substr(dx7, 1, 3) == "T42"|substr(dx7, 1, 3) == "T43"|substr(dx8, 1, 3) == "T40" |substr(dx8, 1, 3) == "T41" |substr(dx8, 1, 3) == "T42"|substr(dx8, 1, 3) == "T43"|substr(dx9, 1, 3) == "T40" |substr(dx9, 1, 3) == "T41" |substr(dx9, 1, 3) == "T42"|substr(dx9, 1, 3) == "T43"|substr(dx10, 1, 3) == "T40" |substr(dx10, 1, 3) == "T41" |substr(dx10, 1, 3) == "T42"|substr(dx10, 1, 3) == "T43"|substr(dx11, 1, 3) == "T40" |substr(dx11, 1, 3) == "T41" |substr(dx11, 1, 3) == "T42"|substr(dx11, 1, 3) == "T43"|substr(dx12, 1, 3) == "T40" |substr(dx12, 1, 3) == "T41" |substr(dx12, 1, 3) == "T42"|substr(dx12, 1, 3) == "T43"|substr(dx13, 1, 3) == "T40" |substr(dx13, 1, 3) == "T41" |substr(dx13, 1, 3) == "T42"|substr(dx13, 1, 3) == "T43"|substr(dx14, 1, 3) == "T40" |substr(dx14, 1, 3) == "T41" |substr(dx14, 1, 3) == "T42"|substr(dx14, 1, 3) == "T43"|substr(dx15, 1, 3) == "T40" |substr(dx15, 1, 3) == "T41" |substr(dx15, 1, 3) == "T42"|substr(dx15, 1, 3) == "T43"|substr(dx16, 1, 3) == "T40" |substr(dx16, 1, 3) == "T41" |substr(dx16, 1, 3) == "T42"|substr(dx16, 1, 3) == "T43"|substr(dx17, 1, 3) == "T40" |substr(dx17, 1, 3) == "T41" |substr(dx17, 1, 3) == "T42"|substr(dx17, 1, 3) == "T43"|substr(dx18, 1, 3) == "T40" |substr(dx18, 1, 3) == "T41" |substr(dx18, 1, 3) == "T42"|substr(dx18, 1, 3) == "T43"|substr(dx19, 1, 3) == "T40" |substr(dx19, 1, 3) == "T41" |substr(dx19, 1, 3) == "T42"|substr(dx19, 1, 3) == "T43"|substr(dx20, 1, 3) == "T40" |substr(dx20, 1, 3) == "T41" |substr(dx20, 1, 3) == "T42"|substr(dx20, 1, 3) == "T43"
(263,711 observations deleted)

. tab sex

```
      sex |      Freq.     Percent        Cum.
----------+-----------------------------------
```

```
          |          1        0.05        0.05
        1 |      1,009       46.91       46.95
        2 |      1,141       53.05      100.00
----------+-----------------------------------
    Total |      2,151      100.00
```

```
. encode chrgs,gen(charges)

. total charges

Total estimation                  Number of obs = 2,151

------------------------------------------------------------------
             |      Total   Std. err.     [95% conf. interval]
-------------+----------------------------------------------------
     charges |    2244410    27576.5       2190331     2298489

. generate MajorIns=.

. replace MajorIns=1 if ppay ==1

.
. replace MajorIns=2 if ppay ==2

. replace MajorIns=3 if ppay ==6 | ppay ==7

. bysort MajorIns : egen Sum = total(charges)

. tab MajorIns Sum

. keep if substr(dx1, 1, 4) == "T404" |substr(dx1, 1, 5) == "T4362" |substr(dx2, 1, 4) == "T404"
|substr(dx2, 1, 5) == "T4362"|substr(dx3, 1, 4) == "T404" |substr(dx3, 1, 5) == "T4362"|substr(dx4, 1, 4)
== "T404" |substr(dx4, 1, 5) == "T4362"|substr(dx5, 1, 4) == "T404" |substr(dx5, 1, 5) ==
```

"T4362"|substr(dx6, 1, 4) == "T404" |substr(dx6, 1, 5) == "T4362"|substr(dx7, 1, 4) == "T404" |substr(dx7, 1, 5) == "T4362"|substr(dx8, 1, 4) == "T404" |substr(dx8, 1, 5) == "T4362"|substr(dx9, 1, 4) == "T404" |substr(dx9, 1, 5) == "T4362"|substr(dx10, 1, 4) == "T404" |substr(dx10, 1, 5) == "T4362"|substr(dx11, 1, 4) == "T404" |substr(dx11, 1, 5) == "T4362" |substr(dx12, 1, 4) == "T404" |substr(dx12, 1, 5) == "T4362"|substr(dx13, 1, 4) == "T404" |substr(dx13, 1, 5) == "T4362"|substr(dx14, 1, 4) == "T404" |substr(dx14, 1, 5) == "T4362"|substr(dx15, 1, 4) == "T404" |substr(dx15, 1, 5) == "T4362"|substr(dx16, 1, 4) == "T404" |substr(dx16, 1, 5) == "T4362"|substr(dx17, 1, 4) == "T404" |substr(dx17, 1, 5) == "T4362"|substr(dx18, 1, 4) == "T404" |substr(dx18, 1, 5) == "T4362"|substr(dx19, 1, 4) == "T404" |substr(dx19, 1, 5) == "T4362"|substr(dx20, 1, 4) == "T404" |substr(dx20, 1, 5) == "T4362"

. bysort txtzip : generate n = _N
. tab txtzip n