

□

Healthcare Data Analytics and Data Mining

# Cluster Analysis

Group 3

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



# Content

- 1 Introduction
- 2 Data Processing
- 3 Results and Analysis
- 4 Conclusion



# Introduction

# Introduction

- Clustering a certain cost categories of the inpatient hospital diagnosis related groups (DRGs)
- Examine the Calinski-Harabasz F-statistics of different clusters
- Analysis of common properties of the DRGs using inpatient data and revenue data
- Some scenarios and prediction



Data Processing

# Data Processing

## Filter data

Only maintaining important DRGs between 20 and 977 and dropping the revenue charges which are less than \$100.

## Create new variable

Creating a new variable which is the combination of Operating Room and Anesthesiology.

## Plug in the names

Converting the code of DRG and PCCR to the names and making a cross table with the DRGs in the row and mean value of the PCCRs.

## Turn NA to zero

Turning all empty cells to zero and used these values for as costs for clustering.

# Data Processing

	Anesthesiology <dbl>	Angiocardiography <dbl>	ASC <dbl>	Audiology <dbl>	Bacteriology and Microbiology <dbl>	Birthing Center <dbl>
Abortion w D&C, aspiration curettage or hysterotomy	803.4263	0.000	0.000	0.0000	0.0000	0.000
Abortion w/o D&C	330.9900	0.000	0.000	0.0000	669.5000	3557.410
Acute & subacute endocarditis w CC	0.0000	0.000	0.000	0.0000	379.0000	0.000
Acute & subacute endocarditis w MCC	596.8000	0.000	914.000	0.0000	1728.0625	0.000
Acute adjustment reaction & psychosocial dysfunction	0.0000	0.000	0.000	0.0000	492.0000	0.000
Acute ischemic stroke w use of thrombolytic agent w CC	0.0000	0.000	0.000	0.0000	0.0000	0.000
Acute ischemic stroke w use of thrombolytic agent w MCC	0.0000	0.000	0.000	0.0000	0.0000	0.000
Acute ischemic stroke w use of thrombolytic agent w/o CC/MCC	0.0000	0.000	0.000	0.0000	0.0000	0.000
Acute leukemia w/o major O.R. procedure w CC	1899.8433	0.000	0.000	0.0000	0.0000	0.000
Acute leukemia w/o major O.R. procedure w MCC	846.5440	0.000	6037.605	0.0000	0.0000	0.000
Acute leukemia w/o major O.R. procedure w/o CC/MCC	0.0000	0.000	0.000	0.0000	0.0000	0.000
Acute major eye infections w CC/MCC	0.0000	0.000	0.000	0.0000	0.0000	0.000
Acute major eye infections w/o CC/ MCC	0.0000	0.000	0.000	0.0000	0.0000	0.000
Acute myocardia infarction, discharged alive w/o CC/MCC	409.5000	0.000	0.000	0.0000	207.0000	0.000
Acute myocardial infarction, discharged alive w CC	296.6600	3326.680	0.000	0.0000	300.6667	0.000
Acute myocardial infarction, discharged alive w MCC	332.2675	0.000	0.000	0.0000	708.2000	0.000
Acute myocardia infarction, expired w CC	0.0000	0.000	0.000	0.0000	242.0000	0.000
Acute myocardial infarction, expired w MCC	0.0000	0.000	0.000	0.0000	382.8000	0.000
Acute myocardial infarction, expired w/o CC/MCC	0.0000	0.000	0.000	0.0000	0.0000	0.000
Admit for renal dialysis	0.0000	0.000	0.000	0.0000	0.0000	0.000
Adrenal & pituitary procedures w CC/ MCC	3404.8300	0.000	12714.110	0.0000	0.0000	0.000
Adrenal & pituitary procedures w/o CC/MCC	2916.3909	0.000	14055.800	0.0000	0.0000	0.000
Aftercare w CC/MCC	730.2400	0.000	0.000	426.2100	256.0000	0.000
Aftercare w/o CC/MCC	1767.5900	0.000	0.000	0.0000	0.0000	0.000
Aftercare, musculoskeletal system & connective tissue w CC	1485.2175	0.000	4963.920	0.0000	329.6667	0.000
Aftercare, musculoskeletal system & connective tissue w MCC	0.0000	0.000	0.000	0.0000	209.0000	0.000
Aftercare, musculoskeletal system & connective tissue w/o CC/MCC	1517.9857	0.000	0.000	0.0000	314.0000	0.000
AIKD lead & generator procedures .....	1728.6150	0.000	0.000	0.0000	0.0000	0.000
AIKD Lead Procedures	0.0000	0.000	0.000	0.0000	0.0000	0.000
Alcohol/drug abuse or dependence w/o rehabilitation therapy w MCC	0.0000	0.000	948.000	0.0000	852.0000	0.000
Alcohol/drug abuse or dependence w/o rehabilitation therapy w/o MCC	596.8000	0.000	1729.920	0.0000	316.0000	0.000
Alcohol/drug abuse or dependence, left ama	~ 0.0000	~ 0.000	~ 0.000	~ 0.0000	~ 0.0000	0.000
Allergic reactions w MCC .....	.....	.....	.....	.....	.....	0.000
Amputat of lower limb for endocrine,nutrit,& meta	1054.2063	0.000	0.000	0.0000	1505.5625	0.000
Amputat of lower limb for endocrine,nutrit,& meta	1212.0000	0.000	0.000	0.0000	0.0000	0.000

687 rows of DRGs, 54 PCCRs and 1 combined PCCR



## Results and Analysis

# Results and Analysis

## F-statistics

We first try to cluster the cost data into 2,3,4 and 5 clusters and examine the Calinski-Harabasz f-statistics to see the best clustering of cost.

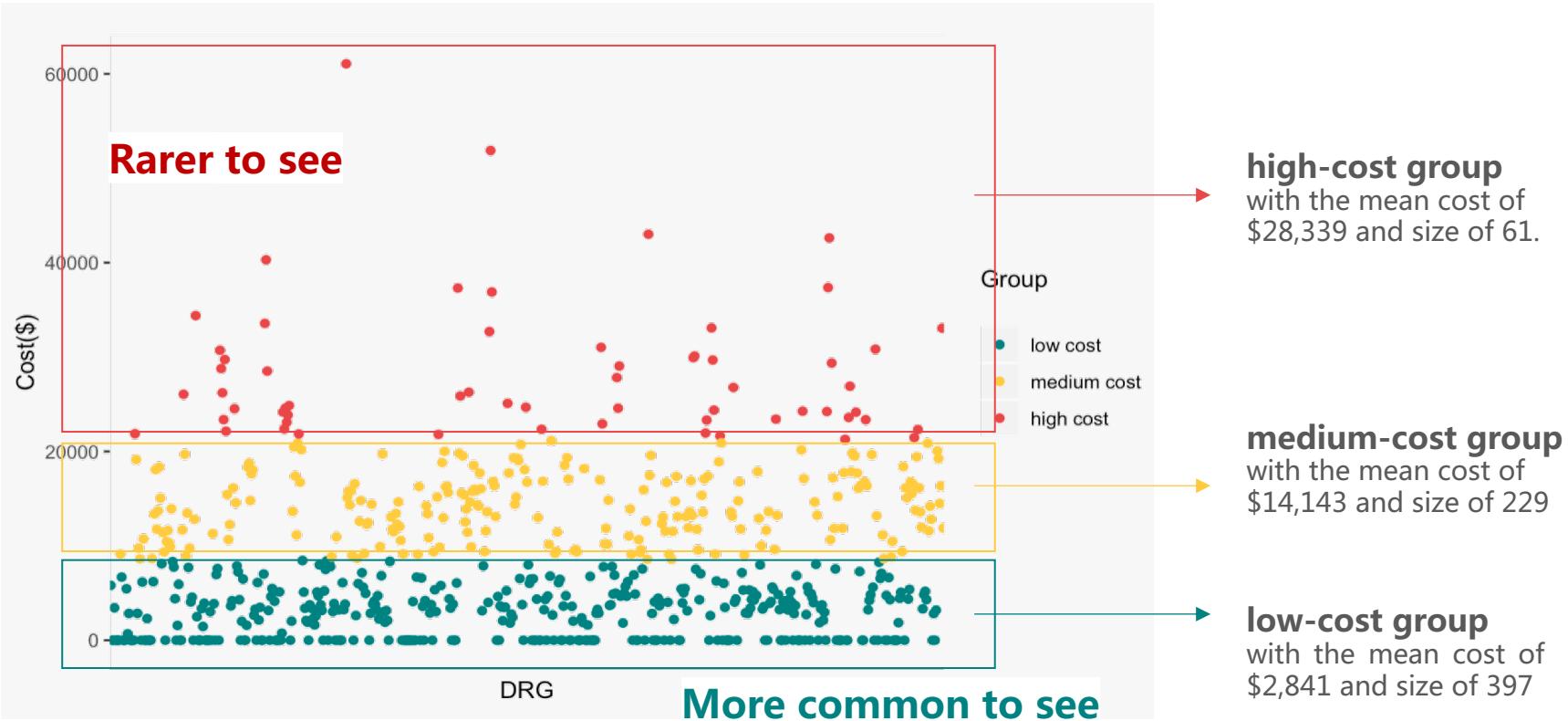
**Table 1 Calinski-Harabasz f-statistics of different numbers of clusters**

Number of clusters	f-statistics
2	1432.768
3	1700.194
4	1934.762
5	2134.944

**5 clusters may be the best clustering of costs**

# Results and Analysis

## Solution with 3 clusters



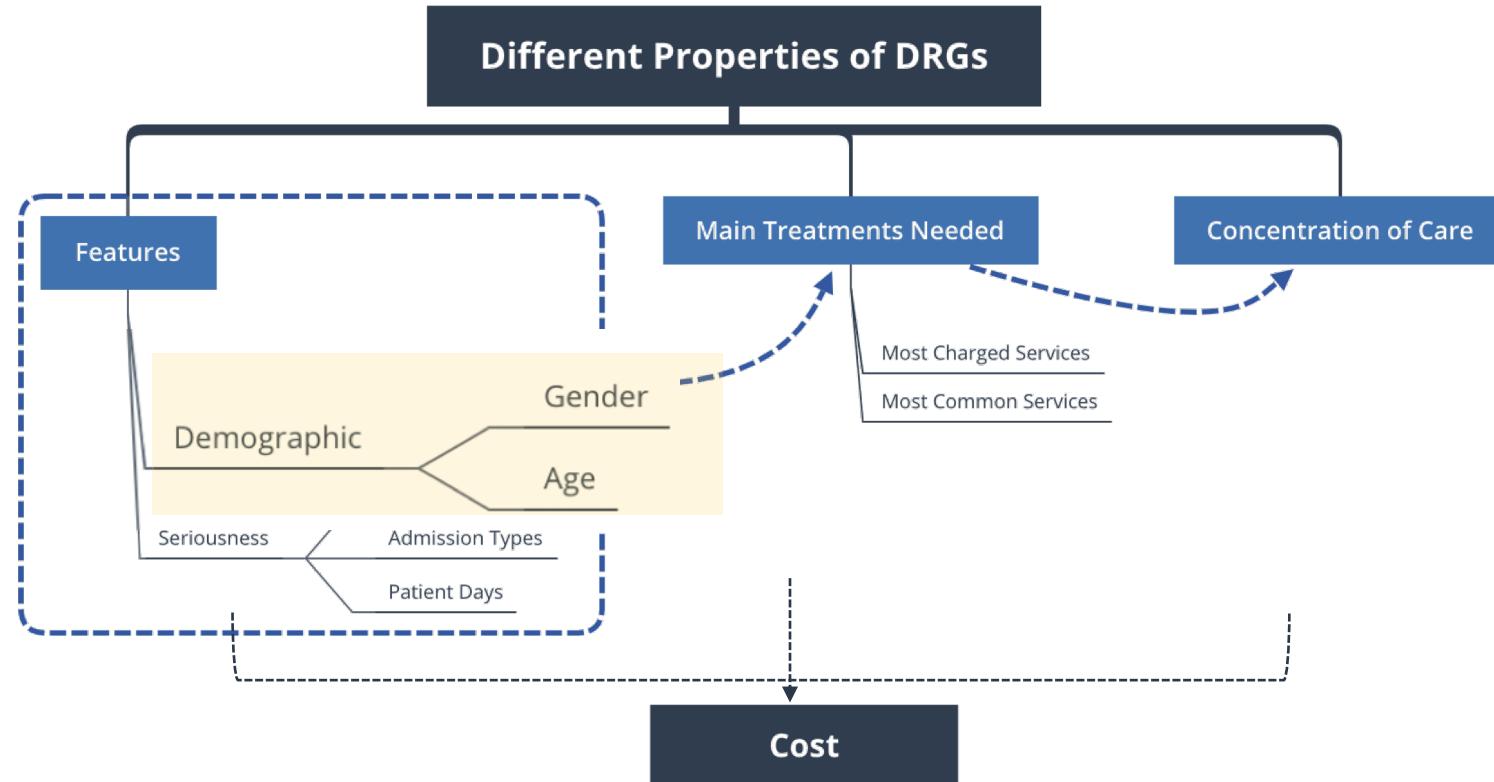
# Results and Analysis

## Features behind



# Results and Analysis

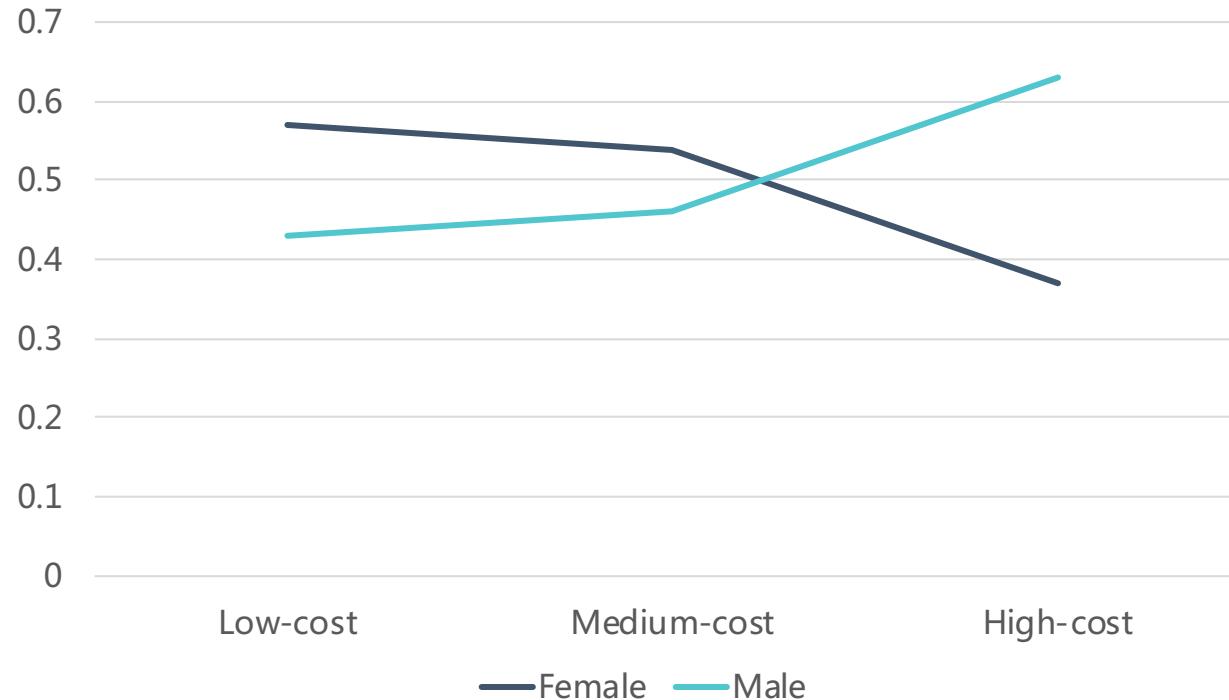
## Analysis logic map



# Results and Analysis

## The demographic of DRGs

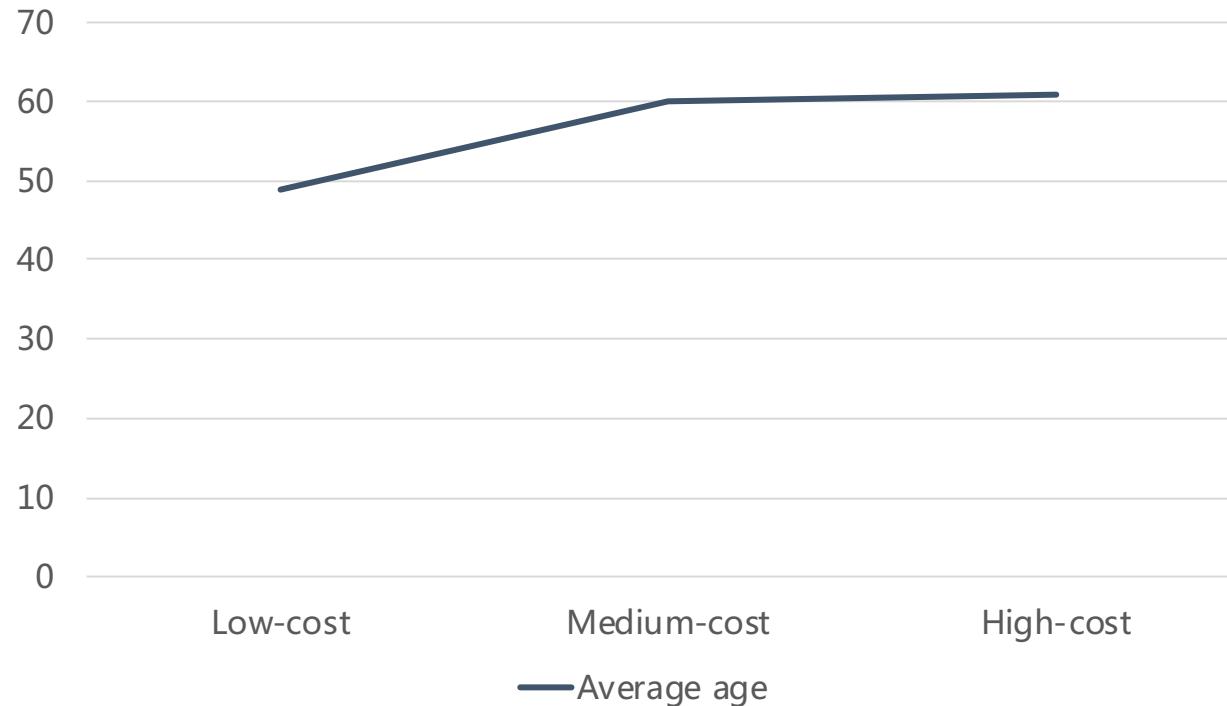
Gender and DRGs



# Results and Analysis

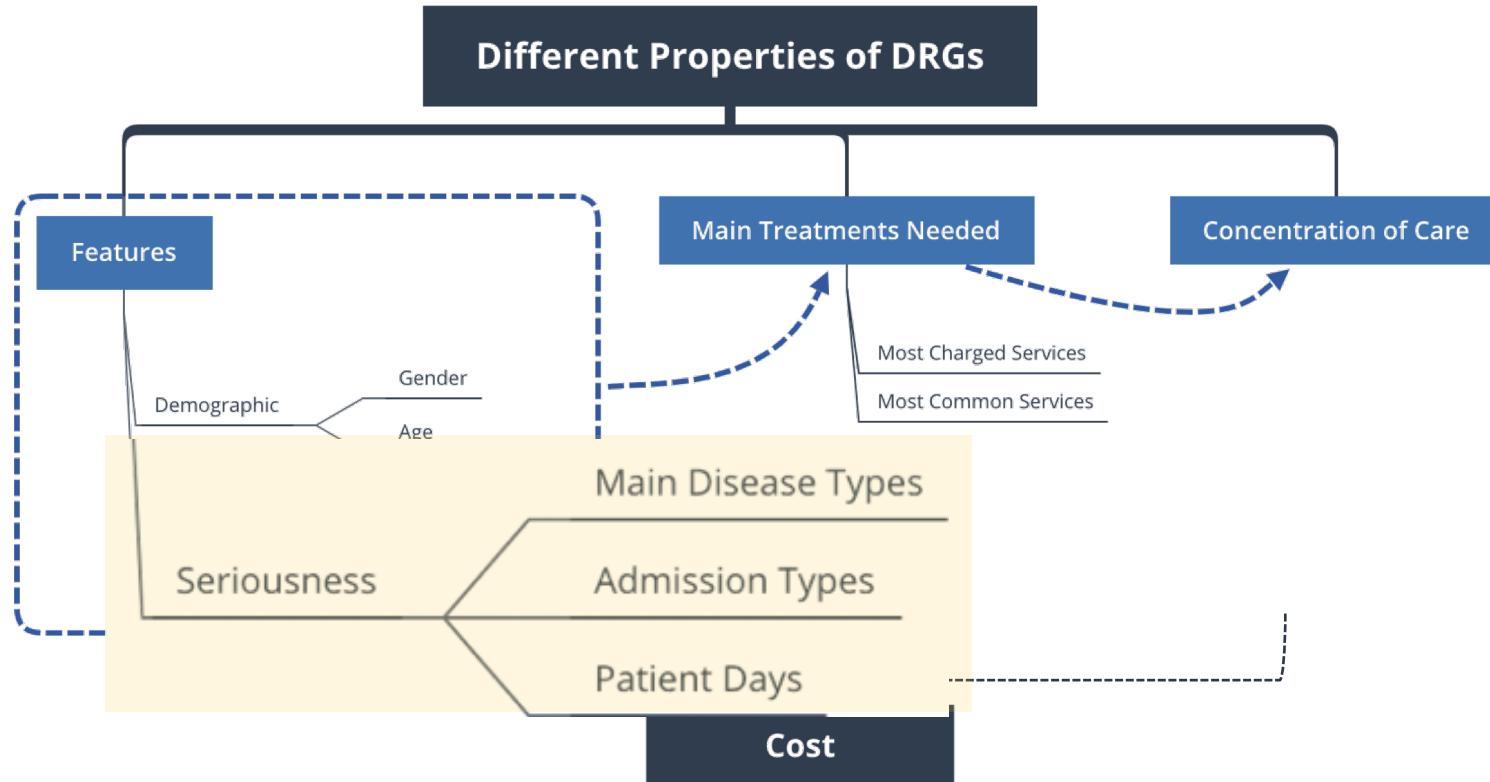
## The demographic of DRGs

Average age and DGRs



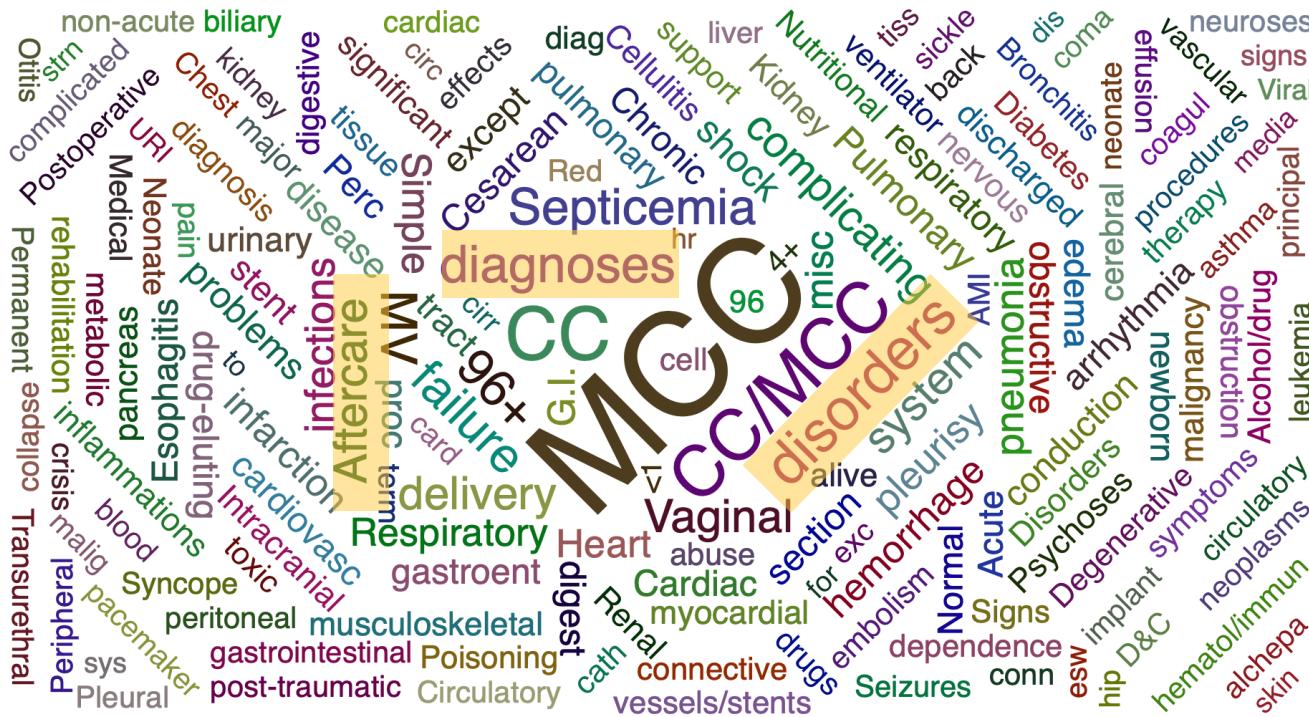
# Results and Analysis

## Analysis logic map



## Results and Analysis

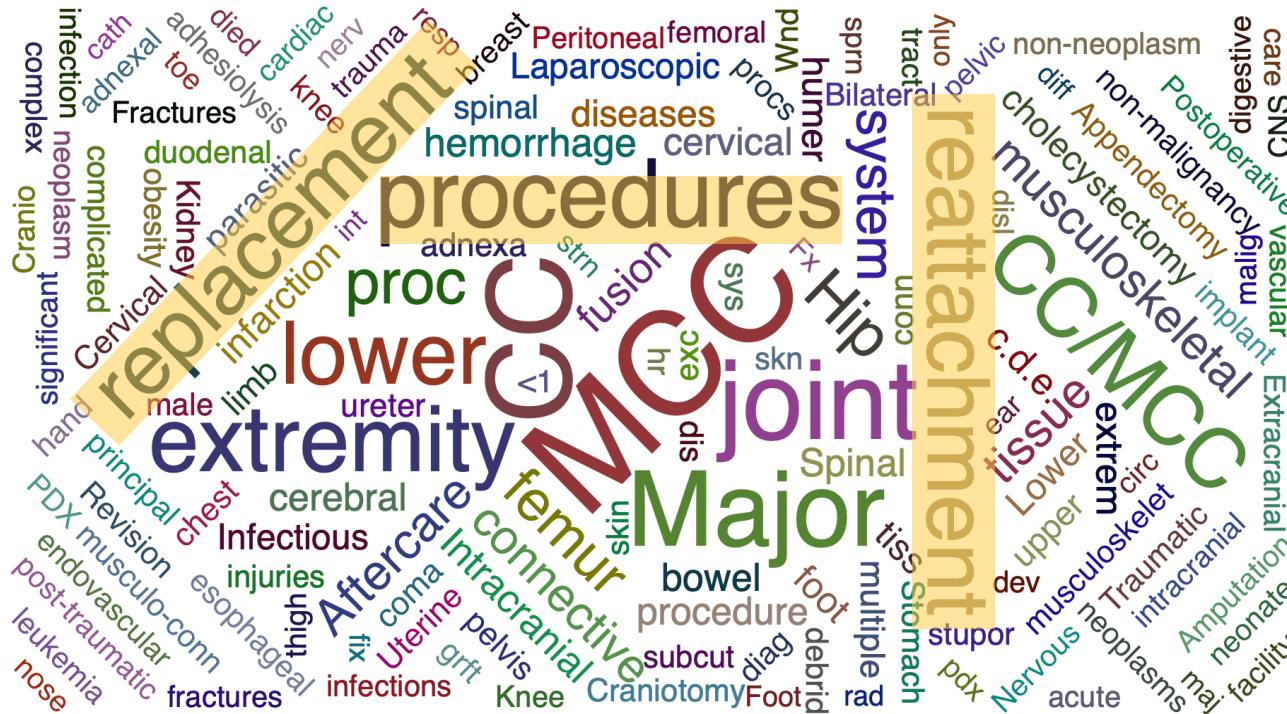
## The seriousness of DRGs



# World cloud of Low-cost group

# Results and Analysis

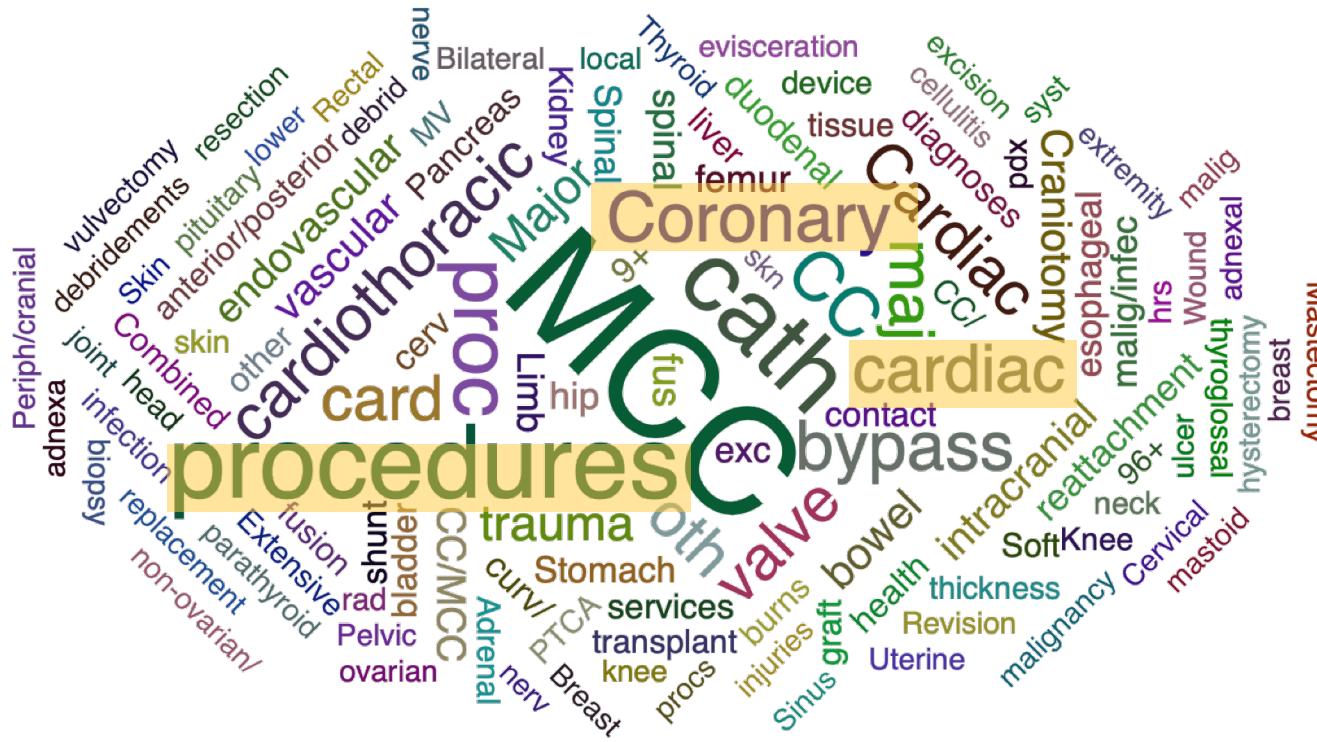
## The seriousness of DRGs



# World cloud of Medium-cost group

## Results and Analysis

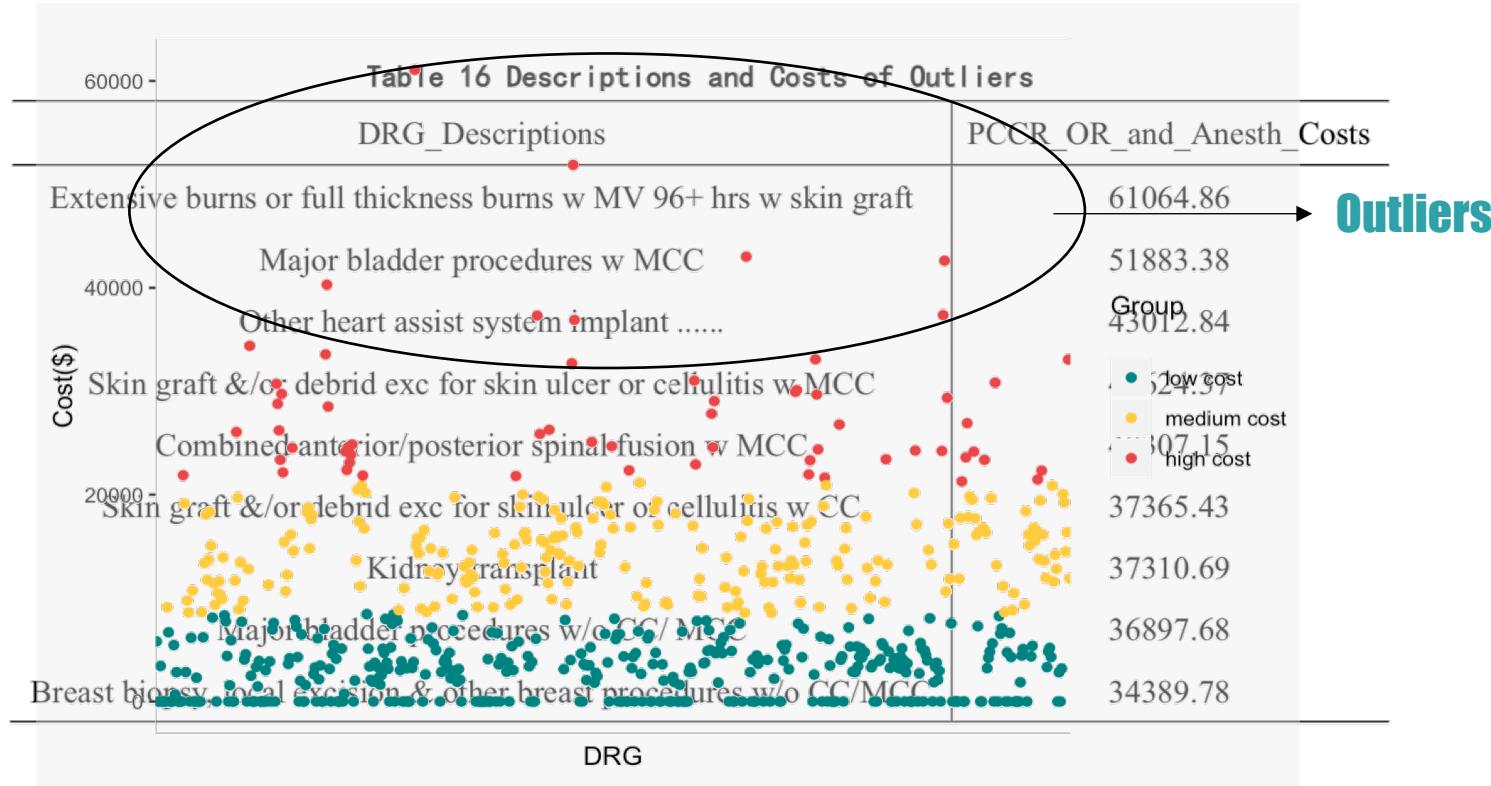
## The seriousness of DRGs



## World cloud of High-cost group

# Results and Analysis

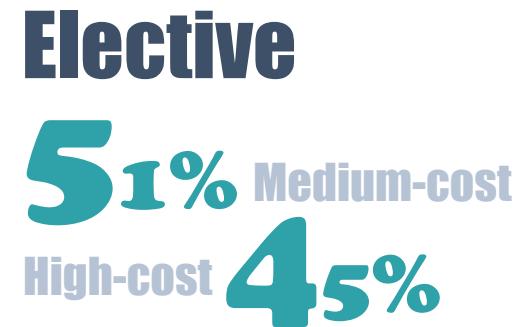
## The seriousness of DRGs



# Results and Analysis

- The seriousness of DRGs

Admission Types



# Results and Analysis

## The seriousness of DRGs

Patients Days



Low-cost



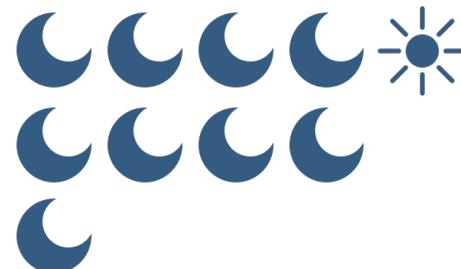
**4.3 days**

Medium-cost



**4.7 days**

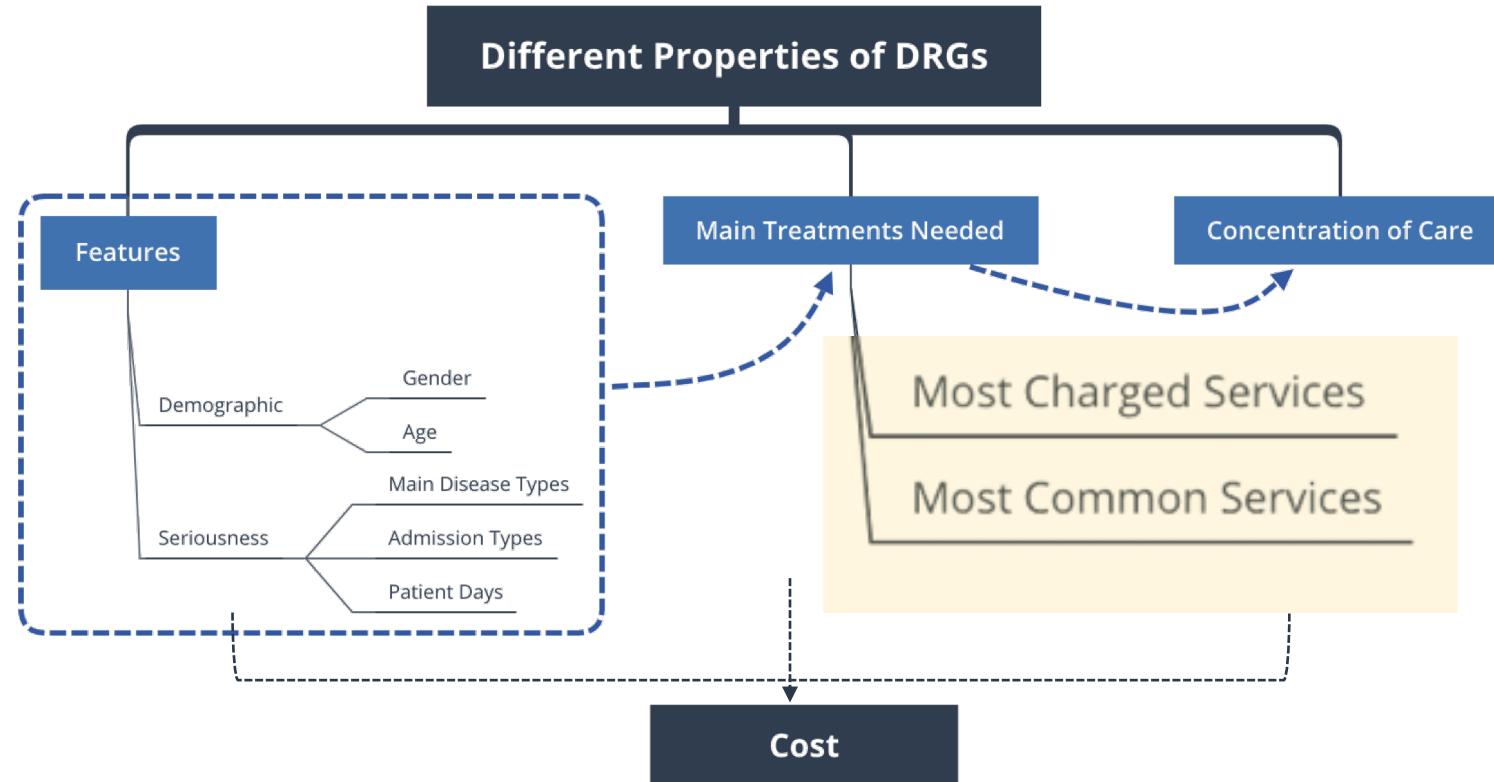
High-cost



**9.5 days**

# Results and Analysis

## Analysis logic map



# Results and Analysis

## The main treatments of DRGs

Top 5 medical services in terms of total charges and total units

Low-cost



Medium-cost



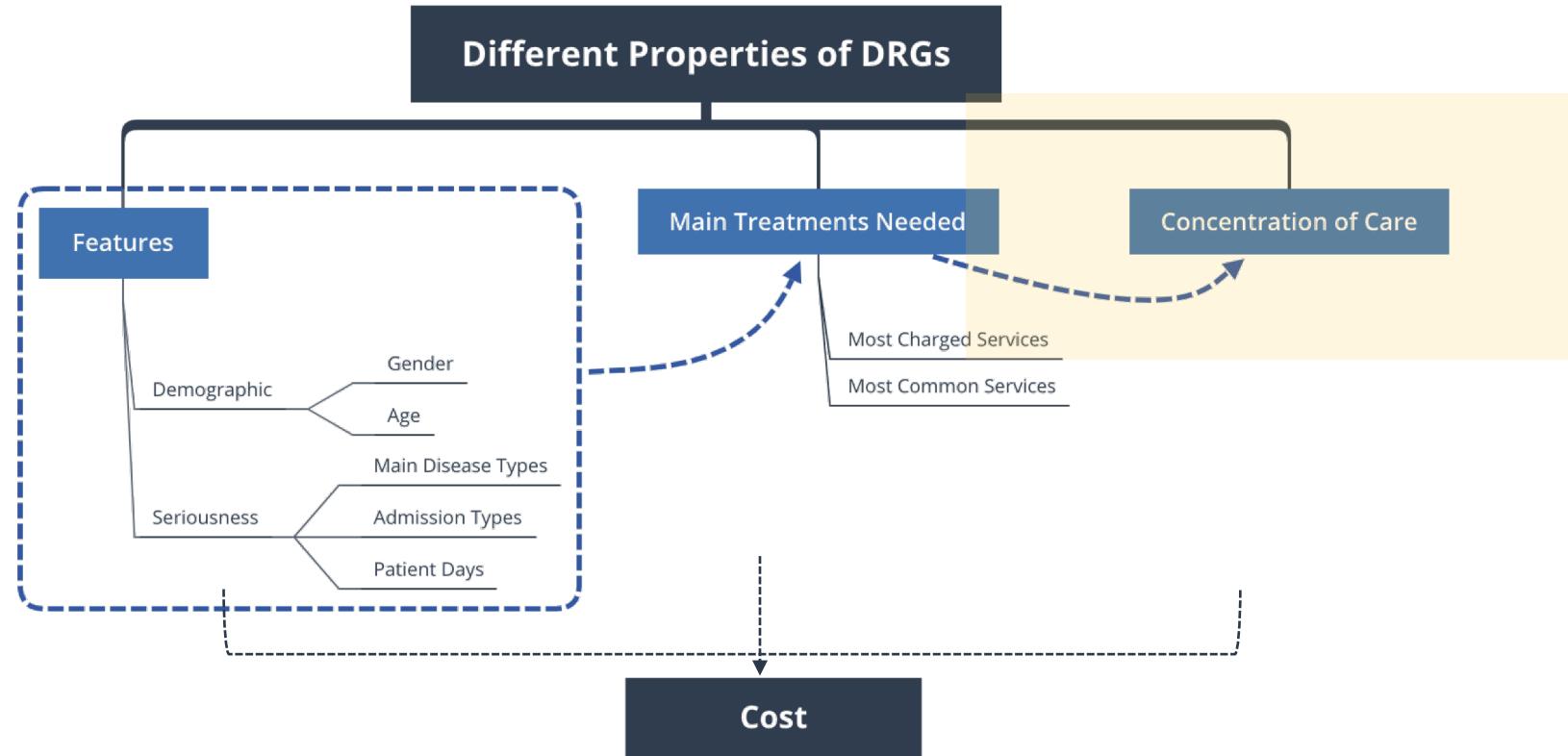
Need the most advanced and complex treatment which may not be done by all hospitals.

High-cost



# Results and Analysis

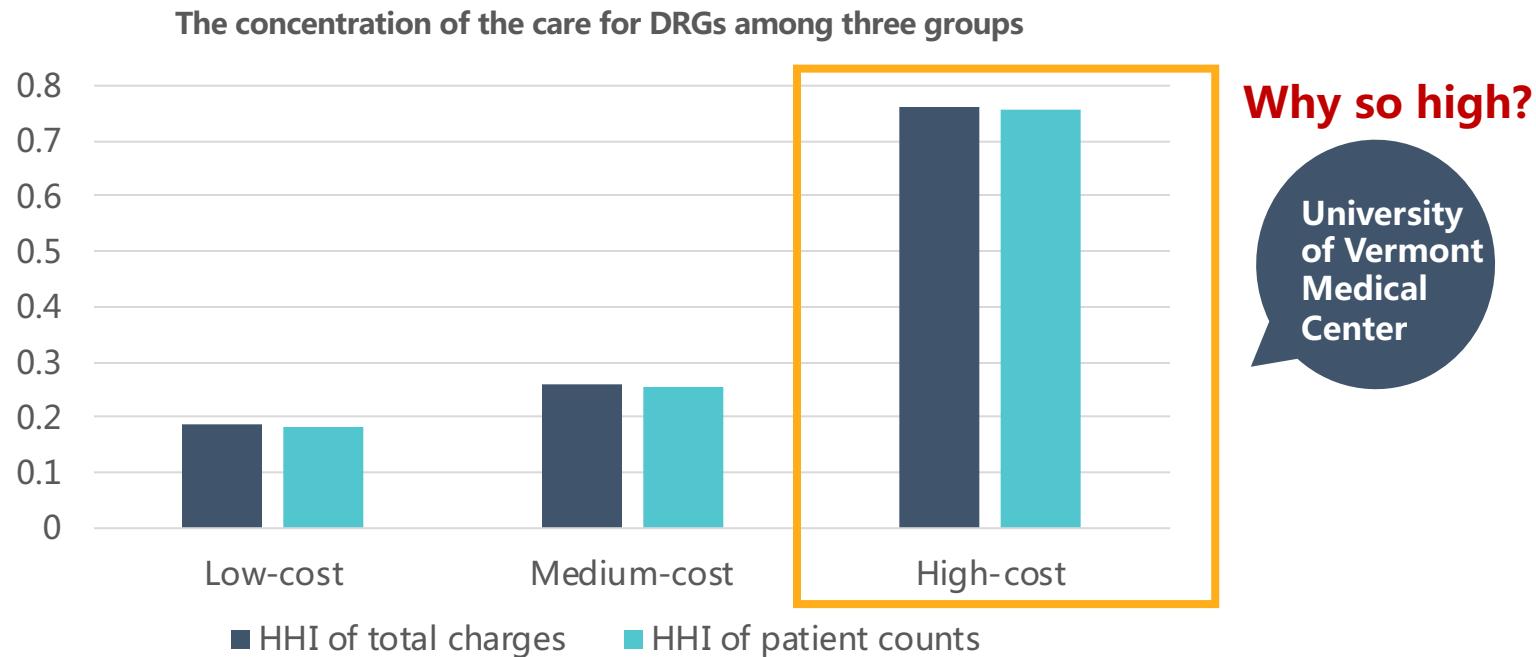
## Analysis logic map



# Results and Analysis

## The concentration of the care for DRGs

The HHI index of market shares of total charges and patient counts of the hospitals.



# Results and Analysis

## University of Vermont Medical Center



- A huge patient number

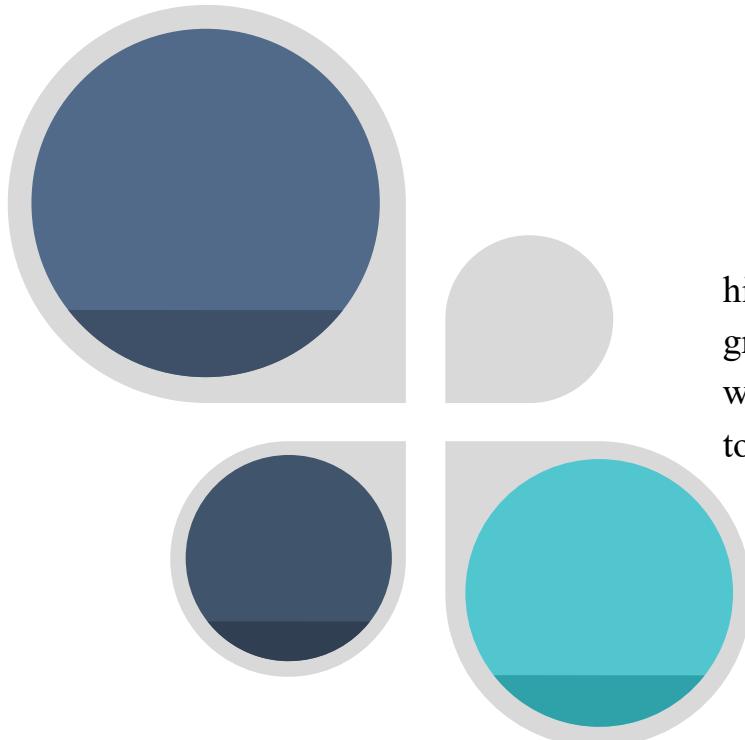
- Regional referral center

- Community hospital

- Advanced technology and techniques

- The medical center has **comprehensive surgical services** and imaging equipment, especially in fields of heart, brain, and tiny blood vessels.

# Results and Analysis



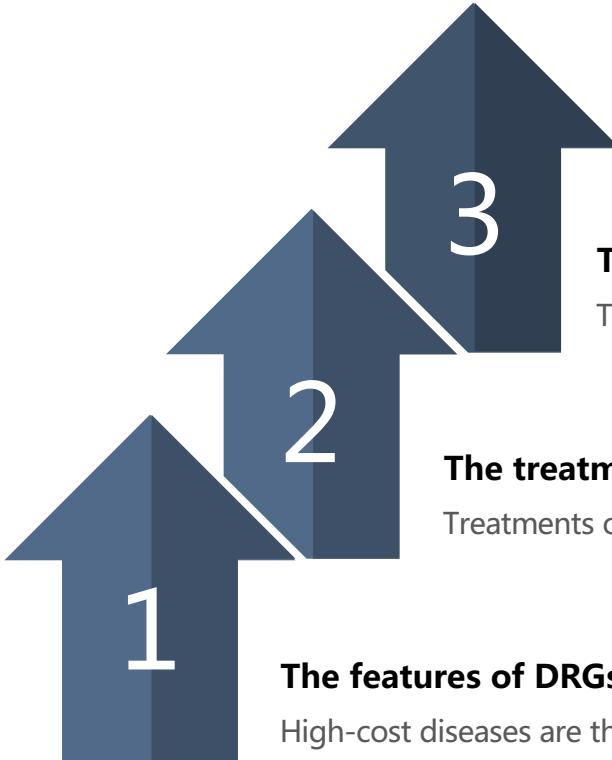
The HHI index also explains the reason why the cost of high-cost group is significantly higher than the low-cost group. When a medical service is in a state of monopoly, whether because of the **technology** or **availability**, according to the principle of market economy, its cost will be higher.



Conclusion

# Conclusion

These features are reflected by the costs.



## **The features of DRGs**

High-cost diseases are the most severe.

## **The treatments needed by DRGs**

Treatments of low-cost diseases are more general, which can be done by most hospitals

## **The concentration of care for DRGs**

The care of high-cost diseases is the most concentrated.

# Conclusion

Some scenarios and prediction

A lady...

A lady gave birth to a baby in the hospital...



A man...

A man broke his bones, had an operation and was hospitalized for a couple days.



An old man...

An old man 's heart function degenerated, affecting his basic life. He had a heart operation in the University of Vermont Medical Center .





Thanks! 2019.12.04