

IS4250 HEALTHCARE ANALYTICS

Project Report

Group 20:

Wang Tianqi (A0105708J)

Cao Shuwei ()

Introduction 0.5

What is community-onset infection? It is defined as an infection which can be diagnosed within the 48 hours of admission.

Cases can be further classified into two categories: healthcare-associated and community acquired. Briefly, the patients of healthcare-associated infection should have recent contact with the healthcare system. For example: patients may have another history of admission in the previous 90 days, or patients who used to live in a nursing home or long-term care facility. While other cases that do not meet the above criteria can be defined as community-acquired infection. To put it in simple words, healthcare-associated means you may get the infection from your previous medication, while community acquired means you get the infection outside the healthcare setting.

Do you know healthcare-associated infection is a significant cause of death in U.S. Here are some data. In 2002, the estimated number of healthcare-associated infections in U.S is approximately 1.7 million, among which there are totally around 99 thousand death cases, and around one third of them are caused by bloodstream infection also named BSI

And one type of bloodstream infection is gram-native bloodstream infection, which caused by gram-negative bacteria. Compared with gram-positive, gram-negative bacteria is resistant to multiple drugs and increasingly resistant to most available antibiotics, so its more difficult to cure.

Many US government agencies like CDC keep monitoring the cases of healthcare-associated gram-negative bloodstream infections and collecting the data on the antibiotic resistance patterns.

Research Objectives 0.5

First, the study hopes to determine the incidence rates of both healthcare-associated and community-acquired gram-negative BSI by age, gender and year, second, to compare their causative pathogens and antimicrobial resistance, and lastly to compare the short and

long-term mortality rates as well as identify risk factors. So that we can optimize the empiric antimicrobial therapy for healthcare-associated gram-negative BSI

Data Collection 0.5

This research was conducted at Olmsted County, Minnesota State of U.S. and 733 unique patients with community-onset Gram-negative BSI were identified among the county residents from January 1998 to December 2007, and 306 are healthcare associated cases and 427 are community-acquired cases

And the study used the laboratory databases at two local medical centers of this county to identify all those patients during the study period, and a local medical records-linkage system named REP was used to exclude those residents who lived outside the county during the period and excluded patients who had recurrent episodes of BSI to ensure that cases were independent.

Statistical Analysis 1

As for the statistical methods used for analysis, descriptive statistics were used to summarize the data, such as medians and interquartile range. Chi-squared test was used to evaluate association between categorical variables and Wilcoxon rank-sum test used to test for differences in medians across continuous variables.

To identify the association between variables and causative pathogens first, we define age, gender, source of infection, site of acquisition as variables, and then an univariate model was build for each variable and if the p value was less than 0.1, then multivariate logistic regression model was used to make adjustment for other variables, to reduce their influence. The same approach was applied to identify association between variables and antimicrobial

resistance. Odds ratios with 95% Confidence interval were calculated to indicate the strength of association.

The Kaplan–Meier method was used to estimate the 28-day and 1-year mortality rates following healthcare-associated and community-acquired Gram-negative BSI. Patients were followed from the date of infection until death or last healthcare encounter. Patients who lost to follow up were censored. Cox Proportional hazard regression was used to identify association between variables and mortality, and hazard ratio with 95% confidence interval were presented to indicate strength of association.

Results & Findings 2

Firstly, from demographic and clinical features, In community onset BSI, 41.7% was healthcare-associated and 58.3% was community-acquired.

Healthcare-associated: median age was 73 and 54.9% were female.

Community-acquired: median age was 66 and 59.3% were female.

Primary source of infection: both were the urinary tract. But we can find that urinary tract was more common cause of CA than HCA, and more HCA are caused by unknown source.

Secondly, from the epidemiology perspective, two types of BSI's incidence rates increase with age in both female and male subjects. When we analyze from the perspective of time, we found that there were no obvious differences of incidence rate between years. We can conclude there was no improvement in the control of infection of community onset BSI.

Thirdly, from microbiology, *Escherichia coli* was the predominant cause of both situations

Compared two types of BSI, *P. aeruginosa* was a more common cause of

healthcare-associated BSI than those with community-acquired BSI (9.2% vs 2.8%).

Similarly, healthcare-associated site of acquisition was also predictive of BSI due to AmpC β -lactamase-producing Enterobacteriaceae, including *Enterobacter*, *Citrobacter*, and *Serratia* species. Using the logistic regression model, it shows clearly that *P.aeruginosa* odds ratio of

HCA is nearly as three times as odds ratio of CA. And Enterbacteriaceae odds ratio of HCA is as two times as odds ratio of CA.

Fourthly, Among two types of BSI, 8.2% and 3.7% respectively were resistant to fluoroquinolones ; 4.4% and 2.7% respectively were resistant to third-generation cephalosporins. Using the logistic regression model, we can found that fluoroquinolone resistance of HCA is as two times as such resistance of CA. And Third-generation cephalosporin resistance of HCA is as 1.5 times as such resistance of CA.

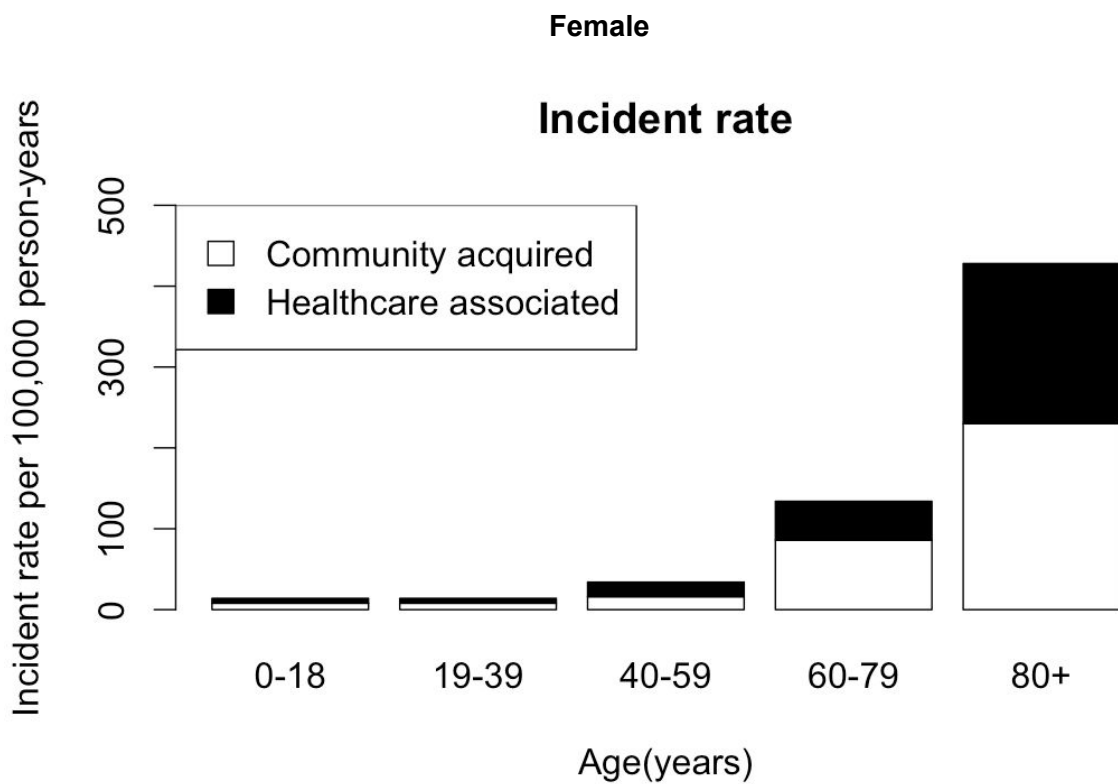
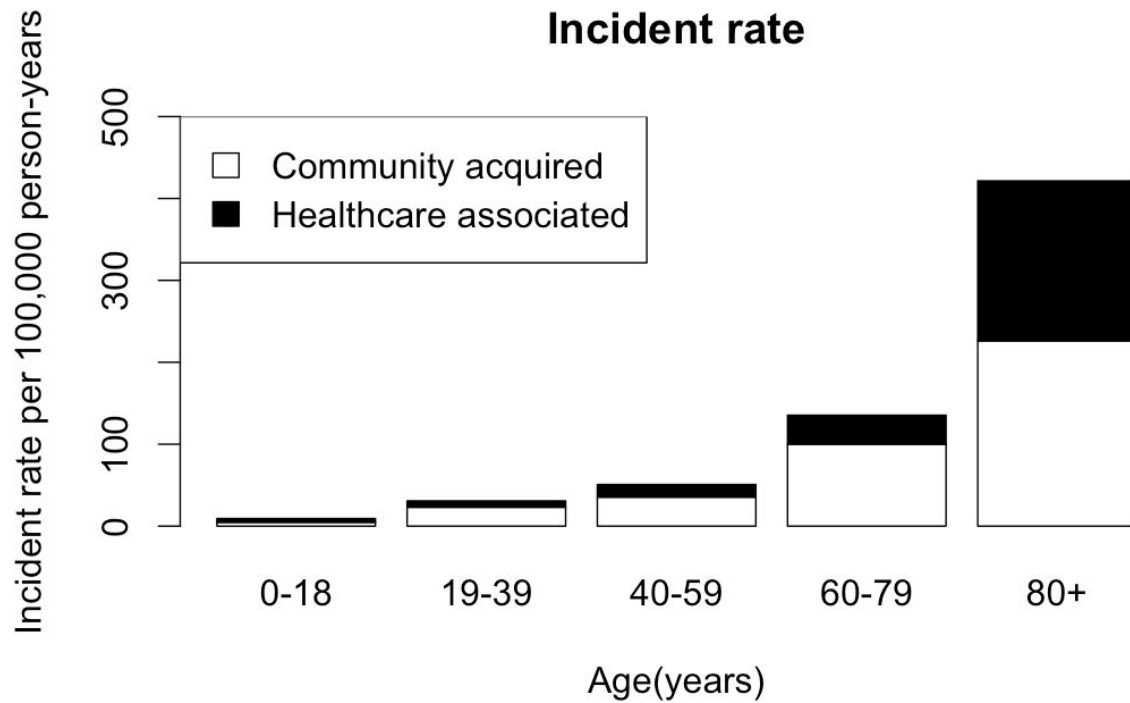
There was a linear increase in fluoroquinolone resistance among healthcare-associated Gram-negative bloodstream isolates between 1998 and 2007 . So we may induce fluoroquinolone is not as efficient as other antibiotics.

Lastly, compared mortality we see that Patients with healthcare-associated Gram-negative BSI had a higher 28-day and 1-year mortality. Linear regression model also shows the same result. So HCA may be more dangerous than CA.

Simulation of Research Experiment (R) 3

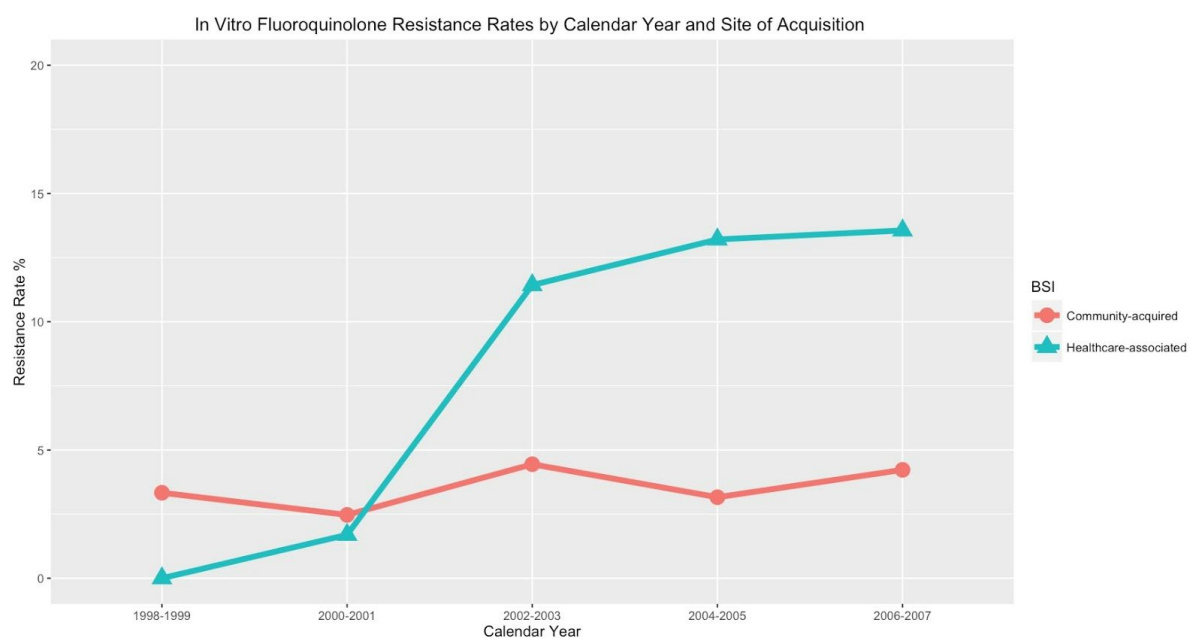
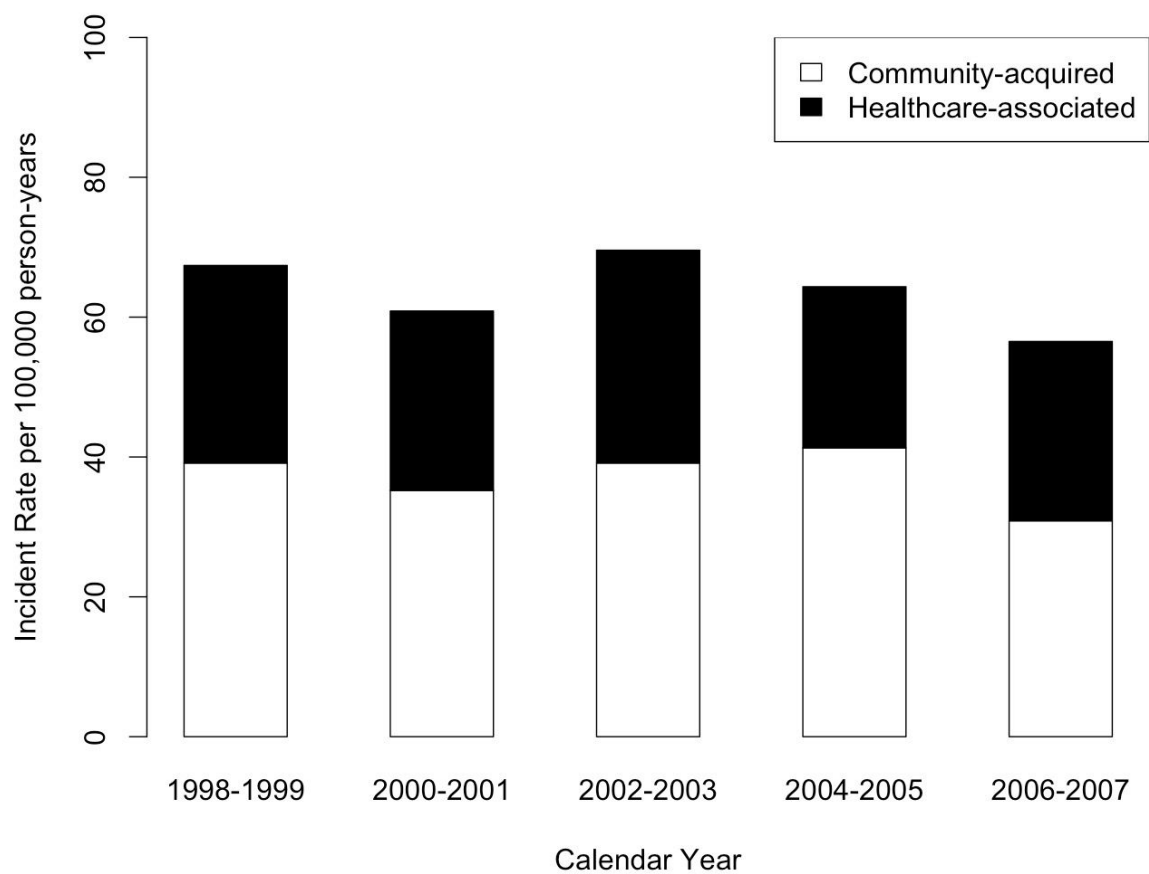
(Replicate one experimental plot with R or Python using data provided in the

experiment or via simulation)



Male

Incident Rate by Calendar Year and Site of Acquisition



Contributions 0.5

(Describe, in the student's own words, the contributions that the paper make to the field of health)

Challenges 0.5

(Discuss issues, challenges and limitations relating the project described in the paper)

Limitations 0.5

Then let us think about the strengths and limitations of this research. As for strengths, our research is population based design which decreased the possibility of referral bias affecting the results. Then we have large sample size and our prolonged follow up is facilitated by REP project. There are also some limitations of this research. Firstly, the population of Olmsted County consists mainly of middle class whites, so we have similar population characteristic. Then the data was from only one geographical area. Lastly, we did not provide data on chronic underlying medical conditions and previous exposure to antimicrobial agents.

Conclusions 0.5

In conclusion, Patients with healthcare-associated Gram-negative BSI have substantially increased short- and long-term mortality rates compared with community-acquired cases. Besides, the causative pathogens for healthcare-associated cases are quite different than that of community-acquired cases. Antimicrobial resistance for healthcare-associated infection is also higher, especially for fluoroquinolone.

Therefore, the study emphasize the importance of identifying healthcare-associated acquisition of gram-negative bsi, so that the empirical therapy, can be optimized, such as include antimicrobial agents with anti-pseudomonal activity, and discourage the use of fluoroquinolone monotherapy for treatment.

References 0.5

Appendix

R code:
1st plot

```

2 |library(dplyr)
3 |library(ggplot2)
4 |dataNew = read.csv("/Users/jim/Downloads/IS4250 Mock Data.csv")
5
6 |BSI = dataNew[,2]
7 |gender = dataNew[,3]
8 |age = dataNew[,4]
9
10 |HA = dataNew[dataNew$BSI=="HA",]
11 |CA = dataNew[dataNew$BSI=="CA",]
12
13 |HA_F = HA[HA$Gender=="F",]
14 |HA_M = HA[HA$Gender=="M",]
15 |CA_F = CA[CA$Gender=="F",]
16 |CA_M = CA[CA$Gender=="M",]
17
18 |HA_F_18 = HA_F[HA_F$Age>=0&HA_F$Age<=18,]
19 |HA_F_39 = HA_F[HA_F$Age>=19&HA_F$Age<=39,]
20 |HA_F_59 = HA_F[HA_F$Age>=40&HA_F$Age<=59,]
21 |HA_F_79 = HA_F[HA_F$Age>=60&HA_F$Age<=79,]
22 |HA_F_80 = HA_F[HA_F$Age>=80,]
23
24 |HA_M_18 = HA_M[HA_M$Age>=0&HA_M$Age<=18,]
25 |HA_M_39 = HA_M[HA_M$Age>=19&HA_M$Age<=39,]
26 |HA_M_59 = HA_M[HA_M$Age>=40&HA_M$Age<=59,]
27 |HA_M_79 = HA_M[HA_M$Age>=60&HA_M$Age<=79,]
28 |HA_M_80 = HA_M[HA_M$Age>=80,]
29
30 |CA_F_18 = CA_F[CA_F$Age>=0&CA_F$Age<=18,]
31 |CA_F_39 = CA_F[CA_F$Age>=19&CA_F$Age<=39,]
32 |CA_F_59 = CA_F[CA_F$Age>=40&CA_F$Age<=59,]
33 |CA_F_79 = CA_F[CA_F$Age>=60&CA_F$Age<=79,]
34 |CA_F_80 = CA_F[CA_F$Age>=80,]
35
36 |CA_M_18 = CA_M[CA_M$Age>=0&CA_M$Age<=18,]
37 |CA_M_39 = CA_M[CA_M$Age>=19&CA_M$Age<=39,]
38 |CA_M_59 = CA_M[CA_M$Age>=40&CA_M$Age<=59,]
39 |CA_M_79 = CA_M[CA_M$Age>=60&CA_M$Age<=79,]
40 |CA_M_80 = CA_M[CA_M$Age>=80,]
41
42 |HA_18_F_Rate = nrow(HA_F_18)/65000*100000
43 |HA_39_F_Rate = nrow(HA_F_39)/65000*100000
44 |HA_59_F_Rate = nrow(HA_F_59)/65000*100000
45 |HA_79_F_Rate = nrow(HA_F_79)/65000*100000
46 |HA_80_F_Rate = nrow(HA_F_80)/65000*100000
47
48 |HA_18_M_Rate = nrow(HA_M_18)/50000*100000
49 |HA_39_M_Rate = nrow(HA_M_39)/50000*100000
50 |HA_59_M_Rate = nrow(HA_M_59)/50000*100000
51 |HA_79_M_Rate = nrow(HA_M_79)/50000*100000
52 |HA_80_M_Rate = nrow(HA_M_80)/50000*100000

```

```

53
54 CA_18_F_Rate = nrow(CA_F_18)/65000*100000
55 CA_39_F_Rate = nrow(CA_F_39)/65000*100000
56 CA_59_F_Rate = nrow(CA_F_59)/65000*100000
57 CA_79_F_Rate = nrow(CA_F_79)/65000*100000
58 CA_80_F_Rate = nrow(CA_F_80)/65000*100000
59
60 CA_18_M_Rate = nrow(CA_M_18)/50000*100000
61 CA_39_M_Rate = nrow(CA_M_39)/50000*100000
62 CA_59_M_Rate = nrow(CA_M_59)/50000*100000
63 CA_79_M_Rate = nrow(CA_M_79)/50000*100000
64 CA_80_M_Rate = nrow(CA_M_80)/50000*100000
65
66 #Female
67 Values = matrix(c(CA_18_F_Rate,CA_39_F_Rate,CA_59_F_Rate,CA_79_F_Rate,
68   CA_80_F_Rate,HA_18_F_Rate,HA_39_F_Rate,HA_59_F_Rate,HA_79_F_Rate,HA_80_F_Rate),
69   nrow=2,ncol=5,byrow=TRUE)
70 colors = c("white","black")
71 ages = c("0-18","19-39","40-59","60-79","80+")
72 regions = c("Community acquired","Healthcare associated")
73
74 bar = barplot(Values,main="Incident rate",names.arg=ages,xlab="Age(years)",
75   ylim=c(0,500),ylab="Incident rate per 100,000 person-years",col=colors)
76
77 # Add the legend to the chart.
78 legend("topleft", regions, cex=1, fill=colors)
79
80 #Male
81 Values2 = matrix(c(CA_18_M_Rate,CA_39_M_Rate,CA_59_M_Rate,CA_79_M_Rate,
82   CA_80_M_Rate,HA_18_M_Rate,HA_39_M_Rate,HA_59_M_Rate,HA_79_M_Rate,HA_80_M_Rate),
83   nrow=2,ncol=5,byrow=TRUE)
84
85 bar2 = barplot(Values2,main="Incident rate",names.arg=ages,xlab="Age(years)",
86   ylim=c(0,500),ylab="Incident rate per 100,000 person-years",col=colors)
87
88 # Add the legend to the chart.
89 legend("topleft", regions, cex=1, fill=colors)
90
91

```

2nd plot

```

94 library(ggplot2)
95 BSI_data=read.csv("/Users/WangTianQi/Downloads/IS4250 Mock Data.csv")
96
97 attach(BSI_data)
98 y9899=0
99 y0001=0
100 y0203=0
101 y0405=0
102 y0607=0
103 for(i in 1:733)
104 {if(BSI_data[i,"Year"] == 1998 | BSI_data[i,"Year"] == 1999) y9899=y9899+1
105 else if(BSI_data[i,"Year"] == 2000 | BSI_data[i,"Year"] == 2001) y0001=y0001+1
106 else if(BSI_data[i,"Year"] == 2002 | BSI_data[i,"Year"] == 2003) y0203=y0203+1
107 else if(BSI_data[i,"Year"] == 2004 | BSI_data[i,"Year"] == 2005) y0405=y0405+1
108 else y0607=y0607+1
109 }
110 y9899HA=0
111 y0001HA=0
112 y0203HA=0
113 y0405HA=0
114 y0607HA=0
115 for(i in 1:733)
116 {if((BSI_data[i,"Year"] == 1998 | BSI_data[i,"Year"] == 1999) & BSI_data[i,"BSI"] == "HA")
117     y9899HA=y9899HA+1
118 else if((BSI_data[i,"Year"] == 2000 | BSI_data[i,"Year"] == 2001) & BSI_data[i,"BSI"] == "HA")
119     y0001HA=y0001HA+1
120 else if((BSI_data[i,"Year"] == 2002 | BSI_data[i,"Year"] == 2003) & BSI_data[i,"BSI"] == "HA")
121     y0203HA=y0203HA+1
122 else if((BSI_data[i,"Year"] == 2004 | BSI_data[i,"Year"] == 2005) & BSI_data[i,"BSI"] == "HA")
123     y0405HA=y0405HA+1
124 else if((BSI_data[i,"Year"] == 2006 | BSI_data[i,"Year"] == 2007) & BSI_data[i,"BSI"] == "HA")
125     y0607HA=y0607HA+1
126 }
127
128 ratio=1/(2*115000)*100000
129 year_all=numeric(10)
130 dim(year_all)=c(2,5)
131 year_all[1,]=c(y9899-y9899HA, y0001-y0001HA, y0203-y0203HA, y0405-y0405HA, y0607-y0607HA)*ratio
132 year_all[2,]=c(y9899HA, y0001HA, y0203HA, y0405HA, y0607HA)*ratio
133 rowNames=c("CA", "HA")
134 row.names(year_all)=rowNames
135 colNames=c("1998-1999", "2000-2001", "2002-2003", "2004-2005", "2006-2007")
136 colnames(year_all)=colNames
137 bar=barplot(year_all, main="Incident Rate by Calendar Year and Site of Acquisition",
138     names.arg = colNames, xlab="Calendar Year", ylab="Incident Rate per 100,000 person-years",
139     ylim=c(0,100), space=1, col=c("white","black"))
140 legend("topright",c("Community-acquired","Healthcare-associated"), cex=1, fill=c("white","black"))
141

```

3rd plot


```

142 resist9899HA=0
143 resist9899CA=0
144 resist0001HA=0
145 resist0001CA=0
146 resist0203HA=0
147 resist0203CA=0
148 resist0405HA=0
149 resist0405CA=0
150 resist0607HA=0
151 resist0607CA=0
152 for(i in 1:733)
153 {
154   if((BSI_data[i,"Year"] == 1998 | BSI_data[i,"Year"] == 1999) & BSI_data[i,"BSI"] == "HA"
155       & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist9899HA=resist9899HA+1
156   if((BSI_data[i,"Year"] == 1998 | BSI_data[i,"Year"] == 1999) & BSI_data[i,"BSI"] == "CA"
157       & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist9899CA=resist9899CA+1
158   if((BSI_data[i,"Year"] == 2000 | BSI_data[i,"Year"] == 2001) & BSI_data[i,"BSI"] == "HA"
159       & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0001HA=resist0001HA+1
160   if((BSI_data[i,"Year"] == 2000 | BSI_data[i,"Year"] == 2001) & BSI_data[i,"BSI"] == "CA"
161       & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0001CA=resist0001CA+1
162   if((BSI_data[i,"Year"] == 2002 | BSI_data[i,"Year"] == 2003) & BSI_data[i,"BSI"] == "HA"
163       & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0203HA=resist0203HA+1
164   if((BSI_data[i,"Year"] == 2002 | BSI_data[i,"Year"] == 2003) & BSI_data[i,"BSI"] == "CA"
165       & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0203CA=resist0203CA+1
166   if((BSI_data[i,"Year"] == 2004 | BSI_data[i,"Year"] == 2005) & BSI_data[i,"BSI"] == "HA"
167       & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0405HA=resist0405HA+1
168   if((BSI_data[i,"Year"] == 2004 | BSI_data[i,"Year"] == 2005) & BSI_data[i,"BSI"] == "CA"
169       & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0405CA=resist0405CA+1
170   if((BSI_data[i,"Year"] == 2006 | BSI_data[i,"Year"] == 2007) & BSI_data[i,"BSI"] == "HA"
171       & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0607HA=resist0607HA+1
172   if((BSI_data[i,"Year"] == 2006 | BSI_data[i,"Year"] == 2007) & BSI_data[i,"BSI"] == "CA"
173       & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0607CA=resist0607CA+1
174 }
175 resistance_all=numeric(10)
176 dim(resistance_all)=c(2,5)
177 resistance_all[1,]=c(resist9899CA/(y9899-y9899HA), resist0001CA/(y0001-y0001HA),
178                     resist0203CA/(y0203-y0203HA), resist0405CA/(y0405-y0405HA), resist0607CA/(y0607-y0607HA))*100
179 resistance_all[2,]=c(resist9899HA/y9899HA, resist0001HA/y0001HA, resist0203HA/y0203HA,
180                     resist0405HA/y0405HA, resist0607HA/y0607HA)*100
181 row.names(resistance_all)=rowNames
182 colnames(resistance_all)=colNames
183
184 type1=c("CA","CA","CA","CA","CA","HA","HA","HA","HA","HA")
185 year1=c("1998-1999", "2000-2001", "2002-2003", "2004-2005", "2006-2007",
186         "1998-1999", "2000-2001", "2002-2003", "2004-2005", "2006-2007")
187 value1=c(c(resist9899CA/(y9899-y9899HA), resist0001CA/(y0001-y0001HA),
188            resist0203CA/(y0203-y0203HA), resist0405CA/(y0405-y0405HA), resist0607CA/(y0607-y0607HA))*100,
189          c(resist9899HA/y9899HA, resist0001HA/y0001HA, resist0203HA/y0203HA, resist0405HA/y0405HA, resist0607HA/y0607HA)*100)
190 df_test=data.frame(type1,year1,value1)
191 ggplot(df_test, aes(x=year1, y=value1, colour=type1, group=type1))
192   +geom_line(size=2)+ggtitle("In Vitro Fluoroquinolone Resistance Rates by Calendar Year and Site of Acquisition ")
193   +xlab("Calendar Year")+ylab("Resistance Rate %")+ylim(0,20)+geom_point(aes(shape=type1),size=5)
194   +scale_color_discrete(name="BSI", labels=c("Community-acquired","Healthcare-associated"))
195   +scale_shape_discrete(name="BSI", labels=c("Community-acquired","Healthcare-associated"))
196

```