IS4250 HEALTHCARE ANALYTICS

Project Report

Discussion on "Impact of healthcare-associated acquisition on community-onset Gram-negative bloodstream infection: a population-based study" by M. N. Al-Hasan, J. E. Eckel-Passow and L. M. Baddour

Group 20:

Wang Tianqi (A0105708J)

Cao Shuwei (A0105728E)

1. Introduction

The researchers of paper "Impact of healthcare-associated acquisition on community-onset Gram-negative bloodstream infection: a population-based study" are a group of people from University of Kentucky Medical Centre and Mayo Clinic, Rochester, MN, USA. They performed an experiment to examine the influence of healthcare-associated acquisition on pathogen distribution, antimicrobial resistance, short-term and long-term mortality of community-onset Gram-negative ***bloodstream infections (BSI)***. Our group would like to study on how these researchers conducted the experiment and what statistical methods they applied to analyze the data and generate the results. In addition, our group would discuss the challenges, limitations and contributions of this research paper from our own perspective. Last but not least, we would try to replicate a few experimental plots in the research paper with R via simulation.

Before going to the next section, we would like to introduce some background information about the medical fields that the research paper studied on.
Community-onset infection is defined as an infection diagnosed within the first 48 h of hospitalization (Kang, C. et al., 2012). Episodes of community-onset infection can be further classified into ***healthcare-associated infection (HCA)*** and ***community-acquired infection (CA)***. Patients of healthcare-associated (HCA) infection should meet the following criteria: a history of a >48-h hospital admission in the previous 90 days; hemodialysis, intravenous medication, home wound care in the previous 30 days; or residence in a nursing home or long-term care facility (Kang, C. et al., 2012). Cases that do not meet these criteria are considered to be community-acquired (CA).

In 2002, the estimated number of HCAs in U.S. hospitals was approximately 1.7 million, and the estimated deaths associated with HCAs in U.S. hospitals were 98,987, of which 30,665 were for bloodstream infections (BSI) (Klevens, R. M. et al., 2007). One type of BSI is ***Gram-negative BSI***, which is caused by Gram-negative bacteria. Compared with Gram-positive bacteria, Gram-negative bacteria is resistant to multiple drugs and increasingly resistant to most available antibiotics ("Gram-positive vs Gram-negative Bacteria", n.d.). U.S. government agencies including CDC's National Healthcare Safety Network (NHSN) are

capturing information on antibiotic resistance patterns in Gram-negative bacteria in healthcare settings and they have found that the percentage of Gram-negatives that are resistant to drugs is increasing ("Gram-negative Bacteria Infections in Healthcare Settings", 2011).

2. Research Objectives

The researchers of the paper identified a large population-based cohort of patients with community-onset Gram-negative BSI in Olmsted County, Minnesota, and their research objectives are as follows:

1. To determine the incidence rates of HCA and CA Gram-negative BSI by age, sex, and calendar year,
2. examine the influence of HCA acquisition on pathogen distribution, fluoroquinolone and third-generation cephalosporin resistance among bloodstream isolates, and compare it with that of CA Gram-negative BSI,
3. and compare the 28-day and 1-year all-cause mortality rates in patients with HCA and CA Gram-negative BSI and identify risk factors.

Based on the differences in results, the researchers aimed to provide some recommendations for empiric antimicrobial therapy in patients with healthcare-associated Gram-negative BSI. Our group would like to emphasize that this research's focus was not looking for the root causes of healthcare-associated Gram-negative BSI.

3. Data Collection

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 Gram-negative BSI cases and 427 are community-acquired cases. And the study used the microbiology laboratory databases at two local medical institutions, Medical Center in Rochester, Minnesota, and the Olmsted Medical Center to identify all those patients during the study period. Besides, the Rochester Epidemiology Project (REP), an unique medical records-linkage system that encompasses care delivered to residents of Olmsted County, was also used to exclude those residents who lived outside the county during the period and exclude patients who had recurrent episodes of BSI to ensure that cases were independent.

4. Statistical Analysis

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

To identify the relationship between variables and causative pathogens, first, age, gender, source of infection, site of acquisition, etc. were defined as variables, and then an univariate model would be built for each variable, and if the p-value was less than 0.1, multivariate logistic regression model would be built as well to make adjustment for other variables and reduce their influence. The p-value of the second model would be compared with 0.05 to determine if there was significant association. The same approach was applied to identify relationship between variables and antimicrobial resistance. Odds ratios with 95% Confidence Interval (CI) 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 HCA and CA 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% CI were presented to indicate strength of association.

5. Results & Findings

Based on the statistical analysis of the collected data, the researchers compared the differences in results between HCA and CA Gram-negative BSI. The results and findings will be discussed from five perspectives.

5.1 Demographic and Clinical Features

Variable	Healthcare-associated *N*=306	Community-acquired *N*=427	*p* value*
Age: median (IQR)	73 (51–83)	66 (45–79)	<0.001
Female sex, *n* (%)	168 (54.9)	253 (59.3)	0.24
Primary source, *n* (%)			<0.001
Urinary tract	166 (54.2)	281 (65.8)	
Gastrointestinal tract	29 (9.5)	61 (14.3)	
Respiratory tract	26 (8.5)	22 (5.2)	
Skin and soft tissue	10 (3.3)	9 (2.1)	
Central venous catheter-related	10 (3.3)	0 (0)	
Other	3 (1.0)	4 (0.9)	
Unknown	62 (20.3)	50 (11.7)	

Table 5.1.1 Clinical characteristics of patients with Gram-negative BSI based on site of acquisition

Firstly, from the perspective of demographic and clinical features, among those community-onset Gram-negative BSIs, healthcare-associated Gram-negative BSIs account for 41.7%, while community-acquired Gram-negative BSIs account for 58.3%. In terms of age and gender, the median age of HCA patients was 73, and 54.9% of them were female subjects. In contrast, the median age of CA patients was 66 and 59.3% were female subjects. Regarding to the primary source of infection for these two types of Gram-negative BSI, Table 5.1.1 shows that urinary tract was the common primary source of infection, but urinary tract was more common for CA cases than HCA cases. Moreover, the source of infection for a large percentage of HCA Gram-negative BSI cases still remained unknown, which should drew the attention and concern from medical fields.

5.2 Epidemiology

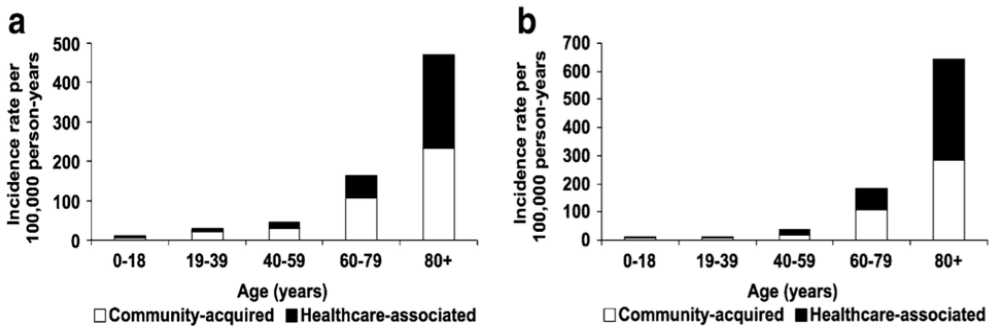

Figure 5.2.1 Incidence rate by age and site of acquisition in female (a) and male (b) subjects

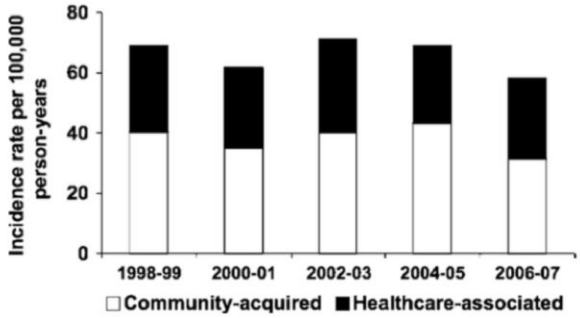

Figure 5.2.2 Age and sex adjusted incidence rate by calendar year and site of acquisition

Secondly, we are going to discuss results from the epidemiology perspective. As Figure 5.2.1 shows, it is clear that the incidence rates for these two types of Gram-negative BSI was increasing with age in both female and male subjects, which means that senior people were the common patients who might be infected easily. In terms of time trend, as shown in Figure 5.2.2, there were no significant differences of incidence rates among years, which indicates that there was not much improvement in the control of community-onset Gram-negative bloodstream infection.

5.3 Microbiology

Variable	Healthcare-associated *N*=306	Community-acquired *N*=427	*p* value*
Microbiology, *n* (%)			<0.001
Escherichia coli	145 (47.4)	271 (63.5)	
Klebsiella pneumoniae	45 (14.7)	48 (11.2)	
Pseudomonas aeruginosa	28 (9.2)	12 (2.8)	
Proteus mirabilis	13 (4.2)	11 (2.6)	
Enterobacter species	20 (6.5)	8 (1.9)	
Citrobacter species	6 (2.0)	4 (0.9)	
Serratia species	5 (1.6)	6 (1.4)	
Other	44 (14.4)	67 (15.7)	

Table 5.3.1 Microbiology results based on site of acquisition

Variable	Univariate model		Multivariate model	
	OR (95% CI)	*p* value	OR (95% CI)	*p* value
Pseudomonas aeruginosa				
Site of acquisition: HCA vs CA	3.48 (1.78–7.22)	<0.001	3.14 (1.59–6.57)	0.001
AmpC-producing *Enterobacteriaceae*				
Site of acquisition: HCA vs CA	2.56 (1.42–4.75)	0.002	2.23 (1.21–4.21)	0.01
Fluoroquinolone resistance				
Site of acquisition: HCA vs CA	2.33 (1.21–4.61)	0.01	2.19 (1.13–4.38)	0.02
Third-generation cephalosporin resistance				
Site of acquisition: HCA vs CA	1.66 (0.73–3.84)	0.22	–	–

Table 5.3.2 Predictors of the causative pathogen and antimicrobial resistance

Thirdly, from microbiology perspective, the result in Table 5.3.1 shows that *Escherichia coli* was the predominant cause of both HCA and CA Gram-negative BSI. In addition, comparing these two types of Gram-negative BSI, *P. aeruginosa* was a more common cause of HCA Gram-negative BSI than those with CA (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. With the logistic regression model, it shows clearly in Table 5.3.2 that *P.aeruginosa* odds ratio of HCA is nearly as three times as odds ratio of CA, and *Enterobacteriaceae* odds ratio of HCA is as two times as odds ratio of CA.

5.4 Antimicrobial Resistance

Variable	Healthcare-associated *N*=306	Community-acquired *N*=427	*p* value*
Fluoroquinolone resistance[a]	24/293 (8.2)	15/406 (3.7)	0.01
Third-generation cephalosporin resistance[a]	13/293 (4.4)	11/405 (2.7)	0.22

Table 5.4.1 Antimicrobial resistance results based on site of acquisition

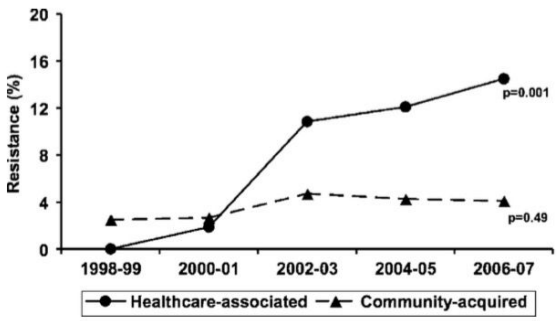

Figure 5.4.2 Fluoroquinolone resistance rate by calendar year and site of acquisition

Fourthly, in terms of antimicrobial resistance, Table 5.4.1 reflects that among two types of Gram-negative BSI, 8.2% and 3.7% respectively were resistant to fluoroquinolones, and 4.4% and 2.7% respectively were resistant to third-generation cephalosporins. Using the logistic regression model, Table 5.3.2 shows that fluoroquinolone resistance of HCA is as two times as such resistance of CA. 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

(Figure 5.4.2). Therefore, we induce that fluoroquinolone is not as efficient as other antibiotics.

5.5 Mortality

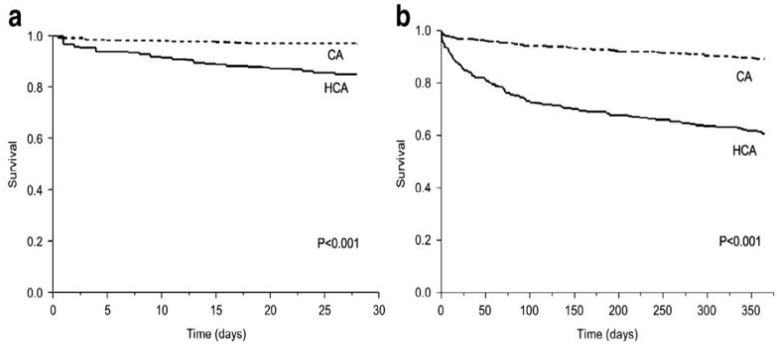

Figure 5.5.1 Kaplan-Meier 28-day (a) and 1-year (b) survival curves by site of acquisition

Variable	Univariate model		Multivariate model	
	HR (95% CI)	*p* value	HR (95% CI)	*p* value
28-day mortality				
Age (per year)	1.02 (1.01–1.03)	0.003	1.02 (1.01–1.03)	0.003
Sex: male vs female	1.35 (0.82–2.23)	0.23	–	–
Year of diagnosis (per year)	0.94 (0.86–1.03)	0.18	–	–
Source of infection: urinary tract vs other	0.27 (0.15–0.45)	<0.001	0.27 (0.16–0.47)	<0.001
Site of acquisition: HCA vs CA	4.61 (2.64–8.53)	<0.001	3.73 (2.13–6.93)	<0.001
1–year mortality				
Age (per year)	1.03 (1.02–1.04)	<0.001	1.03 (1.02–1.03)	<0.001
Sex: male vs female	1.45 (1.06–1.97)	0.02	1.23 (0.90–1.68)	0.19
Year of diagnosis (per year)	0.95 (0.90–1.00)	0.06	0.93 (0.88–0.99)	0.02
Source of infection: urinary tract vs other	0.34 (0.24–0.46)	<0.001	0.34 (0.24–0.47)	<0.001
Site of acquisition: HCA vs CA	4.37 (3.12–6.22)	<0.001	3.67 (2.61–5.24)	<0.001

Table 5.5.2 Factors associated with 28-day and 1-year all-cause mortality

Lastly, comparing mortality in Figure 5.5.1, we can see that patients with healthcare-associated Gram-negative BSI had lower survival rates on both 28-day and 1-year survival curves. In addition, in the Table 5.5.2 linear regression model also shows the same result (i.e., the Hazard Ratios for 28-day mortality is 3.73 and for 1-year mortality is 3.67, meaning HCA Gram-negative BSI cases have more than 3 times of mortality as that of CA cases). To conclude, HCA seems to be much more dangerous than CA, which requires the extensive attention from field of health.

6. Simulation of Research Experiment

Since full dataset is not provided in the research paper, our group decided to try to replicate a few experimental plots in the paper with R via simulation. First, we prepared a mock up dataset based on the results from figures and tables in the paper. For example, from Table 5.1.1, we know that there are totally 306 HCA Gram-negative BSI patients and 168 of them are female, and there are 427 CA patients totally and 253 of them are female. Since this research was a population-based study, so we looked up the population data of Olmsted County from the website of United States Census Bureau Besides, where only data after 2010 presents. To close to the experiment, we chose the population of 2010, 144,248, as baseline, and after eliminating the effect of population growth as well as deducting a small portion of residents who lived outside the county during study period, our group decided to use 115,000 as the final number. From 2010 census, female persons accounted for 51.1% of total population, so we chose 65,000 as the total number of female residents and 50,000 as the total number of male residents for our simulation.

Based on Figure 5.2.1, we calculated the percentage of 5 age groups among the chosen 733 patients, which was used to calculate the exact number of each age group in our simulation. Figure 5.2.2 was used as reference to decide the number of patients based on calendar year. Based on results in Table 5.4.1 and Figure 5.4.2, we decided the number of cases which are resistant to Fluoroquinolone based on calendar year and site of acquisition. Some of the mock up data is shown in the Table 6.1 below. The full dataset can be found in our group's GitHub repository.

	A	B	C	D	E	F	G	H
1	AGE	0-19	20-39	40-59	60-79	80+		Total
2	HA (F)	3	5	10	23	127	168	
3	HA (M)	3	3	8	24	100	138	306
4	CA (F)	3	15	23	65	147	253	
5	CA (M)	4	4	8	43	115	174	427
6								

	A	B	C	D	E	F	G
1	Year of Infection	98-99	00-01	02-03	04-05	06-07	Total
2	HA	65	59	70	53	59	306
3	CA	90	81	90	95	71	427
4							733
5	Fluoroquinolone Resistance						
6	HA	0	1	8	7	8	24
7	CA	3	2	4	3	3	15
8							

Table 6.1 Partial mock up data

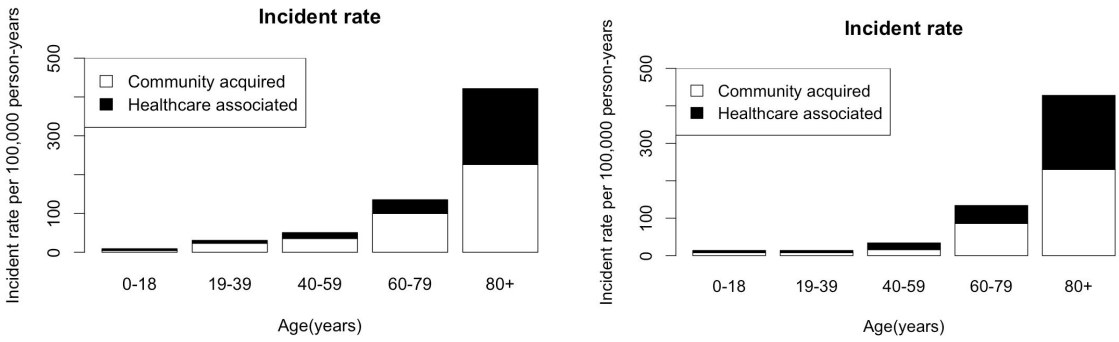

Figure 6.2 Incident rate for female (left) and male (right) subjects by age group and site of acquisition

Figure 6.2 shows the incident rate for female and male subjects by different age groups and types of Gram-negative BSI, which aims to replicate the results in Figure 5.2.1.
Figure 6.3 below shows the incident rate by 5 time periods and site of acquisition, which is a replica of Figure 5.2.2.

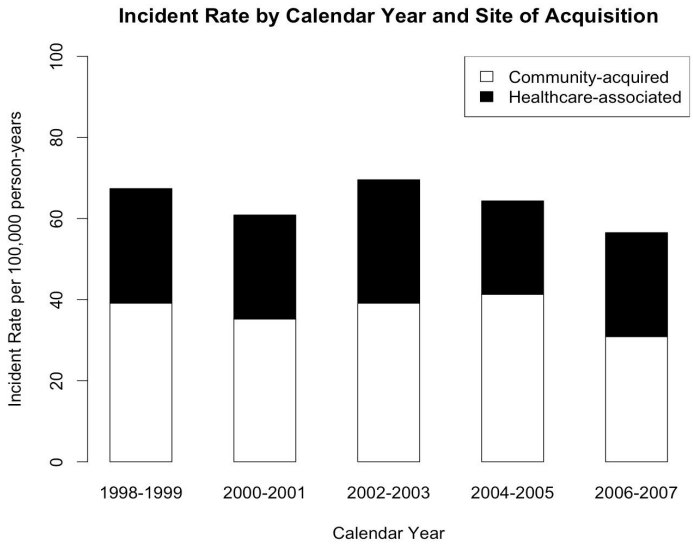

Figure 6.3 Incident rate by calendar year and site of acquisition

Figure 6.4 shows the linear increase in fluoroquinolone resistance among healthcare-associated Gram-negative bloodstream isolates between 1998 and 2007, which is the same result as Figure 5.4.2.

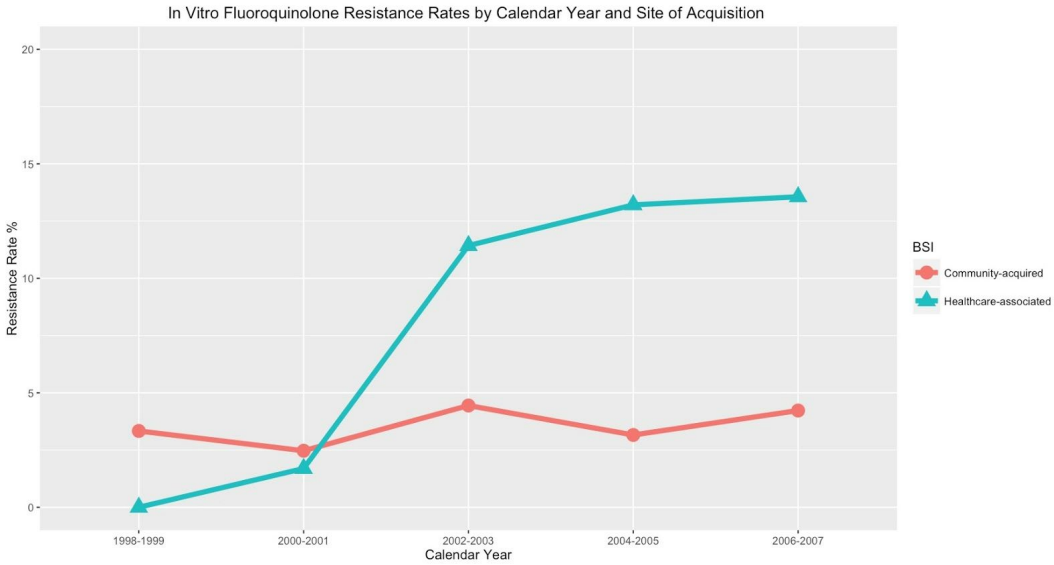

Figure 6.4 Fluoroquinolone resistance rate by calendar year and site of acquisition

One thing we have to mention here is that due to lack of information in the paper, there do exist minor differences between our simulation results and the research results. In addition, our group was not able to perform statistical tests such as linear regression and correlation tests using our mock up data, because of subjectiveness of our data selection. Even though it is possible to perform the tests, the potential results could be extremely meaningless.

7. Contributions

The key findings from the research paper made much contribution to the study of community-onset Gram-negative bloodstream infection.

Firstly, the primary source of infection was identified as urinary tract for both HCA and CA Gram-negative BSI, which can remind patients and medical staff to be careful with infection when performing urinary surgery. Next, the predominant causes and etiology of Gram-negative BSI were revealed in the research, which is helpful information for developing effective drugs. Besides, the statistical analysis also proved that HCA acquisition is increasingly resistant to Fluoroquinolone. This will discourage doctors using fluoroquinolone monotherapy for treatment. In addition, the differences of influence on pathogen distribution between HCA and CA Gram-negative BSI could be used as helpful references for doctors to suit the remedy to the case.

In conclusion, the research provides much information about how to act appropriately to treatment of community-onset Gram-negative BSIs, especially healthcare-associated acquisition cases.

8. Challenges & Limitations

The primary challenge of the research is collecting the population data from Olmsted county during ten years. It is hard to follow up a patient for ten years and it is easy to lose track during the period. The valid definition as well as the reliable and valid identification of populations that can be studied is the most important issue in population-based studies. Besides, this study would require large amount of reference data from the Census Bureau, such as total population, female percentage, age group percentage, etc. for each year during the 10-year study period. If the data for some year was mistaken or inaccurate, it would significantly impact the the credibility of results. As mentioned early, REP, the medical system was used to filter out residents who lived outside the county, so the problem would be how to decide which residents to be removed from dataset, and what should be the criteria. Similarly, whether two local medical centres could cover all medical records of residents, i.e., the representativeness of selection remains controversial.

As for limitations of this research the population of Olmsted County consists mainly of middle class whites, so the study results could be only generalized to communities with similar population characteristics and could not accurately reflect the situation across U.S. Then the data was from only one geographical area, so the results may not provide generalizable applications to another area with different features. Last but not least, the researchers did not provide data on patients' chronic underlying medical conditions and previous exposure to antimicrobial agents, which could lead to inaccuracy of findings.

9. Conclusions

In conclusion, Patients with healthcare-associated Gram-negative BSI have significantly increased short-term 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, and there is a large portion of unknown causes. Antimicrobial resistance for healthcare-associated infection is much higher and the situation is becoming worse, especially for fluoroquinolone.

Therefore, the study emphasizes the importance of identifying healthcare-associated acquisition among those with community-onset Gram-negative BSI, so that empirical antimicrobial therapy could be optimized. Recommendations include using antimicrobial agents with anti-pseudomonal activity and without the ability to induce AmpC β-lactamase Production causing exacerbation, and discouraging the use of fluoroquinolone monotherapy for treatment of healthcare-associated Gram-negative BSI due to its strong resistance.

10. References

1. *Al-Hasan, M. N., Eckel-Passow, J. E., & Baddour, L. M. (2012). Impact of healthcare-associated acquisition on community-onset gram-negative bloodstream infection: A population-based study. European Journal of Clinical Microbiology and Infectious Diseases, 31(6), 1163*
2. *Gram-negative Bacteria Infections in Healthcare Settings. (2011). Retrieved April 03, 2016, from http://www.cdc.gov/hai/organisms/gram-negative-bacteria.html*
3. *Gram-positive vs Gram-negative Bacteria. (n.d.). Retrieved April 03, 2016, from http://www.diffen.com/difference/Gram-negative_Bacteria_vs_Gram-positive_Bacteria*
4. *Kang, C., Wi, Y. M., Lee, M. Y., Ko, K. S., Chung, D. R., Peck, K. R.. . Song, J. (2012). Epidemiology and risk factors of community onset infections caused by extended-spectrum β-lactamase-producing escherichia coli strains. Journal of Clinical Microbiology, 50(2), 312-317*
5. *Klevens, R. M., Edwards, J. R., Richards, C. L., Horan, T. C., Gaynes, R. P., Pollock, D. A., & Cardo, D. M. (2007). Estimating Health Care-Associated Infections and Deaths in U.S. Hospitals, 2002. Public Health Reports, 122(2), 160–166.*
6. *US Census Bureau. Olmsted County QuickFacts, http://quickfacts.census.gov. Accessed 1 April 2016*

Appendix

A1. R code used to replicate experimental plots

```r
library(dplyr)
library(ggplot2)
dataNew = read.csv("/Users/jim/Downloads/IS4250 Mock Data.csv")

BSI = dataNew[,2]
gender = dataNew[,3]
age = dataNew[,4]

HA = dataNew[dataNew$BSI=="HA",]
CA = dataNew[dataNew$BSI=="CA",]

HA_F = HA[HA$Gender=="F",]
HA_M = HA[HA$Gender=="M",]
CA_F = CA[CA$Gender=="F",]
CA_M = CA[CA$Gender=="M",]

HA_F_18 = HA_F[HA_F$Age>=0&HA_F$Age<=18,]
HA_F_39 = HA_F[HA_F$Age>=19&HA_F$Age<=39,]
HA_F_59 = HA_F[HA_F$Age>=40&HA_F$Age<=59,]
HA_F_79 = HA_F[HA_F$Age>=60&HA_F$Age<=79,]
HA_F_80 = HA_F[HA_F$Age>=80,]

HA_M_18 = HA_M[HA_M$Age>=0&HA_M$Age<=18,]
HA_M_39 = HA_M[HA_M$Age>=19&HA_M$Age<=39,]
HA_M_59 = HA_M[HA_M$Age>=40&HA_M$Age<=59,]
HA_M_79 = HA_M[HA_M$Age>=60&HA_M$Age<=79,]
HA_M_80 = HA_M[HA_M$Age>=80,]

CA_F_18 = CA_F[CA_F$Age>=0&CA_F$Age<=18,]
CA_F_39 = CA_F[CA_F$Age>=19&CA_F$Age<=39,]
CA_F_59 = CA_F[CA_F$Age>=40&CA_F$Age<=59,]
CA_F_79 = CA_F[CA_F$Age>=60&CA_F$Age<=79,]
CA_F_80 = CA_F[CA_F$Age>=80,]

CA_M_18 = CA_M[CA_M$Age>=0&CA_M$Age<=18,]
CA_M_39 = CA_M[CA_M$Age>=19&CA_M$Age<=39,]
CA_M_59 = CA_M[CA_M$Age>=40&CA_M$Age<=59,]
CA_M_79 = CA_M[CA_M$Age>=60&CA_M$Age<=79,]
CA_M_80 = CA_M[CA_M$Age>=80,]

HA_18_F_Rate = nrow(HA_F_18)/65000*100000
HA_39_F_Rate = nrow(HA_F_39)/65000*100000
HA_59_F_Rate = nrow(HA_F_59)/65000*100000
HA_79_F_Rate = nrow(HA_F_79)/65000*100000
HA_80_F_Rate = nrow(HA_F_80)/65000*100000

HA_18_M_Rate = nrow(HA_M_18)/50000*100000
HA_39_M_Rate = nrow(HA_M_39)/50000*100000
HA_59_M_Rate = nrow(HA_M_59)/50000*100000
HA_79_M_Rate = nrow(HA_M_79)/50000*100000
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
```

```r
library(ggplot2)
BSI_data=read.csv("/Users/WangTianQi/Downloads/IS4250 Mock Data.csv")

attach(BSI_data)
y9899=0
y0001=0
y0203=0
y0405=0
y0607=0
for(i in 1:733)
{if(BSI_data[i,"Year"] == 1998 | BSI_data[i,"Year"] == 1999) y9899=y9899+1
else if(BSI_data[i,"Year"] == 2000 | BSI_data[i,"Year"] == 2001) y0001=y0001+1
else if(BSI_data[i,"Year"] == 2002 | BSI_data[i,"Year"] == 2003) y0203=y0203+1
else if(BSI_data[i,"Year"] == 2004 | BSI_data[i,"Year"] == 2005) y0405=y0405+1
else y0607=y0607+1
}
y9899HA=0
y0001HA=0
y0203HA=0
y0405HA=0
y0607HA=0
for(i in 1:733)
{if((BSI_data[i,"Year"] == 1998 | BSI_data[i,"Year"] == 1999) & BSI_data[i,"BSI"] == "HA")
    y9899HA=y9899HA+1
else if((BSI_data[i,"Year"] == 2000 | BSI_data[i,"Year"] == 2001) & BSI_data[i,"BSI"] == "HA")
    y0001HA=y0001HA+1
else if((BSI_data[i,"Year"] == 2002 | BSI_data[i,"Year"] == 2003) & BSI_data[i,"BSI"] == "HA")
    y0203HA=y0203HA+1
else if((BSI_data[i,"Year"] == 2004 | BSI_data[i,"Year"] == 2005) & BSI_data[i,"BSI"] == "HA")
    y0405HA=y0405HA+1
else if((BSI_data[i,"Year"] == 2006 | BSI_data[i,"Year"] == 2007) & BSI_data[i,"BSI"] == "HA")
    y0607HA=y0607HA+1
}

ratio=1/(2*115000)*100000
year_all=numeric(10)
dim(year_all)=c(2,5)
year_all[1,]=c(y9899-y9899HA, y0001-y0001HA, y0203-y0203HA, y0405-y0405HA, y0607-y0607HA)*ratio
year_all[2,]=c(y9899HA, y0001HA, y0203HA, y0405HA, y0607HA)*ratio
rowNames=c("CA", "HA")
row.names(year_all)=rowNames
colNames=c("1998-1999", "2000-2001", "2002-2003", "2004-2005", "2006-2007")
colnames(year_all)=colNames
bar=barplot(year_all, main="Incident Rate by Calendar Year and Site of Acquisition",
    names.arg = colNames, xlab="Calendar Year", ylab="Incident Rate per 100,000 person-years",
    ylim=c(0,100), space=1, col=c("white","black"))
legend("topright",c("Community-acquired","Healthcare-associated"), cex=1, fill=c("white","black"))
```

```r
resist9899HA=0
resist9899CA=0
resist0001HA=0
resist0001CA=0
resist0203HA=0
resist0203CA=0
resist0405HA=0
resist0405CA=0
resist0607HA=0
resist0607CA=0
for(i in 1:733)
{
  if((BSI_data[i,"Year"] == 1998 | BSI_data[i,"Year"] == 1999) & BSI_data[i,"BSI"] == "HA"
    & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist9899HA=resist9899HA+1
  if((BSI_data[i,"Year"] == 1998 | BSI_data[i,"Year"] == 1999) & BSI_data[i,"BSI"] == "CA"
    & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist9899CA=resist9899CA+1
  if((BSI_data[i,"Year"] == 2000 | BSI_data[i,"Year"] == 2001) & BSI_data[i,"BSI"] == "HA"
    & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0001HA=resist0001HA+1
  if((BSI_data[i,"Year"] == 2000 | BSI_data[i,"Year"] == 2001) & BSI_data[i,"BSI"] == "CA"
    & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0001CA=resist0001CA+1
  if((BSI_data[i,"Year"] == 2002 | BSI_data[i,"Year"] == 2003) & BSI_data[i,"BSI"] == "HA"
    & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0203HA=resist0203HA+1
  if((BSI_data[i,"Year"] == 2002 | BSI_data[i,"Year"] == 2003) & BSI_data[i,"BSI"] == "CA"
    & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0203CA=resist0203CA+1
  if((BSI_data[i,"Year"] == 2004 | BSI_data[i,"Year"] == 2005) & BSI_data[i,"BSI"] == "HA"
    & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0405HA=resist0405HA+1
  if((BSI_data[i,"Year"] == 2004 | BSI_data[i,"Year"] == 2005) & BSI_data[i,"BSI"] == "CA"
    & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0405CA=resist0405CA+1
  if((BSI_data[i,"Year"] == 2006 | BSI_data[i,"Year"] == 2007) & BSI_data[i,"BSI"] == "HA"
    & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0607HA=resist0607HA+1
  if((BSI_data[i,"Year"] == 2006 | BSI_data[i,"Year"] == 2007) & BSI_data[i,"BSI"] == "CA"
    & BSI_data[i,"Fluoroquinolone.Resistance"] == "Y") resist0607CA=resist0607CA+1
}
resistance_all=numeric(10)
dim(resistance_all)=c(2,5)
resistance_all[1,]=c(resist9899CA/(y9899-y9899HA), resist0001CA/(y0001-y0001HA),
    resist0203CA/(y0203-y0203HA), resist0405CA/(y0405-y0405HA), resist0607CA/(y0607-y0607HA))*100
resistance_all[2,]=c(resist9899HA/y9899HA, resist0001HA/y0001HA, resist0203HA/y0203HA,
    resist0405HA/y0405HA, resist0607HA/y0607HA)*100
row.names(resistance_all)=rowNames
colnames(resistance_all)=colNames

type1=c("CA","CA","CA","CA","CA","HA","HA","HA","HA","HA")
year1=c("1998-1999", "2000-2001", "2002-2003", "2004-2005", "2006-2007",
    "1998-1999", "2000-2001", "2002-2003", "2004-2005", "2006-2007")
value1=c(c(resist9899CA/(y9899-y9899HA), resist0001CA/(y0001-y0001HA),
    resist0203CA/(y0203-y0203HA), resist0405CA/(y0405-y0405HA), resist0607CA/(y0607-y0607HA))*100,
    c(resist9899HA/y9899HA, resist0001HA/y0001HA, resist0203HA/y0203HA, resist0405HA/y0405HA, resist0607HA/y0607HA)*100)
df_test=data.frame(type1,year1,value1)
ggplot(df_test, aes(x=year1, y=value1,colour=type1, group=type1))
    +geom_line(size=2)+ggtitle("In Vitro Fluoroquinolone Resistance Rates by Calendar Year and Site of Acquisition ")
    +xlab("Calendar Year")+ylab("Resistance Rate %")+ylim(0,20)+geom_point(aes(shape=type1),size=5)
    +scale_color_discrete(name="BSI", labels=c("Community-acquired","Healthcare-associated"))
    +scale_shape_discrete(name="BSI",labels=c("Community-acquired","Healthcare-associated"))
```