THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENTS OF BIOLOGY AND HEALTH POLICY AND ADMINISTRATION

QUANTITATIVE SPATIAL ANALYSIS OF NOSOCOMIAL METHICILLIN-RESISTANT *STAPHYLOCOCCUS AUREUS* (MRSA) AND *CLOSTRIDIUM DIFFICILE* IN PENNSYLVANIA HOSPITALS

WILTON PRESCOTT SMITH
SPRING 2016

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Biology
with interdisciplinary honors in Biology and Health Policy and Administration

Reviewed and approved* by the following:

Ottar Bjørnstad
Distinguished Professor of Entomology and Biology
Thesis Supervisor

Sarah Assmann
Waller Professor of Biology
Honors Adviser

Rhonda BeLue
Associate Professor of Health Policy and Administration
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

Nosocomial infections, also known as hospital-acquired infections, are infections caused by pathogens contracted in the hospital environment. It is possible that nosocomial infections are greater than a hospital-level problem, as factors about a hospital's catchment population and geographic region could confer epidemiological risk (1). This research examines population and spatial effects on infection rates of Methicillin-Resistant *Staphylococcus aureus* (MRSA) and *Clostridium difficile* (*C. diff*) in 165 Pennsylvania hospitals. Infection rates were obtained from the CDC's Hospital Compare dataset, and standardized based on patient days (2). Standardized infection rates were regressed against hospital size, as well as county-level socioeconomic data from the Area Health Resource File (AHRF) to assess catchment population trends (3). Hospitals were also compared by spatial autocorrelation analysis, and by local indicators of spatial association (LISA) (4,5). Finally, infection rates of MRSA and *C. diff* were compared by regression and spatial cross-correlation. This analysis finds that the size of a hospital has a significant positive effect on both the presence of infection and the rate of infection in that hospital. Concerning spatial risk, clustering of MRSA occurs in urban areas, with significant positive spatial autocorrelation observed out to a distance of 43 km. However, spatial analysis of *C. diff* shows inconclusive evidence of regional influence. Further, the infection rates of MRSA and *C. diff* are significantly correlated on a hospital-by-hospital basis, but are not correlated spatially. Thus, the conclusion from this research is that both hospital-level and regional-level factors likely contribute to these nosocomial infection rates.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

There are a number of people whom I must thank for their assistance in this project. First I would like to thank my thesis supervisor, Dr. Ottar Bjørnstad, for his patient guidance and kind assistance throughout this project. Our work together has helped me discover new passions and develop invaluable skills; for both of which I am deeply grateful.

Next I would like to thank both of the honors advisors who helped me with my thesis. Thank you to Dr. Rhonda BeLue, my advisor in Health Policy and Administration, who originally pointed me toward this topic, and helped me every step of the way; from project design, to locating data, all the way up to final review and editing. I would also like to thank Dr. Sarah Assmann, my Biology honors advisor, for her years of guidance and for her kind review of this thesis.

I would also like to thank all of my friends, my sister Amanda, and my girlfriend Lindsay for their constant support and encouragement throughout my time at Penn State. I'd also like to thank my extended family, especially my Nana and Pop Pop, who have been so supportive of my education. Lastly, I must thank my mother and father, to whom I owe everything. I am where I am today because of your unending love and sacrifice.

Chapter 1

Introduction

Background

Nosocomial infections are a persistent problem in the healthcare community that jeopardize patient health and lead to billions of dollars in waste every year. A point-prevalence study in 2012 showed that up to 3 percent of all patients in a hospital at a given time are there because of a hospital-acquired infection (HAI) (6). Some estimates project that that HAIs are responsible for upwards of $30 billion in healthcare costs annually in the United States (7).

HAIs are not a new problem. There are studies of nosocomial infections in children's wards dating all the way back to before WWII (8,9). This problem will not be going away any time soon, either. Pathogens have evolved to transmit to new hosts, and hospitals provide a dense population of susceptible targets. Even more disconcerting is that these pathogens are getting better at their job because of the evolution of antibiotic and antimicrobial resistance (10).

In 1988, the CDC published specific definitions for the different type of nosocomial infections. They include surgical wound infections, primary bloodstream infections, pneumonia, urinary tract infections, bone and joint infections, cardiovascular system infections, central nervous system infections, ear eyes, nose, throat, and mouth infections, gastrointestinal tract infections, lower respiratory tract infections (excluding pneumonia), reproductive tract infections, skin and soft tissue infections, and systemic infections (11). Of these categories, the most common are pneumonia, surgical site infections, urinary tract infections, and primary bloodstream infections (12).

Infectious Bacteria

There are a number of different bacteria that commonly cause HAIs. Some of the most prevalent include *Staphylococcus aureus, Enterococcus* spp., *Acinetobacter* spp., and *Clostridium difficile* (13). This study uses data for Methicillin-Resistant *Staphylococcus aureus* (MRSA) and *Clostridium difficile (C. diff)* (2).

MRSA is one of, if not the single most common bacteria causing HAIs (13). In the hospital setting, MRSA is a cause of bloodstream infections, surgical site infections, and pneumonia (14). It is spread by contact with a contaminated wound, surface, or hands of a healthcare worker (HCW) (14). *Staphylococcus aureus* can live dormant or without showing any symptoms for long periods of time; because of this up to 1 in 3 people have *Staphylococcus aureus* colonized in their nasal cavity, and up to 1 in 50 have MRSA (14). Because MRSA is resistant to many common antibiotics, it leaves care providers with limited options to fight the infection (15).

Clostridium difficile is a bacterium that colonizes the intestines and causes gastrointestinal system infections (16). *C. diff* spreads in a fecal-to-oral route, and is most commonly transmitted through the hands of an HCW who has touched a contaminated surface or instrument (16). Those at highest risk for contracting *C. diff* are the elderly and those on bacterial antibiotics who are receiving medical care (16). A big problem with control of *C. diff* is that while hand-washing does kill the bacteria, alcohol-based hand sanitizers do not. This means that "clean" hands may still be colonized; and that many hospital employees must change their typical sanitization practice before providing care to a patient with C. diff (17).

Infection Management

As both infections are unique, rampant, and dangerous, the CDC has published toolkits for healthcare facilities to use to reduce the spread of MRSA and *C. diff* (18,19). Infection control measures center around proper hygiene, reducing contact with contaminated surfaces and individuals, and education of all HCW's about the pathogens (18,19). Electronic presentations containing the core prevention strategies for each bacterium can be found in Appendix A.

The primary focus of HAI control within a hospital is hand washing practice. It has been shown that there is a causal link between hand washing and risk of infection (20). Increasing hand washing compliance among HCW's is thus a main goal of infection management protocols. Other management strategies include reducing skin-to-skin contact with the use of gloves, gowns, nose and eye protection; and proper handing of equipment that comes into contact with infected patients (21).

On a larger scale, the CDC is just one of many partners from the Department of Health and Human Services that have joined forces in an attempt to control HAI incidence across the country. The goals of this alliance are to improve hospital policies and compliance with them through increasing incentive systems and regulatory oversight. (22) Furthermore, the incidence of hospital acquired infections is monitored at both the state and federal level through reporting to the National Healthcare Safety Network (NHSN) (23). Data reported through the NHSN is the primary source of data for this analysis (2).

Quality of Care

There are a number of factors, both endogenous and exogenous, that could contribute to the quality of care received at a specific hospital. With regard to HAIs, high quality of care is represented by low infection rates, and vice versa.

First, it is possible that infection risk could vary with the size of a hospital. The primary mode of HAI transmission is from the hands of a HCW to a patient (14,16). Every time a HCW comes in contact with a patient, it presents the chance for an infection to be passed. Thus, larger hospitals with more patients will have more HCW-patient contacts, and will have more chances for HAI transmission. In addition, hospital size may affect the resources available, such as extra rooms for proper warding and quarantine of infections. Previous research has shown that hospital-acquired pneumonia incidence is three times higher in hospitals where rooms have double occupancy than in hospitals with single occupancy (24).

Second, it is possible that the probability of an infection could be linked to socioeconomic status. It has been shown that greater economic status has been linked to greater access to healthcare, especially in children (25). It is thus possible that hospitals that serve poor communities will serve more sick and susceptible populations, equating to higher infection rates.

There also may be a rural vs. urban component of these infection rates. Research has indicated that rural areas have differential access to healthcare. Compared to urban areas, they have fewer community health centers, sometimes lacking sufficient access to care (26). This may mean that rural hospitals similarly serve more sick and susceptible populations, and thus exhibit higher infection rates.

One possible way to measure the quality of care for a community is to look at the accreditation of its hospitals. The Joint Commission will accredit hospitals based on the quality

of care they provide, also known as JCAHO Accreditation (27). Looking at the number of JCAHO accredited hospitals serving an area is one metric that could be used to evaluate the quality of care received by that area. Furthermore, in 2012, Mumford et al. studied 77 hospitals in Australia and found a negative correlation between hospital *S. aureus* infection rates and hospital accreditation scores (28). Thus, it could be possible to estimate quality of care through HAI incidence analysis.

Epidemiology

Nosocomial infections could also be affected by spatial epidemiology. The spatial dynamics of many different diseases have been studied extensively, suggesting many different models of spatial spread, including waves and clusters (29,30). It is possible that these spatial effects, which play roles in the spread of community-acquired diseases, also have an effect on hospital-acquired diseases. If more patients are entering the hospital colonized with an infection, it increases the chances of contaminating HCWs, and thus increases chances of transmitting disease to an uncolonized patient in the hospital (31). Both MRSA and *C. diff* were once considered only nosocomial infections; however, it has been shown that it is possible to acquire both of the bacteria in the community as well, even for patients without predisposing risk factors (32,33). Thus, this phenomenon could apply to the spread of these diseases.

By this logic, hospitals that are close together, and thus share the same catchment population, would be expected to experience similar community “pressures” and show similar infection rates. In addition, hospitals located in urban areas, where disease is given more

opportunity to spread to new hosts, would be expected to have greater community "pressures" and higher infection rates (34).

Purpose

Thus, since it is possible that HAIs may be a problem greater than hospital control measures, the purpose of this research is to identify what factors about a hospital, its population, and its location increase the risk of nosocomial infection. This purpose is achieved through investigation of three specific research questions.

First, what factors about a hospital and its population affect its infection rates? Using patient days at each hospital, we can create a proxy to compare hospital size across the state. Do smaller or larger hospitals have higher infection rates? Looking at a hospital's catchment population, are there any socioeconomic indicators of high infection rates?

Second, is there a spatial component to the spread of HAIs? Using geographic coordinates, we can compare hospitals regionally and identify spatial trends. Are hospitals that are close in proximity to each other more correlated in their infection rates than those that are far apart? What happens when hospitals share the same catchment population?

Finally, is there a difference between patterns of MRSA and *C. diff* incidence? The two diseases have different control measures and different modes of transmission; do they spread differently? Do they spread differently within a hospital, on a regional level, or both?

Chapter 2

Method

To address the three main research questions outlined above, hospital infection data for both MRSA and *C. diff* were merged with geographic and socioeconomic data. Epidemiological statistical analyses were then performed, specifically population-based and spatial analyses of infection risk. These analyses were performed individually for each pathogen, and a cross-comparison was performed between the two. All analyses were performed using the statistical software R (35).

Sample

The analysis was performed across a sample of hospitals (n=165) from the state of Pennsylvania. Hospital-level data, including location, patient days, and infection counts, was acquired from the Hospital Compare Dataset developed by the Centers for Disease Control (CDC) and collected through the National Healthcare Safety Network (NHSN) (2). County-level data, including population statistics and socioeconomic indicators, was acquired from the Area Health Resource File made available by the Bureau of Health Workforce and American Medical Association (3). Hospital-level and county-level data were merged using Federal Information Processing Standard (FIPS) code for the county location of each hospital.

To achieve a standardized infection rate (SIR) across all hospitals, the total number of observed infections per-hospital, per-disease, was divided by the number of patient days per-hospital, and multiplied by 1000:

$$SIR = [(Observed\ Cases) \div (Patient\ Days)] \times 1000$$

The SIR thus represents the total number of disease cases per 1000 patient days at each hospital. SIR data was plotted geographically to observe general trends and guide the ensuring research questions.

Population Analysis

The methods in this section outline the approach to answer the first research question, namely, what factors about a hospital and its population affect its infection rates?

In order to determine what factors about a hospital and its catchment population contribute to infection rates, multiple regression analysis was performed. The SIR for both MRSA and *C. diff* were regressed against the variables listed in Table 1.

Table 1. Population variables tested in multiple regression analysis

Hospital-Level	County-Level
Patient Days	Population density per square mile
	Per-capita income
	Median household income
	Number of persons living in poverty
	Percent of persons living in poverty
	Percent of hospitals with JCAHO accreditation

Next, stepwise selection of variables using Akaike information criterion (AIC) was performed to determine the most parsimonious model leading to differences in hospital infection rates (36). The resulting model was regressed again to determine significance using a two-sided t-test at $p=0.05$.

The effect of hospital size on the probability of an infection in a hospital was then investigated. To study presence or absence of an infection in a hospital, SIR data were converted to discrete categorical data. All hospitals with an SIR = 0 (infection absent) were assigned a value of 0, and all hospitals with an SIR > 0 (infection present) were assigned a value of 1. Logistical regression with a logit link function and binomial error was performed for patient days versus presence/absence data of each infection (37).

Finally, a spatial regression of MRSA was performed using the nlme package in R (38). Spatial and non-spatial regression models of the parsimonious function were compared using likelihood ratio tests (39).

Spatial Analysis

The methods in this section outline the approach to answer the second research question, is there a spatial component to the spread of nosocomial infections?

Following population analysis, the effects of regional influence were investigated. Each hospital was analyzed using its SIR and the latitude and longitude coordinates of its location. All analyses were performed in R using the ncf package (40).

First, spatial autocorrelation analysis was performed for each disease using correlograms with 10 km distance classes. Significance was calculated based on 1000 random resamples of the data (4). Significance was measured using a two-sided test at p=0.05. Spatial dependence was measured using Moran's I (41).

Second, local indicators of spatial association (LISA) analysis was performed for each disease using a neighborhood size of 25 km. Significance was calculated based on 1000 random resamples of the data (5). Significance was measured using a two-sided test at p=0.05.

Disease Cross-Comparison

The methods in this final section address the third and final research question, is there a difference between patterns of MRSA and *C. diff* incidence?

The patterns of the two diseases were compared using both a linear regression and spatial cross-correlation. Linear regression analysis was performed first, regressing the MRSA and *C. diff* SIR for each hospital Significance was measured using a two-sided t-test at p=0.05 (42).

Spatial autocorrelation was then performed using the MRSA and *C. diff* SIR for each hospital, as well as the latitude and longitude coordinates of each hospital for location. As in the analysis above, the distance increment was set at 10 km, and significance was calculated based on 1000 random resamples of the data, using a two-sided test at p=0.05 (4).

Chapter 3

Results

Depicted below are the original geographic outputs of SIR data that were used to gauge general trends in the data and guide the research questions. Visual observation displays some evidence of higher infection rates in urban areas than in rural areas, as shown by observations in the Philadelphia and Pittsburgh regions. All R output summaries are provided in Appendix B.

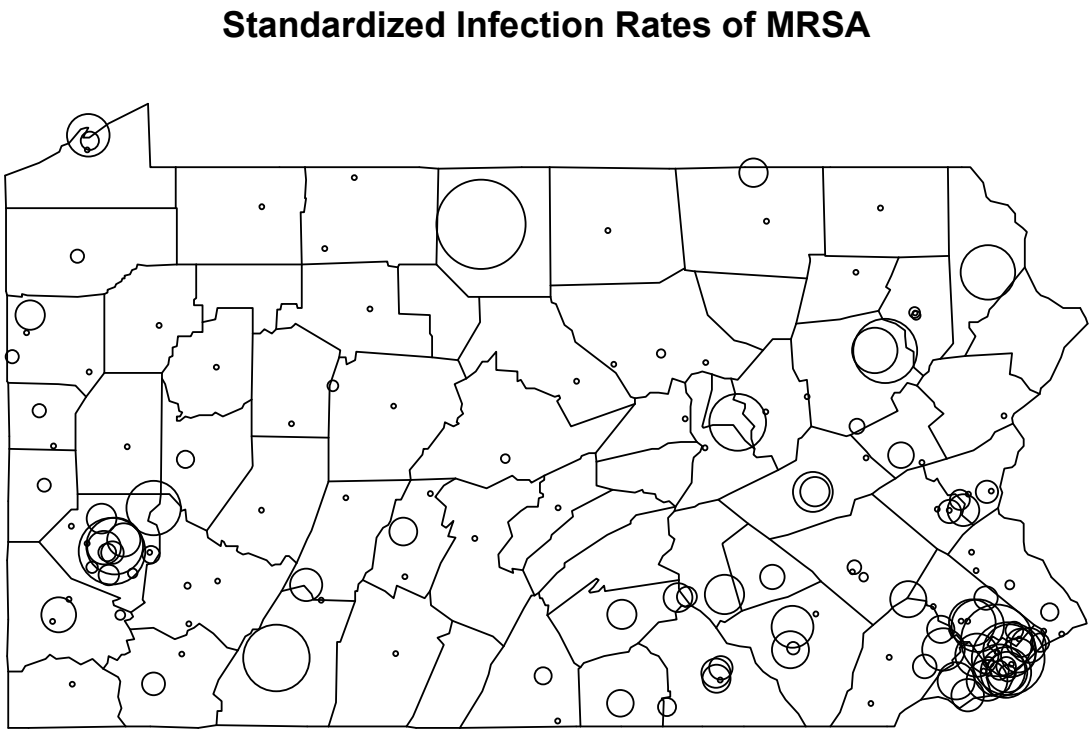

Figure 1. Map of MRSA SIR across Pennsylvania hospitals

In the map above, each circle represents one hospital, and the size of the circle is determined by the SIR for that hospital for MRSA. The circles are scaled in reference to the largest observation, which has a radius of 0.2 inches.

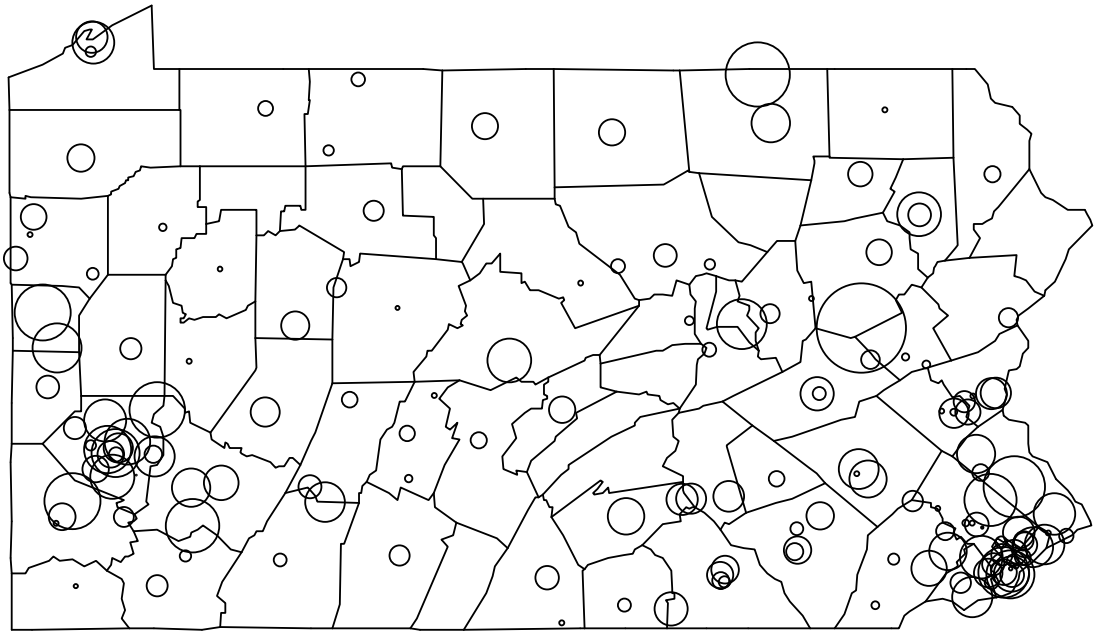

Figure 2. Map of *C. diff* SIR across Pennsylvania hospitals

In the map above, each circle represents one hospital, and the size of that circle is determined by the SIR for that hospital for C. diff. The circles are scaled in reference to the largest observation, which has a radius of 0.2 inches.

Population Analysis

The results of this section address the first research question, exploring the effects of a hospital's population on its infection risk.

The results of the multiple regression of MRSA SIR against the variables outlined in Table 1 yielded one significant positive interaction: patient days (coefficient = 3.129e-09, standard error = 5.842e-08, t = 5.356, df = 148, $p < 0.001$). A total of three variables yielded positive correlations (patient days, population density, per capita income) while the remaining

four returned negative correlations (median household income, percent of persons in poverty, number of persons in poverty, percent of hospitals with JCAHO accreditation). However, the only statistically significant correlation observed was with patient days.

Similarly, the results of the multiple regression of *C. diff* SIR against the variables outlined in Table 1 also yielded one significant interaction: patient days (coefficient = 2.709e-06, standard error = 5.742e-07, t = 4.718, df = 146, $p < 0.001$). A total of six variables yielded positive correlations (patient days, per capita income, median household income, percent of persons in poverty, number of persons in poverty, percent of hospitals with JCAHO accreditation) while only one returned a negative correlation (population density). However, the only statistically significant correlation observed was with patient days.

The MRSA stepwise selection using AIC eliminated all but two variables to produce a parsimonious model for SIR. The remaining variables were patient days (coefficient = 3.133e-07, standard error = 5.486e-08, t = 5.711, df = 153, $p < 0.001$) and population density (coefficient = 2.474e-06, standard error = 9.417e-07, t = 2.627, df = 153, $p = 0.01$).

The *C. diff* stepwise selection using AIC eliminated all variables except for patient days (coefficient = 2.744e-06, standard error = 5.163e-07, t = 5.315, df = 152, $p < 0.001$).

At this point, the SIR for each infection was plotted against the most significant interaction found: patient days. The resulting graphs are depicted below. In figure 3 and figure 4, it is observed that there are a number of hospitals where the SIR is exactly zero.

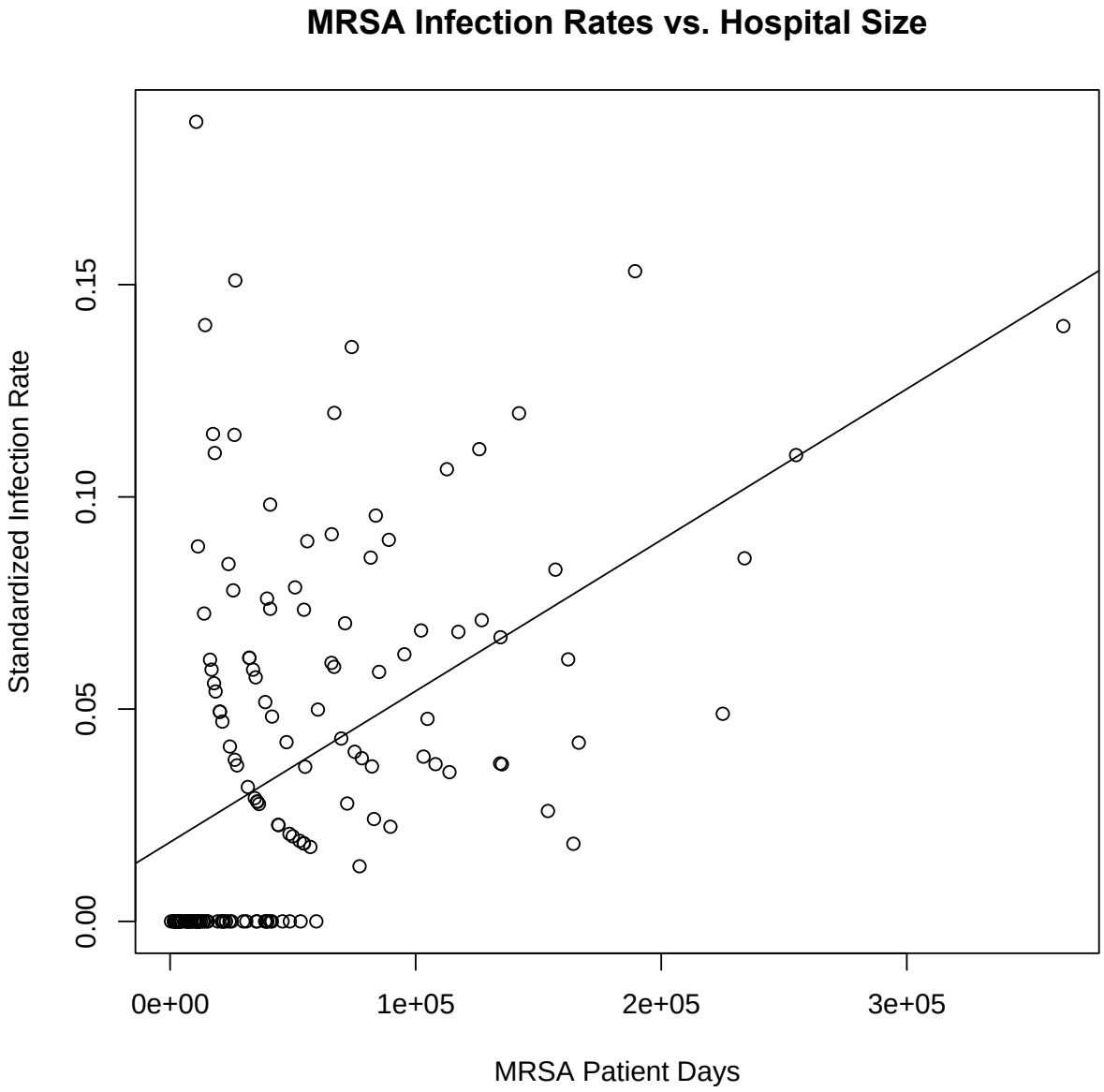

Figure 3. Regression of MRSA SIR vs. patient days

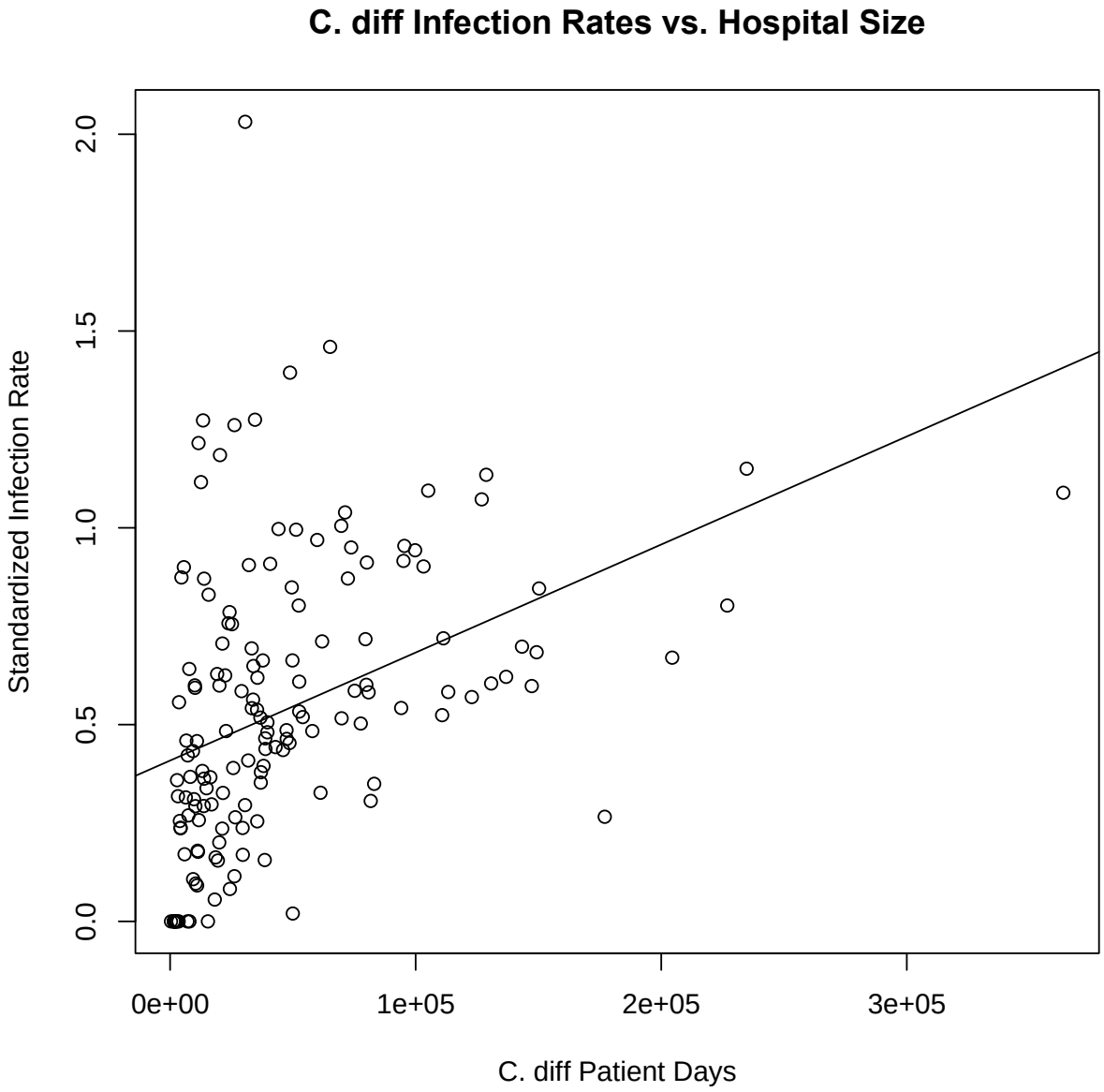

Figure 4. Regression of *C. diff* SIR vs. patient days

After organization into discrete categories, the SIR of all of the hospitals were organized into a box plot for each of the infections. Figures 5 and 6 both indicate a general trend that larger hospitals are generally more likely to have an infection present, or vice versa.

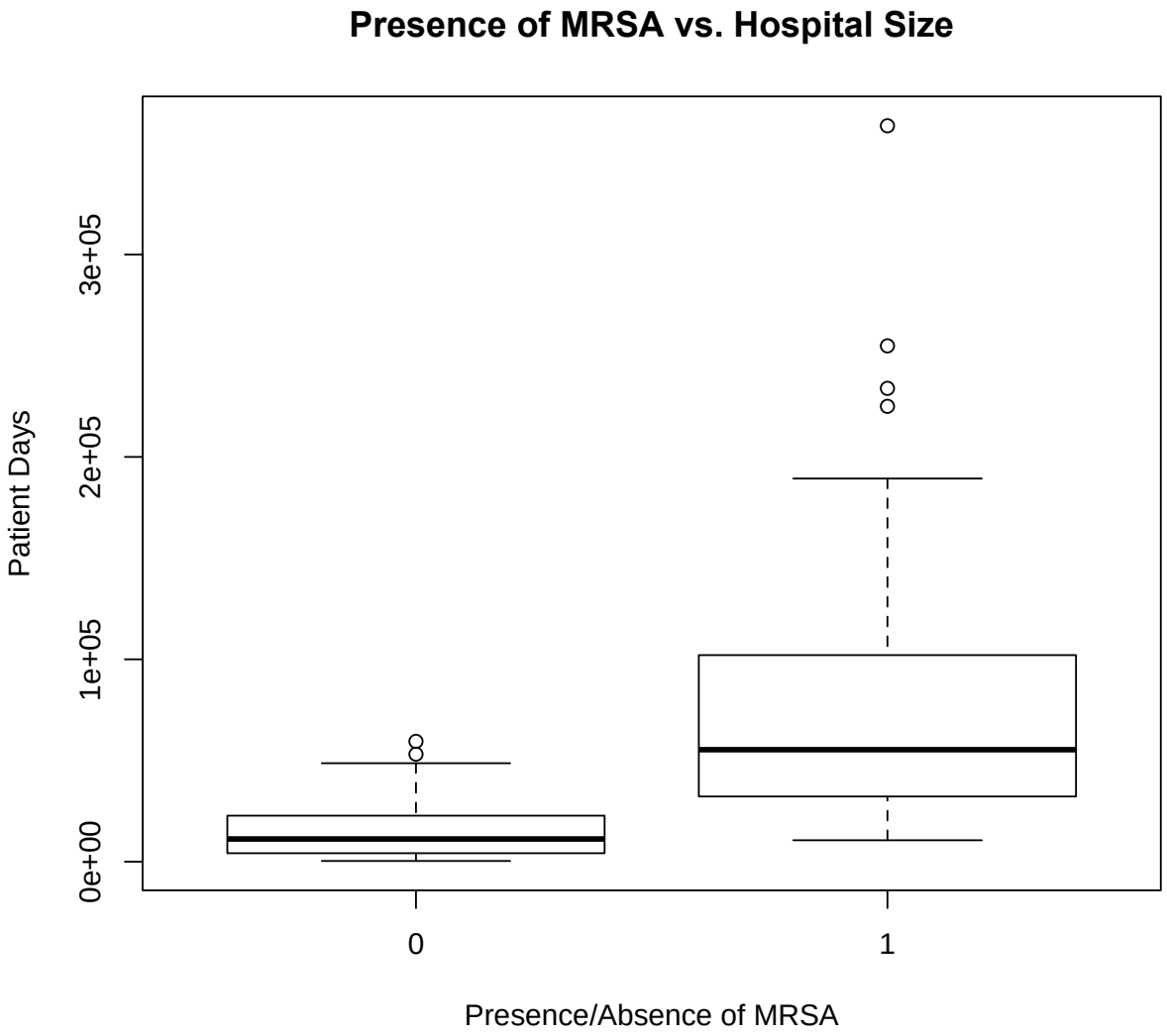

Figure 5. Hospital sizes divided by the presence or absence of an MRSA infection

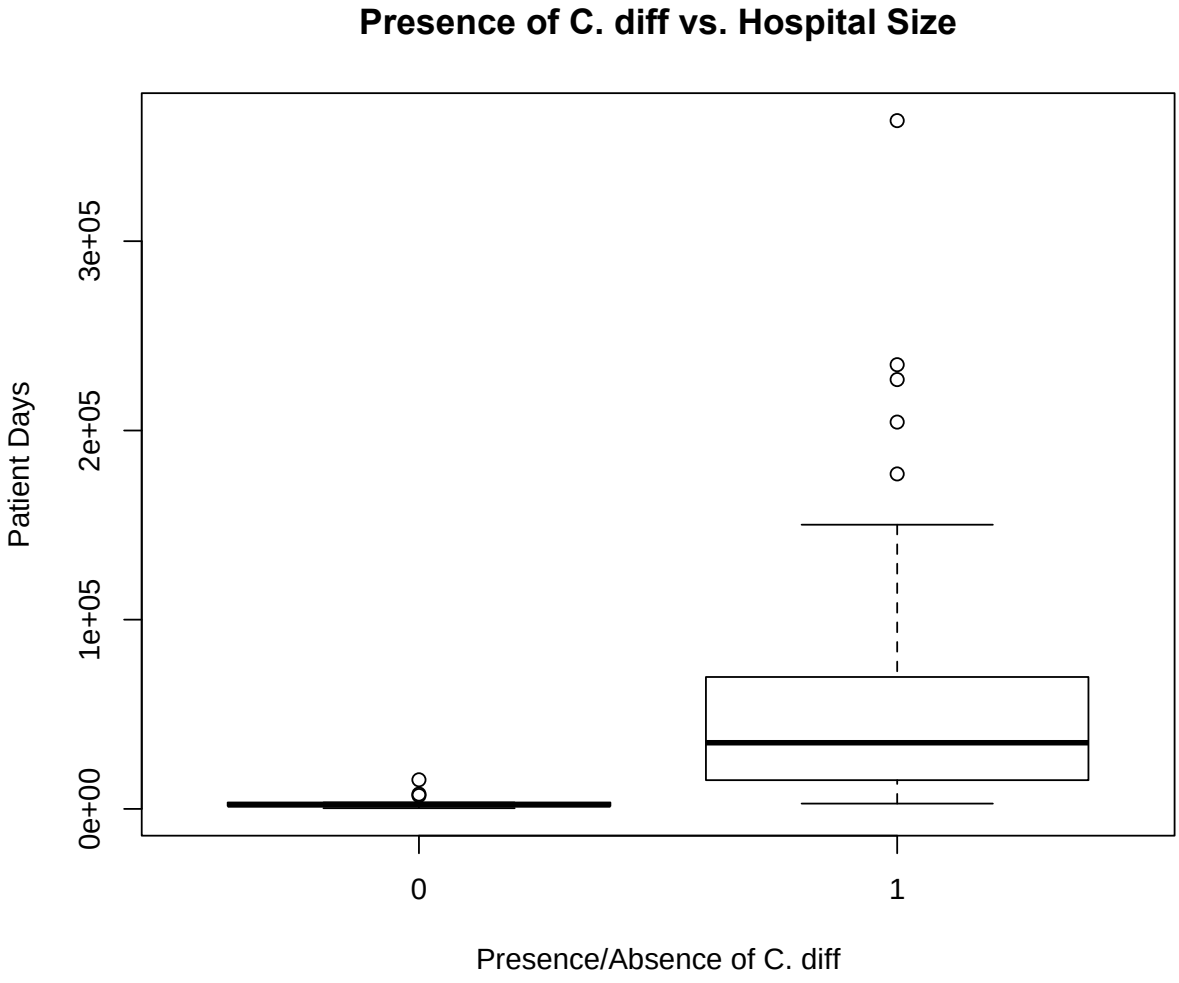

Figure 6. Hospital sizes divided by the presence or absence of a *C. diff* infection

The results of the logistic regression of this categorical data are depicted below in figures 7 and 8. The best-fitting line shows a clear sigmoidal curve of the probability of the presence of an infection as hospital size increases. Figures 7 and 8 show that there are more hospitals with *C. diff* present than with MRSA present. In addition, the probability of a *C. diff* infection increases at a lower hospital size than the probability of a MRSA infection.

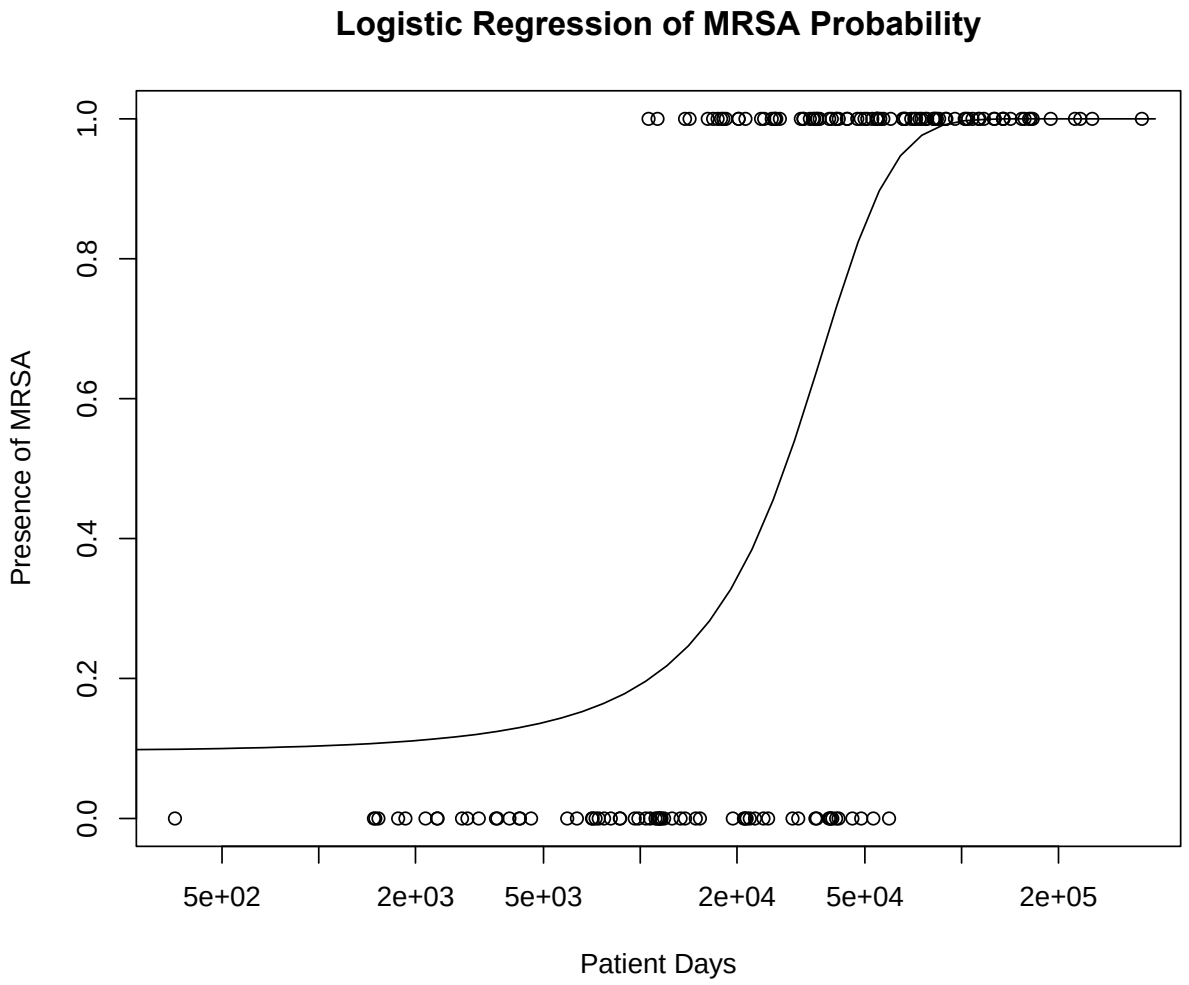

Figure 7. Logistic Regression of MRSA infection probability with increasing hospital size

In the figure above, the line plotted represents the maximum likelihood estimation of the logistic regression model, calculated by $y = (e^{(a+bx)}) / (1+e^{(a+bx)})$*, where a and b are the coefficients from the logistic regression.*

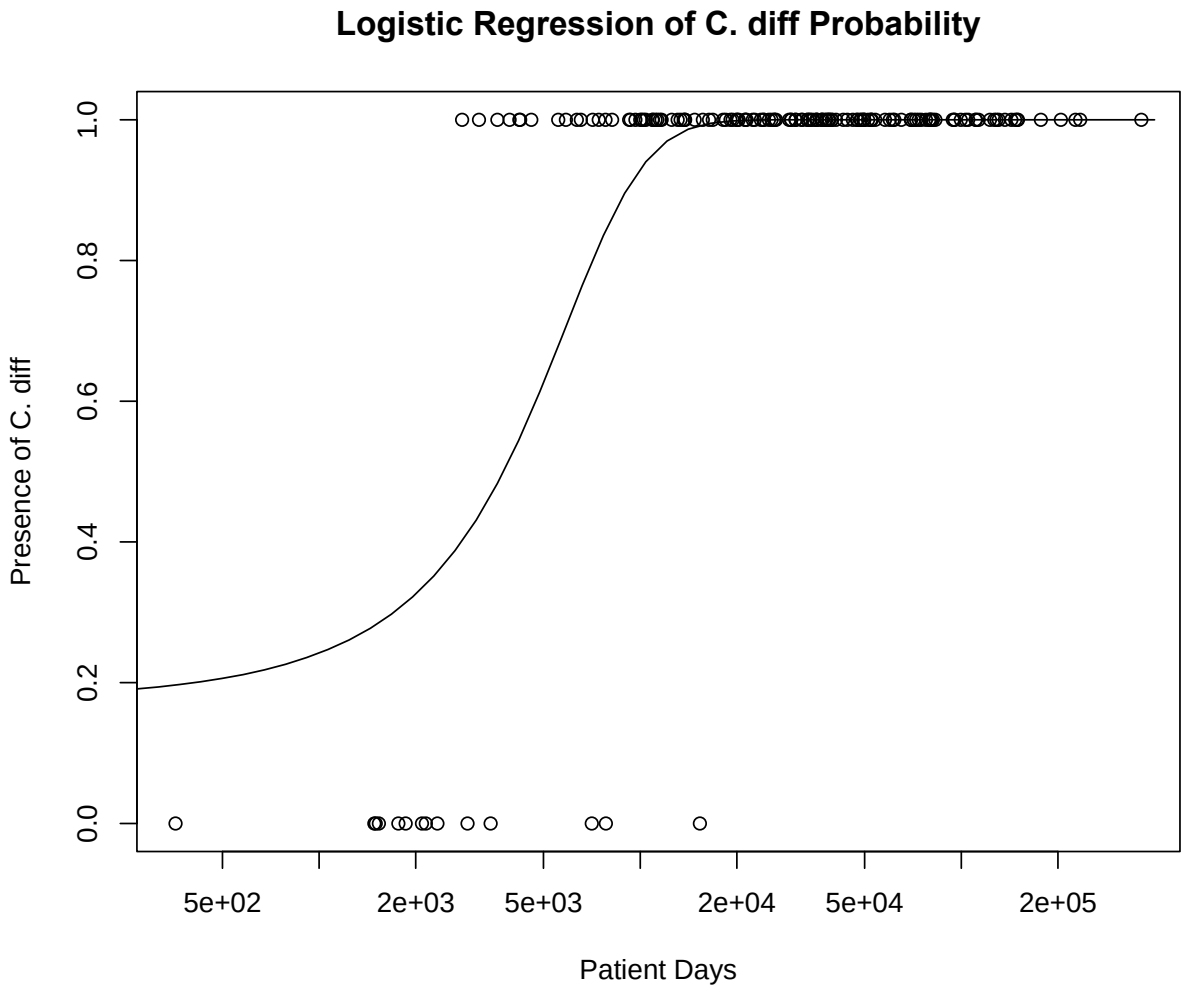

Figure 8. Logistic Regression of *C. diff* infection probability with increasing hospital size

In the figure above, the line plotted represents the maximum likelihood estimation of the logistic regression model, calculated by $y = (e^{(a+bx)}) / (1+e^{(a+bx)})$*, where a and b are the coefficients from the logistic regression.*

Finally, spatial regression analysis for MRSA showed that the interactions for both patient days (coefficient = 3.30e-07, standard error = 4.7e-08, t = 6.986, df = 153, $p < 0.001$) and population density (coefficient = 2.862e-06, standard error = 1.022e-06, t = 2.801, df = 153, p = 0.006) remain significant even when accounting for regional influences. Furthermore, likelihood

ratio testing showed that parsimonious multiple regression and spatial regression were not significantly different (ratio = 3.537, p = 0.17).

Spatial Analysis

The results of this section address the second research question, namely, is there a spatial component to the spread of HAIs?

Spatial autocorrelation analysis found significant positive autocorrelation for MRSA out to the third distance class (10 km classes). Overall, the autocorrelation yielded an intercept of x = 43.154, indicating general positive autocorrelation out to a distance of 43.154 km. The correlations of the first four distance classes are listed in Table 2 and visualized in figure 9.

Table 2. MRSA spatial autocorrelation results

Distance Class	Mean of Class	Correlation	P-Value
1	5.90 km	0.2712	0.004
2	14.97 km	0.1634	0.008
3	25.22 km	0.1064	0.021
4	35.10 km	0.0357	0.217

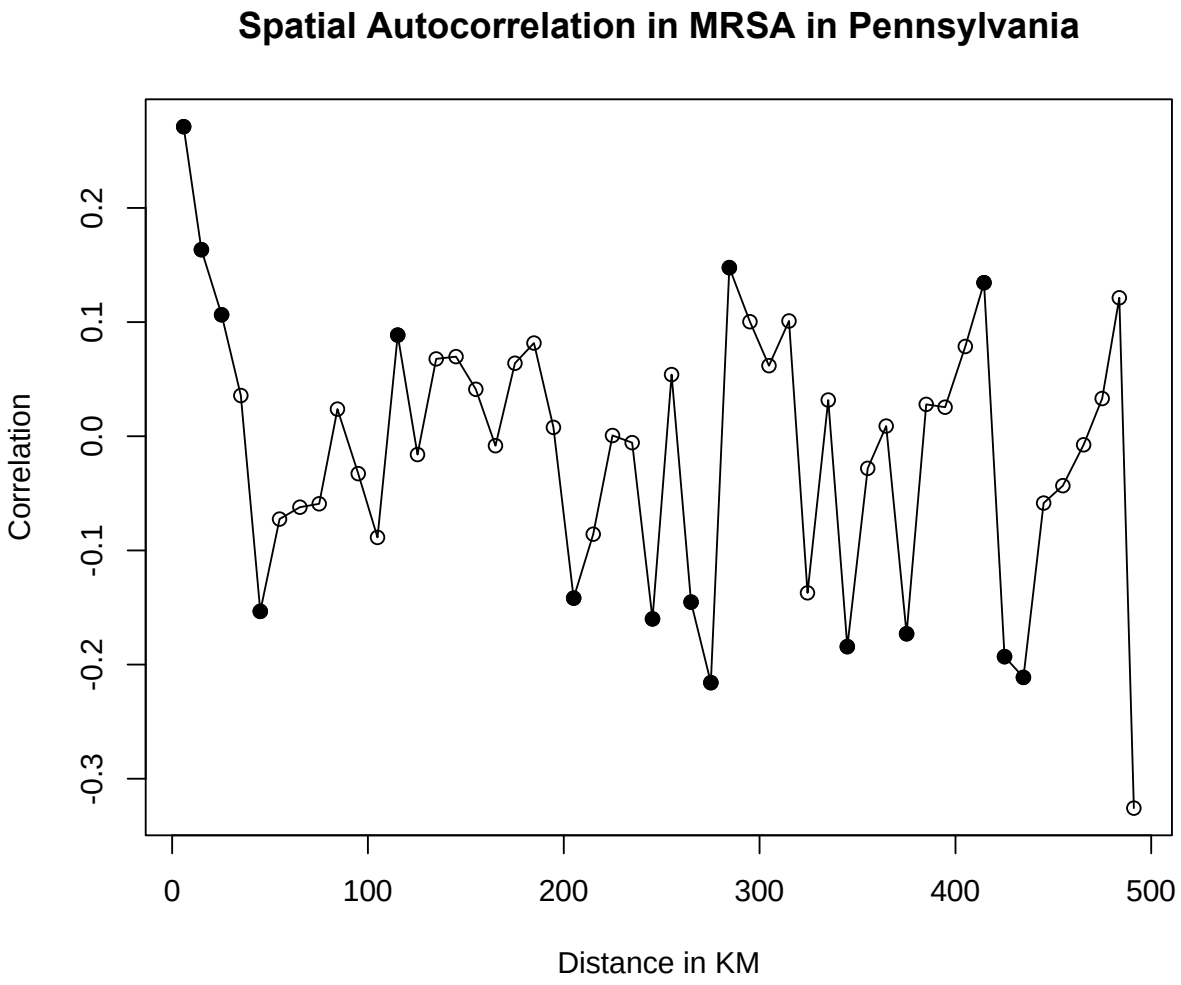

Figure 9. MRSA spatial autocorrelation graph

In the graph above, each circle represents a distance class of the spatial autocorrelation analysis. Filled circles represent statistically significant distance classes. This analysis is only interested in the local autocorrelation, thus the significant interactions that appear after the original positive autocorrelation is lost can be disregarded.

However, spatial autocorrelation analysis did not find significant autocorrelation for *C. diff*. Without any significant distance classes, the autocorrelation yielded an intercept of x = 0.

The correlations of the first four distance classes are listed in Table 3 and visualized in figure 10.

Table 3. *C. diff* spatial autocorrelation results

Distance Class	Mean of Class	Correlation	P-Value
1	5.97 km	-0.0219	0.441
2	14.97 km	-0.0004	0.443
3	25.26 km	0.0351	0.215
4	35.08 km	-0.0481	0.213

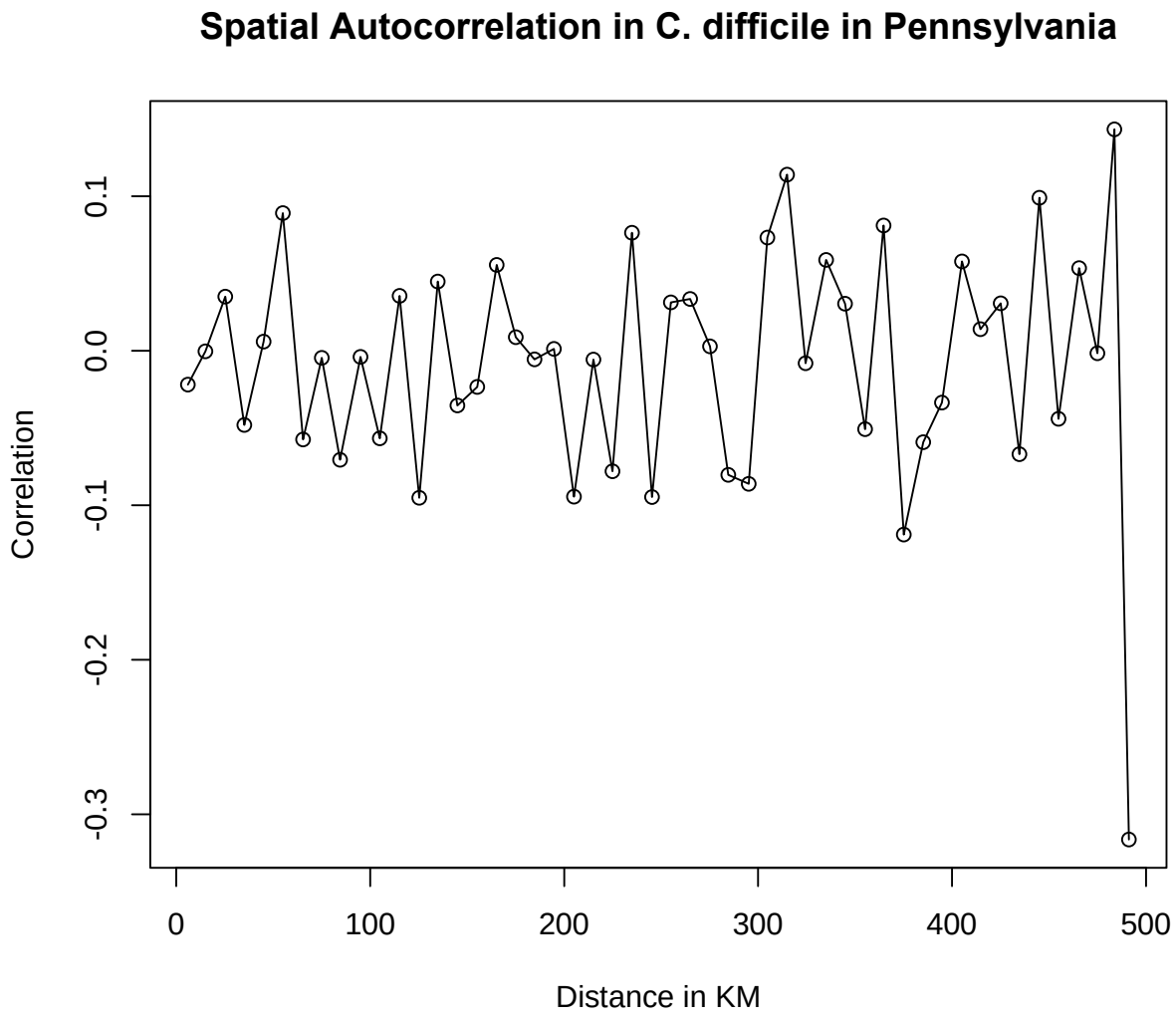

Figure 10. *C. diff* spatial autocorrelation graph

In the graph above, each circle represents a distance class of the spatial autocorrelation analysis. Filled in circles represent statistically significant distance classes.

Following spatial autocorrelation analysis, local hotspots were identified using local indicators of spatial association (LISA). Figure 9 displays the LISA analysis for MRSA, while figure 10 shows the LISA analysis of *C. diff*.

Qualitative observations suggest that MRSA hotspots exist in both of the major cities of Pennsylvania (Philadelphia and Pittsburgh), while most of the rural areas are generally cold spots for the disease. There are a few instances of hospitals with higher-than-average infection rates in rural areas, however only one of these instances is classified as statistically significant.

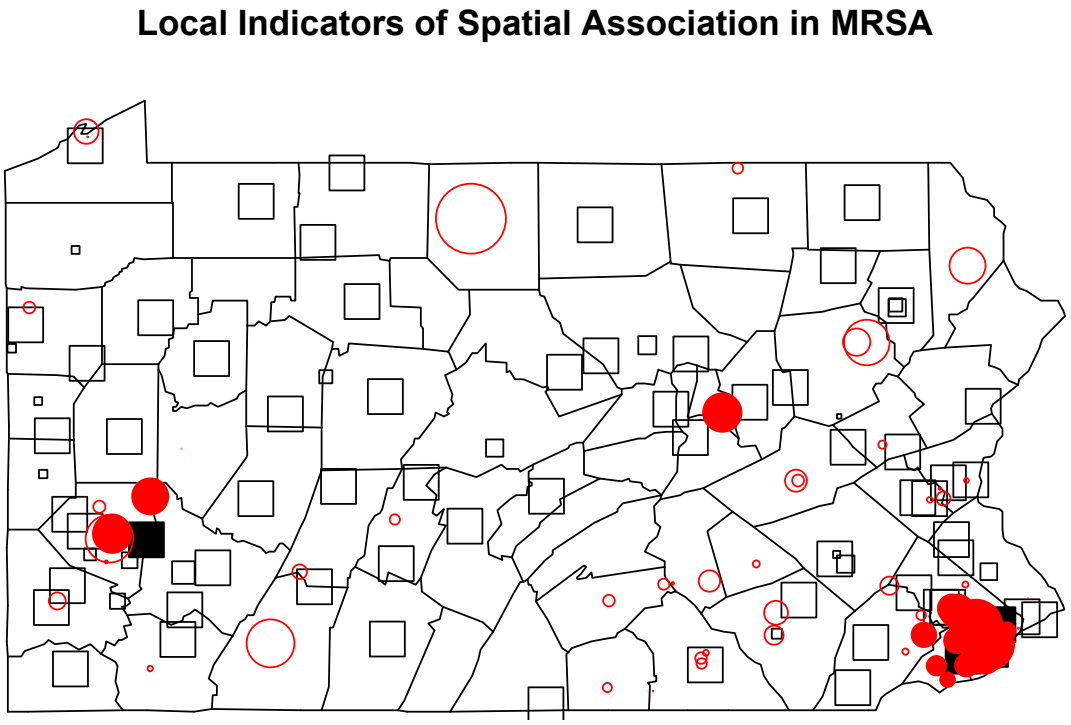

Figure 11. LISA plot of MRSA infection rates

In the graph above, hospitals with above-average infection rates are depicted by red circles, while hospitals with below-average rates are indicated by black squares. The size of the shape is indicative of distance from the mean infection rate. The circles are scaled in reference to the largest observation, which has a radius of 0.2 inches. Finally, shapes that are colored-in represent hospitals that exhibit significant autocorrelation with the other hospitals in their neighborhood (25 km).

Unlike MRSA, observation of *C. diff* yields inconclusive results. There are no significant interactions found in the city of Philadelphia, and a smaller number of significant interactions found in Pittsburgh than in the rural and suburban areas that comprise the rest of the state.

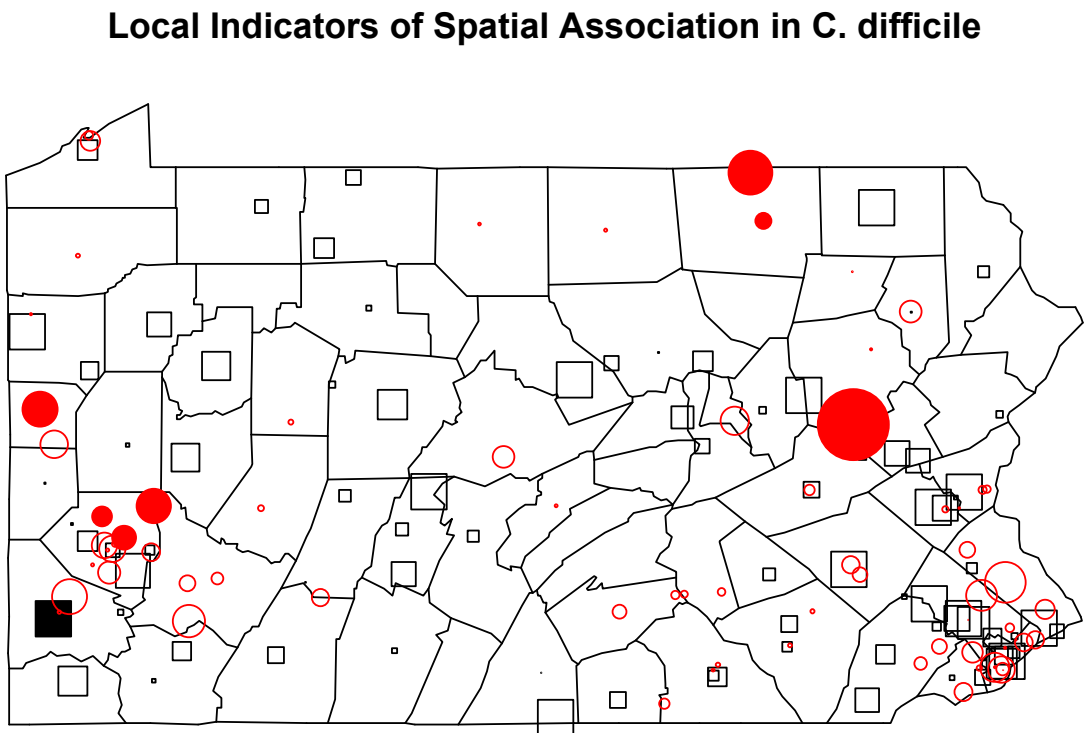

Figure 12. LISA plot of *C. diff* infection rates

In the graph above, hospitals with above-average infection rates are depicted by red circles, while hospitals with below-average rates are indicated by black squares. The size of the shape is indicative of distance from the mean infection rate. The circles are scaled in reference to the largest observation, which has a radius of 0.2 inches. Finally, shapes that are colored-in

represent hospitals that exhibit significant autocorrelation with the other hospitals in their neighborhood (25 km).

Disease Cross-Comparison

The results of this final section concern the final research question, investigating if there is a difference between the patterns of MRSA and *C. diff* incidence.

Linear regression of MRSA and *C. diff* infection rates yielded a statistically significant strong positive correlation between the two infections (coefficient = 0.0329, standard error = 0.0088, $t = 3.732$, $df = 152$, $p < 0.001$).

However, spatial cross-correlation between the two diseases yielded no statistically significant local cross-correlation, thus incidence is correlated at the hospital but not regional level. The correlations of the first four distance classes are listed in Table 4 and visualized in figure 13.

Table 4. Spatial cross-correlation results

Distance Class	Mean of Class	Correlation	P-Value
1	5.97 km	-0.0192	0.391
2	14.97 km	0.0304	0.217
3	25.26 km	0.0324	0.195
4	35.08 km	0.0986	0.009

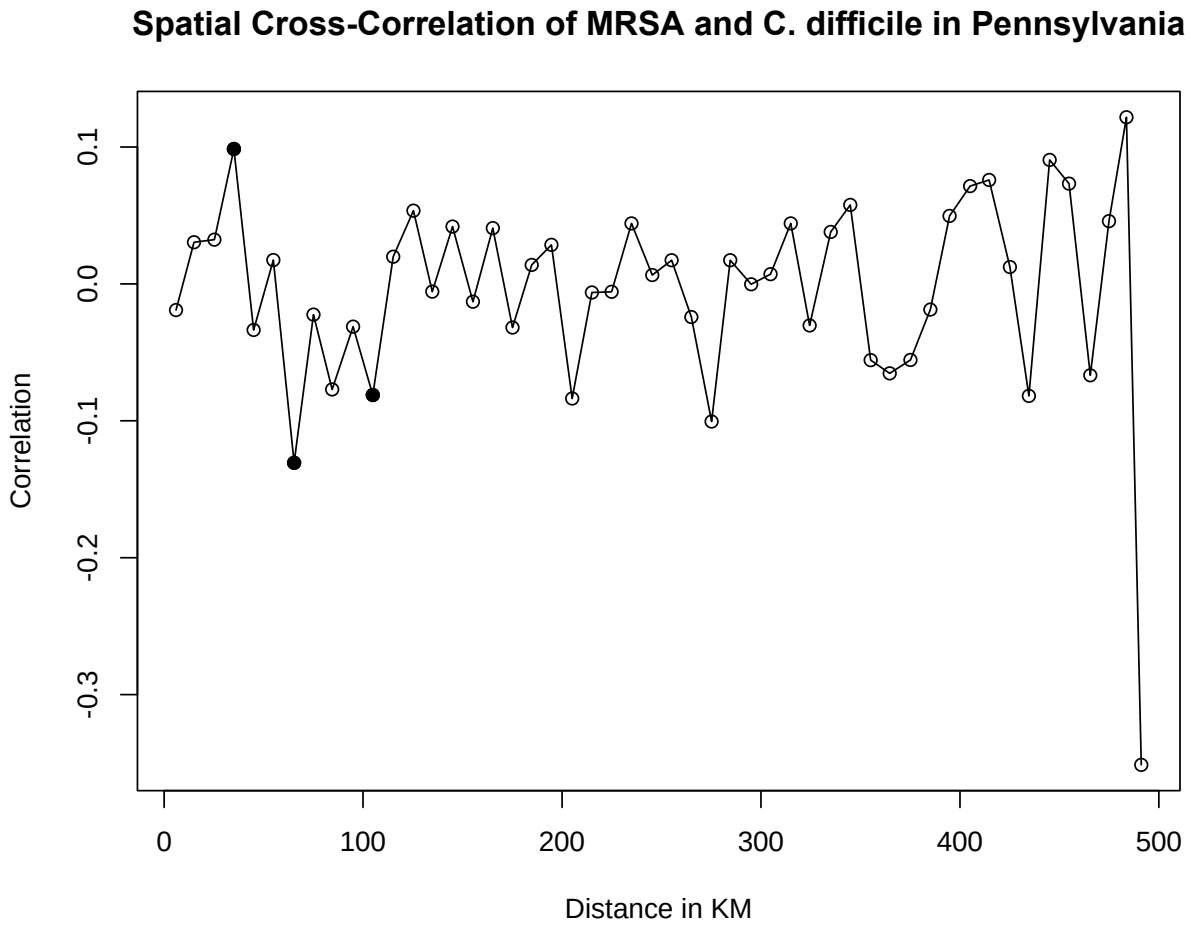

Figure 13. Spatial cross-correlation graph

In the graph above, each circle represents a distance class of the spatial autocorrelation analysis. Filled in circles represent statistically significant distance classes. This analysis is only interested in local cross-correlation, which is not present here. Thus while individual points may be statistically significant, they can be disregarded.

Chapter 4

Discussion

Analysis of Findings

This analysis investigates if there are any characteristics about a hospital's population or region that contribute to its incidence rates of HAIs. The data suggest that there are both hospital-level and regional-level influences in nosocomial infection rate, especially for MRSA.

Analysis of MRSA revealed that there are two statistically significant factors that positively contribute to infection rates: hospital patient days and county population density. As described in the introduction, patient days is used as a proxy for hospital size, thus larger hospitals are more likely to have higher infection rates. This statement is also reinforced by the logistic regression of presence/absence of MRSA in hospitals. Figures 5 and 7 show that hospitals with an infection present are generally larger than hospitals with infection absent; and that larger hospitals have a higher probability of an infection present.

The current literature is ambiguous on the effect of hospital size on infection rates. Our findings support both the aforementioned Mumford et al. study and another study of Australian hospitals by McLaws et al., who found that larger hospitals exhibit higher infection rates (28,43). However, a 2015 study with a slightly different sample found that larger hospitals had a lower incidence of post-surgical infections than smaller hospitals (44).

MRSA infection rates were also significantly affected by county population density. This suggests that hospitals in urban areas generally have higher infection rates than hospitals in less

urban, or rural areas. This statement is also supported by the hospital days finding. Exploratory regression of hospital days versus population density indicates that the two variables are significantly positively correlated (coefficient = 5.078, standard error = 1.321, t = 3.843, df = 154, $p < 0.001$). Logically, this makes sense, as hospitals in more densely populated areas need to serve larger communities; in addition, cities are often home to large tertiary and quaternary level care facilities, that perform more surgeries and treat more trauma cases.

Interestingly, none of the remaining county-level socioeconomic measures were correlated with MRSA SIR. One explanation could be that members of high socioeconomic status and low socioeconomic status receive the same level of care when it comes to nosocomial infection management. However, a 2014 study on hospital-acquired complications found that there were significant disparities in care for minorities and patients with Medicaid insurance (45). Thus, a more likely explanation is that county-level measures of socioeconomic data are just not effective at measuring individual patients, since there is a spectrum of wealthy to impoverished residents in each county. Both urban and rural areas have their share of affluent residents, as well as families who are at or below the poverty line (46).

Concerning MRSA, there is a significant component of SIR that can be attributed to spatial influences. Spatial autocorrelation analysis indicated that the infection rates of hospitals out to 43 kilometers apart are positively correlated with one another. After this point, there are random positive and negative autocorrelations, indicating that spatial influence is lost at greater distances.

LISA analysis with 25-kilometer neighborhoods showed that most of this spatial clustering occurs in urban environments. A majority of the significant interactions were found in the cities of Philadelphia and Pittsburgh, while most of the rural hospitals showed no spatial

interaction. However, it should be noted that for many rural hospitals, there were no other hospitals within 25 km for comparison. Thus, it is not possible to draw absolute conclusions on the lack of rural spatial clustering.

The existence of a spatial influence on infection risk is supported by findings in previous literature. In 2010, Simor et al. found that multiple strains of community-acquired MRSA have become endemic in hospital settings as well, supporting the idea that community spread can result in an increase in nosocomial incidence (47). In addition, genotyping analysis enabled a separate 2010 study to reconstruct spatiotemporal dispersion of hospital-acquired strains of MRSA across Central European hospitals, finding that strains of the ST225 clone of MRSA have spread spatially among hospitals within Central Europe (48).

Finally, the evidence of spatial influence on infection rates led us to use spatial regression to check for spurious conclusions due to spatial dependence. The spatial regression models confirmed that there are still significant interactions between MRSA SIR and hospital size, as well as population density, even when accounting for possible confounding spatial influence. The likelihood ratio test of the spatial linear regression and non-spatial parsimonious linear regression concluded that the two models are not significantly different, and we thus accept the findings of the multiple regression analysis.

Concerning *C. diff*, only one factor about a hospital's population was found to be significant: hospital patient days. Once again, this is a proxy for hospital size. Thus, figures 6 and 8 combine to illustrate that hospitals with an infection present generally are larger than hospitals with infection absent; and that larger hospitals have a higher probability of an infection present.

None of the socioeconomic data showed a strong correlation with the infection rates of *C. diff.* As stated before with MRSA, this is likely because the county-level measure is not precise enough to measure the specific care given to individuals of each class.

Unlike MRSA, there is no significant component of the movement of *C. diff* that can be attributed to spatial influences. The spatial autocorrelation analysis found no spatial dependence in *C. diff* according to Moran's I (41). Infection rates among hospitals that are close together are just as random as for hospitals that are far apart.

LISA analysis supports the findings of the spatial autocorrelation. Looking at figure 12, Philadelphia, the largest city in the state, lacks a single significant spatial hotspot, while a few other hot spots are dispersed across rural areas of the state. This is in stark contrast to the map for MRSA, where Philadelphia is covered in spatial hotspots. As stated before, in rural areas, there are some hospitals that lack another facility within their 25-km neighborhood, and thus we cannot draw absolute conclusions on all locations.

Unlike the case for MRSA, a search of the current literature on *C. diff* did not find any results that could either support or refute the findings of this analysis. Since there was no evidence of a regional influence on the infection rates of *C. diff*, spatial regression analysis was not performed.

Finally, analysis to compare the difference between the spread of the two infections yielded interesting results. While the rates of MRSA and *C. diff* were correlated by a regression analysis, they were not significantly correlated by spatial cross-correlation analysis. This means that within a facility, the SIR's of the two infections are correlated. However, when considering SIR from a regional perspective, this correlation is absent.

Perhaps this lack in regional correlation could be due to the fact that *C. diff* incidence does not show any evidence of significant spatial patterns. But there is another possible interpretation. This result could indicate that there are more differences in management of the same infection type at different hospitals (i.e. Hospital 1 MRSA vs. Hospital 2 MRSA) than management of different infection types at the same hospital (i.e. Hospital 1 MRSA vs. Hospital 1 *C. diff*). This notion is supported in the current literature. A 2016 study of New York hospitals showed significant differences of *C. diff* infection rates between hospitals, supporting the idea that there are large differences in the management of the same infection at different care centers (49). Thus, while some our findings indicate a greater-than-hospital level problem, this result indicates that there remain hospital-level factors that are contributing to nosocomial infection incidence.

Another observed difference between the two diseases is that there are more hospitals with a MRSA infection absent than hospitals with a *C. diff* infection absent. In addition, the probability of the presence of an infection increases at a lower threshold for *C. diff* than for MRSA. Thus, while MRSA shows more spatial dependence, hospitals display more susceptibility to a *C. diff* infection than a MRSA infection.

It is unknown why one of the diseases shows spatial dependence while the other does not. One hypothesis could be the differential spread of MRSA and *C. diff*. Diseases will spread most effectively where there is the highest population of susceptible individuals (50). Since MRSA colonizes in wounds and surgical sites, hospitals with large populations of these patients will be at the highest risk (14). Referencing the correlation from above, larger hospitals are found in more densely populated areas. Furthermore, cities are more often home to large tertiary and quaternary care centers that perform more surgeries and treat more trauma cases.

Limitations and Future Directions

There are several potential limitations of this study. The first is that the population analysis in this study relied on county-level data. Pooling such a large number of residents results in an average that is not representative of all of its neighborhoods, and could potentially overlook serious gaps in care between socioeconomic classes.

A further data limitation is that the Hospital Compare dataset counts infections and patient days for an entire year, then presents the totals all at once. Thus, while the hospital data had excellent spatial resolution, they lacked a temporal component. This analysis is thus a cross-sectional analysis.

One factor that hindered the spatial analysis was that there were many rural care centers analyzed that did not have other hospitals within a certain range to be compared to. Thus, both the spatial autocorrelation and LISA analyses may fail to include potential community effects in these rural areas simply because of a lack of other hospitals for comparison.

Along the same lines, there are many facilities that only have one or two other hospitals in their neighborhood for comparison. Using figure 12 as an example, in northeast Pennsylvania, there are two hospitals that show significant autocorrelation and infection rates above the mean SIR. However, each of those hospitals is the only facility in the other hospital's neighborhood. Thus, statistical analysis sees that all of the hospitals in their neighborhood have higher-than-mean infection rates, and labels this as a significant interaction. But, this "interaction" could just be two hospitals with high infection rates that happen to be within 25 km of each other. This example brings up the point that while this analysis finds many statistically significant results, not all of these may be practically important results.

Another limitation is that while this analysis considers infection rate data to compare regionally, it does not include infection management data (such as hand washing data) on a hospital-by-hospital basis that could account for facility differences (18,19,20). It would be very interesting to integrate such information into the findings about the differences between the two infections at the facility level.

Lastly, the analyses may have been somewhat limited by their sample. While the state of Pennsylvania has a large population and a broad spectrum from urban to rural environments, it is possible that the study could benefit from more urban examples than just the two cities of Philadelphia and Pittsburgh. Many of the conclusions drawn from LISA analysis were taken from the effects observed in Philadelphia, which is just one city.

There are a number of future questions that can be addressed following this research, many of which arise from the aforementioned limitations of this study. First, it would be beneficial to follow up this study with a more comprehensive analysis of a larger geographic area. Perhaps a region of the country, or the entire country could be the basis of a future analysis. Next, it would be enlightening to align the findings of this report with findings on the efficacy of treatment protocols on infection rates at individual facilities. The integration of these findings could be beneficial for future targeting of problem hospitals or problem regions.

Finally, it could be beneficial to build mathematical models for the dynamics of nosocomial infections. At this crucial intersection between community-acquired and healthcare-acquired disease, it could be beneficial to model this complex movement of disease to increase our ability to understand, manage, and predict outbreaks of nosocomial infection.

Conclusion

In conclusion, the spread of nosocomial infection has both hospital-level and regional-level inputs. The size of a hospital has a significant effect on that facility's infection rates for both MRSA and *C. diff*. At the same time, spatial dependence is also observed for MRSA. Finally, it has been shown that while MRSA and *C. diff* do not co-vary spatially, their infection rates correlate on a facility-by-facility basis, indicating the importance of individual hospital efforts. Together, these findings show that nosocomial infections are a multi-faceted problem, requiring both hospital and regional levels of management to reduce their risk.

Appendix A

CDC Infection Prevention Measures

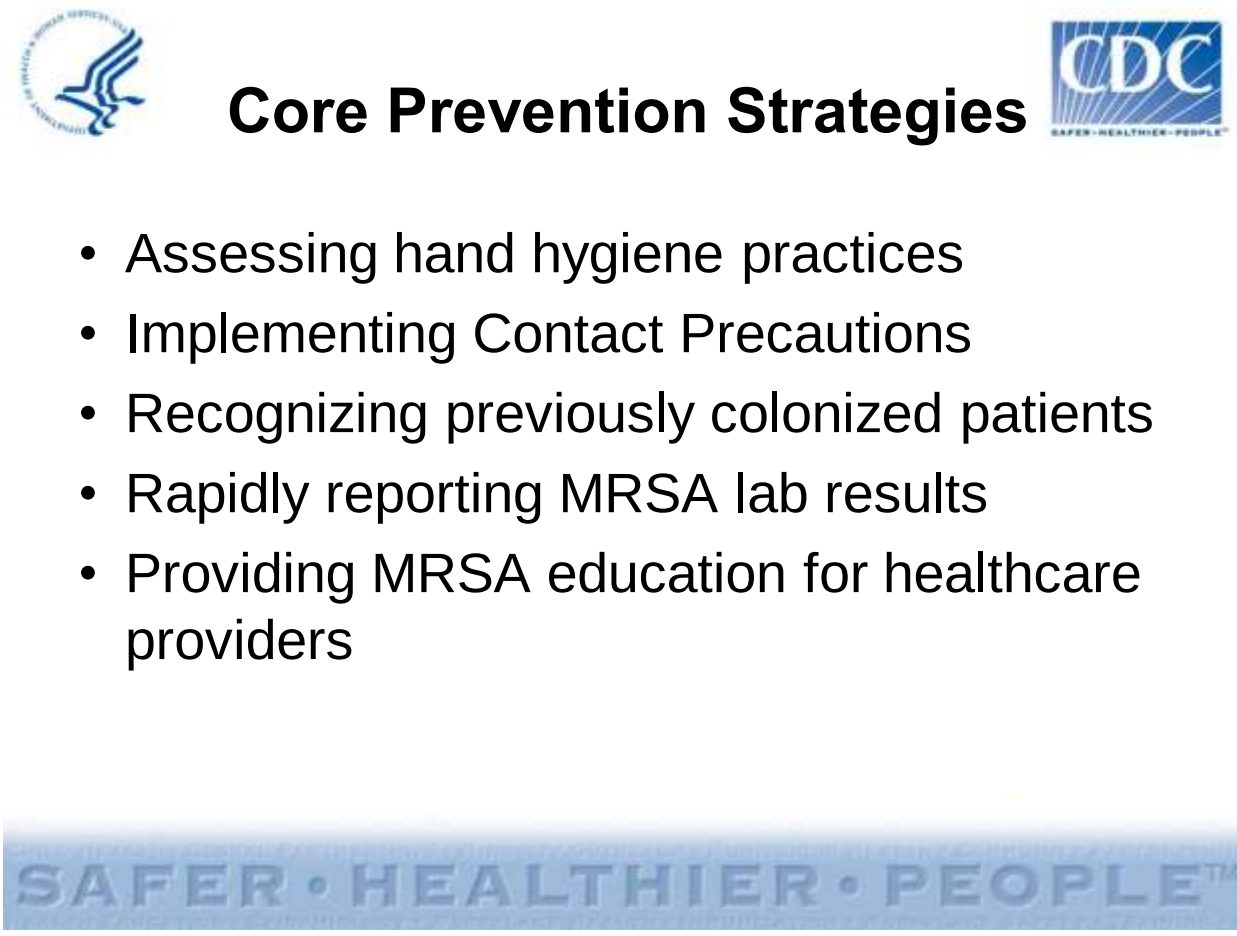

Figure 14. MRSA Prevention Techniques Published by the CDC (18)

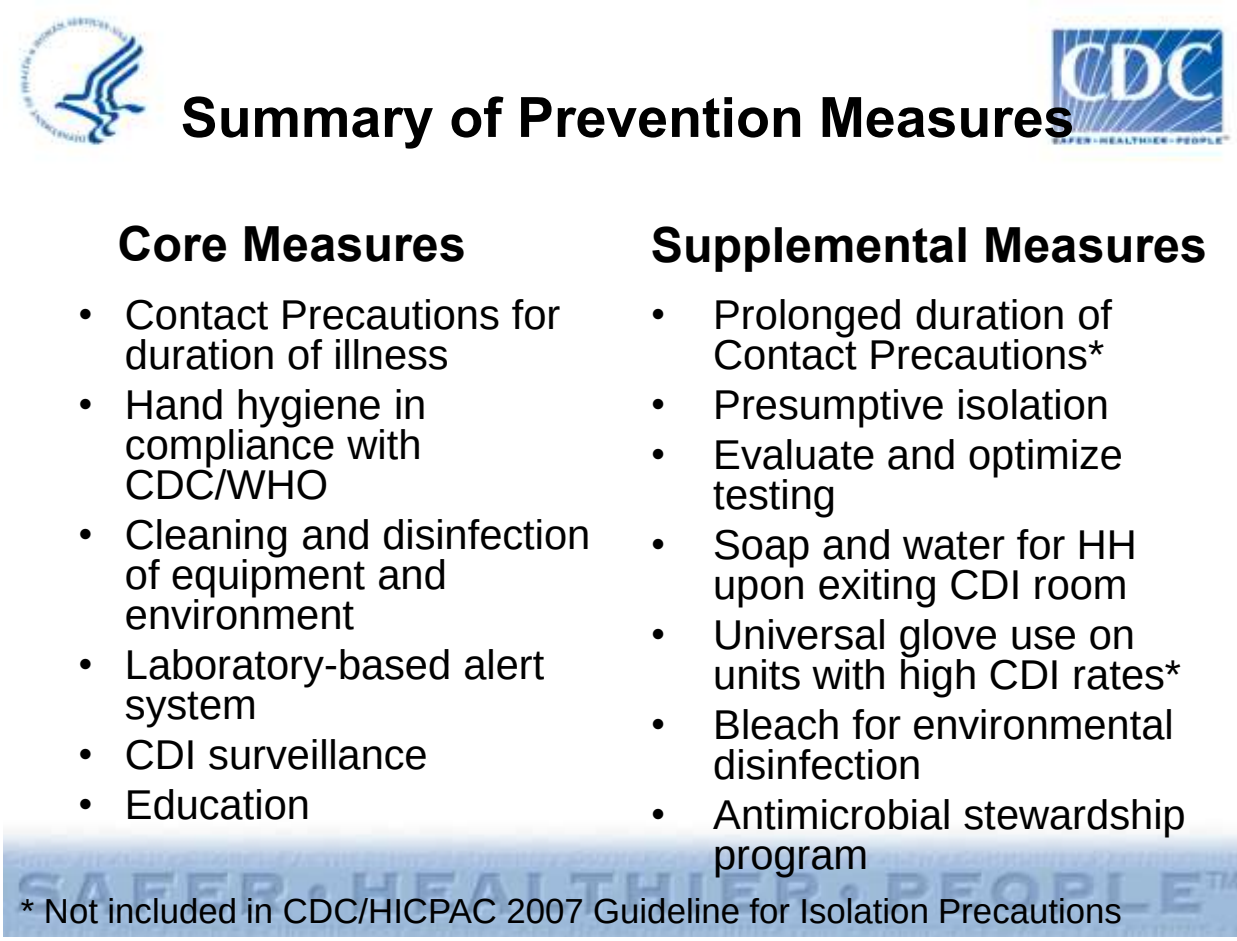

Figure 15. *C. diff* Prevention Techniques Published by the CDC (19)

Appendix B

R Summaries for Regression Analyses

Multiple regression analysis of MRSA standardized infection rates and socioeconomic correlates. (Question 1, Population Analysis)

```
summary(mrsa.SES)

##
## Call:
## lm(formula = M1000 ~ MRSAdays + Pop.Density + PerCap.Income +
##     Median.Hshld + PctPoverty + NumPoverty + PctJCAHO, data = PAdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.05140 -0.02258 -0.01681  0.01865  0.16992
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.679e-02  6.031e-02   0.278    0.781
## MRSAdays       3.129e-07  5.842e-08   5.356 3.19e-07 ***
## Pop.Density    3.035e-06  4.282e-06   0.709    0.480
## PerCap.Income  2.643e-07  8.214e-07   0.322    0.748
## Median.Hshld  -1.610e-08  1.149e-07  -0.140    0.889
## PctPoverty    -1.387e-04  2.267e-03  -0.061    0.951
## NumPoverty    -1.520e-07  8.398e-07  -0.181    0.857
## PctJCAHO      -2.356e-03  1.148e-02  -0.205    0.838
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.03686 on 148 degrees of freedom
##   (9 observations deleted due to missingness)
## Multiple R-squared:  0.2584, Adjusted R-squared:  0.2234
## F-statistic: 7.368 on 7 and 148 DF,  p-value: 1.382e-07
```

Multiple regression analysis of C. diff standardized infection rates and socioeconomic correlates. (Question 1, Population Analysis)

```
summary(Cdiff.SES)

##
## Call:
## lm(formula = C1000 ~ Cdiffdays + Pop.Density + PerCap.Income +
##     Median.Hshld + PctPoverty + NumPoverty + PctJCAHO, data = PAdata)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.60515 -0.21706 -0.06855  0.19817  1.49582
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    4.389e-02  5.569e-01   0.079    0.937
## Cdiffdays      2.709e-06  5.742e-07   4.718 5.53e-06 ***
## Pop.Density   -4.594e-05  3.958e-05  -1.161    0.248
## PerCap.Income  4.255e-06  7.615e-06   0.559    0.577
## Median.Hshld   6.885e-07  1.063e-06   0.648    0.518
## PctPoverty     1.244e-02  2.096e-02   0.593    0.554
## NumPoverty     6.441e-07  7.772e-06   0.083    0.934
## PctJCAHO       2.098e-03  1.060e-01   0.020    0.984
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3397 on 146 degrees of freedom
##   (11 observations deleted due to missingness)
## Multiple R-squared:  0.1696, Adjusted R-squared:  0.1297
## F-statistic: 4.259 on 7 and 146 DF,  p-value: 0.0002617
```

Stepwise elimination of socioeconomic correlates for MRSA using AIC. The resulting function is the parsimonious function. (Question 1, Population Analysis)

```
step(mrsa.SES)

## Start:  AIC=-1022.02
## M1000 ~ MRSAdays + Pop.Density + PerCap.Income + Median.Hshld +
##     PctPoverty + NumPoverty + PctJCAHO
##
##                 Df Sum of Sq     RSS      AIC
## - PctPoverty     1  0.000005 0.20107 -1024.02
```

```
## - Median.Hshld   1  0.000027 0.20109 -1024.00
## - NumPoverty     1  0.000045 0.20111 -1023.99
## - PctJCAHO       1  0.000057 0.20112 -1023.98
## - PerCap.Income  1  0.000141 0.20120 -1023.91
## - Pop.Density    1  0.000683 0.20175 -1023.49
## <none>                       0.20106 -1022.02
## - MRSAdays       1  0.038968 0.24003  -996.39
##
## Step:  AIC=-1024.02
## M1000 ~ MRSAdays + Pop.Density + PerCap.Income + Median.Hshld +
##     NumPoverty + PctJCAHO
##
##                 Df Sum of Sq     RSS      AIC
## - Median.Hshld   1  0.000027 0.20109 -1026.00
## - NumPoverty     1  0.000043 0.20111 -1025.99
## - PctJCAHO       1  0.000053 0.20112 -1025.98
## - PerCap.Income  1  0.000152 0.20122 -1025.90
## - Pop.Density    1  0.000847 0.20192 -1025.36
## <none>                       0.20107 -1024.02
## - MRSAdays       1  0.038972 0.24004  -998.38
##
## Step:  AIC=-1026
## M1000 ~ MRSAdays + Pop.Density + PerCap.Income + NumPoverty +
##     PctJCAHO
##
##                 Df Sum of Sq     RSS      AIC
## - NumPoverty     1  0.000021 0.20111 -1027.98
## - PctJCAHO       1  0.000071 0.20117 -1027.94
## - PerCap.Income  1  0.000125 0.20122 -1027.90
## <none>                       0.20109 -1026.00
## - Pop.Density    1  0.004945 0.20604 -1024.21
## - MRSAdays       1  0.039962 0.24106  -999.72
##
## Step:  AIC=-1027.98
## M1000 ~ MRSAdays + Pop.Density + PerCap.Income + PctJCAHO
##
##                 Df Sum of Sq     RSS     AIC
## - PctJCAHO       1  0.000082 0.20120 -1029.9
## - PerCap.Income  1  0.000227 0.20134 -1029.8
## <none>                       0.20111 -1028.0
## - Pop.Density    1  0.008987 0.21010 -1023.2
## - MRSAdays       1  0.041181 0.24230 -1000.9
##
## Step:  AIC=-1029.92
## M1000 ~ MRSAdays + Pop.Density + PerCap.Income
##
##                 Df Sum of Sq     RSS     AIC
## - PerCap.Income  1  0.000151 0.20135 -1031.8
## <none>                       0.20120 -1029.9
```

```
## - Pop.Density    1  0.009015 0.21021 -1025.1
## - MRSAdays       1  0.041349 0.24255 -1002.8
##
## Step:  AIC=-1031.8
## M1000 ~ MRSAdays + Pop.Density
##
##               Df Sum of Sq     RSS     AIC
## <none>                      0.20135 -1031.8
## - Pop.Density  1  0.009080 0.21043 -1026.9
## - MRSAdays     1  0.042919 0.24427 -1003.7

##
## Call:
## lm(formula = M1000 ~ MRSAdays + Pop.Density, data = PAdata)
##
## Coefficients:
## (Intercept)     MRSAdays  Pop.Density
##   1.656e-02    3.133e-07    2.474e-06
```

Stepwise elimination of socioeconomic correlates for C. diff using AIC. The resulting function is the parsimonious function. (Question 1, Population Analysis)

```
step(Cdiff.SES)
```

```
## Start:  AIC=-324.76
## C1000 ~ Cdiffdays + Pop.Density + PerCap.Income + Median.Hshld +
##     PctPoverty + NumPoverty + PctJCAHO
##
##                 Df Sum of Sq    RSS     AIC
## - PctJCAHO       1   0.00005 16.847 -326.76
## - NumPoverty     1   0.00079 16.848 -326.76
## - PerCap.Income  1   0.03602 16.883 -326.44
## - PctPoverty     1   0.04064 16.888 -326.39
## - Median.Hshld   1   0.04842 16.896 -326.32
## - Pop.Density    1   0.15544 17.003 -325.35
## <none>                       16.847 -324.76
## - Cdiffdays      1   2.56816 19.416 -304.92
##
## Step:  AIC=-326.76
## C1000 ~ Cdiffdays + Pop.Density + PerCap.Income + Median.Hshld +
##     PctPoverty + NumPoverty
##
##                 Df Sum of Sq    RSS     AIC
## - NumPoverty     1   0.00082 16.848 -328.76
## - PerCap.Income  1   0.03598 16.883 -328.44
```

```
## - PctPoverty     1   0.04159 16.889 -328.38
## - Median.Hshld   1   0.05090 16.898 -328.30
## - Pop.Density    1   0.15540 17.003 -327.35
## <none>                       16.847 -326.76
## - Cdiffdays      1   2.59621 19.444 -306.69
##
## Step:  AIC=-328.76
## C1000 ~ Cdiffdays + Pop.Density + PerCap.Income + Median.Hshld +
##     PctPoverty
##
##                 Df Sum of Sq    RSS     AIC
## - PctPoverty     1   0.05352 16.902 -330.27
## - Median.Hshld   1   0.05496 16.903 -330.26
## - PerCap.Income  1   0.07566 16.924 -330.07
## - Pop.Density    1   0.19151 17.040 -329.02
## <none>                       16.848 -328.76
## - Cdiffdays      1   2.59552 19.444 -308.69
##
## Step:  AIC=-330.27
## C1000 ~ Cdiffdays + Pop.Density + PerCap.Income + Median.Hshld
##
##                 Df Sum of Sq    RSS     AIC
## - PerCap.Income  1   0.02243 16.924 -332.06
## - Median.Hshld   1   0.10005 17.002 -331.36
## - Pop.Density    1   0.14658 17.048 -330.94
## <none>                       16.902 -330.27
## - Cdiffdays      1   2.60312 19.505 -310.21
##
## Step:  AIC=-332.06
## C1000 ~ Cdiffdays + Pop.Density + Median.Hshld
##
##                Df Sum of Sq    RSS     AIC
## - Median.Hshld  1   0.10247 17.027 -333.13
## - Pop.Density   1   0.14859 17.073 -332.72
## <none>                      16.924 -332.06
## - Cdiffdays     1   2.70431 19.628 -311.24
##
## Step:  AIC=-333.13
## C1000 ~ Cdiffdays + Pop.Density
##
##               Df Sum of Sq    RSS     AIC
## - Pop.Density  1    0.0809 17.108 -334.40
## <none>                     17.027 -333.13
## - Cdiffdays    1    3.1918 20.218 -308.68
##
## Step:  AIC=-334.4
## C1000 ~ Cdiffdays
##
##             Df Sum of Sq    RSS     AIC
```

```
## <none>                   17.108 -334.40
## - Cdiffdays  1    3.1798 20.287 -310.15

##
## Call:
## lm(formula = C1000 ~ Cdiffdays, data = PAdata)
##
## Coefficients:
## (Intercept)    Cdiffdays
##   4.087e-01    2.744e-06
```

Logistic regression of presence/absence data for MRSA. (Question 1, Population Analysis)

```
summary(mrsa.log)

##
## Call:
## glm(formula = catMRSA ~ MRSAdays, family = binomial, data = PAdata)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.26514  -0.62025   0.02053   0.55836   1.79771
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.237e+00  4.156e-01  -5.384 7.30e-08 ***
## MRSAdays     7.939e-05  1.373e-05   5.782 7.37e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 212.55  on 155  degrees of freedom
## Residual deviance: 123.06  on 154  degrees of freedom
##   (9 observations deleted due to missingness)
## AIC: 127.06
##
## Number of Fisher Scoring iterations: 7
```

Logistic regression of presence/absence data for C. diff. (Question 1, Population Analysis)

```
summary(mrsa.log)
```

```
##
## Call:
## glm(formula = catMRSA ~ MRSAdays, family = binomial, data = PAdata)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.26514  -0.62025   0.02053   0.55836   1.79771
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.237e+00  4.156e-01  -5.384 7.30e-08 ***
## MRSAdays     7.939e-05  1.373e-05   5.782 7.37e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 212.55  on 155  degrees of freedom
## Residual deviance: 123.06  on 154  degrees of freedom
##   (9 observations deleted due to missingness)
## AIC: 127.06
##
## Number of Fisher Scoring iterations: 7
```

Spatial regression of parsimonious model for MRSA risk. (Question 1, Population Analysis)

```
MRSAspatial=gls(M1000~MRSAdays+Pop.Density, data=PAdata,
corr=corSpatial(form=~Longitude+Latitude, type='exponential', nugget=T),
na.action = na.omit)
summary(MRSAspatial)
```

```
## Generalized least squares fit by REML
##   Model: M1000 ~ MRSAdays + Pop.Density
##   Data: PAdata
##         AIC       BIC   logLik
##   -519.1521 -500.9695 265.5761
##
## Correlation Structure: Exponential spatial correlation
##  Formula: ~Longitude + Latitude
##  Parameter estimate(s):
##        range       nugget
```

```
## 1.035606e-02 4.707759e-09
##
## Coefficients:
##                   Value   Std.Error  t-value p-value
## (Intercept) 0.015784163 0.003869455 4.079170  0.0001
## MRSAdays    0.000000330 0.000000047 6.985607  0.0000
## Pop.Density 0.000002862 0.000001022 2.801155  0.0057
##
##  Correlation:
##             (Intr) MRSAdy
## MRSAdays    -0.490
## Pop.Density -0.297 -0.181
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -1.5354382 -0.6005975 -0.4708485  0.5116159  4.6461416
##
## Residual standard error: 0.03637621
## Degrees of freedom: 156 total; 153 residual
```

Non-spatial regression of parsimonious model for MRSA risk. (Question 1, Population Analysis)

```
MRSAnspatial=gls(M1000~MRSAdays+Pop.Density, data=PAdata, na.action =
na.omit)
summary(MRSAnspatial)
```

```
## Generalized least squares fit by REML
##   Model: M1000 ~ MRSAdays + Pop.Density
##   Data: PAdata
##         AIC       BIC   logLik
##   -519.6146 -507.4929 263.8073
##
## Coefficients:
##                   Value   Std.Error  t-value p-value
## (Intercept) 0.016559561 0.003992517 4.147650  0.0001
## MRSAdays    0.000000313 0.000000055 5.710786  0.0000
## Pop.Density 0.000002474 0.000000942 2.626718  0.0095
##
##  Correlation:
##             (Intr) MRSAdy
## MRSAdays    -0.568
## Pop.Density -0.200 -0.296
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -1.4287815 -0.5923411 -0.4759785  0.5280542  4.6425237
```

```
##
## Residual standard error: 0.03627672
## Degrees of freedom: 156 total; 153 residual
```

Likelihood ratio testing of spatial and non-spatial parsimonious models for MRSA risk.

(Question 1, Population Analysis)

```
anova(MRSAspatial, MRSAnspatial)
```

```
##              Model df       AIC       BIC   logLik   Test  L.Ratio p-value
## MRSAspatial      1  6 -519.1521 -500.9695 265.5761
## MRSAnspatial     2  4 -519.6146 -507.4929 263.8073 1 vs 2 3.537486  0.1705
```

Spatial autocorrelation analysis of MRSA standardized infection rates. (Question 2, Spatial Analysis)

```
correlogM
```

```
## $n
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
## 162 252 308 319 301 324 301 379 346 376 339 370 328 348 327 400 321 244
##  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
## 273 253 232 237 246 205 227 193 232 216 218 199 229 219 199 230 214 205
##  37  38  39  40  41  42  43  44  45  46  47  48  49  50
## 262 290 267 300 236 277 190 157 101  75  69  75  16   3
##
## $mean.of.class
##          1          2          3          4          5          6
##   5.900423  14.973336  25.220960  35.095489  45.031338  54.907822
##          7          8          9         10         11         12
##  65.268720  75.060862  84.461315  95.008819 104.875276 115.197803
##         13         14         15         16         17         18
## 125.306052 134.824067 144.940366 155.144256 165.111019 175.075015
##         19         20         21         22         23         24
## 184.783039 194.731517 205.080271 215.018058 224.902043 234.936022
##         25         26         27         28         29         30
## 245.343340 255.031323 264.959576 275.141206 284.546951 295.088461
##         31         32         33         34         35         36
## 304.806148 314.987433 324.484232 335.066313 344.836722 355.167321
##         37         38         39         40         41         42
```

```
## 364.636954 375.095745 385.050135 394.747207 405.062815 414.637548
##         43         44         45         46         47         48
## 425.091214 434.688302 445.056994 454.834671 465.487473 474.914357
##         49         50
## 483.663817 491.152659
##
## $correlation
##             1             2             3             4             5
##  0.2712317128  0.1634323411  0.1064009896  0.0357309327 -0.1534054294
##             6             7             8             9            10
## -0.0725722093 -0.0621976306 -0.0591666005  0.0237550626 -0.0326439373
##            11            12            13            14            15
## -0.0884982902  0.0886710512 -0.0160718590  0.0677471282  0.0697946582
##            16            17            18            19            20
##  0.0412276739 -0.0082451689  0.0639690367  0.0816036957  0.0077997117
##            21            22            23            24            25
## -0.1417360424 -0.0857296483  0.0006716193 -0.0055673545 -0.1599512483
##            26            27            28            29            30
##  0.0540161122 -0.1452750150 -0.2157804676  0.1476860182  0.1003522280
##            31            32            33            34            35
##  0.0618132714  0.1009671717 -0.1372093572  0.0317810647 -0.1842588225
##            36            37            38            39            40
## -0.0280425318  0.0090140299 -0.1729501872  0.0279362639  0.0254064300
##            41            42            43            44            45
##  0.0786899276  0.1344760750 -0.1929206916 -0.2111434335 -0.0585278169
##            46            47            48            49            50
## -0.0433273586 -0.0074454931  0.0330680784  0.1212375386 -0.3256685186
##
## $x.intercept
## (Intercept)
##    43.15429
##
## $p
##  [1] 0.003996004 0.007992008 0.020979021 0.216783217 0.007992008
##  [6] 0.108891109 0.138861139 0.138861139 0.254745255 0.284715285
## [11] 0.058941059 0.021978022 0.431568432 0.075924076 0.072927073
## [16] 0.137862138 0.489510490 0.118881119 0.075924076 0.388611389
## [21] 0.017982018 0.098901099 0.431568432 0.486513487 0.010989011
## [26] 0.200799201 0.017982018 0.001998002 0.012987013 0.066933067
## [31] 0.138861139 0.054945055 0.031968032 0.282717283 0.004995005
## [36] 0.394605395 0.399600400 0.002997003 0.266733267 0.286713287
## [41] 0.088911089 0.009990010 0.015984016 0.008991009 0.258741259
## [46] 0.347652348 0.464535465 0.324675325 0.233766234 0.215784216
##
## $call
## [1] "correlog(x = Mdata$Longitude, y = Mdata$Latitude, z = Mdata$M1000, "
## [2] "    increment = 10, resamp = 1000, latlon = TRUE, na.rm = TRUE, "
## [3] "    quiet = FALSE)"
##
```

```
## attr(,"class")
## [1] "correlog"
```

Spatial autocorrelation analysis of C. diff standardized infection rates. (Question 2, Spatial Analysis)

```
correlogC

## $n
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
## 159 252 304 315 298 320 293 367 336 365 334 360 318 334 303 371 297 241
##  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
## 265 250 230 235 242 199 223 188 227 215 214 196 225 216 197 224 210 197
##  37  38  39  40  41  42  43  44  45  46  47  48  49  50
## 247 275 255 290 234 274 190 157 101  75  69  75  16   3
##
## $mean.of.class
##          1          2          3          4          5          6
##   5.965552  14.973336  25.259526  35.081028  45.023220  54.909635
##          7          8          9         10         11         12
##  65.325323  75.025045  84.443063  95.004474 104.883251 115.172357
##         13         14         15         16         17         18
## 125.297345 134.788100 144.968312 155.185487 165.184548 175.064883
##         19         20         21         22         23         24
## 184.726323 194.693819 205.065515 215.025299 224.943499 234.938926
##         25         26         27         28         29         30
## 245.317460 255.034876 264.945081 275.124793 284.570077 295.095568
##         31         32         33         34         35         36
## 304.814663 314.954609 324.460401 335.028789 344.842139 355.147635
##         37         38         39         40         41         42
## 364.686991 375.108562 385.111500 394.732028 405.068975 414.649213
##         43         44         45         46         47         48
## 425.091214 434.688302 445.056994 454.834671 465.487473 474.914357
##         49         50
## 483.663817 491.152659
##
## $correlation
##             1             2             3             4             5
## -0.0218701223 -0.0003924976  0.0350618491 -0.0480983103  0.0059286685
##             6             7             8             9            10
##  0.0890557219 -0.0574824122 -0.0046019189 -0.0706297496 -0.0039640007
##            11            12            13            14            15
## -0.0566972340  0.0354586203 -0.0951707765  0.0447615110 -0.0354207243
##            16            17            18            19            20
```

```
## -0.0234589922  0.0555498994  0.0088216581 -0.0055427418  0.0011624498
##            21            22            23            24            25
## -0.0944605733 -0.0057090629 -0.0780452247  0.0763786055 -0.0946513770
##            26            27            28            29            30
##  0.0312664901  0.0335085756  0.0027841209 -0.0803232507 -0.0861823573
##            31            32            33            34            35
##  0.0732100299  0.1140084991 -0.0081214460  0.0588107352  0.0304454543
##            36            37            38            39            40
## -0.0507387249  0.0810886309 -0.1189234487 -0.0591332992 -0.0335892542
##            41            42            43            44            45
##  0.0577904977  0.0139766805  0.0307170052 -0.0669815683  0.0989944860
##            46            47            48            49            50
## -0.0441319875  0.0534651816 -0.0015872490  0.1432507865 -0.3163472077
##
## $x.intercept
## [1] 0
##
## $p
##  [1] 0.44055944 0.44255744 0.21578422 0.21278721 0.39160839 0.05594406
##  [7] 0.17982018 0.45654346 0.11688312 0.48251748 0.15184815 0.17082917
## [13] 0.04995005 0.16583417 0.28971029 0.35364635 0.12087912 0.38261738
## [19] 0.47952048 0.46853147 0.08391608 0.48751249 0.12287712 0.11488511
## [25] 0.08391608 0.29470529 0.26873127 0.44955045 0.12987013 0.11288711
## [31] 0.12087912 0.02897103 0.48351648 0.15384615 0.27572428 0.26073926
## [37] 0.07192807 0.02697303 0.19080919 0.31068931 0.15684316 0.33066933
## [43] 0.28271728 0.20479520 0.12487512 0.34865135 0.23976024 0.49350649
## [49] 0.20579421 0.17482517
##
## $call
## [1] "correlog(x = Cdata$Longitude, y = Cdata$Latitude, z = Cdata$C1000, "
## [2] "    increment = 10, resamp = 1000, latlon = TRUE, na.rm = TRUE, "
## [3] "    quiet = FALSE)"
##
## attr(,"class")
## [1] "correlog"
```

Spatial cross-correlation analysis of MRSA and C. diff standardized infection rates. (Question 3, Disease Cross-Comparison)

```
correlogCompare
```

```
## $n
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
## 318 504 608 630 596 640 586 734 672 730 668 720 636 668 606 742 594 482
```

```
##  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
## 530 500 460 470 484 398 446 376 454 430 428 392 450 432 394 448 420 394
##  37  38  39  40  41  42  43  44  45  46  47  48  49  50
## 494 550 510 580 468 548 380 314 202 150 138 150  32   6
##
## $mean.of.class
##          1          2          3          4          5          6
##   5.965552  14.973336  25.259526  35.081028  45.023220  54.909635
##          7          8          9         10         11         12
##  65.325323  75.025045  84.443063  95.004474 104.883251 115.172357
##         13         14         15         16         17         18
## 125.297345 134.788100 144.968312 155.185487 165.184548 175.064883
##         19         20         21         22         23         24
## 184.726323 194.693819 205.065515 215.025299 224.943499 234.938926
##         25         26         27         28         29         30
## 245.317460 255.034876 264.945081 275.124793 284.570077 295.095568
##         31         32         33         34         35         36
## 304.814663 314.954609 324.460401 335.028789 344.842139 355.147635
##         37         38         39         40         41         42
## 364.686991 375.108562 385.111500 394.732028 405.068975 414.649213
##         43         44         45         46         47         48
## 425.091214 434.688302 445.056994 454.834671 465.487473 474.914357
##         49         50
## 483.663817 491.152659
##
## $correlation
##             1             2             3             4             5
## -0.0191898482  0.0304015167  0.0323760171  0.0985722010 -0.0336235143
##             6             7             8             9            10
##  0.0174394221 -0.1307166642 -0.0224892565 -0.0771160207 -0.0312074347
##            11            12            13            14            15
## -0.0811817407  0.0198667139  0.0534805556 -0.0056699125  0.0418221794
##            16            17            18            19            20
## -0.0129636606  0.0407030964 -0.0318491898  0.0137919970  0.0285191671
##            21            22            23            24            25
## -0.0837440483 -0.0063351026 -0.0057038189  0.0441966681  0.0064065536
##            26            27            28            29            30
##  0.0173483495 -0.0241735792 -0.1004781760  0.0172747278 -0.0002524151
##            31            32            33            34            35
##  0.0071256835  0.0441910408 -0.0304007445  0.0378847435  0.0576359464
##            36            37            38            39            40
## -0.0556517790 -0.0653136187 -0.0555154006 -0.0188150666  0.0496649269
##            41            42            43            44            45
##  0.0713558371  0.0759124864  0.0122954446 -0.0817618218  0.0905544671
##            46            47            48            49            50
##  0.0732451744 -0.0667099706  0.0458703371  0.1216689435 -0.3512587836
##
## $x.intercept
## [1] 0
```

```
##
## $p
##  [1] 0.390609391 0.216783217 0.194805195 0.008991009 0.205794206
##  [6] 0.336663337 0.002997003 0.266733267 0.035964036 0.208791209
## [11] 0.019980020 0.273726274 0.084915085 0.465534466 0.117882118
## [16] 0.375624376 0.161838162 0.245754246 0.387612388 0.248751249
## [21] 0.036963037 0.465534466 0.480519481 0.185814186 0.442557443
## [26] 0.369630370 0.327672328 0.025974026 0.327672328 0.479520480
## [31] 0.419580420 0.174825175 0.298701299 0.205794206 0.123876124
## [36] 0.148851149 0.068931069 0.107892108 0.367632368 0.100899101
## [41] 0.057942058 0.027972028 0.386613387 0.083916084 0.101898102
## [46] 0.176823177 0.183816184 0.223776224 0.192807193 0.173826174
##
## $call
## [1] "correlog(x = CompareData$Longitude, y = CompareData$Latitude, "
## [2] "    z = CompareData$M1000, w = CompareData$C1000, increment = 10, "
## [3] "    resamp = 1000, latlon = TRUE, na.rm = TRUE, quiet = FALSE)"
##
## $corr0
## [1] 0.2897498
##
## attr(,"class")
## [1] "correlog"
```

Linear regression model of MRSA and C. diff standardized infection rates. (Question 3, Disease Cross-Comparison)

```
summary(compare.reg)
```

```
##
## Call:
## lm(formula = M1000 ~ C1000, data = CompareData)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.06012 -0.02847 -0.01265  0.02345  0.15055
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 0.018294   0.005704   3.207 0.001635 **
## C1000       0.032865   0.008805   3.732 0.000267 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
## Residual standard error: 0.03966 on 152 degrees of freedom
## Multiple R-squared:  0.08395,    Adjusted R-squared:  0.07793
## F-statistic: 13.93 on 1 and 152 DF,  p-value: 0.0002675
```

BIBLIOGRAPHY

1. Smith, D. L., Dushoff, J., Perencevich, E. N., Harris, A. D., & Levin, S. A. (2004). Persistent colonization and the spread of antibiotic resistance in nosocomial pathogens: resistance is a regional problem. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(10), 3709-3714.
2. U.S. Department of Health & Human Services. Centers for Medicare & Medicaid Services. "Hospital Compare—A Quality Tool for Adults, Including People with Medicare" [accessed on March 29, 2016]. Retrieved from http://www.medicare.gov/hospitalcompare/search.html
3. US Department of Health and Human Services, Health Resources and Services Administration. *Area Health Resources Files. 2013-2014.* Rockville, MD: US Department of Health and Human Services, Health Resources and Services Administration, Bureau of Health Workforce; 2014. Retrieved from http://sodapop.pop.psu.edu/arf1.html
4. Epperson, B.K. (1993) Recent advances in correlation studies of spatial patterns of genetic variation. *Evolutionary Biology*, 27, 95-155.
5. Anselin, L. (1995) Local indicators of spatial association - LISA. *Geographical Analysis*, 27:93-115.
6. XING, G. X., LU, W. L., & LI, Q. (2012). Prevalence of nosocomial infections. Chinese Journal of Nosocomiology, 24, 033.

7. Scott, R. D. (2009). The direct medical costs of healthcare-associated infections in US hospitals and the benefits of prevention.
8. Lichtenstein, A. (1935). Nosocomial Infections in Children's Hospitals and Institutions. Our Means for combating these Infections. *Acta Paediatrica,17*(S1), 36-49.
9. Wright, J. (1940). Nosocomial infections in children's wards. *Journal of Hygiene*, *40*(06), 647-672.
10. Struelens, M. J. (1998). The epidemiology of antimicrobial resistance in hospital acquired infections: problems and possible solutions. BMJ: British Medical Journal, 317(7159), 652.
11. Garner, J. S., Jarvis, W. R., Emori, T. G., Horan, T. C., & Hughes, J. M. (1988). CDC definitions for nosocomial infections, 1988. *American journal of infection control*, *16*(3), 128-140.
12. Weinstein, R. A., Gaynes, R., Edwards, J. R., & National Nosocomial Infections Surveillance System. (2005). Overview of nosocomial infections caused by gram-negative bacilli. *Clinical infectious diseases*, *41*(6), 848-854.
13. Inweregbu, K., Dave, J., & Pittard, A. (2005). Nosocomial infections. *Continuing Education in Anaesthesia, Critical Care & Pain*, *5*(1), 14-17.
14. Centers for Disease Control and Prevention. (2014). *General Information About MRSA in Healthcare Settings*. Retrieved from http://www.cdc.gov/MRSA/healthcare/index.html

15. Centers for Disease Control and Prevention. (2013). *MRSA Information for Inpatient Clinicians and Administrators*. Retrieved from http://www.cdc.gov/MRSA/healthcare/clinicians/index.html
16. Centers for Disease Control and Prevention. (2015). *Clostridium difficile Infection Information for Patients*. Retrieved from http://www.cdc.gov/hai/organisms/cdiff/Cdiff-patient.html
17. Centers for Disease Control and Prevention. (2012). *Frequently Asked Questions about Clostridium difficile for Healthcare Providers*. Retrieved from http://www.cdc.gov/HAI/organisms/cdiff/Cdiff_faqs_HCP.html
18. Jernigan, J., & Kallen, A. (2010). *Methicillin-Resistant Staphylococcus aureus (MRSA) Infections* [Activity C: ELC Prevention Collaboratives]. Retrieved from Centers for Disease Control and Prevention website: http://www.cdc.gov/
19. Gould, C., & McDonald, C. (2009). *Clostridium difficile (CDI) Infections Toolkit*[Activity C: ELC Prevention Collaboratives]. Retrieved from Centers for Disease Control and Prevention website: http://www.cdc.gov/
20. Larson, E. (1988). A causal link between handwashing and risk of infection? Examination of the evidence. *Infection Control*, *9*(01), 28-36.
21. Centers for Disease Control and Prevention. (2014). *Precautions to Prevent the Spread of MRSA.* Retrieved from http://www.cdc.gov/mrsa/healthcare/clinicians/precautions.html
22. The Office of Disease Prevention and Health Promotion. (2013). *The National Action Plan to Prevent Health Care-Associated Infections: Road Map to Elimination.* Retrieved from http://health.gov/hcq/prevent-hai-action-plan.asp

23. Centers for Disease Control and Prevention. (2014) *Monitoring Infections.* Retrieved from http://www.cdc.gov/hai/HHS-HAI-Toolkit/tracking/monitoring-infections.html

24. Munier, E., Bénet, T., Régis, C., Lina, B., Morfin, F., & Vanhems, P. (2016). Hospitalisation in double-occupancy rooms and the risk of hospital-acquired influenza: a prospective cohort study. *Clinical Microbiology and Infection*.

25. Salkever, D. S. (1975). Economic class and differential access to care: comparisons among health care systems. *International journal of health services*, *5*(3), 373-395.

26. Merwin, E., Snyder, A., & Katz, E. (2006). Differential access to quality rural healthcare: Professional and policy challenges. *Family & community health*,*29*(3), 186-194.

27. The Joint Commission. (2016). *Facts about Hospital Accreditation*. Retrieved from http://www.jointcommission.org/accreditation/accreditation_main.aspx

28. Mumford, V., Reeve, R., Greenfield, D., Forde, K., Westbrook, J., & Braithwaite, J. (2015). Is accreditation linked to hospital infection rates? A 4-year, data linkage study of Staphylococcus aureus rates and accreditation scores in 77 Australian acute hospitals. *International Journal for Quality in Health Care*, *27*(6), 479-485.

29. Viboud, C., Bjørnstad, O. N., Smith, D. L., Simonsen, L., Miller, M. A., & Grenfell, B. T. (2006). Synchrony, waves, and spatial hierarchies in the spread of influenza. *science*, *312*(5772), 447-451.

30. Gangnon, R. E., & Clayton, M. K. (2000). Bayesian detection and modeling of spatial disease clustering. *Biometrics*, *56*(3), 922-935.

31. Austin, D. J., Bonten, M. J., Weinstein, R. A., Slaughter, S., & Anderson, R. M. (1999). Vancomycin-resistant enterococci in intensive-care hospital settings: transmission dynamics, persistence, and the impact of infection control programs. *Proceedings of the National Academy of Sciences*, *96*(12), 6908-6913.
32. Gorak, E. J., Yamada, S. M., & Brown, J. D. (1999). Community-acquired methicillin-resistant Staphylococcus aureus in hospitalized adults and children without known risk factors. *Clinical Infectious Diseases*, *29*(4), 797-800.
33. Kyne, L., Merry, C., O'Connell, B., Keane, C., & O'Neill, D. (1998). Community-acquired Clostridium difficile infection. *Journal of Infection*, *36*(3), 287-288.
34. Morse, S. S. (2001). *Factors in the emergence of infectious diseases* (pp. 8-26). Palgrave Macmillan UK.
35. R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/
36. Venables, W. N. and Ripley, B. D. (2002) Modern Applied Statistics with S. New York: Springer (4th ed).
37. Hastie, T. J. and Pregibon, D. (1992) Generalized linear models. Chapter 6 of Statistical Models in S eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.
38. Pinheiro J, Bates D, DebRoy S, Sarkar D and R Core Team (2016). nlme: Linear and Nonlinear Mixed Effects Models_. R package version 3.1-125, http://CRAN.R-project.org/package=nlme

39. Chambers, J. M. and Hastie, T. J. (1992) Statistical Models in S, Wadsworth & Brooks/Cole.

40. Ottar N. Bjornstad (2015). ncf: Spatial Nonparametric Covariance Functions. R package version 1.1-6. http://CRAN.R-project.org/package=ncf

41. Moran, P. A. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17-23.

42. Wilkinson, G. N. and Rogers, C. E. (1973) Symbolic descriptions of factorial models for analysis of variance. Applied Statistics, 22, 392–9.

43. McLaws, M. L., Gold, J., King, K., Irwig, L. M., & Berry, G. (1987). The prevalence of nosocomial and community-acquired infections in Australian hospitals. *The Medical Journal of Australia*, *149*(11-12), 582-590.

44. Allareddy, V., Das, A., Lee, M. K., Nalliah, R. P., Rampa, S., Allareddy, V., & Rotta, A. T. (2015). Prevalence, predictors, and outcomes of methicillin-resistant Staphylococcus aureus infections in patients undergoing major surgical procedures in the United States: a population-based study. *The American Journal of Surgery*, *210*(1), 59-67.

45. Wen, T., Attenello, F. J., He, S., Cen, Y., Kim-Tenser, M. A., Sanossian, N., ... & Mack, W. J. (2014). Racial and socioeconomic disparities in incidence of hospital-acquired complications following cerebrovascular procedures.*Neurosurgery*, *75*(1), 43-50.

46. Ravallion, M. (2007). Urban Poverty. *Finance & Development*, *44*(3), Retrieved from http://www.imf.org/external/pubs/ft/fandd/2007/09/ravalli.htm

47. Simor, A. E., Gilbert, N. L., Gravel, D., Mulvey, M. R., Bryce, E., Loeb, M., ... & Campbell, J. (2010). Methicillin-resistant Staphylococcus aureus colonization or infection in Canada: national surveillance and changing epidemiology, 1995–2007. *Infection Control & Hospital Epidemiology*, *31*(04), 348-356.

48. Nübel, U., Dordel, J., Kurt, K., Strommenger, B., Westh, H., Shukla, S. K., ... & Balloux, F. (2010). A timescale for evolution, population expansion, and spatial spread of an emerging clone of methicillin-resistant Staphylococcus aureus. *PLoS Pathog*, *6*(4), e1000855.

49. Aquina, C. T., Probst, C. P., Becerra, A. Z., Hensley, B. J., Iannuzzi, J. C., Noyes, K., ... & Fleming, F. J. (2016). High Variability in Nosocomial Clostridium difficile Infection Rates Across Hospitals After Colorectal Resection. *Diseases of the Colon & Rectum*, *59*(4), 323-331.

50. Kermack, W. O., & McKendrick, A. G. (1927, August). A contribution to the mathematical theory of epidemics. In *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences* (Vol. 115, No. 772, pp. 700-721). The Royal Society.

Academic Vita
of
Wilton Prescott Smith

wiltonpsmith@gmail.com

Education

The Pennsylvania State University – Class of 2016
University Park, PA
Schreyer Honors College
Eberly College of Science
B.S. Biology – Neuroscience option
Minors: Psychology, Neuroscience
Cedar Cliff High School – Class of 2012
Camp Hill, PA

Relevant Coursework

The Pennsylvania State University
University Park, PA
Biological Sciences: Medical Embryology, Mammalian Physiology, Histology, Neuroscience, Neurobiology, Honors Molecular and Cell Biology, Honors Biology: Function and Development of Organisms, Honors Biology: Populations and Communities, Honors General Biology
Physical Sciences: Honors General Biochemistry I, General Biochemistry II, Organic Chemistry I, II, and Lab, Physics: Mechanic, Physics: Electricity and Magnetism, General Chemistry I, II, and Lab, Single-Variable Calculus, Intermediate Statistics
Social and Behavioral Sciences: Clinical Neuropsychology, Physiological Psychology, Neurological Bases of Human Behavior, Cognitive Psychology, Research Methods in Psychology, Introduction to Psychology, Introduction to Sociology

Research Experience

Undergraduate Honors Thesis – 2015-2016
University Park, PA
Quantitative spatial analysis of nosocomial Methicillin-Resistant *Staphylococcus aureus (MRSA)* and *Clostridium difficile* in Pennsylvania hospitals
Poster Presentation – Spring 2016
University Park, PA
Presented a poster on honors thesis research at the Health Policy and Administration Senior Showcase

Teaching Experience

Teaching Assistant – Fall 2014 – Spring 2016
University Park, PA

Spring 2016 – Biology 240W Laboratory
Function and Development of Organisms – Freshman Research Initiative
Fall 2015 – Biology 110S Laboratory
Basic Concepts and Biodiversity – Freshman Research Initiative
Fall 2014 – Biology 230M Laboratory
Honors Molecular and Cell Biology

Work Experience

Research Analyst – Orthopedic Institute of Pennsylvania – 2015
Camp Hill, PA
Summer Intern - Member's 1st Federal Credit Union – 2014
Mechanicsburg, PA
Food Expeditor – Arooga's Grille House and Sports Bar – 2013-2014
Camp Hill, PA

Volunteer Service

Atlas Benefiting the Penn State Dance Marathon – 2012-2016
University Park, PA

Volunteer for the Penn State Dance Marathon, providing financial and emotional support to children and families battling pediatric cancer. Served as Donor and Alumni Relations Chair (2014-2015) and Donor and Alumni Relations Administrative Captain (2013-2014)

Penn State Lion Scout – 2013-2016
University Park, PA

Student liaison and volunteer for the Penn State Office of Undergraduate Admissions. Served as Alumni Relations Chair (2015-2016)

Schreyer Honors Orientation – 2013-2015
University Park, PA

Mentor and Orientation Leader for the three-day orientation serving incoming Schreyer scholars

Penn State Dance Marathon – 2013-2014
University Park, PA

Finance Committee Member for the world's largest student-run philanthropy

Awards and Honors

Schreyer Honors College Academic Excellence Scholarship – 2012-2016
University Park, PA