

# Spatiotemporal Visualization and Modeling of Nosocomial Infections

---

A Thesis

Presented to  
the faculty of the School of Engineering and Applied Science  
University of Virginia

---

in partial fulfillment  
of the requirements for the degree

Master of Science

by

Ketki Vilankar

December

2016

APPROVAL SHEET

The thesis  
is submitted in partial fulfillment of the requirements  
for the degree of  
Master of Science

K. K. Vilankar

AUTHOR

The thesis has been read and approved by the examining committee:

Laura Barnes

Advisor

Donald Brown

Amy Mathers

Hyojung Kang

Accepted for the School of Engineering and Applied Science:

James H. Ayer

Dean, School of Engineering and Applied Science

December

2016

# Spatiotemporal Visualization and Modeling of Nosocomial Infections

A Thesis

Presented to

The Faculty of the  
School of Engineering and Applied Science  
University of Virginia

In Partial Fulfillment  
of the Requirements for the Degree

Master of Science in Systems and Information Engineering

by

Ketki Vilankar

September 17, 2016

# APPROVAL SHEET

The Thesis

is submitted in partial fulfillments of the requirements

for the degree of Master of Science

---

(Author)

The Thesis has been read and approved by the examining committee:

---

Dr. Laura Barnes (Signature) (Advisor)

---

Dr. Donald Brown (Signature) (Committee Member)

---

Dr. Amy Mathers (Signature) (Committee Member)

---

Dr. Hyojung Kang (Signature) (Committee Member)

Accepted for the School of Engineering and Applied Science:

---

Dean, School of Engineering and Applied Science

September

2016

# ABSTRACT

The Centers for Disease Control and Prevention (CDC) estimates that more than two million people contract antibiotic-resistant infections every year, and at least 23,000 die as a result of these infections in the U.S. alone. Traditionally, tracking hospital outbreaks with drug-resistant pathogens focuses on transmission chains of infected or colonized patients as the reservoir for organisms to be transferred to new patients via healthcare workers, but it has become increasingly recognized that non-patient reservoirs within the hospital may play a larger role than previously realized in acting as a niche for the transmission of drug-resistant pathogens. Non-patient sources for pathogen acquisition may require incorporating environmental culture data into existing transmission models. However, the number of risk factors, potential interactions and inherent complexity of the data continue to increase, and thus, exploratory analysis is required to aid in knowledge discovery. Interactive visualization of these data over space and time enables exploration and hypothesis generation to better inform transmission models. This thesis presents an interactive visualization system for the analysis of spatiotemporal environmental and patient data to aid in understanding nosocomial infection. Interactive dashboards allow users to view patient movement through hospital environments while overlaying multivariate environmental microbiological data as it evolves over time. Furthermore, a multivariate logistic regression model is constructed to understand the factors associated with sink contamination. The results show that temporal factors, including the presence of infected patients in the past 14 days and use of interventions in the past 7 days, and spatial factors, including the presence of infected patients in adjacent rooms and the presence of contaminated sinks in adjacent rooms, are significant factors in sink contamination.

# ACKNOWLEDGEMENTS

I would like to thank individuals and organizations that have assisted me throughout this research and my time at University of Virginia. Thank you Prof. Laura Barnes for giving me the opportunity to work on this thesis. I am grateful to her for her continuous support, insightful thoughts and patient guidance throughout this work. Her high expectations pushed me beyond limits known to me and were instrumental in successful completion of this work. Next I wish to thank my committee chair, Dr. Brown for his valuable statistical guidance and constructive comments. Most importantly he served as a role model and inspiration. I also want to thank Dr. Amy Mathers for her contribution of time, ideas and domain expertise throughout this study. I thank Dr. Hyojung Kang for always being there when I needed advice on my thesis and appreciate her thoughtful comments and feedback on literature review.

I also want to thank following people for their helpful discussions and suggestions: John Ainsworth, Josh Baumann, Ian Clark, Chao Dai, Marika Grabowski, Julia Lensing, Jennifer Lobo, Wendy Repass and Antoine Rigaut. Thanks to Ken Scully for providing data. Special thanks to Coulter Foundation and University of Virginia for funding support.

# CONTENTS

<b>APPROVAL SHEET</b>	<b>i</b>
<b>Abstract</b>	<b>ii</b>
<b>Acknowledgements</b>	<b>iii</b>
<b>Contents</b>	<b>iv</b>
<b>List of Tables</b>	<b>vi</b>
<b>List of Figures</b>	<b>vii</b>
<b>List of Abbreviations</b>	<b>viii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Literature Review</b>	<b>3</b>
2.1 CRE Transmission and Risk Factors . . . . .	3
2.1.1 Patient Risk Factors . . . . .	3
2.1.2 Environmental Risk Factors . . . . .	3
2.1.3 Logistic Regression Modeling in Infection Transmission . . . . .	4
2.2 Spatiotemporal Data Visualization . . . . .	6
<b>3 Problem, Objective, and Contribution</b>	<b>11</b>
3.1 Problem Description . . . . .	11
3.2 Objective . . . . .	11
3.3 Metrics . . . . .	12
3.3.1 Spatiotemporal Visualization Metrics . . . . .	12
3.3.2 Infection Risk Modeling . . . . .	12
3.4 Contribution . . . . .	12
<b>4 Data Collection and Transformation</b>	<b>13</b>
4.1 Environmental Data . . . . .	13
4.2 Patient Data . . . . .	15
4.3 Geospatial data . . . . .	16
4.4 Data Transformations . . . . .	17

<b>5</b>	<b>Spatiotemporal Visualization</b>	<b>20</b>
5.1	Visualization Design Requirements . . . . .	20
5.2	Visualization Interface Components . . . . .	21
5.3	Visualization Dashboards . . . . .	22
<b>6</b>	<b>Modeling Methodology and Results</b>	<b>26</b>
6.1	Variable Selection . . . . .	26
6.2	Results . . . . .	28
6.2.1	Logistic Regression Modeling . . . . .	28
6.2.2	Random Forest Model . . . . .	31
<b>7</b>	<b>Conclusion, Limitations and Future Work</b>	<b>35</b>
7.0.1	Conclusion . . . . .	35
7.0.2	Limitations and Future Work . . . . .	37



## LIST OF TABLES

6.1	Variable Selection for Sink Risk Modeling . . . . .	27
6.2	Comparison of Models . . . . .	28
6.3	Parameter Estimates and Significance . . . . .	29
6.4	Odds Ratio with Confidence Intervals . . . . .	29
6.5	Analysis of Deviance Table . . . . .	30
6.6	Logistic Regression Model Confusion Matrix . . . . .	31
6.7	Logistic Regression Model Results . . . . .	31
6.8	Random Forest Confusion Matrix . . . . .	32
6.9	Logistic Regression and Random Forest Model Metrics . . . . .	33

# LIST OF FIGURES

2.1	Charles Minard's map of Napoleon's disastrous Russian Campaign of 1812. . . . .	6
2.2	Adaptation of Dr. John Snow's 1854 London cholera epidemic map. (Image from: <a href="http://www.mystorybook.com/books/28272">http://www.mystorybook.com/books/28272</a> ) . . . . .	7
4.1	Environmental Sampling Summary . . . . .	14
4.2	Imputation example for room sinks 5187 and 5195 using midpoint method	15
4.3	Positive patient distribution by floors and units . . . . .	16
4.4	Geospatial mapping example for the 3 <sup>rd</sup> floor . . . . .	17
5.1	Spatiotemporal Visualization Dashboard . . . . .	23
5.2	Patient stay and same room sink . . . . .	24
5.3	Environment history dashboard showing sampling and imputation . .	25
6.1	ROC Curve for Logistic Regression Model . . . . .	32
6.2	Random Forest - Variable Importance . . . . .	33
6.3	ROC Curves for Logistic Regression and Random forest models . . .	34

# LIST OF ABBREVIATIONS

CDC	Centers for Disease Control and Prevention
CRE	Carbapenem-Resistant Enterobacteriaceae
KPC	<i>Klebsiella pneumoniae</i> Carbapenemase
E. Coli	<i>Escherichia coli</i>
ICP	Infection Control Practitioners
STBICU	Surgical and Trauma Burn Intensive Care Unit
MICU	Medical Intensive Care Unit
ED	Emergency Department
PERI	Perianesthesia Facility
IMAGEJ	Image Processing and Analysis in Java
GIS	Geographical Information System
P	Same Room Patient Status
P <sub>lag7</sub>	Positive Patient in last 7 days
P <sub>lag14</sub>	Positive Patient in last 14 days
P <sub>lag21</sub>	Positive Patient in last 21 days
P <sub>lag28</sub>	Positive Patient in last 28 days
P <sub>1</sub>	Patient Status in adjacent rooms that share common plumbing
P <sub>2</sub>	Patient Status in adjacent rooms that do not share common plumbing
S <sub>1</sub>	Sink Status in adjacent rooms that share common plumbing
S <sub>2</sub>	Sink Status in adjacent rooms that do not share common plumbing
I <sub>3</sub>	Intervention in the last 3 days
I <sub>5</sub>	Intervention in the last 5 days
I <sub>7</sub>	Intervention in the last 7 days
S <sub>lag30</sub>	Sink status in last 30 days

# CHAPTER 1

## INTRODUCTION

Antimicrobial-resistant pathogens have been on the rise for several years. Carbapenem-resistant Enterobacteriaceae (CRE) is a gram-negative bacteria that has shown resistance to antibiotics. In 2014, the Centers for Disease Control and Prevention (CDC) in its study on antibiotic-resistant threats in the United States reported that approximately 140,000 infections occur in the United States that are associated with Enterobacteriaceae, and 9,300 of these infections are caused by CRE. This study also reported that approximately 600 deaths are caused each year due to infections by the two most common types of bacteria carbapenem-resistant *Klebsiella pneumoniae* and *Escherichia coli* (E. coli) [1]. Since 2012, there have been over 200 hospitals in 48 states that have encountered KPC-producing CRE, including the deadly infection in Bakersfield, California [2].

CRE transmission from patient-to-patient through the hands of healthcare personnel is the main transmission route in healthcare settings. However, the application of the traditional patient-to-patient route of nosocomial infection transmission does not fully account for the observed cases [3]. Environmental reservoirs, particularly sinks, may play a major role in transmission [4–6]. It is therefore important to understand the risk factors associated with the spread of this infection and to develop an approach to prevent it.

This study is conducted at a major medical center in the United States that has been tracking a low level trend of CRE infections for the past eight years. There have been timely recordings of cultures from hospital rooms from various environmental sites, including patient sinks, patient bathrooms, hoppers, toilets and others. Appropriate intervention procedures have occasionally been implemented based on the environmental sampling results, but these interventions have not proven to be very effective as a long-term solution. Thus, there is a need to have appropriate tools and models in place that can track CRE infection in a timely manner and provide risk factors that could prevent the occurrence of outbreaks. Spatiotemporal data visualization can serve as an effective tool in exploring complex datasets that are otherwise extremely cumbersome to analyze. The aim of this work is to determine potential elements that are responsible for sink contamination.

## CHAPTER 2

### LITERATURE REVIEW

#### 2.1 CRE Transmission and Risk Factors

##### 2.1.1 Patient Risk Factors

The transmission of CRE infection occurs when a non-colonized person comes into direct contact with an infected or colonized patient, through intermediate carriers such as healthcare workers or through contact with contaminated environmental reservoirs such as sinks and toilets, among others. Healthcare settings are the most common locations for the spread of CRE [2]. Previous studies [7, 8] from the US and Israel have shown that the primary risk factors for patient acquisition include ICU stays, long-term hospitalization, transplantation and antibiotics. However, these risk factors did not completely explain infection transmission. Evidence indicating that environmental reservoirs are a source of infection transmission were found in previous studies [4–6].

##### 2.1.2 Environmental Risk Factors

A study from Spain [4] described an outbreak due to multidrug-resistant *Klebsiella oxytoca* in an ICU where damp environmental reservoirs were linked to bacterial transmission. Samples collected from sinks drainpipes and traps showed that only one

storage sink, which had its drainpipes connected to two other sinks, was found to be positive. The connecting drainpipes were also found to be positive. Furthermore, this study showed that the outbreak was completely eradicated after replacing the horizontal drainage system that connected the two impacted sinks. In conclusion, this study stated that wet environmental reservoirs should be considered when strictly applied traditional control measures are not efficacious.

A study from France [6] found that sinks were frequently contaminated in ICUs as a result of their use in disposing of patient bodily fluids and were a potential source of extended-spectrum beta lactamase-producing *Enterobacteriaceae* (ESBLE), thus increasing risk in the environment of patients as a consequence of the splash-back effect. Recent research from Wolf et al. [5] demonstrated that sinks acted as a source of infection by verifying that the ESBLEs recovered from patients were identical to those that had been previously recovered from sinks. The outbreak described in a Colombian [9] study found that the likely cause of the infections was the improper design of sinks; in particular, the joints of the sinks to the walls were not sealed, leading to facilitation of colonization.

### **2.1.3 Logistic Regression Modeling in Infection Transmission**

A previous study [10] that considered spatial variables as factors in explaining the efficacy of infection control measures in preventing the transmission of multidrug-resistant tuberculosis used exposure to specific infected patients in a logistic regression model. This study found that exposure distance is a significant predictor of nosocomial transmission. The logistic regression approach has also been used in some studies [11, 12] to identify patient risk factors for KPC transmission in hospital settings. Research by Papadimitriou-Olivgeris et al. [11] focused on patient

characteristics, diagnosis and procedures and determined that prior ICU stay, duration of previous hospitalization, diagnosis of chronic obstructive pulmonary disease, carbapenem administration and beta-lactamase administration were significant risk factors. Similarly, Tuon et al. [12] also considered procedures such as mechanical ventilation and found that urinary catheter devices and central venous catheter devices were significant risk factors, along with advanced age and antibiotic exposure to ciprofloxacin.

Previous work performed using the same dataset as our study but for a different time period (Jan-April 2014) focused on modeling the nosocomial transmission of carbapenem-resistant bacteria in 2014 [13] using logistic regression and random forest models to determine important risk factors for CRE transmission. Both models showed that distance to the infected room was one of the significant predictors. One of the models showed that the proximity to sinks is important in predicting infection. However, the model was constructed using a limited data and time range and requires further research into the role of sinks in infection transmission. The most recent work performed in this area included the development of imputation methods [14] that would allow extrapolating the time series of sink status to resolve any gaps between two time periods. This work also found that the cumulative presence of positive patients in the same room as a sink, distance from the bed to sink and sink design are significant predictors of sink positivity.

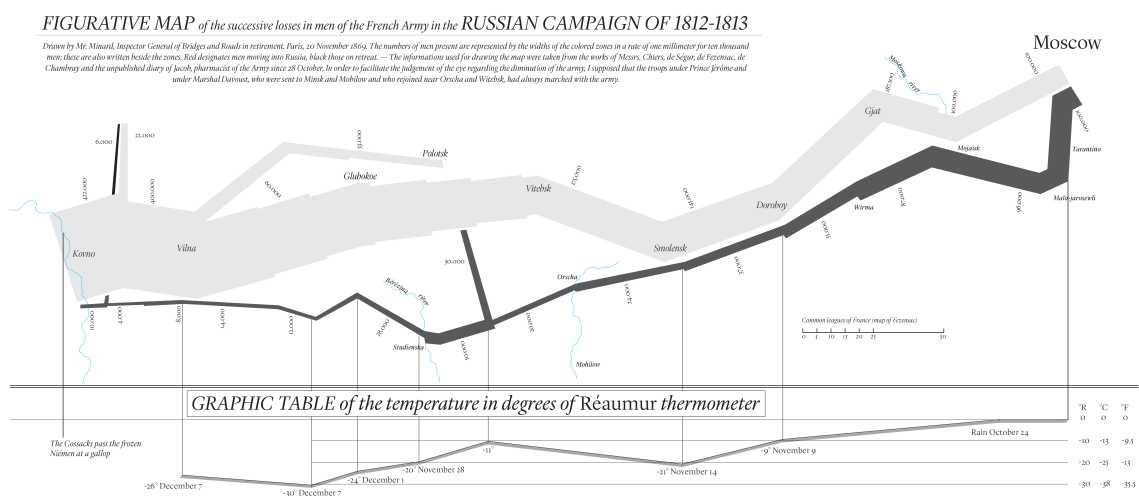
Studies [4–6, 9] have demonstrated that sinks play a role in infection transmission but did not highlight additional environmental risk factors responsible for sink contamination. This thesis improves on the understanding of sink contamination by highlighting important variables. This study is similar to [13, 14] in terms of the modeling approach used, but it differs in the level of spatiotemporal variables from



environmental data. In addition to examining the presence of positive patients in the same room, this study also considers the status of neighboring rooms and sinks as potential risk factors.

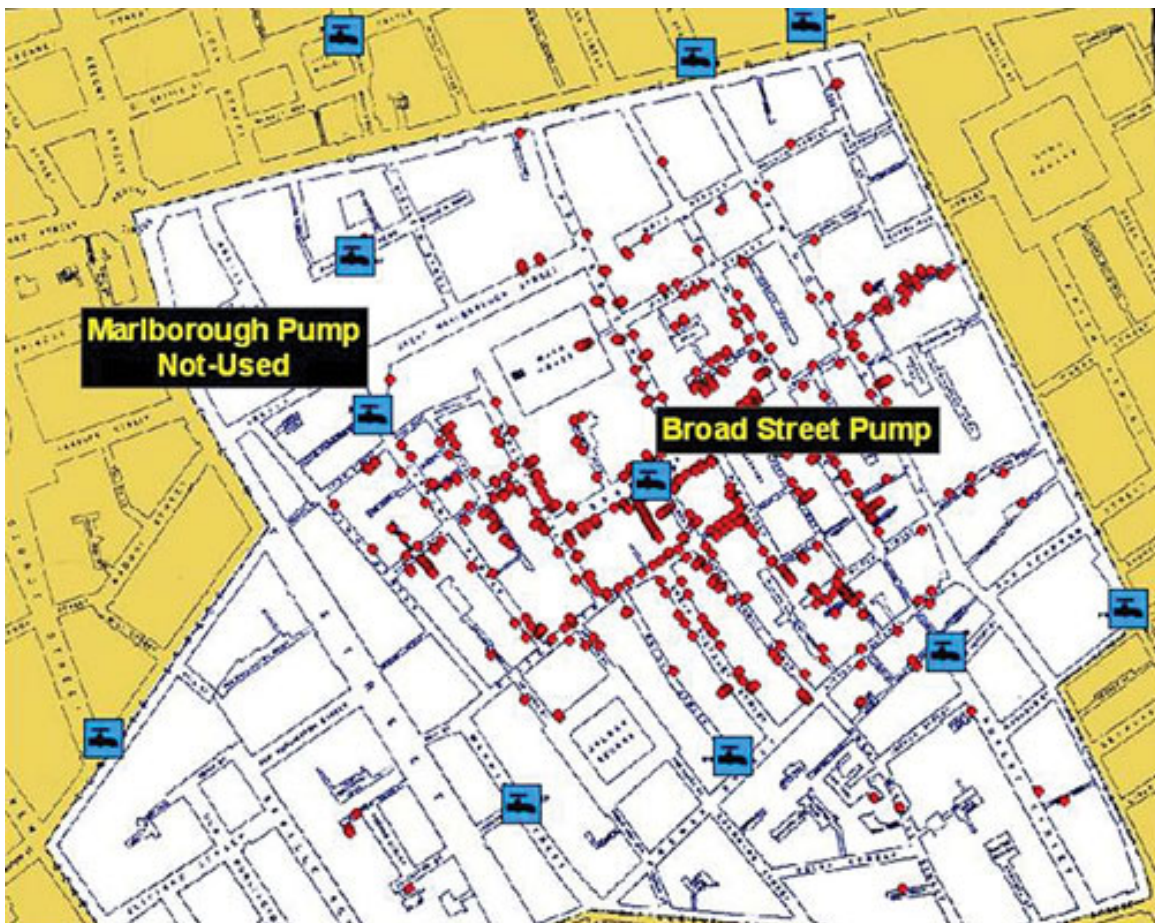
## 2.2 Spatiotemporal Data Visualization

Visualization is important when analyzing spatiotemporal data because it can help humans discover complex relationships in the data. Formally, spatiotemporal data are defined as high-dimensional records of data in which different dimensions can be classified into three components: spatial (geographical coordinates), temporal (time stamps) and attributes (patient status) [15]. The objective behind analyzing this type of data is to study a process in the context of both space and time. To effectively explore multivariate datasets and to be able to arrive at the potential list of variables that may not be otherwise easily identifiable, visualization techniques help by using the power of the human visual system to decode visual representations of data for analysis.



**Figure 2.1:** Charles Minard’s map of Napoleon’s disastrous Russian Campaign of 1812.

There is a long history of using visualization techniques in various fields, including in epidemiological or medical studies. One of the classic examples of visualization techniques comes from Charles Minard's representation of the fate of Napoleon's army on a campaign against Russia, as shown in Figure 2.1. The visual is noted for its representation of six types of data in two dimensions - number of Napoleon's troops (denoted by thick band), distance traveled, its location relative to dates, latitude and longitude, direction of the army's movement and temperature at various points in time. The visualization was developed primarily with the objective of communication or storytelling.



**Figure 2.2:** Adaptation of Dr. John Snow's 1854 London cholera epidemic map.

(Image from: <http://www.mystorybook.com/books/28272>)

Another notable early work on visualization came from John Snow, who is con-

sidered by many to be the pioneer in disease mapping. He depicted the 1854 cholera epidemic in London’s Broad Street region, as shown in Figure 2.2. In this map, deaths due to cholera are denoted by dots, and the area’s eleven water pumps are labeled. Snow noted a spatial clustering of cases around one particular water pump on Broad Street. This outbreak was a result of a contaminated handle on the water pump, which was later removed; consequently, the outbreak quickly subsided. Thus, the purpose of his visualization was to be able to visualize the outbreak and determine its cause.

With the advent of new technologies, diverse data collections and complex epidemiology challenges, recent research works have focused on advanced interactive visualizations [16–20]. Wongsuphasawat et al. [16] presented a Lifeflow visualization that summarizes point-based sequence of events and time gaps between events in an aggregated view to identify patterns. They further illustrated the usefulness of Lifeflow through a case study of patient transfers, in which they examined the sequence of all patients coming to the hospital, particularly those for medical emergencies. Common patterns of patient transfers (arrival-ER-floor-discharge) were easily identified through the visualization. Excluding common patterns from visualization allowed them to focus on irregular patterns, which included patients that left without being seen or patients that transferred from a higher level of care to a lower level of care and back to a higher level of care (arrival-ER-floor-ICU), thus highlighting areas where quality control were needed. Monroe et al. [17] further improved the visualization by including interval-based events to provide additional information on point-based events. They showed that the sequences of patient transfers, such as stroke  $\rightarrow$  admitted to hospital  $\rightarrow$  diagnosed, can be further enhanced by displaying additional interval information during the event, such as started medication A  $\rightarrow$  stroke  $\rightarrow$  ended medication A, in the dashboard. This helped to primarily focus on drug-related studies,

including prescription administration and medical interaction.

Geographic information system (GIS) is the common approach employed by studies [18–20] for designing spatiotemporal visualization. One of the recent studies on this subject was conducted by Ali et al. [18], in which they constructed a visual analytics decision support system to detect spatiotemporal disease outbreaks and manage epidemic responses. The visualizations utilized choropleth and heat maps to show instances of infectious disease prevalence across regions on a map (GIS). Kistemann et al. [19] investigated a *Salmonella* outbreak in Germany at a university hospital where the source of infection could not be detected using microbiological methods. GIS helped to trace all the events that occurred prior to the outbreak. For instance, they were able to identify how many cases of infection occurred at specific locations, on which days and times, paths taken by food carriers and narrow down the specific item on the lunch menu (i.e., vanilla pudding) that caused the infection. AvRuskin et al. [20] developed STIS (space time information system) to analyze and visualize data from a bladder cancer case control study with the objective of identifying the range of factors contributing to bladder cancer incidence in Michigan. They reconstructed individual arsenic exposure by incorporating spatiotemporal data, including residential mobility and drinking water habits, into a visualization that provided dynamic views in the form of tables, graphs and maps that show changing environmental variables through time. The map view utilizes GIS to display animated views of arsenic-emitting industries founded, operating and going out of business through time along with individuals changing places of residence, providing a historical view of areas with high arsenic concentrations and incidence of cases. The visualization thus highlights the environmental variables that play a role in the incidence of bladder cancer cases, which are further analyzed for statistical significance.

While previous studies [16, 17] focus on highlighting and summarizing temporal events, our thesis goes further by adding a spatial element to the visualization, such as hospital units, floors and patient rooms. Additionally, previous studies [18–20] utilize GIS for depicting spatiotemporal visualization, whereas this thesis uses detailed views of floor plans to project patient and environment status.

## CHAPTER 3

### PROBLEM, OBJECTIVE, AND CONTRIBUTION

#### 3.1 Problem Description

There have been ongoing multi-species CRE outbreaks at a major medical center in the United States. The conventional method of infection transmission through patient to patient contact does not fully account for all the cases of patient acquisition. Samples from sinks (drain and p-trap) have been found to be positive with CRE. Despite aggressive interventions, the bacterial strain continued to reappear, making it unclear whether other environmental or patient-specific risk factors could be involved in continued sink contamination.

#### 3.2 Objective

The objective of this research is to identify significant risk factors for understanding sink contamination through interactive visualization incorporating multidimensional patient and environmental data. We aim to accomplish this goal through the use of predictive models that incorporate spatial and temporal risk factors.

## **3.3 Metrics**

### **3.3.1 Spatiotemporal Visualization Metrics**

Evaluating a visualization is a difficult task, but there are a few guidelines that one can follow based on the research of [21]. One of the methods that the authors suggest is a case study approach, where the visualization is tested on a new dataset that is distinct from the one that was used for visualization. A similar method is employed to test visualization based on the most recent data and to determine whether it could perform all the animations and display dashboards without failing.

### **3.3.2 Infection Risk Modeling**

Models will be evaluated using area under the receiver operating curve (AUC), accuracy, sensitivity (number of positive cases correctly identified) and specificity (number of negative cases correctly identified).

## **3.4 Contribution**

This dissertation has analytically investigated risk factors responsible for contamination of non-patient reservoirs in hospital settings using interactive spatiotemporal visualization. We construct a model that shows impact of surrounding environmental risk factors on sink contamination. The results not only contribute to the open medical literature by demonstrating interaction of spatial and temporal risk factors on sink contamination but also inform infection control professionals at the hospital to design intervention strategies appropriately. Additionally, the visualization approach used in this thesis to explore heterogeneous datasets can be extended to examine any other nosocomial infection and inform transmission models.

## CHAPTER 4

# DATA COLLECTION AND TRANSFORMATION

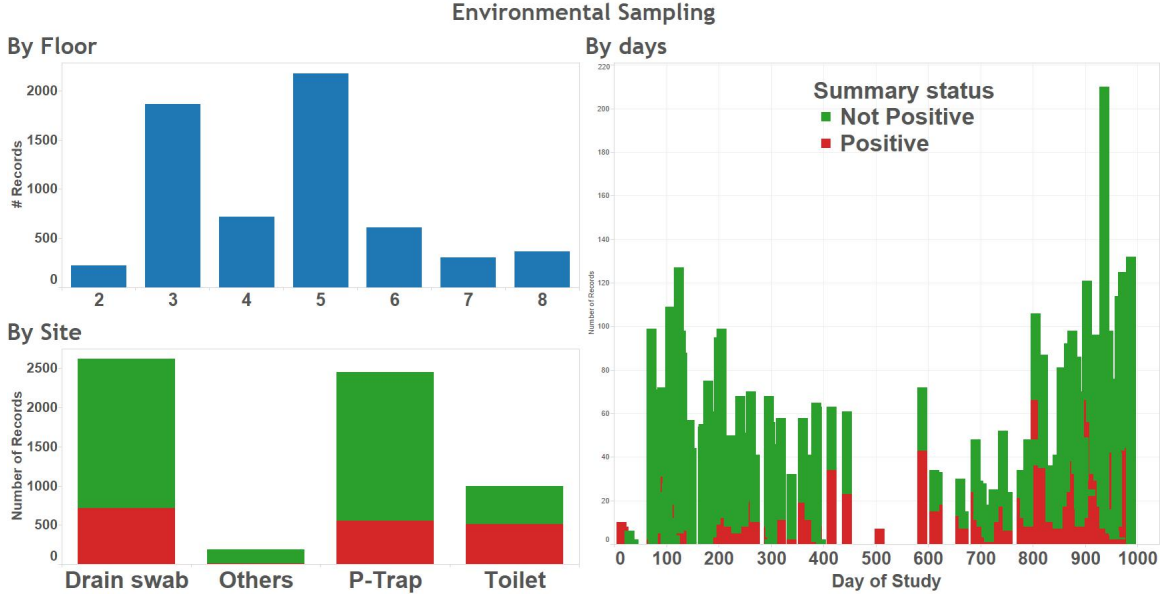
The data sources for this study can be broadly divided into three categories: environmental data, patient data, and geospatial data. The time period for which the data were extracted ranged from Sep 2013 to Feb 2016. For the purpose of our analysis, we selected the minimum observation interval to be one day for both environmental and patient datasets because that is the smallest interval for which records exist.

### 4.1 Environmental Data

Environmental data were collected by a team of infection control staff and students who periodically tested the status of sites such as sinks, washrooms, toilets, hoppers, alcoves, locker rooms, and kitchens within patient and staff rooms across floors 2 to 8 in the hospital for the presence of CRE. The time period for environmental samples considered for this study aligns with the patient time period mentioned above. The procedure for collecting clinical samples consists of using a swab along the interior wall of a sink drain. To collect sink p-trap samples, water samples were collected using IV tubing lowered through the drain hole. The samples were then processed in a clinical microbiological laboratory to determine whether the specimen tested positive for CRE. Figure 4.1 summarizes the data. As shown in this figure, although samples were collected from all floors, the majority of samples were collected from



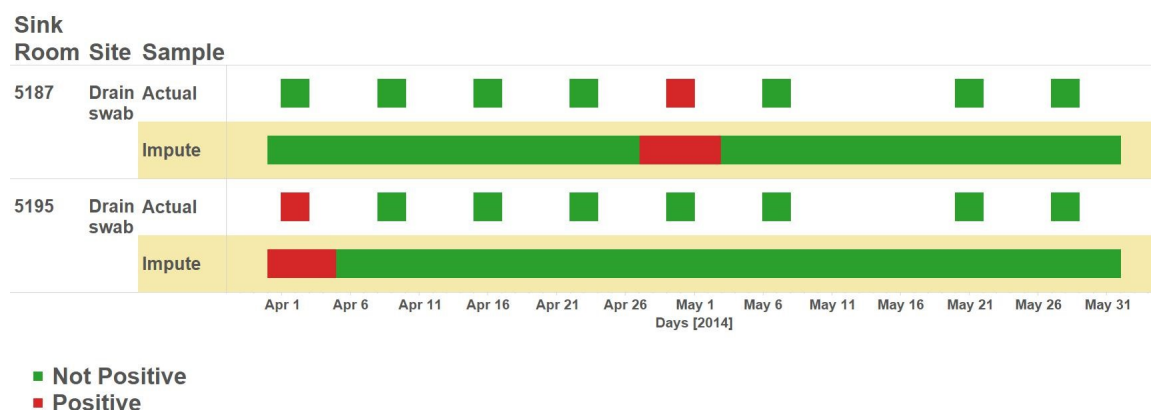
the 3rd and 5th floors, which is why these floors are used for modeling.



**Figure 4.1:** Environmental Sampling Summary

Figure 4.1 also indicates that the two most common sites where sampling was performed were drains and p-traps. Previous studies [13, 14] have focused on using sink status as one of the environmental variables for modeling purposes. We decided to use the two most common environment sampling sites, namely, drain and p-trap, for our visualization purposes because these sites could provide additional information regarding sink contamination or patient acquisition. As shown in Figure 4.1, the environmental sampling was not performed on a regular basis. There are noticeable gaps in the sampling during the days of the study. We decided to use the mid-point imputation method described in the work of [14] to ensure that the series is consistent. However based on feedback obtained from ICPs, we made an adjustment to the imputation methodology to adjust cases of consecutive positive-negative-positive culture sequence found in one month to convert to positive-positive-positive. The understanding behind the adjustment was that since the period is too short the negative observation in between the two positives could have been a case of environmental sampling error.

Figure 4.2 presents an example of the imputation method for sample sinks. The top row of every sink in the figure represents when samples were actually collected, whereas the second row represents the imputed series. As shown in this figure, the samples for sink 5187 were collected on April 23<sup>rd</sup> and April 30<sup>th</sup>. The sink drain sampling was negative on April 23<sup>rd</sup>, whereas it was positive on April 30<sup>th</sup>. The imputation series for that sink on the second row shows that half days between the two are filled with green (negative status), whereas the latter half are imputed as positive.

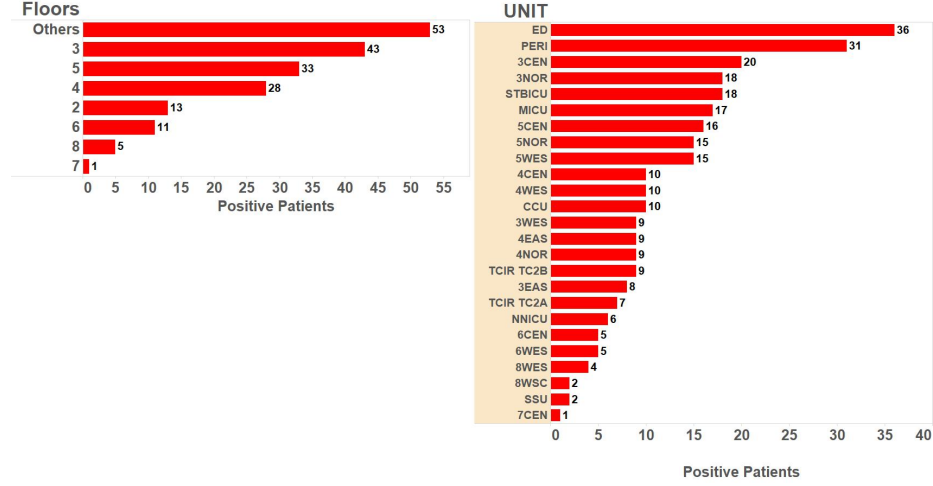


**Figure 4.2:** Imputation example for room sinks 5187 and 5195 using midpoint method

## 4.2 Patient Data

Based on the time period from September 2013 to February 2016, we identified 98 patients across the 2<sup>nd</sup> to 6<sup>th</sup> floors that were found to be CRE positive after 48 hours of admission. Figure 4.3 shows the summary of CRE-positive patients distributed by floors and units. Note that 'Others' include any patients that visited the emergency department (ED), perianesthesia facility (PERI) or SSU.

It is observed that most positive patients in the study visited the ED or anesthesia facility at least once during their stay, which is why we observe that the number of



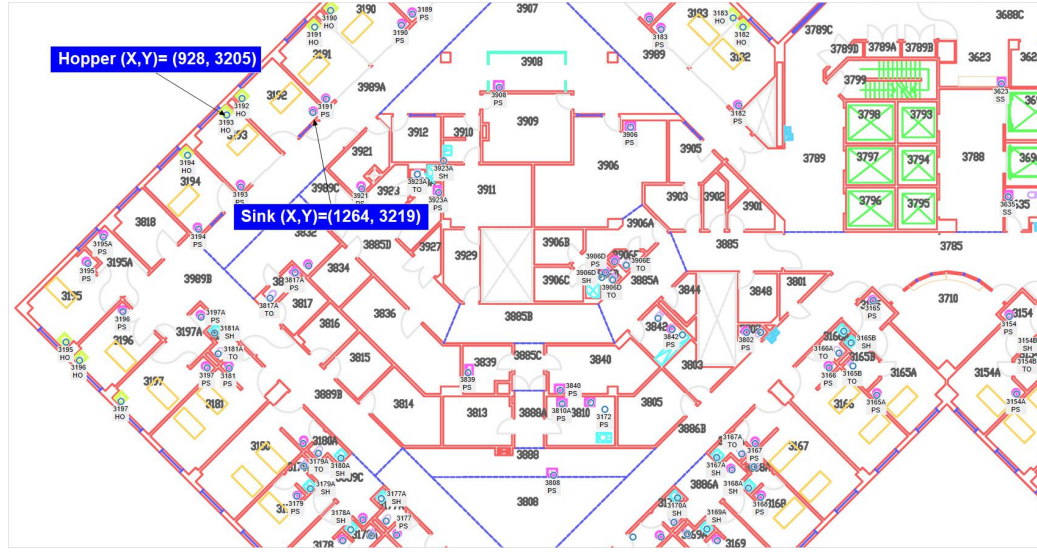
**Figure 4.3:** Positive patient distribution by floors and units

patients attributed to the 'Others' floor is higher than other floors. We observe that a large number of patients remain either on the 3<sup>rd</sup> floor or 5<sup>th</sup> floor, where the STBICU and MICU units are located. Based on the units chart, its also evident that positive patients in this study stayed more at 3 Central (3CEN) and 3 North(3NOR) units than STBICU and MICU.

### 4.3 Geospatial data

Geospatial data consist of floor plans of the hospital that have details on the location of every room along with the details of the environmental sites, including patient beds, staff rooms, alcoves, sinks, bathrooms, showers and toilets. To generate geospatial location coordinates for each room and environmental site of a floor, the floor plan image was loaded in an image processing tool called ImageJ [22]. For the tool to generate X and Y coordinates of any location on the floor plan, the user needs to click on the location of the site after the image is loaded. The user then needs to copy the generated location into a spreadsheet alongside the respective room and sink names. The mapping of the floors (3<sup>rd</sup> - 8<sup>th</sup>) of the hospital was completed in a similar manner. A sample layout for a floor after completion of X,Y mapping is

shown in Figure 4.4.



**Figure 4.4:** Geospatial mapping example for the 3<sup>rd</sup> floor

The floor plan image along with the geospatial data is loaded into the Tableau software [23]. Each of the dots highlighted on the floor plan in Figure 4.4 is a result of the X, Y coordinate mapping performed in ImageJ. Sample points on Figure 4.4 are highlighted to show the coordinates for the hoppers and sinks for rooms 3193 and 3192.

## 4.4 Data Transformations

To arrive at the variables for our analysis, we created the following transformations based on the existing variables in the dataset:

- **CRE\_after\_48hours:** This variable is created based on the difference between the first positive date and the admission date. If the difference between the two dates is greater than 2 days, then the variable receives a value of 1; otherwise, it is coded as 0.
- **Extending patient series to daily:** The patient bed transfer table has one record

for every admitted patient in the hospital that moves into a room. The variables `in_dttm` and `out_dttm` denote a patient’s entry and exit from the room, respectively. Because our modeling method uses daily records for every patient, we extended the series for every patient to have a daily interval.

- Mapping status of adjoining rooms to every patient room: For each room in the patient dataset, we included the status of four adjacent rooms along with their lags. The reasoning for obtaining the variables for four adjoining rooms was that a nurse typically would cover or serve four patients in adjoining rooms at a time and there could be a possibility that she might act as a carrier. The following variables were created: `adj_pat1.status`, `adj_pat2.status`, `adj_pat3.status`, and `adj_pat4.status`, which represented the status of the patient in the first adjacent, second adjacent, room next to first adjacent and room next to second adjacent room, respectively, during the same time as the patient under consideration.
- Lag variables for patient: To determine whether the status of the last patient(s) played any role in the infection transmission, we included lag variables, namely, `last_pat_pos_7`, `last_pat_pos_14`, `last_pat_pos_21`, and `last_pat_pos_28`, which represent whether there was at least one positive patient in the past 7, 14, 21 and 28 days in the same room before a patient came to the room. Similarly, we also introduced additional variables for adjacent rooms and next to adjacent rooms: `adj_pat1.status_lag7`, `adj_pat1.status_lag14`, `adj_pat1.status_lag21` and `adj_pat1.status_lag28`. Immediate adjacent rooms have prefixes `adj1` and `adj2`, whereas the rooms next to adjacent have a prefixes `adj3` and `adj4`.
- Mapping status of adjoining room environmental variables: For each room in the patient dataset, we included the status of four adjacent environment sites for sinks, including distinct variables indicating drain and p-trap statuses.

The purpose is to test whether the same/adjacent patient or adjacent environmental attributes play a role in sink contamination. The following variables were created: `adj_dr1_status`, `adj_dr2_status`, `adj_drt3_status`, and `adj_dr4_status`, which represented the status of the drain sink in the first adjacent, second adjacent, room next to first adjacent and room next to second adjacent room, respectively, during the same time as the patient under consideration. Similarly, `adj_pt1_status`, `adj_pt2_status`, `adj_pt3_status`, and `adj_pt4_status` represent variables for adjoining room sinks. We also derived a rolled up list of variables to include adjacent sinks (rather than drain and p-trap) for the convenience of using them if we did not find any significant difference between the drain and p-trap attributes.

- Lag variables for sinks: Lag variables for sinks were also created similar to room variables to determine whether such variables play a role in sink contamination. The sink lag variables are as follows: `adj_sink_status_lag7`, `adj_sink_status_lag14`, `adj_sink_status_lag14` and `adj_sink_status_lag28`.

# CHAPTER 5

## SPATIOTEMPORAL VISUALIZATION

### 5.1 Visualization Design Requirements

To be able to effectively explore multivariate data, we identify a set of objectives that the spatiotemporal visualization should be able to achieve. Because our study focuses on understanding how adjacent environmental sites around a patient change over the course of time, we need a design that can integrate patient and environmental microbiological data and be able to show us views and summary statistics over a period of time and by different attribute slices. Some of the questions that the visualization should be able to address are as follows: What is the status of rooms and environmental sites for a particular day? How are the positive patients located across units? What is the status of the environmental sites (particularly sinks/drains) prior to the patient becoming infected? Where are frequent environmental samplings occurring? Is sampling uniform across rooms?

Additionally, the visualization should also provide the user with the ability to focus on any unit and examine the status of the patient and environment for a range of time. This will provide a good view for comparing the events across units and detecting any differences. There has not been much focus on the type of sink elements, namely, drain or p-trap, and the role (if any) that they play in patient acquisition. One of the requirements of our visualization is that it should be capable of providing a distinct

view of drain and p-trap sampling status and portray the summary statistics and floor plan overlay view of these along with patient timelines.

Finally, the visualization should provide an integrated view of the patient room with its environmental sites, adjoining room patients and their status, and adjoining environmental sites along with their status. The user should have the ability to select any patient and look at her entire history of transfers along with the status of the rooms before he arrived, time series of adjacent bed status during his period of stay, and time series of environmental status for same and adjacent rooms during his period of stay, which will be helpful in arriving at the hypothesis that the same and adjacent room/environmental sites are playing a role in patient acquisition.

## 5.2 Visualization Interface Components

The key components that a user is able to interact with in this system are as follows:

- Floor plans: Addition of floor plans to the visualization provides spatial context to the user. It also provides a canvas on which the user is able to see queries.
- Time selection: It allows the user to specify temporal constraints for patient and environmental data. This is in the form of a slider that a user can choose to drag two ends to indicate the 'start' and 'end' times.
- Unit: It allows the user to specify any unit(s) of the hospital and display visualization or summary charts.
- Status: The user can choose to see only infected patients/environmental sites
- Source type: Users can select drain or p-trap source types.
- Species: Allows the user to choose species for any environmental site.



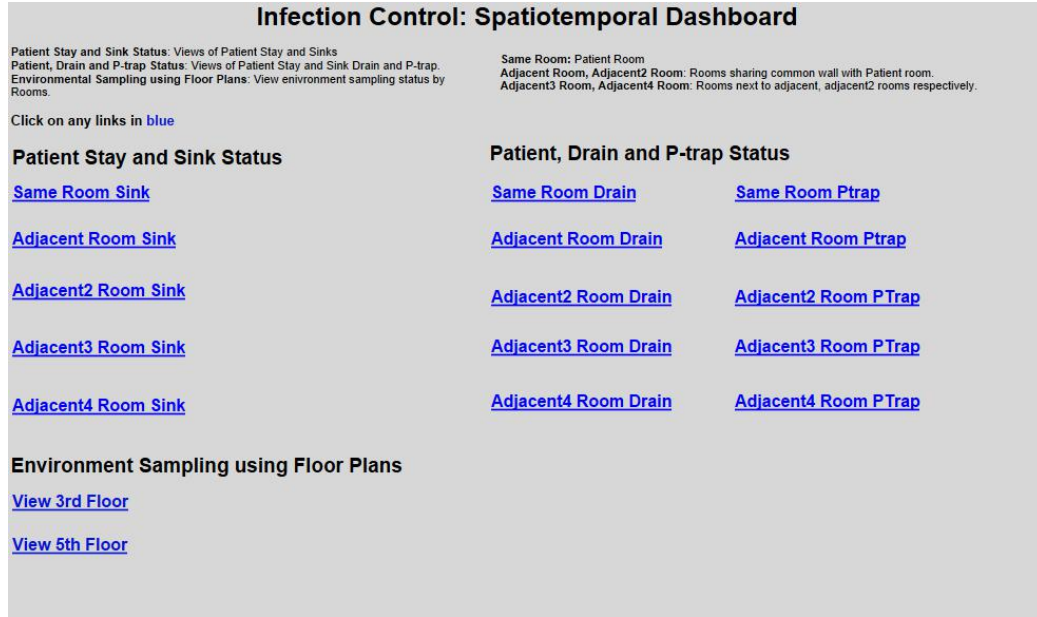
## 5.3 Visualization Dashboards

Dashboards are an important component of spatiotemporal visualization. Dashboards created in Tableau software [23] serve the purpose of integrating different views on a single page that are tied together by interface components. We show some of the dashboards created in this study for exploring data and arriving at the hypothesis. For a new user to navigate through the dashboards easily, we recommend a 15-20 min training time.

- Spatiotemporal Visualization Menu: Figure 5.1 shows the main screen of visualization with options that allow users to select views specific to the analysis. It has three distinct sections for users to focus their analysis. Patient stay and sink status has views related to patient stay and sinks for patient rooms and adjacent rooms. Patient, drain and p-trap status provides users with views of patient room and discrete views of sink drain and sink p-trap status for patient stay.

Environmental sampling using floor plans provides a view of environment samplings performed in different rooms at different time periods.

- Patient stay and sink status: The objective of this dashboard, along with other dashboards located in the patient stay and sink status section in the main menu, is to enable easy exploration and visualization of instances when a patient had a positive sink in the same room or adjacent rooms during his risk period (time between last negative and first positive). This view allows users to view patient stay history, last negative date of patient, date when he/she first tested positive and the species found in the patient. The user can select any patient id from the menu on the right to access the chart and information. The color of the

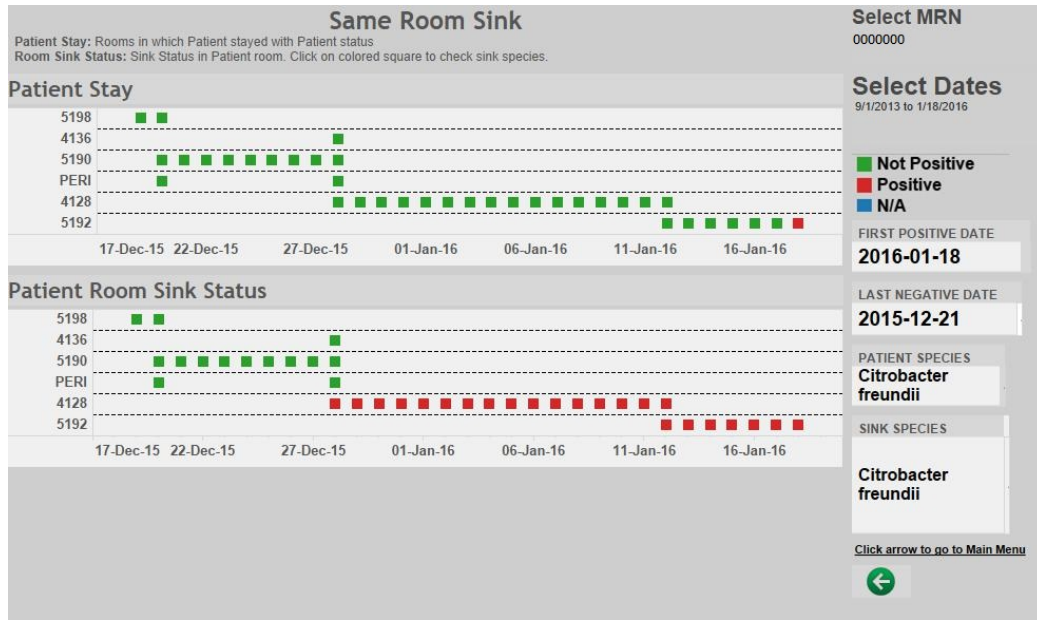


**Figure 5.1:** Spatiotemporal Visualization Dashboard

squares indicates the status of the patient or sink.

This dashboard also has an integrated view of the status of the sink, adjacent sink or next to adjacent sink depending on the option that the user selects on the main screen. Figure 5.2 shows the status of the same room sink for a sample patient. The red dots indicate that the patient became infected. The patient to be analyzed is on the top row, while the charts underneath are the statuses of other patients. In this case, it appears that the time period between the last negative specimen collected from the patient and first positive date, there was a positive patient in the same room.

- Environment sampling using floor plans: For the purpose of this study, the objective of the two dashboards in this section is to enable exploration of environmental variables relevant for sink modeling. The question that it helps one to understand is how different are sink statuses in the same and adjacent rooms and whether adjacent sinks could be potential risk factors. It also provides a view of how consistent sampling is in different rooms. There are two floor op-

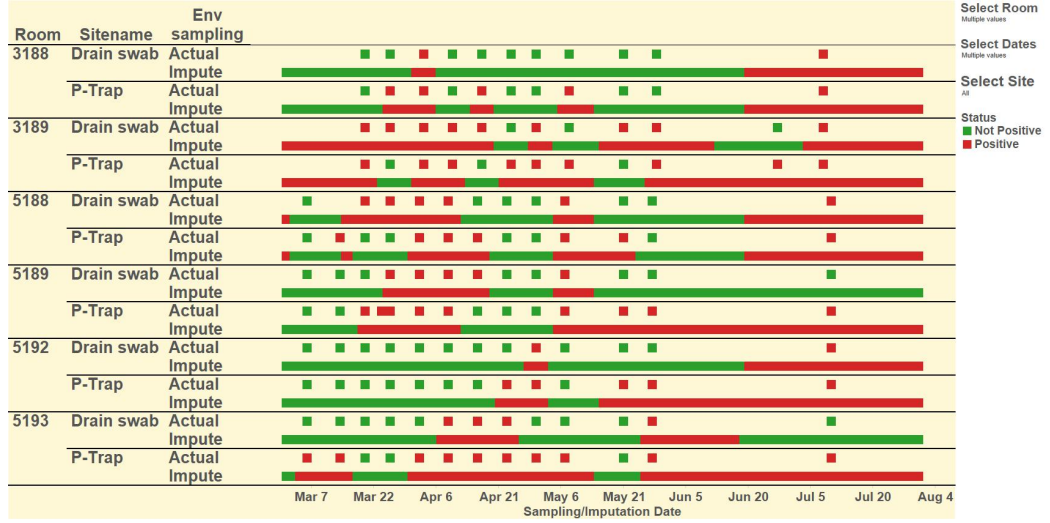


**Figure 5.2:** Patient stay and same room sink

tions for this section that the user can select from the main menu. The user is presented with a floor plan screen after clicking on any floor option. The dots on the screen represent sinks in the respective rooms. The user can click on any of the dots to navigate to the environment dashboard that shows the sampling history for that room.

Figure 5.3 shows the environmental sampling view of a few STBICU and MICU rooms for the period between Mar and Aug 2014 split by drain and p-trap. This dashboard can also be used in conjunction with patient dashboards to further analyze whether the positive drain or p-trap status that is shown is a result of actual sampling or imputation. The user controls on this screen allow for filtering by drain or p-trap status, date range, sampling or imputation view. For each room, the dashboard shows two rows per sampling site. In the figure, it can be observed that for room 3188, the actual sampling is on the top row and the green or red squares represent the status of sampling on that date. The bars corresponding to the imputed sampling series are on the second row with

the label 'Impute'. The calculation of imputed series is based on the midpoint method recommended in the most recent work [14].



**Figure 5.3:** Environment history dashboard showing sampling and imputation

- Patient, drain and p-trap status: The objective of this dashboard is to enable exploration of drain and p-trap sampling status with patient stay and to determine whether either or both of them could be potential variables in determining patient positivity. There are several dashboards in this section that show drain or p-trap history for the same room, adjacent rooms or outer adjacent rooms. A group of four rooms around a patient room (2 rooms on either side) are chosen to be on the dashboard main menu with the understanding that these could be a potential group or cluster that could explain patient risk factors.

## CHAPTER 6

# MODELING METHODOLOGY AND RESULTS

### 6.1 Variable Selection

The environment sampling dataset used for this analysis was converted into a daily time series of all sinks in STBICU and MICU for the period September 2013 through February 2016. Thus, each day represented a sampling datapoint that showed the status of a sink along with the predictor variables. The sink status has a value of "1" for positive and "0" for negative. To arrive at a potential list of variables for modeling consideration, we used the spatiotemporal visualization module (described in Chapter 5). We also considered the variables used in the previous research on the subject [14]. A complete list of the variables considered for the purposes of modeling is presented in Table 6.1.

We performed stepwise regression for variable selection using all the variables from Table 6.1 and examined the Akaike information criterion (AIC) results. We used AIC since it provides an objective way to choose the most parsimonious. The final model based on this method retains 9 of the 13 variables. We summarize the model results based on the full model and reduced model obtained from stepwise regression in Table

**Table 6.1:** Variable Selection for Sink Risk Modeling

Variable	Range	Description
Same Room Patient Status (P)	0=Not Positive; 1=Positive	Indicates whether a Patient was Positive on that day
Positive Patient in last 7 days( $P_{\text{lag}7}$ )	0=Not Positive; 1=Positive	Indicates whether any Patient was Positive in last 7 days
Positive Patient in last 14 days( $P_{\text{lag}14}$ )	0=Not Positive; 1=Positive	Indicates whether any Patient was Positive in last 14 days
Positive Patient in last 21 days( $P_{\text{lag}21}$ )	0=Not Positive; 1=Positive	Indicates whether any Patient was Positive in last 21 days
Positive Patient in last 28 days( $P_{\text{lag}28}$ )	0=Not Positive; 1=Positive	Indicates whether any Patient was Positive in last 28 days
Adjacent Patient status( $P_1$ )	0=Not Positive; 1=Positive	Indicates whether Patient in adjacent rooms that share common plumbing with next room was Positive
Adjacent Patient status( $P_2$ )	0=Not Positive; 1=Positive	Indicates whether Patient in adjacent rooms that <u>do not</u> share common plumbing with next room was Positive
Adjacent Room sink status( $S_1$ )	0=Not Positive; 1=Positive	Indicates whether adjacent rooms that share common plumbing were Positive
Adjacent Room sink status( $S_2$ )	0=Not Positive; 1=Positive	Indicates whether adjacent rooms that <u>do not</u> share common plumbing were Positive
Intervention in the last 3 days ( $I_3$ )	0=Not Occurred; 1=Occurred	Indicates whether Intervention occurred in the same room in last 3 days
Intervention in the last 5 days ( $I_5$ )	0=Not Occurred; 1=Occurred	Indicates whether Intervention occurred in the same room in last 5 days
Intervention in the last 7 days ( $I_7$ )	0=Not Occurred; 1=Occurred	Indicates whether Intervention occurred in the same room in last 7 days
Sink Status in last 30days ( $S_{\text{lag}30}$ )	0=Not Positive; 1=Positive	Indicates whether same room Sink was Positive in last 30 days

6.2.

Model	Model Variables	AIC
Full Model	All variables	22918
Final Model from Stepwise approach	Selected Same room and adjacent room variables	22912

**Table 6.2:** Comparison of Models

The stepwise regression model with 9 out of the 13 variables has a lower AIC compared to the full model. Hence, we select this model for further evaluation.

## 6.2 Results

### 6.2.1 Logistic Regression Modeling

We performed a logistic regression using variables from final model in section 6.1. The response variable represents sink status which has value 0(Negative) or 1(Positive). We perform regression using 10-fold cross validation on the dataset. The results from 10-fold cross validation are averaged to produce a single estimate. One of the advantage of this methods is that all observations are used from the dataset for training and validation and each observation is used only once. The parameter estimates are shown in Table 6.3.

To interpret the estimates from Table 6.3, we express them in terms of the odds ratio with a 95% confidence interval. The odds ratio [24] compares the relative odds of an event occurring given the exposure of other variables. Thus, for the purposes of this study, the odds ratio will signify the increase in probability of a sink becoming positive given the presence of any variable. To arrive at the confidence interval associated with

	Variable	Estimate	Std Error	z value	p-value
( $x_0$ )	(Intercept)	-1.58	0.04	-39.43	2e-16
(P)	Same room Patient status	0.65	0.14	5.70	1.13e-08
( $P_{lag14}$ )	Positive Patient in last 14 days	0.57	0.07	7.23	4.78e-13
( $P_1$ )	Adjacent room <sub>1</sub> Patient status	0.12	0.10	1.73	0.084
( $P_2$ )	Adjacent room <sub>2</sub> Patient status	0.53	0.10	4.07	4.57e-05
( $S_1$ )	Adjacent room <sub>1</sub> Sink Status	0.59	0.03	16.13	2e-16
( $S_2$ )	Adjacent room <sub>2</sub> Sink Status	0.12	0.03	2.97	2e-03
( $S_{lag30}$ )	Sink status in last 30 days	2.58	0.03	77.89	2e-16
( $I_7$ )	Interventions carried out in last 7 days	-0.69	0.05	-10.89	2e-16

**Table 6.3:** Parameter Estimates and Significance

the odds ratio, we use the bootstrap method with 10,000 replications. Table 6.4 shows

the odds ratios and confidence intervals for the variables.

	Variable	Odds Ratio	95% C.I
(P)	Same Room Patient Status	1.92	1.46 - 2.53
( $P_{lag14}$ )	Positive patient last 14 days	1.78	1.52 - 2.08
( $P_1$ )	Adjacent Room <sub>1</sub> Patient status	1.13	0.92 - 1.40
( $P_2$ )	Adjacent Room <sub>2</sub> Patient Status	1.70	1.37 - 2.09
( $S_1$ )	Adjacent Room <sub>1</sub> Sink Status	1.80	1.68 - 1.93
( $S_2$ )	Adjacent Room <sub>2</sub> Sink Status	1.13	1.03 - 1.24
( $S_{lag30}$ )	Sink status in last 30 days	13.25	12.39 - 14.18
( $I_7$ )	Interventions carried out	0.50	0.45 - 0.56

**Table 6.4:** Odds Ratio with Confidence Intervals

It is observed that the odds of a sink becoming positive increases by almost 13-fold (CI: 12.39-14.18) when the status of the sink was positive in the last 30 days. Intervention has a negative effect on sink positivity. This is consistent with the general understanding that interventions would reduce the odds of sink contamination for a short duration. The odds reduce by 50% if an intervention is performed within the last 7 days. Instances of positive sinks in adjacent rooms<sub>1</sub> which share common plumbing increase the chances of a sink becoming positive by 80% (1.68-1.93). Instances of



positive sinks in adjacent rooms<sub>2</sub> that do not share common plumbing increase the chances of a sink becoming positive by only 13% (1.03-1.24). The presence of a positive patient in the same room increases the odds of a sink becoming positive by 1.92-fold (CI: 1.46-2.53), whereas the presence of any positive patient within the past 14 days increases the odds by 1.78-fold (CI: 1.52-2.08). Furthermore, the presence of a positive patient in an adjacent room<sub>1</sub> which shares common plumbing increases the odds of sink positivity by 1.13-fold (CI: 0.92-1.40), while the presence of a patient in adjacent room that do not share common plumbing increases the odds by 1.70-fold (CI: 1.37-2.09).

We test the overall significance of the model using the likelihood ratio test. This test compares the likelihood of the final model (from stepwise regression) with the likelihood of the null model (intercept only). This is similar to the overall F-test in linear regression. Table 6.5 shows that the p-value is significant, and hence, we conclude that our model fits well.

	Model	Resid. df	Resid. Dev	df	Deviance	p-value
1	Null Model	18726	24983			
2	Final Model	18717	17772	9	7211.2	2.2e-16

**Table 6.5:** Analysis of Deviance Table

From the confusion matrix (Table 6.6), we see that model correctly predicted 84% of the positive cases. The threshold for confusion matrix was arrived at by testing various estimates such that resulting value of sensitivity was higher than specificity without overly penalizing overall accuracy. For our case in point, missing a positive sink is important than having negative sink classified as positive. The threshold for

confusion matrix was found to have an optimum value of 0.4.

		Actual	
		0	1
Predicted	0	7543	2584
	1	2758	13868

**Table 6.6:** Logistic Regression Model Confusion Matrix

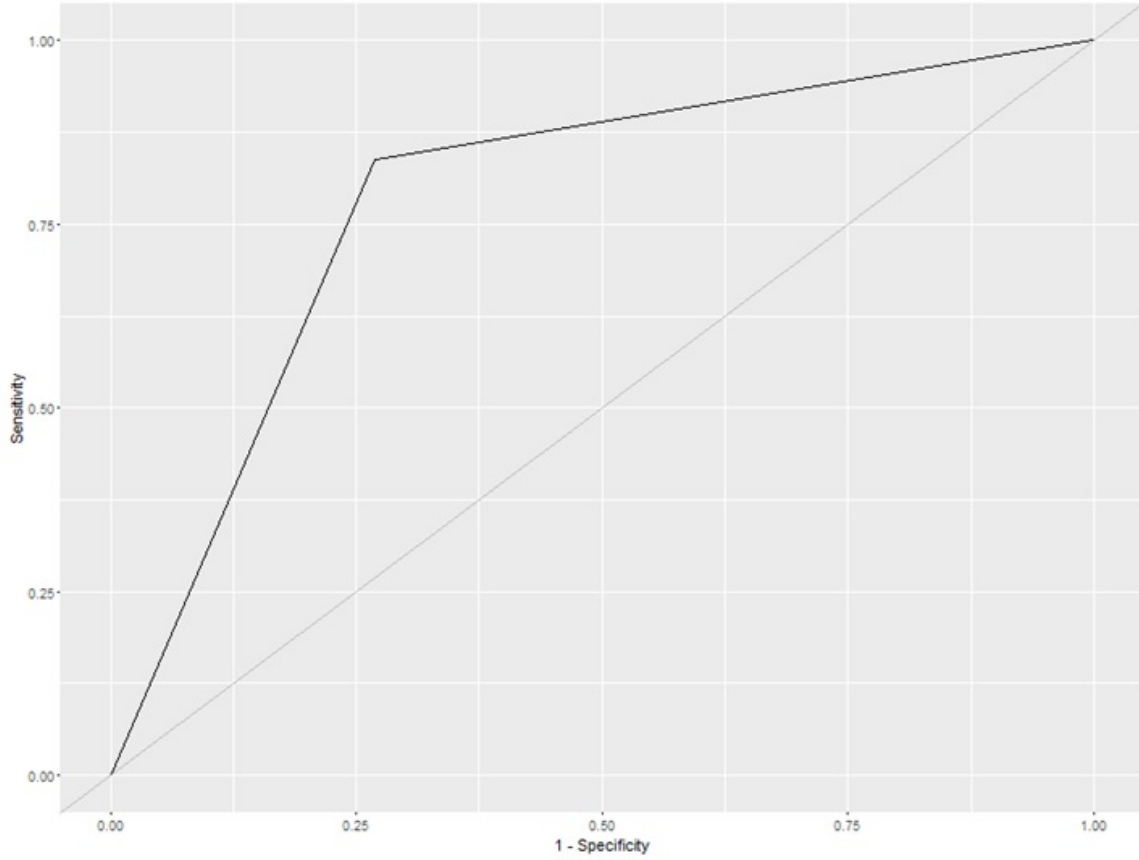
Other model metrics, including area under the curve (AUC), sensitivity and specificity, are presented in Table 6.7, while Figure 6.1 shows the ROC curve for the model.

Metric	Results
AUC	78.8%
Sensitivity	84.3%
Specificity	73.2%

**Table 6.7:** Logistic Regression Model Results

## 6.2.2 Random Forest Model

To compare the variables and accuracy obtained through logistic regression, we construct a random forest model. The model is run with 10-fold cross validation, using 500 trees with a maximum node size of 5. The results of the model are shown in Figure 6.2, which summarizes the importance of variables by MDA (mean decrease accuracy). MDA can be used as a feature selection method because it shows the impact of each variable on the model accuracy. The results show that 7 of the 9 variables included in the logistic regression also appear in Figure 6.2 based on the random forest model.



**Figure 6.1:** ROC Curve for Logistic Regression Model

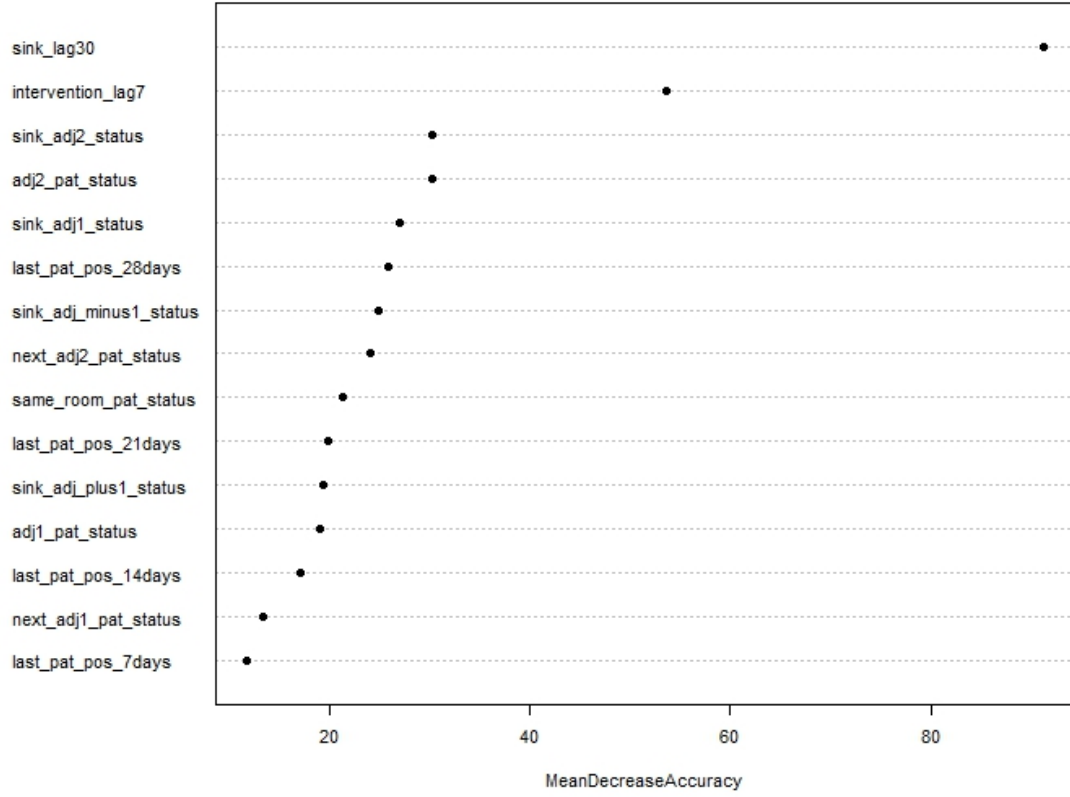
The confusion matrix based on the random forest model is shown in Table 6.8.

		Actual	
		0	1
Predicted	0	7891	2590
	1	2410	13862

**Table 6.8:** Random Forest Confusion Matrix

Table 6.9 summarizes metrics obtained from both models:

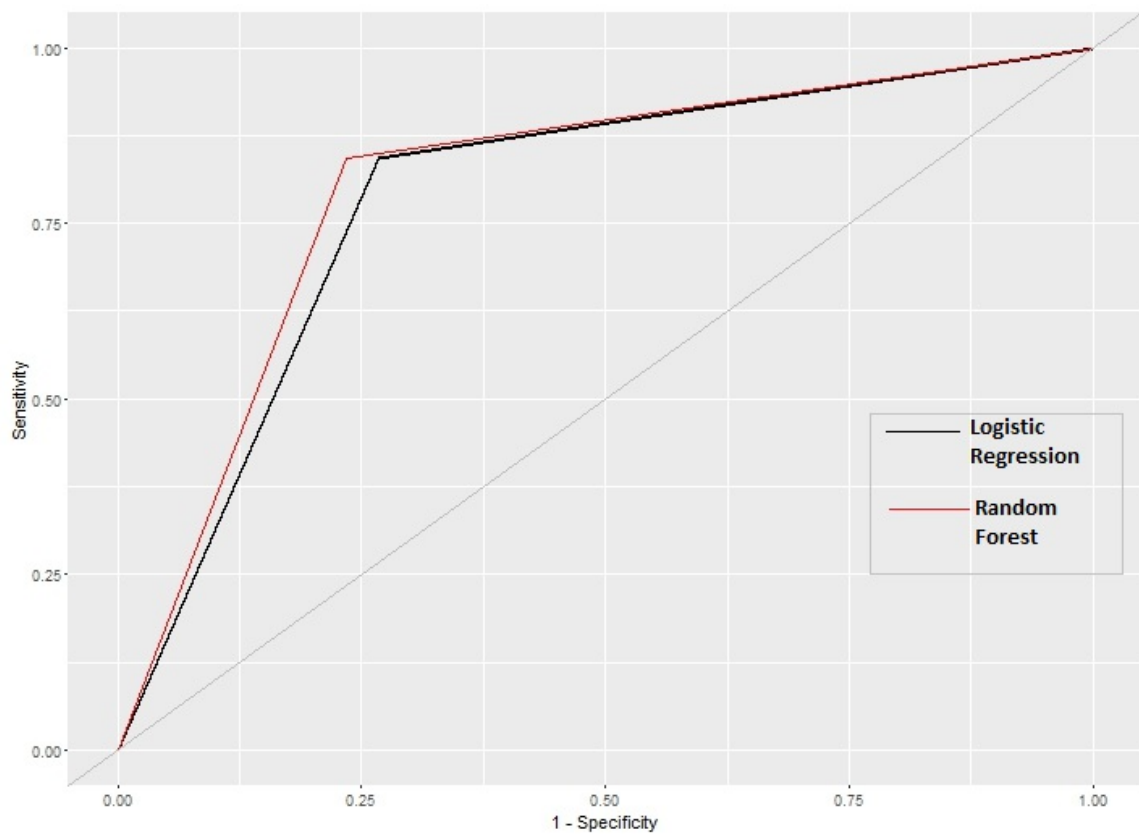
Figure 6.3 shows a ROC curves for Logistic Regression and Random forest model.



**Figure 6.2:** Random Forest - Variable Importance

	Model	Accuracy	AUC	Sensitivity	Specificity
1	Logistic Regression	80%	78.8%	84.3%	73.2%
2	Random Forest	81%	80.4%	84.3%	76.6%

**Table 6.9:** Logistic Regression and Random Forest Model Metrics



**Figure 6.3:** ROC Curves for Logistic Regression and Random forest models

## CHAPTER 7

### CONCLUSION, LIMITATIONS AND FUTURE WORK

#### 7.0.1 Conclusion

The results from logistic regression model indicate that variables such as presence of positive sink in adjacent rooms that share common plumbing, status of the patient in the same room, status of sink in the past 30 days, presence of a positive patient in last 14 days, presence of a positive patient in the adjacent room and interventions performed in the past 7 days are significant risk factors in explaining sink contamination. The findings on the presence of a positive patient in last 14 days, status of sink in the last 30 days and interventions implemented in the last 7 days is consistent with the expert opinion and with previous research [14].

Based on our modeling results, the odds of sink positivity increase by 13 times when the sink was found to be positive in the last 30 days. Additionally, the negative coefficient of the intervention variable is consistent with the understanding that the sink becomes negative when interventions are performed. The odds of a sink becoming positive decreases by a factor of 0.50 based on the logistic regression model.

Furthermore, the presence of any positive patient within the past 14 days resulted in an increase in the odds of sink positivity by 1.78 times. Another patient-specific factor that has a larger impact is the presence of a positive patient in the same room. The probability of a sink becoming positive increases by 1.92-fold when a positive patient is in the same room. Perhaps one of the most interesting factors that we found through our modeling results is the impact of adjacent room sinks that share common plumbing on the positivity of the same room sink. Based on our results, we observe that the presence of a positive sink in adjacent rooms that share common plumbing increases the odds of a sink turning positive by 1.80-fold. This contrasts with the odds of just 1.13-fold for rooms that do not share common plumbing. Previous research [25] has shown that biofilms found in sinks were linked to outbreaks. Some research works [26, 27] have also shown that these biofilms are resistant to traditional disinfectant methods. Our results indicating the significance of adjacent rooms that share common plumbing can be explained by the probable presence of biofilms in sink drain walls, traps or drainpipes connecting the two room sinks. Despite the timely implementation of intervention strategies, it could be likely that the presence of biofilms contributed to the adjacent sinks becoming positive.

Another spatial factor that we found to be significant in the model is the presence of a positive patient in the adjacent room. The model results show that the odds of sink positivity increase by 1.13-1.70 fold when the adjacent rooms have a positive patient. We would expect this to be true given that positive patients in the adjacent room would use the sink (adjacent room sink). Thus, it is likely that they would contaminate the sink of the room that they are staying in.

Results from Random forest model show a slight increase in accuracy (81% vs 80%) compared to Logistic regression. Both models show identical True positive rate which is important in our case since the cost associated with misclassifying a positive sink as negative is higher than falsely classifying negative sink as positive. Additionally, we see a higher specificity based on random forest model compared to logistic regression. As a result, we see that the Area under curve(AUC) is higher for Random forest (Shown in Figure 6.3).

### **7.0.2 Limitations and Future Work**

There has been a fair amount of work devoted to recording of cultures for environment sampling. However, much of this work has been manual and maintained in excel spreadsheets. If there have been any translation errors from the actual environment sampling to recording in the spreadsheet, there could be a shift in the modeling results. Any changes to the outcome of the sampling performed at a later stage and not updated in the environment spreadsheet or anything missing from the environment database could also impact the results.

Furthermore, it was also found out that some of the individual patient records that were provided to us had mixing of patient identifiers and medical record numbers (MRN) for the control population. We have tried our best to reconcile the data at our end to ensure that the PIDs were converted to MRN, but further research should ensure that such cases are addressed.



Our analysis considered only sinks as the environmental reservoirs for modeling purposes. Other environmental attributes, such as toilets, hoppers, and showers, were not considered. Future modeling attempts should consider these attributes to show the extent of infection transmission contribution through such reservoirs.

Our modeling includes interventions as one of the key variables. The information provided to us regarding interventions was at a unit level for specific dates. Hence, we have performed this analysis with the assumption that interventions were implemented for all rooms within units. However, it was later learned that some of the room sinks acted as 'control' sinks and were not subjected to the interventions. For feature selection, we only used AIC as the sole criteria for arriving at the final model. While there are many advantages to using AIC, it also has disadvantages including evaluating only those candidate models that are pre-specified before analysis.

The spatiotemporal visualization tool currently incorporates patient and environment data. We have brought both the datasets together to provide a combined view for analysis. However, the tool does not include provider movement and contact information. Future work on improving the tool would be to integrate the provider dataset. Additionally, we have done our best to provide users with an optimized view of patient transfer history once the user makes a patient selection. However, in cases when a patient has made a large number of visits and transferred through many rooms, the initial view can appear cluttered with a considerable amount of

information. This is found in instances when the patient has made visits multiple times over more than 3 years. One approach to make the tool to display better would be to shorten the visit duration and focus on a smaller time period.

## REFERENCES

- [1] Tom Frieden. Antibiotic resistance threats in united states, 2013. *Technical report*, 2013.
- [2] D Raeeed and D Dorothy. Carbapenem-resistant enterobacteriaceae: Deadly superbugs on the rise. *Epidemiology*, 35(8):978–983, 2014.
- [3] D. Lobo J.-Mathers A. Barnes L., Brown and J. Papin. Superbug tracker. Design of Context-Aware Surveillance System for Nosocomial Outbreaks involving Non-Patient Reservoirs. Technical report, 2015.
- [4] S Vergara-López, MC Domínguez, MC Conejo, A Pascual, and J Rodríguez-Baño. Wastewater drainage system as an occult reservoir in a protracted clonal outbreak due to metallo- $\beta$ -lactamase-producing klebsiella oxytoca. *Clinical Microbiology and Infection*, 19(11):E490–E498, 2013.
- [5] I Wolf, PWM Bergervoet, FW Sebens, HLA Van den Oever, PHM Savelkoul, and WC Van der Zwet. The sink as a correctable source of extended-spectrum  $\beta$ -lactamase contamination for patients in the intensive care unit. *Journal of Hospital Infection*, 87(2):126–130, 2014.
- [6] D Roux, B Aubier, H Cochard, R Quentin, N van der Mee-Marquet, et al. Contaminated sinks in intensive care units: an underestimated source of extended-spectrum beta-lactamase-producing enterobacteriaceae in the patient environment. *Journal of Hospital Infection*, 85(2):106–111, 2013.
- [7] Factor SH-Jenkins SG Calfee DP. Patel G, Huprikar S. Outcomes of carbapenem-resistant klebsiella pneumoniae infection and the impact of antimicrobial and adjunctive therapies. *Infect Control Hosp Epidemiol*, (29):10991106, 2008.
- [8] Navon-Venezia S Schwartz D Leavitt A Carmeli Y. Schwaber MJ, Klarfeld-Lidji S. Predictors of carbapenem-resistant klebsiella pneumoniae acquisition among hospitalized adults and effect of acquisition on mortality. *Antimicrob Agents Chemother*, (52):1028–1033, 2008.
- [9] MP Crespo, N Woodford, A Sinclair, ME Kaufmann, J Turton, J Glover, JD Velez, CR Castaneda, M Recalde, and DM Livermore. Outbreak of carbapenem-resistant pseudomonas aeruginosa producing vim-8, a novel metallo- $\beta$ -lactamase, in a tertiary care center in cali, colombia. *Journal of clinical microbiology*, 42(11):5094–5101, 2004.

- [10] Leonardo A Stroud, Jerome I Tokars, Michael H Grieco, Jack T Crawford, David H Culver, Brian R Edlin, Emilia M Sordillo, Charles L Woodley, Mary Ellen Gilligan, Nancy Schneider, et al. Evaluation of infection control measures in preventing the nosocomial transmission of multidrug-resistant mycobacterium tuberculosis in a new york city hospital. *Infection Control & Hospital Epidemiology*, 16(03):141–147, 1995.
- [11] Matthaios Papadimitriou-Olivgeris, Markos Marangos, Fotini Fligou, Myrto Christofidou, Christina Bartzavali, Evangelos D Anastassiou, and Kriton S Filos. Risk factors for kpc-producing klebsiella pneumoniae enteric colonization upon icu admission. *Journal of antimicrobial chemotherapy*, 67(12):2976–2981, 2012.
- [12] Felipe F Tuon, Jaime L Rocha, Paula Toledo, Lavinia N Arend, Camila H Dias, Talita M Leite, Sergio R Penteado-Filho, Marcelo Pilonetto, and Alexandre P Zavascki. Risk factors for kpc-producing klebsiella pneumoniae bacteremia. *Brazilian Journal of Infectious Diseases*, 16(5):416–419, 2012.
- [13] Jeremy Stern, Sam Hewitt, Michael Guilfoyle, Chetan Mishra, Amy Mathers, Jennifer Lobo, Donald Brown, and Laura Barnes. Modeling nosocomial transmission of carbapenem-resistant bacteria. In *Systems and Information Engineering Design Symposium (SIEDS), 2015*, pages 176–181. IEEE, 2015.
- [14] Julia Lensing. Imputation methods and modeling of environmental reservoirs of nosocomial carbapenem-resistant enterobacteriaceae. Master’s thesis, University of Virginia, Charlottesville, Virginia, 2016.
- [15] Nivan Ferreira. *Visual analytics techniques for exploration of spatiotemporal data*. PhD thesis, POLYTECHNIC INSTITUTE OF NEW YORK UNIVERSITY, 2015.
- [16] Krist Wongsuphasawat and David Gotz. Exploring flow, factors, and outcomes of temporal event sequences with the outflow visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2659–2668, 2012.
- [17] Megan Monroe, Rongjian Lan, Hanseung Lee, Catherine Plaisant, and Ben Shneiderman. Temporal event sequence simplification. *IEEE transactions on visualization and computer graphics*, 19(12):2227–2236, 2013.
- [18] MA Ali, Z Ahsan, M Amin, S Latif, A Ayyaz, and MN Ayyaz. Id-viewer: a visual analytics architecture for infectious diseases surveillance and response management in pakistan. *Public health*, 134:72–85, 2016.
- [19] Thomas Kistemann, Friederike Dangendorf, Ludmila Krizek, Hans-Georg Sahl, Steffen Engelhart, and Martin Exner. Gis-supported investigation of a nosocomial salmonella outbreak. *International journal of hygiene and environmental health*, 203(2):117–126, 2000.

- [20] Gillian A AvRuskin, Geoffrey M Jacquez, Jaymie R Meliker, Melissa J Slotnick, Andrew M Kaufmann, and Jerome O Nriagu. Visualization and exploratory analysis of epidemiologic data using a novel space time information system. *International Journal of Health Geographics*, 3(1):1, 2004.
- [21] Niklas Elmqvist and Ji Soo Yi. Patterns for visualization evaluation. *Information Visualization*, 14(3):250–269, 2015.
- [22] WS Rasband. Imagej: Image processing and analysis in java. *Astrophysics Source Code Library*, 1:06013, 2012.
- [23] Christian Chabot, Chris Stolte, and Pat Hanrahan. Tableau software. *Tableau Software*, 2003.
- [24] Nova Scotia. Explaining odds ratios. *J Can Acad Child Adolesc Psychiatry*, 19: 227, 2010.
- [25] Susy Hota, Zahir Hirji, Karen Stockton, Camille Lemieux, Helen Dedier, Gideon Wolfaardt, and Michael A Gardam. Outbreak of multidrug-resistant pseudomonas aeruginosa colonization and infection secondary to imperfect intensive care unit room design. *Infection Control & Hospital Epidemiology*, 30(01):25–33, 2009.
- [26] Elisabeth Presterl, Miranda Suchomel, Michaela Eder, Sonja Reichmann, Andrea Lassnigg, Wolfgang Graninger, and Manfred Rotter. Effects of alcohols, povidone-iodine and hydrogen peroxide on biofilms of staphylococcus epidermidis. *Journal of Antimicrobial Chemotherapy*, 60(2):417–420, 2007.
- [27] Kelli Buckingham-Meyer, Darla M Goeres, and Martin A Hamilton. Comparative evaluation of biofilm disinfectant efficacy tests. *Journal of microbiological methods*, 70(2):236–244, 2007.