

Imputation Methods and Modeling of Environmental Reservoirs of Nosocomial
Carbapenem-Resistant Enterobacteriaceae

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

Julia C. Lensing

May

2016

APPROVAL SHEET

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

Julia C. Lensing
AUTHOR

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

Donald Brown

Advisor

Laura Barnes

Amy Mathers

Scott Crino

Accepted for the School of Engineering and Applied Science:

James H. Ayler

Dean, School of Engineering and Applied Science

May
2016

Imputation Methods and Modeling of Environmental Reservoirs of Nosocomial Carbapenem-Resistant Bacteria

A Thesis Defense

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 (Systems and Information Engineering)

by

Julia C. Lensing

May 2016

Approval Sheet

This Thesis Defense is submitted in partial fulfillment of the requirements for the degree of
Master of Science (Systems and Information Engineering)

Julia C. Lensing

This Thesis Defense has been read and approved by the Examining Committee:

Prof. Donald Brown, Advisor

Prof. Laura Barnes, Committee Chairperson

Prof. Amy Mathers, Committee Member

Dr. Scott Crino, Committee Member, Red Six Solutions

Accepted for the School of Engineering and Applied Science:

Dean, School of Engineering and Applied Science

May 2016

Abstract

The transmission of hospital-acquired Carbapenem-resistant Enterobacteriaceae (CRE) is a serious and growing problem in hospitals worldwide. Carbapenem - the antibiotic typically used as a last resort to fight bacterial infections - is being rendered useless against some strains of bacteria that cause high rates of mortality among critically ill patients. Previous research of CRE found that traditional patient-to-patient transmission can only account for 20 percent of transmission of this bacteria in hospital settings and it is unknown through what means the remaining 80 percent of cases are transmitted. Recent efforts to better understand other means of transmission found identical genomes of CRE in patient sinks as was found in cultures collected from patients, indicating that environmental reservoirs could be playing a larger role in transmission than was first realized. This study uses clinical data from a major U.S. hospital to evaluate imputation methods for understanding CRE presence in sinks between known cultures and uses the selected imputation method as the response variable in modeling sink positivity over time as a factor of positive patient presence, use of clinical interventions, and sink characteristics. This study found that an ad hoc imputation method based on expert knowledge provided the best representation of sink positivity. Modeling results indicated that the cumulative presence of positive patients in the same room as a sink, the distance from the patient bed to the sink, and sink design are significant predictors of sink positivity.

Contents

Contents	iii
1 Introduction	1
2 Literature Review	3
2.1 Risk Factors	3
2.2 Environmental Reservoirs	3
2.3 Imputation in Epidemiology	4
2.4 Previous Research	4
3 Problem and Objectives	6
3.1 Problem Description	6
3.2 Objective, Metrics, and Potential Impact	6
3.2.1 Objective	6
3.2.2 Metrics	7
3.2.3 Impact	7
4 Data	8
4.1 Data and Cleaning	8
4.1.1 Positive Patient Data	8
4.1.2 Environmental Data	8
4.1.3 Sink and Room Characteristics	10
4.1.4 Interventions	11
5 Imputation Methodology and Results	14
5.1 Cases, Feature Extraction, and Feature Selection	14
5.1.1 Cases	14
5.1.2 Feature Extraction	14
5.1.3 Feature Selection	17
5.2 Methodology	18
5.2.1 Linear Method	19
5.2.2 Midpoint Method	19
5.2.3 Logistic Regression	20
5.2.4 Multiple Imputation - MICE	20
5.2.5 Analysis	21
5.3 Results	22

6	Modeling Methodology and Results	24
6.1	Cases, Feature Extraction, and Feature Selection	24
6.2	Methodology	25
6.3	Results	26
6.3.1	Logistic Regression Modeling	26
6.3.2	Random Forest	29
7	Conclusions, Limitations, and Future Work	32
7.1	Conclusions	32
7.2	Limitations and Future Work	34
7.2.1	Data Collection	34
7.2.2	Imputation Methods	35
7.2.3	Validation Set	36
7.2.4	Modeling	36
	Bibliography	37

Introduction

The spread of hospital-acquired antibiotic-resistant bacteria is a serious and growing problem. The Center for Disease Control and Prevention (CDC) estimate that more than two million people are sickened each year with antibiotic-resistant infections resulting in at least 23,000 deaths[1]. Of the many bacteria that are growing in drug resistance, there is particular concern with gram-negative pathogens as they are becoming increasingly resistant to nearly all drugs, including Carbapenems. Carbapenems are widely considered to be the strongest class of antibiotic in use today and are typically administered as a last resort in treating gram-negative infections[1]. Because of their resistance to the strongest antibiotics available, infections caused by Carbapenem-resistant Enterobacteriaceae (CRE), most commonly found in a patient's gastrointestinal tract, are difficult to treat and can result in death[2]. Though spread of these bacteria is currently limited to hospital facilities, if not controlled, it is likely that these bacteria could begin to spread to the outside community [1].

A large healthcare facility in the United States has been tracking low-level spread of a nosocomial Carbapenem-resistant Enterobacteriaceae pathogen since August of 2007. Approximately 450 patients have been colonized with the CRE pathogen since 2007 and more cases are identified weekly. Traditionally, understanding and tracking of hospital acquired infections focused on infected and colonized patients acting as reservoirs of transmission between uncolonized patients and hospital workers. However, it has become increasingly recognized that non-patient reservoirs within the hospital setting may play a larger role in the transmission of drug resistant pathogens than was first realized[3]. In this case, traditional epidemiological transmission (person-to-person) can only account for 20 to 30 percent of the transmission, and it is unknown how the other 70 to 80 percent of patients are acquiring the CRE pathogen[4].

Since discovery of this low-level occurrence, active surveillance of all patients was instituted

and strict contact precautions were emplaced for patients who revealed a positive culture for the bacteria. Additionally, large-scale cleaning, disinfecting, sterilization, replacement, and limited-use interventions were conducted, but only had temporary success in reducing the spread of the bacteria and not eradicating it as hoped. The rate of transmission has remained steady - without significant growth - over the past 5 years with approximately 70 new cases identified each year. Recent efforts to understand transmission of the waterborne bacteria led hospital epidemiologists to begin sampling wet surfaces including patient sinks, toilets, and hoppers for the CRE pathogen. Sampling found identical genomes in samples of patients as was found in the sinks, indicating that sinks - or other environmental factors- could be acting as reservoirs for transmission[5]. Beginning in September of 2013, nearly 3500 samples of sink drains, sink p-traps, toilets, hoppers, and other wet surfaces have been collected periodically throughout the hospital looking for the presence of CRE. Manpower and financial constraints limited the frequency and volume of samples taken over time. Understanding what is occurring between samples would provide a more complete picture of the lifecycle of this bacteria in sinks and if its' presence in sinks is indicative of a relationship between the sink and positive patients.

Literature Review

2.1 Risk Factors

Independent studies of similar occurrences in hospitals in Brazil and Greece indicated that the Intensive Care Units (ICUs) were the source of most transmission [6][7]. The Brazilian study found that factors associated with severity of illness, such as age and use of mechanical ventilation, were the primary risk factors for acquiring CRE [7]. The study in Greece found that duration of previous hospitalization and the existence of multiple chronic diseases to be primary risk factors[6]. Another outbreak of CRE at the Tisch Hospital at New York Medical Center found 24 infected patients in intensive care units (ICUs) over the course of a year. Similar to other outbreaks, risk factors for infection during this outbreak included prolonged hospital stay, a stay in the ICU, and ventilator usage. Principal nosocomial reservoirs that were reported included contaminated medical equipment, hands of hospital staff, and gastrointestinal tracts of patients[8].

2.2 Environmental Reservoirs

In a study of an outbreak at Dandenong Hospital in Melbourne, Australia, CRE was found in patient and provider hand washing sinks. A series of interventions and cleaning of sinks and sink traps were conducted, but CRE reemerged within days. The study was inconclusive about whether the sinks were definitively acting as the source of transmission, but the resistance of the organism in a wet environment is cause for concern and further investigation of this and other environmental reservoirs[9]. Additionally, a study at Deventer Hospital in the Netherlands conducted a prospective study that cultured sinks and patients over a 20-week period and identified 4 patients who became colonized with identical bacteria as was

previously found in the sinks[10]. This study also discussed the use of a self-disinfecting sink to successfully halt further spread of bacteria. However, in both cases, interventions were only found to be temporarily effective indicating the need for better understanding of the lifecycle of these organism and the role of the environment in their spread.

2.3 Imputation in Epidemiology

There exist many approaches to handling missing data, specific to the public health sciences field, such as ad hoc approaches of last observation carried forward and mean value replacement as well as other more computationally expensive methods like multiple imputation[11][12]. Regardless of method used, there is increased interest in the details of imputation and authors stress the importance of proper description within research in order to better understand the outcomes and to assist other researchers in furthering their use of imputation methods[11]. Though there is much literature on imputation of predictor variables, there is less guidance on the imputation of the dependent variable as we aim to do in this study. Researchers are hesitant to impute response variables as it is like “treating unknown outcomes as though they are known”[13]. However, using only cases with the dependent variable in tact can lead to bias, loss in overall sample size, and ultimately loss of information. In an article about handling missing data, the author stresses that while more robust statistical software is now available to researchers, it is still crucial to utilize expert knowledge of a data set or issue to ensure that the imputation methods and results make sense and are an accurate reflection of what is occurring[14].

2.4 Previous Research

Previous research and analysis of the current data only utilized a small portion of the data, from January of 2014 to April of 2014, to predict the outcomes of whether a patient would or would not acquire CRE based on patient data and some environmental data. Consistent

environmental sampling was only conducted during this short time frame which resulted in the modeling of only 19 of 407 positive cases in the data set. The results of these research efforts indicated that evidence was not conclusive that patient sinks played any role in the transmission of CRE. However, this research did indicate that there was some correlational between the environment and a patient's likelihood of acquiring CRE and indicated the need for further evaluation with a larger dataset of environmental samples and smaller time interval[15].

Problem and Objectives

3.1 Problem Description

Patients at a major healthcare system are acquiring an highly drug-resistant bacteria, CRE, through unknown means. Infections caused by CRE are virtually untreatable by antibiotics and can lead to death among some of the most ill patients. Traditional patient-to-patient transmission cannot account for all cases of transmission. Routine culturing of sinks in the hospital have found CRE bacteria in p-trap and drains of sinks and even the strongest interventions have been unsuccessful in eradicating the bacteria from environmental reservoirs. Limited and inconsistent sampling of environmental reservoirs leave a gap in knowledge of its presence in sinks over time and the relationship of sinks to positive patients.

3.2 Objective, Metrics, and Potential Impact

3.2.1 Objective

The objective of this research is two fold -

- develop an imputation method that provides an accurate representation of bacterial presence in sinks between known samples of data.
- develop a predictive model for whether a sink will test positive for the CRE pathogen based on presence of colonized patients in the same room as the sink, emplacement of clinical interventions in sinks, and the physical characteristics of the sink.

3.2.2 Metrics

- Imputation Methods: Feature selection will be conducted using stepwise regression with Akaike information criterion (AIC). We will use analysis of variance (ANOVA) with chi-squared (χ^2) test to confirm the correct features were selected. Methods will be measured based on their improvement to baseline prediction accuracy using 10-fold cross validation (CV).
- Modeling: Feature selection will be conducted similarly to the method for imputations - stepwise regression with AIC and χ^2 ANOVA. Models will be assessed based on AIC, analysis of diagnostic plots, overall prediction accuracy from 10-fold cross validation, and Receiver operating characteristic (ROC) curve comparison.

3.2.3 Impact

Developing an imputation method for bacterial presence between known samples will provide valuable insight into the growth and movement of the bacteria throughout sinks in the hospital. Additionally, the ability to complete a data set of environmental testing allows for the capability for it to be included in other models, such as patient risk modeling, that will provide a more comprehensive predictive capability. In modeling the presence of CRE in sinks, this research aims to understand what factors contribute to sink positivity. Knowledge gain in both capacities will lead to better understanding the role of environmental reservoirs in the spread of the bacteria and could be used to support changes to hospital policy and procedures that ultimately aid in the containment or eradication of the bacteria from the hospital.

Data

4.1 Data and Cleaning

4.1.1 Positive Patient Data

Positive patient data included a de-identified list of 444 patients from between August 2007 and September 2015 who were tested for CRE and had a positive culture at any time during their admittance to the hospital. Data points included the initial date of admittance, the first date of positive culture, and date, time, and location of all bed movements throughout the hospital for all admittances to the hospital. Bed movement data for a patient was removed for any time period before their positive culture date. Bed movement data was rolled up to the day and room level. If the patient spent any length of time in a room on a given day, then the entire day was considered to be a positive day. Additionally, the data provided included information on what bed in a room that the patient occupied. For this analysis, we considered all beds in a given room to be the same room.

4.1.2 Environmental Data

Environmental data includes 3420 positive and negative swab and liquid samples from sink drains and sink p-traps in patient rooms, bathrooms, anterooms, alcoves, staff bathrooms, locker rooms, kitchens, and staff lounges for floors 3-8 in the hospital. Samples were collected between September of 2013 and September of 2015. Cultures were collected using clinical swabs that are applied to the interior wall of the sink drain. Water samples from p-traps were collected utilizing IV tubing that was lowered through the drain hole to collect water. Samples are processed in a microbiology lab over a 3-day period allowing for growth of

bacteria and further identification of specific organism found. This type of sampling has an estimated 90 percent sensitivity for identifying CRE organisms[4].

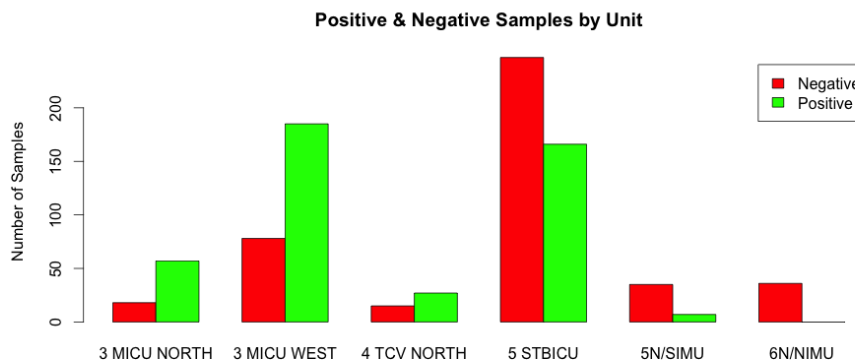


Figure 4.1: Bar plot of total samples by unit, Sample Dataset

One of the main issues with the environmental data provided is inconsistency in the sampling. Samples taken vary by week, by location, by sampling technician, and testing technician. Sampling frequency varied from weekly at some points to only quarterly or singular at others. Additionally, while positive samples had more refinement on their location - patient sink vs. bathroom sink, and p-trap vs. drain - negative samples lacked the same level of detail. Conversations with hospital epidemiology revealed that when samples were taken in a unit - regardless of outcome - all sinks in the unit were sampled and samples for both drain and p-traps were always taken[4]. Using this information and blueprints (as seen in Figures 4.2 and 4.3) we were able to manually impute locations for both positive and negative samples. If at any point we were unable to definitively determine the location from which the sample was taken, the observation was removed from the data set.

For the purpose of this analysis, we examined samples at the sink level. In some cases, a sample from drain of a sink may be positive while the p-trap may be negative, or vice versa. For these cases, if any sample of a sink on a given day was positive, the sink was annotated as a positive sink for that day. Additionally, the focus of this research was to examine the movement of bacteria from patients to sinks. Due to this scope and the lack of



Figure 4.2: Hospital floor plan, 3rd floor, rooms without patient sinks

other supporting data points for other sink types, samples for non-patient sinks, such as staff lounge rooms and kitchens were removed. Samples for which no sink characteristics were provided were also removed. Of the given 3420 test samples, 1907 samples were retained for modeling purposes. Upon consolidating p-trap and drain samples to a single sink result, 967 results remained for 166 patient and patient bathroom sinks over a period of 740 days.

4.1.3 Sink and Room Characteristics

A series of 26 attributes were manually collected describing the physical characteristics of the sink, surrounding countertop area, faucet, exposed piping, and the presence of other wet areas such as showers, hoppers, and toilets. The distance of the patient bed to the sink is

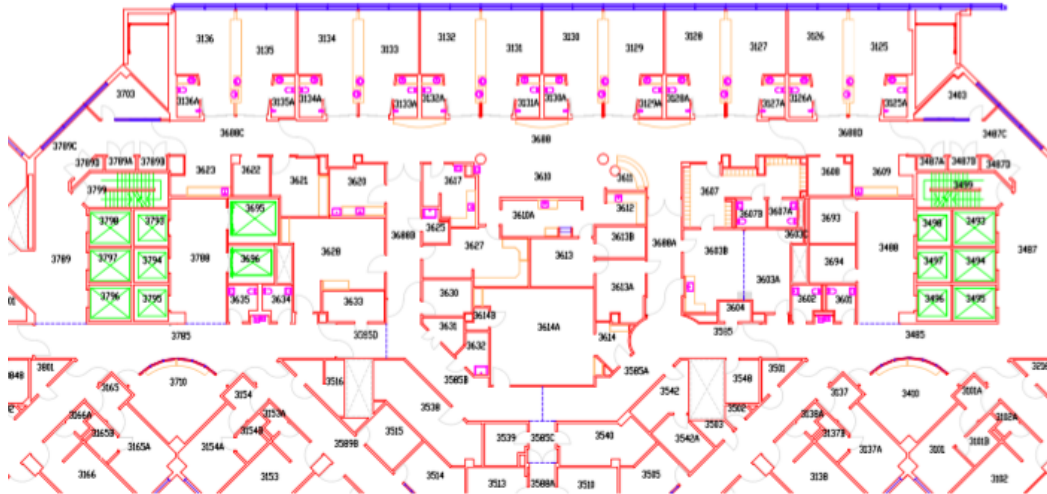


Figure 4.3: Hospital floor plan, 3rd floor, rooms with patient sinks

also provided. A full description of characteristics can be found in Table 4.2

4.1.4 Interventions

Dates and Units of any clinical interventions were provided. It was verified that all sinks in a given unit were held to an intervention during the days given. Interventions included the replacement of pipes, the closure of the unit to all patients, and cleaning and disinfecting interventions as seen in Table 4.1.

Table 4.1: Interventions

Unit	Rooms	Begin Date	End Date	Intervention
5 STBICU	All	12/10/13	2/10/14	Unit closed to patients. All sink p-traps replaced. Overflows removed from sinks. Unit reopened on end date.
4 TCV West	All	12/10/13	2/10/14	Overflows removed from sinks
5 STBICU	All	2/28/14	4/18/14	Weekly sink treatment/flushing with bleach, hydrogen peroxide, and ozone.
3 MICU West	All	3/14/14	4/18/14	Weekly sink treatment/flushing with bleach, hydrogen peroxide, and ozone.
3 MICU West	All	4/25/14	5/10/14	Weekly sink treatment/flushing with bleach, hydrogen peroxide, and ozone.

Table 4.2: Sink Variable Description

Variable	Type	Description	Levels or Units
Bowl.Material	categorical	material of sink bowl	“stainless steel”, “porcelain”, “corian”
Bowl.Shape	categorical	shape of sink bowl	“kidney”, “oval”, “knob”
Faucet.Type	categorical	type of faucet	“auto”, “paddle”
Counter	categorical	presence of counter	“yes”, “no”
Countertop.Material	categorical	material of countertop	“black corian”, “copper”, “corian”
Counter.Length	numeric	length of countertop	inches (24.5 - 71.5)
Aerator	categorical	presence of aerator	“yes”, “no”
Drain.Cover	categorical	presence of drain cover	“yes”, “no”
LidDrain.Type	categorical	type of lid or drain	“empty”, “grate”, “fan”
Overflow	categorical	presence of overflow	“yes”, “no”
Overflow.Location	categorical	location of overflow	“back”, “front”, “no”
Overflow.Number	numeric	number of overflows	count (0-2)
p.trap	categorical	type of p-trap	“normal”, “handicap”, “inaccessible”
F.to.D.Distance	numeric	faucet to drain distance	inches (7.75 to 40)
Bowl.Size.X	numeric	width of sink bowl	inches(12-45)
Bowl.Size.Y	numeric	length of sink bowl	inches(8.5-26.5)
Bowl.Size.Z	numeric	depth of sink bowl	inches(4-24)
P.Trap.Material	categorical	material of p-trap	“stainless steel”, “pvc”, “inaccessible”
Top.to.P.trap.D	categorical	distance from drain to p-trap	inches(0-19.5)
Water.Hit.Drain.	categorical	does water hit the drain directly	“yes”, “no”
Dialysis.Faucet	categorical	presence of dialysis faucet	“yes”, “no”
Toilet	categorical	presence of toilet in room w/sink	“yes”, “no”
Hopper	categorical	presence of hopper in room w/sink	“yes”, “no”
Floor.Drain	numeric	number of floor drains	count(0-2)
Shower	categorical	presence of shower in room w/sink	“yes”, “no”
SinkBedDis	numeric	distance from bed to sink	feet(4-14)

Imputation Methodology and Results

5.1 Cases, Feature Extraction, and Feature Selection

5.1.1 Cases

For the purpose of this analysis, our dataset was converted to a sink-day format that includes a record for each sink and each day during the given time period. We defined our positive (“1”) population as the set of sink-days for which a known positive culture was found as detailed in section 4.1.2. We define our negative (“0”) population as the set of sink-days for which a known negative culture was found. We use the table below to describe the sink-days for both the sample data set and the modeling data set. However, the sample data set uses only known values while the modeling data set uses the full set of imputed data.

Table 5.1: Variables in the Sink-Day Data Format

Variable	Range	Description
Room	unique ID for hospital room	repeated across all records pertaining to a given room
Sink	“PT” “BR”	identified patient or bathroom sink for given room
Day	09-16-2013 to 09-25-2015	repeated across all sinks
Event Indicator	“1” = positive “0” = negative	indicates whether sink tested positive or negative for CRE

5.1.2 Feature Extraction

Using bed movement data, we created variables Lag0, Lag1, Lag2, and Lag3 that annotated whether on the given day, a positive patient moved through the room. The lag number corresponds to the time. For example, Lag0 would annotate time=present. On the given day,

for the given sink, a binary response of “1” was annotated if a positive patient moved through the room. Similarly, Lag3 annotates that 3 days prior to the given date, for the given sink, a binary response of “1” annotated if a positive patient moved through the room. Similarly, CumSum7 and CumSum14 were created to show the cumulative number of days in the past 7 or 14 days respectively that a positive patient was in the given room with the given sink. In a similar fashion as the Lag variables created, Intervention0, Intervention1, Intervention2, and Intervention3 were created to annotate whether at the given time (present day to 3 days prior) if there was an intervention in place in the given sink. Multiple Correspondence Analysis of sink characteristics revealed many levels of multicollinearity. For this reason, all sink characteristics were combined together and 25 unique sinks designs were found. Upon further analysis, it was found that the difference between some sinks were likely due to human error when measuring distances. Therefore, if measurements of a single distance was the only difference between two otherwise identical sinks and were within 1 unit of each other, they were combined to create a single sink design. The distance from the sink to the bed was retained as an individual feature as it was a characteristic of the room. The resulting 16 sink designs can be found in the appendix. Description of extracted variables can be found in Table 5.2.

Table 5.2: Extracted Variable Description

Variable	Type	Description	Levels or Units
Unit	categorical	unit of hospital	“3 MICU North”, “3 MICU West”, “4 TCV North”, “5 STBICU”, “6N/NIMU”
PTBR	categorical	patient (PT) or bathroom sink (BR)	“PT”, “BR”
Lag0	categorical	presence of positive patient for sink/day at time t=0	“0”, “1”
Lag1	categorical	presence of positive patient at time t-1	“0”, “1”
Lag2	categorical	presence of positive patient at time t-2	“0”, “1”
Lag3	categorical	presence of positive patient at time t-3	“0”, “1”
Intervention0	categorical	intervention in place at time t=0	“0”, “1”
Intervention1	categorical	intervention in place at time t-1	“0”, “1”
Intervention2	categorical	intervention in place at time t-2	“0”, “1”
Intervention3	categorical	intervention in place at time t-3	“0”, “1”
CumSum7	numeric	cumulative number of positive patients in 7 days prior	count (0-7)
CumSum14	numeric	cumulative number of positive patients in 14 days prior	count (0-14)
SinkDesign	categorical	sink design based on unique values from sink variable description	25 designs “A-Y”
SinkBedDis	numeric	measurement of distance from sink to bed	feet (4-14)

5.1.3 Feature Selection

We first began feature selection by conducting a stepwise regression, in the forward and backward directions, of the given and extracted features. Stepwise regression, using Akaike information criterion (AIC) as the evaluation measure eliminated the following factors. Factors shown as ~~strikethrough~~ where eliminated and those shown without strikethrough were retained. The beginning AIC for all variables was 609.78 and ending AIC was 596.2. Factors were confirmed by using a sequential chi-squared test of analysis of variance of the addition of evaluation measures. Factors retained by the stepwise regression were statistically significant at a p-value of less than .001 indicating that the sequential inclusion of each additional variable provided information gain to the final model. The chi-squared test was conducted for factors eliminated by stepwise regression and found that the variables were statistically significant at p-values greater than .001 and therefore provided no additional gain to the model. Table 5.3 provides an outline of the resulting features and their resulting chi-squared value.

Table 5.3: Stepwise and Chi-Squared Test Results

Stepwise	Chi-Squared Test Results
Unit	2.2e-16
PTBR	.247
Lag0	.379
Lag1	.218
Lag2	.490
Lag3	.465
Intervention0	.245
Intervention1	.072
Intervention2	3.14e-04
Intervention3	.654
CumSum7	.410
CumSum14	2.2e-16
SinkDesign	1.21e-07
SinkBedDis	.178

Throughout model building, we identified some multicollinearity between Sink Design and

other predictors which would result in no coefficients being given. We determined that all bathroom sinks (BR) in the dataset are of Sink Design “STUV”. Therefore, we removed the variable “PTBR” from further model with the understanding that this distinction was already represented in the Sink Design attribute. Additionally, we determined that Sink Design “BC” is only in Unit 3 MICU West. We did not remove either variable from consideration as there were many other levels in each of these variables. However, we were mindful of this issue when analyzing coefficients in our models.

5.2 Methodology

Missing data is a frequent issue in epidemiological and health sciences research. While most modeling techniques would simply call for the removal of incomplete cases, doing so can cause bias and loss of information[16]. Standard methods for imputation such as k-nearest neighbor, last observation carried forward, or mean value can also lead to bias and are not effective in handling categorical variables[11]. These methods are also typically used to address one of the three types of missing data - missing completely at random, missing at random, and missing not a random. In this case, we see all three types of missing data present in 1 variable. And not just 1 predictor variable, but in this case, the response variable, adding to difficulty of using traditional imputation methods. Additionally, imputation is typically used for filling in a few missing values with the help of many present values. In our case, missing values far outnumber present values. For these reason, we chose to examine 2 ad hoc imputation methods as well as 2 regularly practiced methods.

For the purpose of imputation, we consider each sink to be it’s own time series object. Methods described below are implemented on each sink over the period of 740 days resulting in a value for each sink-day that will be used later in modeling.

5.2.1 Linear Method

The 2nd method we examined was a linear probability. As stated in the background, in most cases, once a sink was tested positive, it stayed positive until some type of intervention was emplaced. Sinks were retested after the intervention and samples were annotated as negative. However, at various time points following a previous negative sample, a sink was determined to have gone positive again. Given this information we used a linear imputation for time between a negative and positive sample annotating the probability of the sink testing positive linearly increasing between 0 and 1. We carried the last observation forward from a positive sample until the day before a negative sample (see plot) to accurately describe emplacement of an intervention. Time between two like observations were filled in with the consistent observation (negative or positive). The result of this method was a numeric probability between 0 and 1.

5.2.2 Midpoint Method

The first method we considered is straightforward. At some point between a negative and positive observation (or a positive and negative observation), the sink transitioned from negative to positive (or positive to negative). Therefore, time between known samples was divided in half and the value of the 1st observation was carried forward to the half way point and the value of the next observation was carried backward to the halfway point creating a step function over the time series. Time between consistent samples (negative to negative or positive to positive) was carried forward. Tails on the front end were carried backward from the 1st known observation and tails on the back end were carried forward from the last observation to the end of the time period. The result of this method was a binary response for all observations.

5.2.3 Logistic Regression

Lastly, using all of the given information, we conducted a logistic regression of the given variables and predicted a response. The model was trained based on the Sample data frame and predicted on the missing values in the full data frame. The response represents the probability (between 0 and 1) of the sink-day being positive based on all variables available. We replaced predicted values with known values.

5.2.4 Multiple Imputation - MICE

Using the MICE (Multivariate Imputation by Chained Equations) package in R, we conducted a univariate multiple imputation of 5 iterations on 5 data sets and the output was a binary response based on a logistic regression of all other known and complete variables[17].

Figure 5.1 depicts the known values (red) and the predicted values with each imputation methods (blue - Linear, green - Midpoint, purple - Logistic Regression, black - Multiple Imputation).

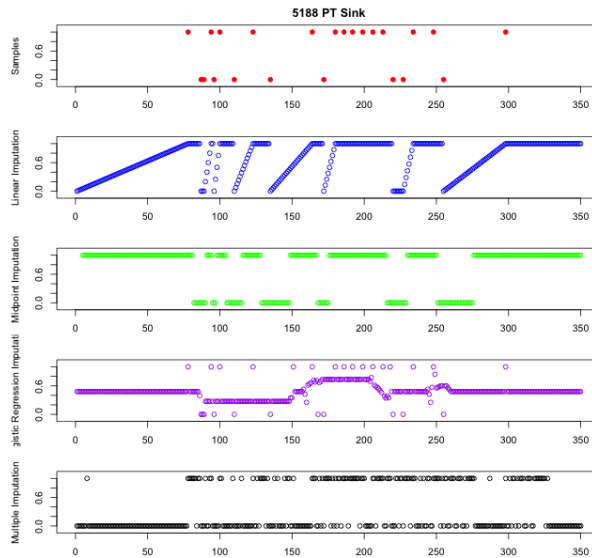


Figure 5.1: Example plot of Imputation Methods for Patient Sink in Room 5188

Table 5.4: Imputation Analysis Variable Description

Variable (X_5)	Description
Mid3	imputed data at time t-3 (3 days prior to date) using Midpoint method
Line3	imputed data at time t-3 (3 days prior to date) using Linear method
LogReg3	imputed data at time t-3 (3 days prior to date) using Logistic Regression method
Mice3	imputed data at time t-3 (3 days prior to date) using MICE method
Mid7	imputed data at time t-7 (7 days prior to date) using Midpoint method
Line7	imputed data at time t-7 (7 days prior to date) using Linear method
LogReg7	imputed data at time t-7 (7 days prior to date) using Logistic Regression method
Mice7	imputed data at time t-7 (7 days prior to date) using MICE method
Mid15	imputed data at time t-15 (15 days prior to date) using Midpoint method
Line15	imputed data at time t-15 (15 days prior to date) using Linear method
LogReg15	imputed data at time t-15 (15 days prior to date) using Logistic Regression method
Mice15	imputed data at time t-15 (15 days prior to date) using MICE method
Past.30	binary ("0","1") for presence of positive sink culture in the past 30 days
Past.60	binary ("0","1") for presence of positive sink culture in the past 60 days
Past.90	binary ("0","1") for presence of positive sink culture in the past 90 days

5.2.5 Analysis

Using the 3 imputation methods, we utilized our Sample data frame consisting of 967 known positive samples to test our imputation methods. We began by evaluating a simple generalized linear model (logistic regression) of the given samples as a response to the extracted and selected features in Table 5.3. g_t annotates the probability of the a positive sink given the set of predictor variables.

$$g_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 \quad (5.1)$$

$$\log \frac{Pr(Y = 1)}{1 - Pr(Y = 1)} = \beta_0 + \beta_1(\text{Unit}) + \beta_2(\text{CumSum14}) + \beta_3(\text{SinkDesign}) + \beta_4(\text{Intervention2}) \quad (5.2)$$

Next, we sequentially add in an imputed value from each method at t-3 (3 days prior to sink-day), t-7 (7 days prior to sink-day), and t-15 (15 days prior to sink-day) to determine

which imputation method provided the most increase in the prediction accuracy of the sink based on 10-fold cross validation.

$$g_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 \quad (5.3)$$

$$\begin{aligned} \log \frac{Pr(Y = 1)}{1 - Pr(Y = 1)} = & \beta_0 + \beta_1(\text{Unit}) + \beta_2(\text{CumSum14}) \\ & + \beta_3(\text{SinkDesign}) + \beta_4(\text{Intervention2}) + \beta_5(\text{Imputed Value}) \end{aligned} \quad (5.4)$$

As a method of comparison, we also created a known variable at 3 different time intervals that annotated (“0” or “1”) whether in the past 30, 60, and 90 days a sink had a positive culture. In the same manner as above, this variable was sequentially included in the model and evaluated based on 10-fold cross validation accuracy.

5.3 Results

We find that regardless of time interval, the Midpoint imputation method provides the most boost to our accuracy of prediction of positive sink compared to the baseline (without any imputed values). However, it is also interesting that the addition of the Past.30 variable (which is not imputed) increased the accuracy of the model significantly as well. Given the improved results, we will include this variable as a predictor in modeling of the imputed dataset.

Table 5.5: Imputation Method Accuracy Results

Method	t-3	t-7	t-15
Linear	82.4%	79.9%	78.1%
Midpoint	98.2%	86.1%	80.5%
Logistic Regression	69.7%	72.0%	70.4%
Mice	71.2%	71.0%	71.3%
Baseline (no imputation)	70.4%		

Table 5.6: Look-back Accuracy

Variable	Accuracy
Past.30	83.9%
Past.60	80.7%
Past.90	79.1%

Though the focus of this portion of this thesis was to determine that best imputation method for environmental sampling, it is also interesting that during feature selection, the model was reduced to only 4 data points - unit, cumulative number of positive patients having stayed in the room 14 days prior prior to the culture, sink design, and whether an intervention had been implemented in the last 2 days. Further discussion of these factors is included in the modeling results section.

Modeling Methodology and Results

6.1 Cases, Feature Extraction, and Feature Selection

Using the Midpoint imputation as the response variable, the non-imputed variable - Past.30, and the original set of features, a new dataset was created for modeling. The dataset consisted of 81,858 complete cases using the same sink-day format as described in Chapter 5. A breakdown of the positive and negative samples can be found in Figure 6.1.

Similar to the method utilized for the Sample dataframe, stepwise regression was used for feature selection and confirmed using sequential testing of the ANOVA using the χ^2 test. As a reference for comparison, both sets of features are shown below. We would expect the features selected to be different between models, else there is no value added in completing an imputation model. The argument for imputing values is to account for any bias that may be eliminated by only using complete cases, as in the Sample dataframe.

Table 6.1: Stepwise and Chi-Squared Test Results

Sample Dataframe		Imputed Dataframe	
Stepwise Results	Chi-Squared Test Results	Stepwise Results	Chi-Squared Test Results
Unit	2.2e-16	Unit	2.2e-16
Lag0	.379	Lag0	2.2e-16
Lag1	.218	Lag1	.623
Lag2	.490	Lag2	.245
Lag3	.465	Lag3	.594
Intervention0	.245	Intervention0	2.2e-16
Intervention1	.072	Intervention1	.072
Intervention2	3.14e-04	Intervention2	.916
Intervention3	.654	Intervention3	2.6e-13
CumSum7	.410	CumSum7	2.2e-16
CumSum14	2.2e-16	CumSum14	2.2e-16
SinkDesign	1.21e-07	SinkDesign	2.2e-16
SinkBedDis	.178	SinkBedDis	2.2e-16
		Past.30	2.2e-16

6.2 Methodology

Using `glm()` package in R, we began with a simple logistic model using all variables and selected variables on the imputed data set. For this model we established the intercept as Unit 6N/NIMU, Sink Design F, and "0" for all other categorical and numerical variables. As shown below in figure 6.1, Unit 6N/NIMU reflects primarily negative samples making it a good base case for comparison. Similarly, SinkDesign "F" has the least number represented in the model, as shown in Figure 6.2, so it was made the base case (intercept)[18].

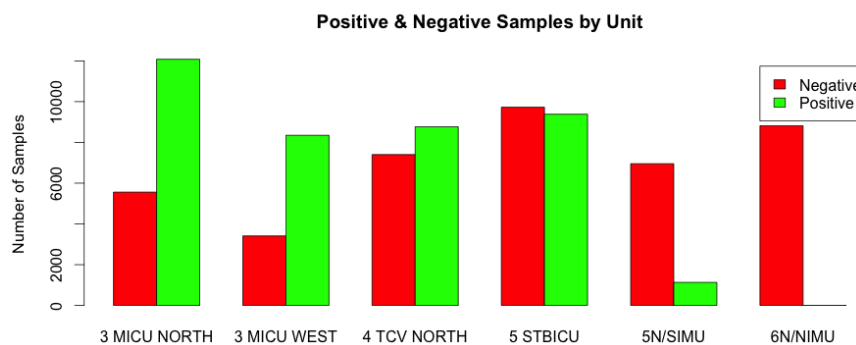


Figure 6.1: Bar plot of samples by Unit, Imputed Dataset

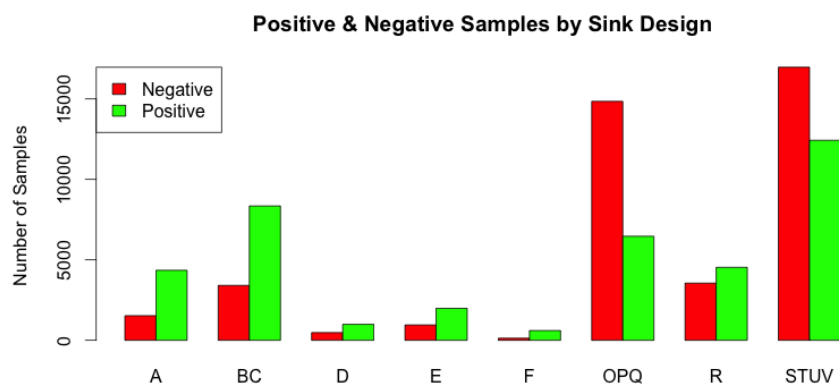


Figure 6.2: Bar plot of samples by Sink Design, Imputed Dataset

We also conducted a stepwise regression of all possible interactions of the selected variables. Lastly, we examined a classification machine learning technique, random forest, to examine feature importance and assess accuracy. We compared all 4 models by looking at AIC (when available) and accuracy of prediction based on 10-fold cross validation.

6.3 Results

6.3.1 Logistic Regression Modeling

Diagnostic plots for models utilizing all variables, stepwise selected variables, and interaction of selected variables are shown below. There is little difference between both the diagnostic plots and accuracy of the model with all variables and the model with variables selected via stepwise regression. This indicates that there are many variables within the current data frame that are not significant predictors of sink positivity and were therefore removed during stepwise regression. Though the overall accuracy of the interactions model only increased by 1%, we can see from the diagnostic plot below that there is significant improvement of the goodness of fit from the model with all variables and the model with variables selected via stepwise regression. Figure 6.5 shows more evenly distributed residual vs. predictor plot and significant improvement to the number of observations located above Cook's distance in both the standardized leverages plot and the total cases plot.

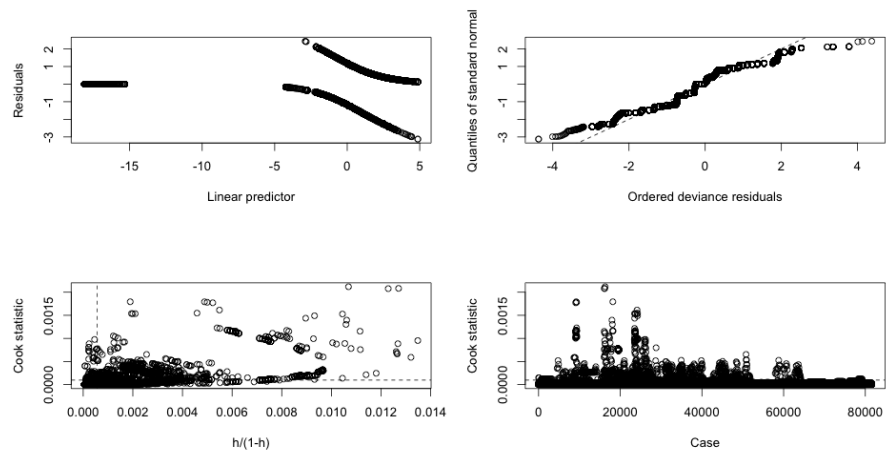


Figure 6.3: diagnostic plots for model with all variables

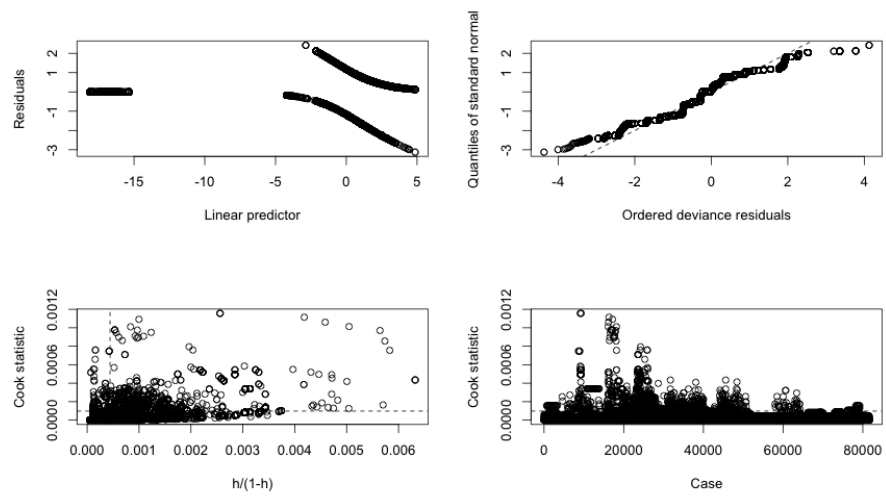


Figure 6.4: diagnostic plots for model with variables from stepwise regression

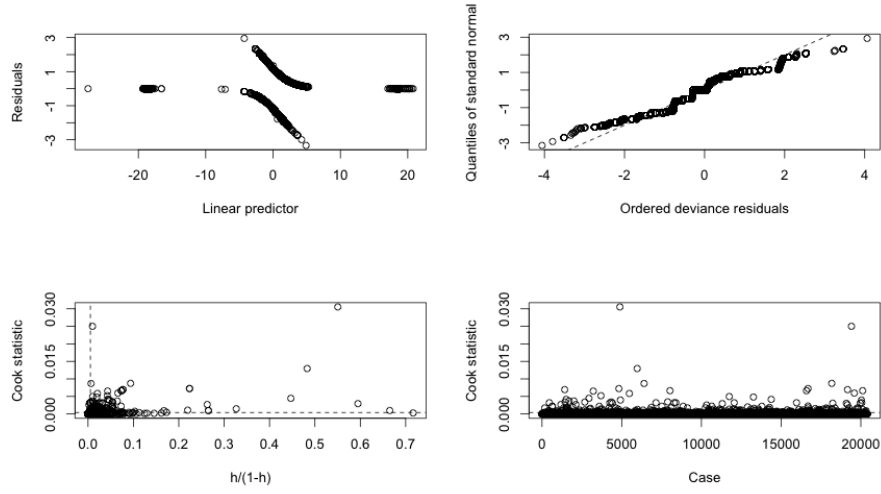


Figure 6.5: diagnostic plots for model with interaction of variables from selected variables

Additionally, the receiver operating characteristic (ROC) curve is shown for all 3 models in Figure 6.6. The ROC curve is a graphical representation of the true positive rate against the false positive rate. As suspected, the ROC curve for the model with all predictor variables is the same as the curve with variables selected from stepwise regression. The ROC curve for the interactions model lays slightly higher than that of the other two models indicating that it performs better in determining positivity of sinks.

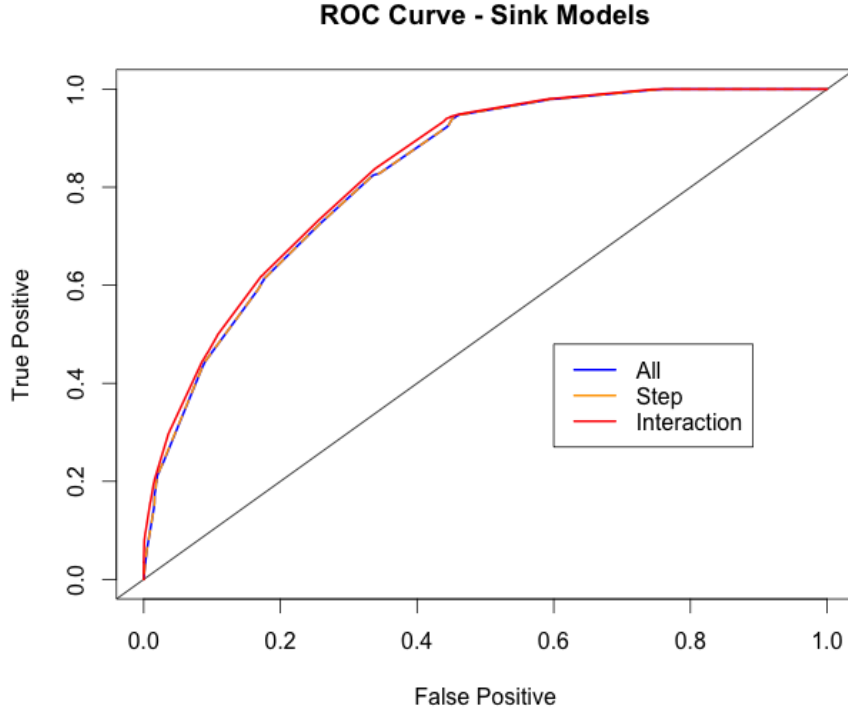


Figure 6.6: ROC Curve for all 3 models

Table 6.2 shows the AIC and accuracy of all 3 models. Accuracy was assessed using 10-fold cross validation. Using both selection criteria (AIC) and accuracy, as well as visual inspection of the diagnostic plots and ROC curves, the interactions model clearly performs the best of the 3 logistic regression techniques explored in this study.

Table 6.2: Model Results

Model	AIC	Accuracy
All Variables Model	79318	74.1%
Stepwise Variables Model	79308	74.2%
Interactions from Stepwise Variables Model	77617	75.2%

6.3.2 Random Forest

As a method of comparison, we also examined a popular classification machine learning techniques to further examine which variables are important in the prediction of sink positivity

and to determine if classification of this nature can produce more accurate results. We trained a random forest model, with 500 trees and max nodes of 3 on 75% of the data and used the remaining 25% as our testing set. The resulting random forest produced an accuracy of 75.1%. Table 6.3 describes the Mean Decrease Accuracy (MDA) factor Mean Decrease Gini (MDG) factor for each variable. MDA is a measure of how much inclusion of the predictor in the model reduces the classification error. MDG is a measure of the average number of times the variable was used to split a tree node. Figure 6.7 is a graphical representation of variables importance based on both MDA and MDG factors as described above.

Table 6.3: Variable Importance for Random Forest

Variable	MDA
Unit	48.0
Lag0	18.9
Lag1	17.2
Lag2	13.1
Lag3	15.3
Intervention0	27.8
Intervention1	20.8
Intervention2	22.2
Intervention3	28.4
CumSum7	23.0
CumSum14	31.5
SinkDesign	30.3
SinkBedDis	18.9
Past.30	40.6
MDA = Mean Decrease Accuracy, a measure of how much inclusion of this predictor in the model reduces classification error	

Variable Importance Random Forest Imputed Dataframe

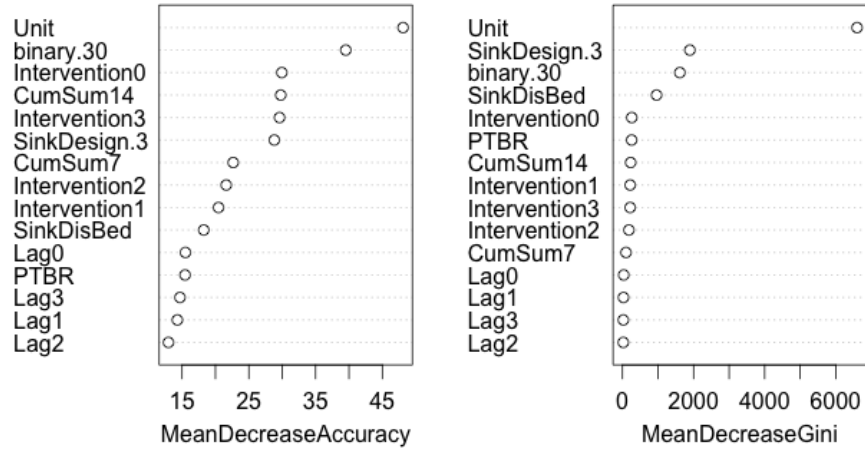


Figure 6.7: Variable Importance for Random Forest

As shown below, our Random Forest Model performed the best in classifying sink positivity with accuracy of 75.1%.

Table 6.4: Model Results

Model	AIC	Accuracy
All Variables Model	79318	74.1%
Stepwise Variables Model	79308	74.2%
Interactions from Stepwise Variables Model	77617	75.2%
Random Forest		75.1%

Conclusions, Limitations, and Future Work

7.1 Conclusions

Though each model identified a slightly different set of variables that were significant or important in the prediction of sink positivity, there are a few factors that remained consistent throughout - unit, an intervention at times between 0 and 3 days before positive culture (Intervention0 or Intervention3), the cumulative sum of patients having been in the room in the 14 day to the positive culture (CumSum14), presence of a positive culture within 30 days prior (Past.30), sink design (SinkDesign), and distance from patient bed to the sink (SinkBedDis). It is consistent with expert knowledge of this issue specific to the hospital that unit would play a major role in the spread of the bacteria. Coefficients for all Units except the intercept (Unit 6N/NIMU) were positive indicating that sinks in any other unit than 6N/NIMU have a significant higher odds of being positive. Resulting negative coefficients for interventions at time zero and time t-2 indicate that the odds of a sink testing positive for the CRE bacteria decrease with the presence of an intervention. This is again consistent with expert knowledge that interventions temporarily decreased the presence of the bacteria, but it would reappear soon after. Additionally, we found that for each unit increase in the distance between the patient bed and sink (measured in feet) there was a 10% decrease in the odds of the sink being positive. We would expect this to be true showing that the further away the patient is from the sink, the less likely that the bacteria is to be in the sink. Perhaps the most interesting finding was the strong correlation between the cumulative number of patients present in the same room as the sink and the odds of the sink testing positive for the bacteria. Our best regression model indicated that the odds of a positive culture increased

15% percent for every day that a patient was present in the room prior to the positive culture. Interestingly, the presence of positive patients 1,2, and 3 days prior to a positive culture had very little effect on the overall adds, it was the cumulative presence of patients that had the most effect. This finding supports the hypothesis that CRE bacteria is transmitting from the patient to the sink. Additionally, with SinkDesign “F” modeled as the intercept, all remaining coefficients for SinkDesign were negative indicating that the odds of the sink being positive were negative compared to that of SinkDesign “F”. Further analysis of the SinkDesign “F” found that this particular sink was associated with the patient sink found in from 5196. It is the only sink of its type in our data set. Further analysis of this SinkDesign and room is warranted to determine if this is simply an outlier or a mode of transmission that should be considered to be more vulnerable to transmission than others.

Throughout the course of this work, many other forms of regression and classification modeling were considered in a effort to increase accuracy and fit of the model. We examined transformation of numerical predictors, mixed effect models, and other ensemble methods like boosting. However, in all cases, accuracy remained consistent with the results presented - around 74 to 75% accuracy. As shown here, even random forest machine learning technique, only made a slight improvement to the regression techniques used. This leads us to believe that we have reached a maximum capacity for predicting the outcome of a positive sink with the current variables used.

Additionally, we found a false negative rate of 13.5% where our models predicted that the sink was not positive when, in fact, it was. Further analysis of these specific cases show that nearly 80% of those cases had no presence of positive patients at any time in the previous 14 days prior to the positive culture, no interventions emplaced in the previous 3 days, and no positive culture in the previous 30 days. These findings indicate that perhaps there is another mode of transmission that is unrelated to the presence of patients in the room. Upon presenting these findings to hospital epidemiology, they confirmed that an additional hypothesis was the movement of the bacteria was retrograde through piping from other

positive sinks. An additional element that could be added to further explore this hypothesis is the distance and structure of piping between sinks.

7.2 Limitations and Future Work

7.2.1 Data Collection

Though much work has been done to collect the most accurate and up-to-date information regarding patient movement and cultures, much of the collection for this specific bacteria is done by hand and is QAQCed (quality assured/quality controlled) by a single entity. With so few positive patients (approximately 450 out of 15,000 patients per year), the addition or subtraction of a single patient from the positive patient pool could have a significant impact on the outcome of modeling.

Additionally, it has been identified through individual patient record review that for some positive patients, incomplete bed movement data was provided. Manpower and time constraints prevent review of all 450 positive patient records, but future analysis will require more refined analysis of patient movement in order to more fully understand the impact of a positive patient on the positivity of sinks.

Though the environmental data provided fine-grain detail of sink characteristics, it did not include other environmental factors that could be evaluated in further modeling. Items such as distance from patient beds to sinks, presence of hand sanitizing dispensers, and other medical equipment are not currently included in the model. This analysis has shown that there is a relationship between the environment, patients, and the spread of this bacteria and therefore warrants further examination of other environmental reservoirs that could be playing a role in its spread.

We would also recommend taking a further look into the interventions that were emplaced in units throughout the time period. Information was provided to us that interventions were emplaced for said units at said times. However, upon speaking with professionals who

emplaced interventions, it was determined that some sinks were annotated as 'control' sinks and no intervention was emplaced. Further detail about which sinks and which interventions were emplaced could assist in further understanding the spread of the bacteria and improving prediction results of the models explored in this research.

Most importantly, we recommend a steady and consistent testing protocol of sinks and environmental reservoirs in the hospital. Routine testing and consistent testing procedures will go a long way in assisting further researchers in understanding the movement of this bacteria through environmental reservoirs and provide for the ability to develop a true validations set for testing of our imputation methods and models. If financial and human resources are limited, I would recommend weekly sampling of at least 3 sinks in each unit, preferably at least one known positive sink and one known negative sink in each unit. I would also recommend including room 5196 in routine sampling.

7.2.2 Imputation Methods

Though the midpoint method proved to be the most accurate reflection of the presence of the bacteria over time, it is limited by the number of samples provided. For some sinks, only 3 total samples were collected in 2 years time making a large gap between samples. With such a large gap between samples, there is a high chance for error in predicting positivity. As discussed above, more routine and frequent sampling would improve not only this methodology, but also the overall understanding of the bacteria's movement over time.

Also, our literature review has revealed many different imputation methods that could be explored that would further refine our understanding of the spread of the bacteria. For the purpose of this analysis, the assumption was made that an intervention of some sort had to have been implemented in order for sink to go from a positive culture to a negative culture and this was reflected in the linear regression imputation. However, in some cases, an intervention could have simply been that the sink was put out of use by patients and staff. Therefore, for these specific cases, a linear or logistic imputation from a positive to a negative

sample may more accurately reflect the probability of a positive sink culture. Additionally, as variables and data are more refined as described above, a reevaluation of a logistic regression of positivity may provide a more accurate picture and ultimately a better prediction.

7.2.3 Validation Set

For the purpose of determining our best imputation method, we used cross validation across the entire Sample dataframe to determine which imputation method provided the most gain to our overall prediction accuracy. Ideally we would have liked to have tested our imputation methods on a completely separate validation set. There is currently work being done on patient risk and we could potentially test various imputation methods in these models to determine which provides the most gain to their modeling. Additionally, this is an ongoing project with continuing data collection. Due to the limitation of data available at the time, our time period only stretches until September 2015. We could use an additional data set with samples and supporting predictors for a time period after September 2015 in order to confirm our findings.

7.2.4 Modeling

There is much that can be done to further this work. It is clear that there is a relationship between the environment and the spread of CRE within the hospital setting. While we have provided convincing evidence that bacteria are moving from patient to sink within the hospital, there is much work that still needs to be done to confirm whether the sink is acting as a transmission reservoir back to the patient. Further analysis of exact genome found in sinks and in patients would further help us to understand not only the movement of the bacteria but the metamorphosis of genomes throughout environmental reservoirs.

Bibliography

- [1] Tom Frieden. Antibiotic resistance threats in the United States, 2013. Technical report, 2013.
- [2] Alexander J. Kallen. Vital Signs : Carbapenem-Resistant Enterobacteriaceae. *Morbidity and Mortality Weekly Report*, 62(9):165–170, 2013.
- [3] S. Vergara-López, M. C. Domínguez, M. C. Conejo, Á Pascual, and J. Rodríguez-Baño. Wastewater drainage system as an occult reservoir in a protracted clonal outbreak due to metallo- β -lactamase-producing *Klebsiella oxytoca*. *Clinical Microbiology and Infection*, 19(11), 2013.
- [4] Amy J. Mathers. Personal interview with Dr. Amy Mathers. September 2015.
- [5] Laura Barnes, Donald Brown, Jennifer Lobo, Amy Mathers, and Jason Papin. Superbug Tracker - Design of Context-Aware Surveillance System for Nosocomial Outbreaks involving Non-Patient Reservoirs. Technical report, 2015.
- [6] Matthaïos 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.
- [7] 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.
- [8] Neil Woodford, Philip M Tierno, Katherine Young, Luke Tysall, Marie-france I Palepou, Elaina Ward, Ronald E Painter, Deborah F Suber, Daniel Shungu, Lynn L Silver, Kenneth Inglima, John Kornblum, and David M Livermore. Outbreak of *Klebsiella pneumoniae* Producing a New Carbapenem- Hydrolyzing Class A -Lactamase , KPC-3 , in a New York Medical Center. *Antimicrobial Agents and Chemotherapy*, 48(12):4793–4799, 2004.
- [9] Despina Kotsanas, W. R P L I Wijesooriya, Tony M. Korman, Elizabeth E. Gillespie, Louise Wright, Kylie Snook, Natalie Williams, Jan M. Bell, Hua Y. Li, and Rhonda L. Stuart. "Down the drain": Carbapenem-resistant bacteria in intensive care unit patients and handwashing sinks. *Medical Journal of Australia*, 198(5):267–269, 2013.

- [10] I. Wolf, P. W M Bergervoet, F. W. Sebens, H. L a Van den Oever, P. H M Savelkoul, and W. C. Van der Zwet. The sink as a correctable source of extended-spectrum β -lactamase contamination for patients in the intensive care unit. *Journal of Hospital Infection*, 87(2):126–130, 2014.
- [11] Jonathan a C Sterne, Ian R White, John B Carlin, Michael Spratt, Patrick Royston, Michael G Kenward, Angela M Wood, and James R Carpenter. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ (Clinical research ed.)*, 338:1–12, 2009.
- [12] Trivellore E Raghunathan. What do we do with missing data? Some options for analysis of incomplete data. *Annual review of public health*, 25:99–117, 2004.
- [13] Rebekah Young and Dr Johnson. 'Imputing the Missing Y's: Implications for Survey Producers and Survey Users. *64th Annual Conference of the American ...*, (1987):6242–6248, 2010.
- [14] James Honaker and Gary King. What to do about missing values in time-series cross-section data. *American Journal of Political Science*, 54(2):561–581, 2010.
- [15] Jeremy Stern, Sam Hewitt, Michael Guilfoyle, Chetan Mishra, Amy Mathers, Jennifer Lobo, and Donald Brown. Modeling Nosocomial Transmission of Carbapenem-Resistant Bacteria. *Systems and Information Engineering Design Symposium (SIEDS)*, 2015, pages 176–181, 2015.
- [16] Nicholas J Horton and Ken P Kleinman. Much ado about nothing: A comparison of missing data methods and software to fit incomplete data regression models. *The American statistician*, 61(1):79–90, 2007.
- [17] Stef Van Buuren and Karin Groothuis-Oudshoorn. Multivariate Imputation by Chained Equations. *Journal Of Statistical Software*, 45(3):1–67, 2011.
- [18] John Maindonald and W. John Braun. Data Analysis and Graphics Using R - an Example-Based Approach (3rd edition). *Analysis*, pages 1–565, 2003.

Table 7.1: Variable Significance by Model

Variables	Stepwise All Variables	Stepwise Interactions
Intercept - 6N/NIMU, SinkDesignF, 0		
Unit - 3 MICU North		
Unit - 3 MIU West		
Unit - 4 TCV North		
Unit - 5 STBICU		
Lag0	***	
Lag1		
Lag2		
Lag2		
Intervention0	***	
Intervention1		
Intervention2		
Intervention3	***	***
CumSum7	*	
CumSum14	***	
SinkDesign "A"	***	***
SinkDesign "BC"		
SinkDesign "D"	***	***
SinkDesign "E"	***	***
SinkDesign "OPQ"	***	
SinkDesign "R"	***	***
SinkDesign "STUV"	***	***
SinkBedDis	***	
Past.30	***	
SinkBedDis: Intervention0		***
SinkBedDis: Intervention3		*
SinkBedDis: CumSum14		*
Intervention0: Intervention3		***
Intervention0: CumSum14		**
Intervention3: Past.30		***
Intervention3: SinkDesign "A"		*
CumSum7: Past.30		***
CumSum14: Past.30		***
CumSum14: SinkDesign "OPQ"		***
"****" = p-value less than .001, "***" = p-value less than .01, "*" = p-value less than .05		

Table 2: Sink Design

Sink Design	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
“A”	C	K	A	Y	B	41.5	Y	N	G	N	N	0	N	15.5	21	13.5	7	SS	10	Y	Y	N	Y	0	N
“BC”	C	K	A	Y	P	24.5	Y	N	G	Y	B	1	N	14	20	13.5	6.5	SS	10	Y	N	N	Y	0	N
“D”	C	K	A	Y	P	29.5	Y	N	G	N	N	0	N	14	20	13.5	6	SS	7.5	Y	Y	N	N	0	N
“E”	C	K	A	Y	P	41.5	Y	N	G	N	N	0	N	15.5	21	13.5	7	SS	10	Y	Y	N	Y	0	N
“F”	C	J	A	Y	P	71.5	Y	N	G	N	N	0	N	14	20	13.5	6	SS	8	Y	Y	N	N	0	N
“GH”	C	O	P	Y	P	29	N	N	G	Y	F	1	N	16.5	21	15	7	SS	8.5	Y	N	N	N	0	N
“I”	C	O	P	Y	P	53	N	N	G	Y	F	1	N	16.5	18.5	15	7.5	SS	10	N	N	N	N	0	N
“J”	C	O	P	Y	P	71	Y	N	G	Y	F	1	N	16.5	18.5	15	7.5	SS	10	N	N	N	N	0	N
“KN”	C	O	P	Y	P	72	N	N	G	Y	F	1	N	16.5	18.5	15	7.5	SS	9	N	N	N	N	0	N
“LM”	C	O	P	Y	P	73	N	N	G	Y	F	1	N	16.5	21	15	7	SS	9.5	Y	N	N	N	0	N
“OPQ”	C	O	P	Y	C	29	N	N	G	Y	F	1	N	13.5	17	13.5	6	I	0	Y	N	N	N	0	N
“R”	C	O	P	Y	C	29.5	N	N	G	Y	F	1	H	13.5	17	13.5	6.5	SS	11	Y	N	N	N	0	N
“STUV”	C	O	P	Y	C	59.5	N	N	G	Y	F	1	H	13	15	11	6.5	SS	10.5	Y	N	Y	N	2	Y
“W”	C	R	P	Y	C	46.5	N	Y	F	N	N	0	N	21	18	16	8	SS	10.5	N	N	N	N	0	N
“X”	C	R	P	Y	C	73	N	Y	F	N	N	0	N	19	16	18	8.5	SS	12	Y	N	N	N	0	N
“Y”	C	R	P	Y	C	73	N	Y	F	N	N	0	N	20.5	18	16	8.5	SS	7	N	N	N	N	0	N

Legend: 1(C=corian), 2(K=kidney, J=knob, O=oval, R=rectangle), 3(A=Auto, P=paddle), 5(B=black corian, P=copper, C=corian), 9(G=grate, F=front), 11(F=front, B=back, N=no), 13(H=handicap, N=normal), 18(SS=stainless steel, I=inaccessable), else(Y=yes, N=no)