2018-05-31

Tackling the Antibiotic Resistant Bacteria Crisis Using Longitudinal Antibiograms

Monica Tlachac
mltlachac@wpi.edu

Tackling the Antibiotic Resistant Bacteria Crisis Using Longitudinal Antibiograms

by

Monica Lauren Tlachac

A thesis

Submitted to the Faculty

of the

WORCESTER POLYTECHNIC INSTITUTE

in partial fulfillment of the requirements for the

Degree of Master of Science

in

Data Science

May 2018

Approved by

Advisor:

Prof. Elke Rundensteiner

Co-Advisor:

Prof. Jian Zou

Abstract: Antibiotic resistant bacteria, a growing health crisis, arise due to antibiotic overuse and misuse. Resistant infections endanger the lives of patients and are financially burdensome. Aggregate antimicrobial susceptibility reports, called antibiograms, are critical for tracking antibiotic susceptibility and evaluating the likelihood of the effectiveness of different antibiotics to treat an infection prior to the availability of patient specific susceptibility data. This research leverages the Massachusetts Statewide Antibiogram database, a rich dataset composed of antibiograms for 754 antibiotic-bacteria pairs collected by the Massachusetts Department of Public Health from 2002 to 2016. However, these antibiograms are at least a year old, meaning antibiotics are prescribed based on outdated data which unnecessarily furthers resistance. Our objective is to employ data science techniques on these antibiograms to assist in developing more responsible antibiotic prescription practices. First, we use model selectors with regression-based techniques to forecast the current antimicrobial resistance. Next, we develop an assistant to immediately identify clinically and statistically significant changes in antimicrobial resistance between years once the most recent year of antibiograms are collected. Lastly, we use k-means clustering on resistance trends to detect antibiotic-bacteria pairs with resistance trends for which forecasting will not be effective. These three strategies can be implemented to guide more responsible antibiotic prescription practices and thus reduce unnecessary increases in antibiotic resistance.

Keywords: *Antimicrobial Resistance, Antibiograms, Regression, Support Vector Regression, ARIMA, Model Selector, Outlier Detection, Statistical Significance, Clinical Significance*

Acknowledgements: I thank my advisor Professor Elke Rundensteiner for her valuable guidance throughout this entire process. I thank my co-advisor Professor Jian Zou for providing statistical guidance. I thank my medical collaborators and co-authors Kerri Barton from Maine Medical Center, Scott Troppy from Massachusetts Department of Health (MDPH), Kira Beaulac from Tufts Medical Center, and Shira Doron from Tufts Medical Center. Without their contributions and medical expertise, this work would not have been possible. In addition, I would like to thank Albert DeMaria and Monina Klevens from the MDPH. I thank Worcester Polytechnic Institute (WPI) and the DSRG research group for providing an environment for me to develop as data science researcher and receive feedback on my research. Specifically, I would like to thank Tom Hartvigsen for always being there to bounce ideas off of and helping me develop my coding skills. I would also like to specifically mention Olga Poppe, Caitlin Kuhlman, Brendan Foley, Xiao Qin, Tabassum Kakar, and Susmitha Wunnava from the DSRG group. In addition, I thank Sam Ogden from WPI and Matthew Tlachac from University of Minnesota for reviewing code. Lastly, I thank my parents for their support. I thank US Department of Education P200A150306: GAANN Fellowships to Support Data-Driven Computing Research.

Contents

List of Tables

List of Figures

1. INTRODUCTION

1.1. Background

Antibiotic resistant bacteria of clinical significance are becoming increasingly prevalent around the world. The World Health Organization has classified the reported levels of antimicrobial resistance as alarming. Infections due to antibiotic resistant bacteria are more expensive to treat than other bacterial infections, costing the U.S. economy an estimated 20 billion dollars a year in direct healthcare costs, as well as at least that much in additional financial burdens to patients, family members, and society at large for loss of productivity. Patients with antibiotic resistant bacterial infections also experience more devastating health outcomes ranging from extended hospital stays to increased risk of death [5][28]. Conservative estimates from 2013 attribute over two million infections and 23 thousand deaths to antibiotic resistant bacteria per year in the United States [5]. At a global scale, this increases to an estimated $700,000$ deaths from antibiotic resistance each year [17]. Without a deeper understanding of resistance patterns and more informed prescription practices, resistance rates will continue to increase until there is no way to cure some bacterial infections. The consequences of inaction are catastrophic.

The overuse of antibiotics is one of the main causes of antimicrobial resistance [5][27]. Once viewed as life-saving therapies, the role of antibiotics in the public eye has shifted to being thought of as ubiquitous within healthcare. In fact, Antibiotics remain one of the most prescribed human medicines [5]. Unfortunately, antibiotics are not always prescribed responsibly, with up to 50 percent of prescriptions either being unnecessary or ineffective [5][27]. In particular, incorrectly prescribed antibiotics have been shown to contribute to antimicrobial resistance [27].

1.2. Motivation

Antibiotic resistance is tracked using antibiograms, reports that provide the average percent susceptibility of select antibiotics tested against cultures of bacterial infections, called isolates, that are collected from patients in medical facilities. These antibiograms are useful in monitoring resistance trends and empiric antibiotic selection [8][3]. Specifically, at the facility level, these antibiograms are used primarily to guide prescription practices before patient specific laboratory data is available. The Massachusetts Department of Public Health (MDPH) amasses these antibiograms

from state hospitals annually, typically around six months after the sampling has ceased [4]. As consequence, antibiotics are being prescribed based off of outdated resistance knowledge.

Unfortunately, outdated resistance information facilitates the propagation of ineffective and inappropriate antibiotic use by suggesting antibiotics may be effective when they are not, further increasing resistance. In fact, the Centers for Disease Control and Prevention state that out of the four main actions to fight antimicrobial resistance, improving antibiotic prescribing by reducing the misuse and overuse of antibiotics is the most important of these actions [6]. Thus, it is important to immediately identify antibiotic-bacteria pairs with relevant changes in resistance, detect anomalous resistance trends, and to forecast future susceptibility for a large set of antibiotic-bacteria pairs. The completion of these tasks will provide valuable information for guiding prescription practices.

1.3. Previous Antibiotic Resistance Analysis

The chi-square test [21] is a popular statistical method to apply to antibiogram data as the test has minimal data requirements: the antibiotic resistance of two different sets of isolates collected at two different times. The chi-square test has been used in multiple papers [7][11][22] to determine if there was a statistically significant change in the proportion of antibiotic-resistant bacteria between two sets of time. However, despite the popularity of statistical significance, no prior research covers mathematically determining clinical significance of these changes in antimicrobial susceptibility.

Regression is another popular method to analyze antibiotic-resistant bacteria datasets [1][2]/citeCrnich[10][13]. However, only one of these studies utilizes regression to forecast future susceptibilities. This study [1] predicts susceptibility through 2030 with linear mixed models; there is no validation of the predictions. In other studies, regression is implemented for other pusposes. One of these studies [7] utilizes regression for forecasting but with the amount of antimicrobial infections as the target variable. Another study [2] determines the length of time a patient specific antibiogram is a reliable indicator for the susceptibility of the infection. The other two studies [13][10] use logistic regression to isolate the impact of time on susceptibility changes and the effect of azithromycin exposure to resistance, respectively. All of these studies are restricted both in term of the longitudinal horizon, the geographical scope, and the coverage of antibiotic-bacteria pairs. The papers from [2][7][10] each only focus on one infection, namely, *Pseudomonas aeruginosa, Clostridium difficile*, and Bronchiectasis, respectively. The research by [13] is the least restrictive as it includes multiple bacterial infections from eight Canadian providences, but it is limited by its

data set being composed of only five years of data.

There are not many other analytics projects that model antimicrobial resistance. The Review on Antimicrobial Resistance estimates the number of deaths from a subset of antibiotic resistant infections in 2050 but relies on unprecedented assumptions regarding future resistance rates [17]. In addition to the chi-square test, [22] uses Markov models to determine the likelihood of a change in Escherichia coli antimicrobial susceptibility. While not used in this manner, Markov models may be applied to predict future susceptibilities. Lastly, an HIV study uses probabilistic neural networks to predict resistance to an antiretroviral drug with patient specific data [19].

1.4. Our Contributions

The aforementioned projects fail to leverage antibiogram data to its full potential. Specifically, these studies fail to take advantage of longitudinal data assets to forecast future antibiotic susceptibilities with validation. No techniques such as outlier detection are implemented to improve the forecasting ability. Also, none of these studies cover a way to mathematically determine clinical significance. In addition, these studies only cover limited antibiotic-bacteria pairs and geographical scope.

We address these deficiencies in our research through three main tasks: forecasting the current antimicrobial resistance using historic antibiograms, identifying whether the change in antimicrobial resistance between years is significant, and detecting anomalous antimicrobial that are not predictable. This research leverages the expansive Massachusetts statewide antibiogram (MA-SA) dataset, described in Section 2. Not only does the MA-SA span 15 years and contain location granularity, it contains at least a year of resistance data for over 750 antibiotic-bacteria pairs. Thus, our research has the advantage of not being limited by a minimal number of antibiotic-bacteria pairs, a small geographical scope, or lack of longitudinal tracking. Specifically, we take advantage of this when forecasting current and future antimicrobial susceptibility.

2. The Massachusetts Statewide Antibiogram Dataset

In this research, we leverage a 15-year subset of the Massachusetts Statewide Antibiogram (MA-SA) dataset shared with us by the Massachusetts Department of Public Health (MDPH). MA-SA, curated by the MDPH, contains antibiograms collected from over 50 acute-care hospitals

across the state. From 2002 to 2016, this expansive dataset contains at least a year of susceptibility for 14 bacteria and 85 antibiotics, combining to form 754 antibiotic-bacteria pairs.

Cultures of bacterial infections, called isolates, were collected from inpatients and outpatients at Massachusetts acute-care hospitals. The antibiotic isolates were considered susceptible based on Clinical and Laboratory Standards Institute (CLSI) guidelines. The isolates collected from the same hospital are tested, aggregated, and then reported to the MDPH the following year. The MA-SA is composed of $107,968$ individual data instances that contain the antibiotic, bacteria, number of samples, percent of susceptible samples, hospital, location within the hospital where the samples were collected, and county. A sample of three of these individual data instances, further referred to as individual reports, is seen in Figure 2.

Antibiotic	Bacteria	Year	County	HospitalID	Nsamples	Susceptibility	PatientPop
AMOXICILLIN+CLAVULANATE	Enterobacter aerogenes	2002	Essex	3	19	16	All Patient
CARBENICILLIN	Enterobacter aerogenes	2002	Essex	3	19	79	All Patient
CEFAZOLIN	Enterobacter aerogenes	2002	Essex	3	19	11	All Patient

Figure 1. Example of the data.

The bacteria in MA-SA include *Acinetobacter baumannii, Enterobacter aerogenes, Enterobacter cloacae, Escherichia coli, Klebsiella oxytoca, Klebsiella pneumonia, Pseudomonas Aeruginosa, Serratia marcenscens, Staphylococcus aureus*, and *Stenotrophomas maltophilia*. Distinctions are also made for bacterial subtypes methicillin-sensitive *Staphylococcus aureus* (MSSA) and methicillin-resistant *Staphylococcus aureus* (MRSA) as well as extended-spectrum-*beta*-lactamase-producting(ESLB) *Klebsiella oxytoca* and *Klebsiella pneumonia* [25].

2.1. Actual Mean Susceptibility

The MDPH receives antibiograms containing many reports with different susceptibilities. The actual mean susceptibility A of the reports weighted by the number of isolates, defined in Equation 1, is used to make this information consumable for the user. It is the annual mean susceptibility that is of interest to the MDPH.

$$A = \frac{\sum_{i=1}^{n} s_i t_i}{\sum_{i=1}^{n} t_i} \tag{1}$$

where n is the number of reports, s_i is the susceptibility percent of the ith report, and t_{1i} is the number isolates of the ith report. The actual mean susceptibility A is a percent between 0 and 100.

2.2. Data Views

The Massachusetts Department of Public Health publishes a statewide summary of the antibiotics annually in single year view [4]. This view is often applied at the hospital granularity to be more useful in guiding prescription practices, as seen for Tufts Medical Center in Figure 2 [3].

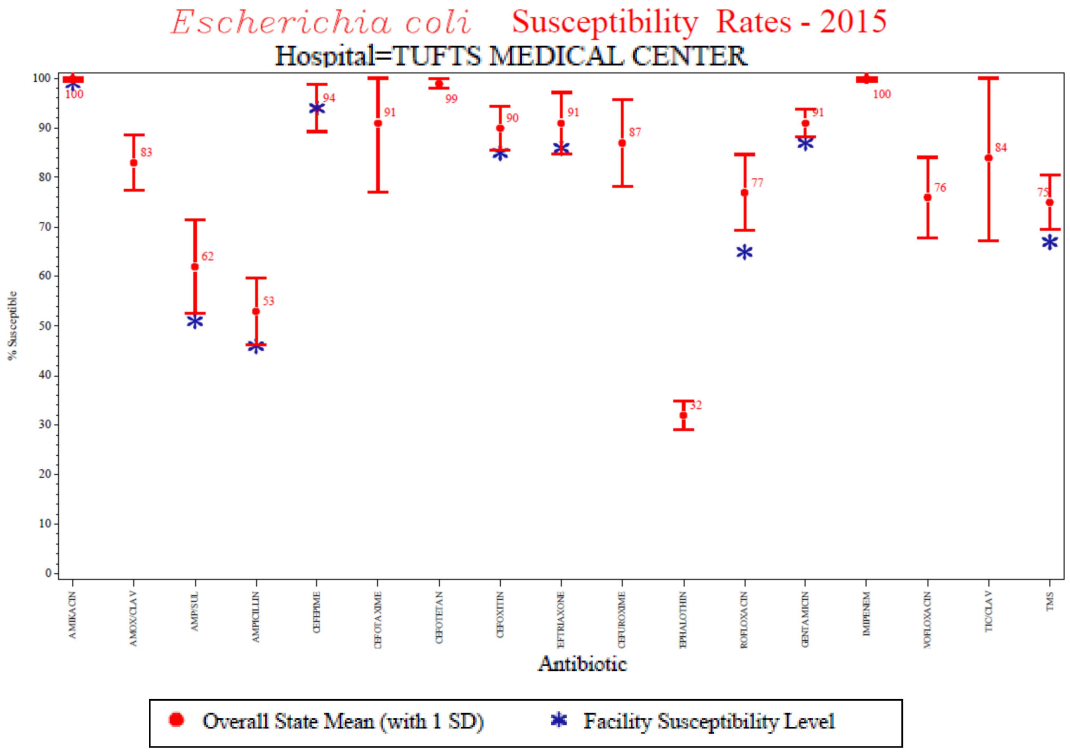

Figure 2. Single year data view.

This view, while it allows for the comparison of the mean susceptibility for multiple antibiotic-bacteria pairs, it does not indicate the resistance trend over time. Thus, we introduce two other views: the longitudinal reports view and the annual mean susceptibility view. Specifically, Figure 3 and Figure 4 show the longitudinal data at the state granularity for *E. coli* and levofloxacin.

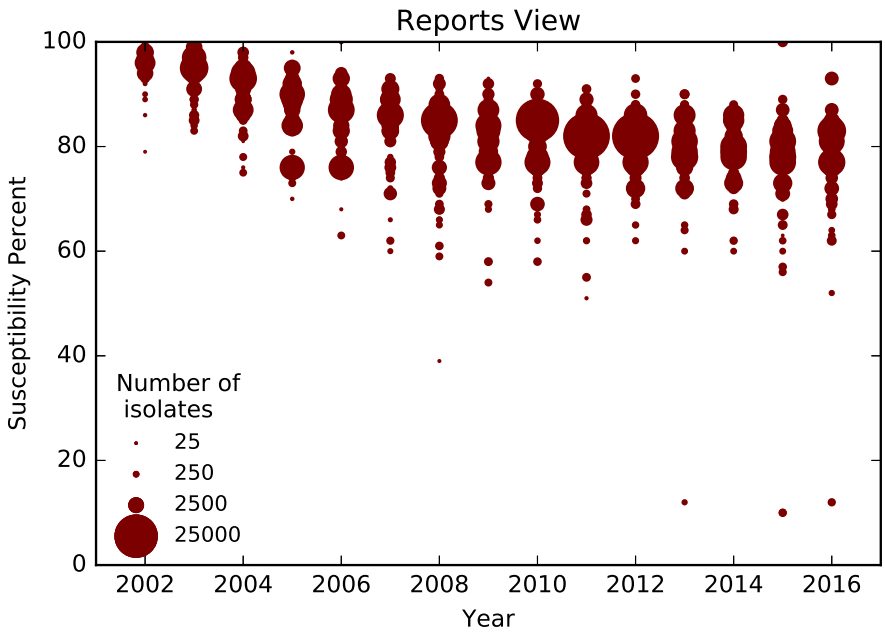

Figure 3. Longitudinal state reports.

2.3. Data Quality

While the MA-SA dataset is impressive in size and scope, data procurement occurring over 15 years and more than 50 hospitals varies in reliability. The most prevalent concern involves the microbiology laboratories' varied response times to implementing CLSI changes. However, policy changes have greatly improved the quality and coverage of antibiograms submitted to the MDPH. For instance, due to these policies the number of duplicate isolates contained within the dataset have consistently decreased over time. Additionally the percent of hospitals submitting data has improved. Though due to original guidelines, some of these hospitals do not submit reports with fewer than 20 to 30 isolates to the MDPH. As data quality has been consistently increasing over time, we have opted for minimal data cleaning implemented based on the specific task requirements.

2.4. Data Subsets

We demonstrate our methods on three distinct subsets of the MA-SA dataset: all antibiotic-bacteria pairs (AP dataset), all clinically relevant antibiotic-bacteria pairs (ACP dataset), and the most clinically relevant antibiotic-bacteria pairs (MCP dataset). The MCP dataset contains *E. coli*,

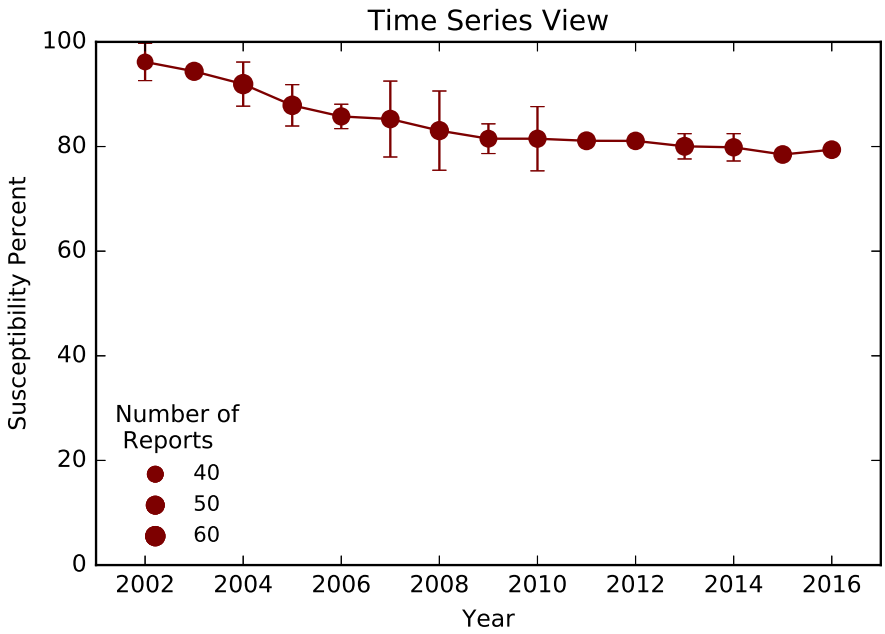

Figure 4. Annual state mean susceptibility.

Klebsiella pneumoniae, and *Klebsiella oxytoca* susceptibility to antibiotics in the carbapenem, cephalosporin, fluroquinoline families. We also consider these subsets at different levels of location granularity resulting in unique antibiotic-bacteria, antibiotic-bacteria-county, and antibiotic-bacteria-hospital combinations.

As previously mentioned, MA-SA contains 754 antibiotic-bacteria pairs, so the AP dataset does as well. In addition, the AP dataset encompasses 4743 antibiotic-bacteria-county combinations and 18363 antibiotic-bacteria-hospital combinations. The ACP dataset contains the 189 antibiotic-bacteria pairs, 2212 antibiotic-bacteria-county combinations, and 10842 antibotic-bacteria-hospital combinations. The MCP dataset contains 63 antibiotic-bacteria pairs, 589 antibiotic-bacteria-county combinations, and 2633 antibiotic-bacteria-hospital combinations.

3. Forecasting

In this section, we will use historic antibiograms to forecast future antibiotic susceptibility. We will validate these forecasts by comparing to the actual mean susceptibility for the forecasted year (2.1). These forecasts can be used to guide prescription practices so antibiotics are not prescribed based off of outdated resistance knowledge.

We will start by comparing the forecasting ability of regression-based methods and variations of our custom model selector Previous Year Prediction Error Reduction (PYPER) on the AP dataset. Then we progress to explore the impact of the clinical significance of the dataset, the location granularity, and the number of years of data on forecasting. We will also explore the prediction quality multiple years into the future. Lastly, we will discuss the reasons why some of our forecasted susceptibilities are not very close to the actual mean susceptibility.

3.1. Global Methodology

The global methodology is when the same method is utilized to predict the next year's mean susceptibility for every antibiotic-bacteria-location (A-B-L) combination. Each antibiotic-bacteria-location combination for a particular subset of MA-SA is modeled separately using the same method [26].

In particular, our global predictive methodology, seen in Figure 5, is a four step process for each subset of MA-SA: (1) select parameters, (2) fit model, (3) make prediction, (4) evaluate prediction. Step 1 involves selecting the parameters to build the model such as the A-B-L combination, the method M, the target year Y, and the number of years of historic data H. Step 2 involves using the selected method M to establish a model D for the data using the historic years of data H. Step 3 involves using that model D to predict the mean susceptibility P for target year Y. Lastly, Step 4 involves utilizing evaluation metrics to measure the effectiveness of the model D in predicting current susceptibility percent for the A-B-L combination by comparing the predicted mean susceptibility P to the actual mean susceptibility A in target year Y. This methodology is repeated for each A-B-L combination in the dataset [26].

To ensure that the results of our methodology are not year specific, we implement a sliding window mechanism that enables us to repeat the methodology for multiple target years [**?**]. Specifically, we have chosen the three most recent years in the data as our target years: 2014, 2015, and 2016. The number of historic years of data H selected in methodology Step 1 is restricted by this sliding window mechanism. As 2014 is the earliest year in our target years and 2002 is the earliest year in MA-SA, the maximum number of years of historic data H is 12. Thus, the historic years of data H consist of twelve years $y_{-12}, ..., y_{-1}$ of susceptibility data. This historic susceptibility H is utilized to predict the subsequent susceptibility percent, otherwise known as the susceptibility percent for target year Y [26].

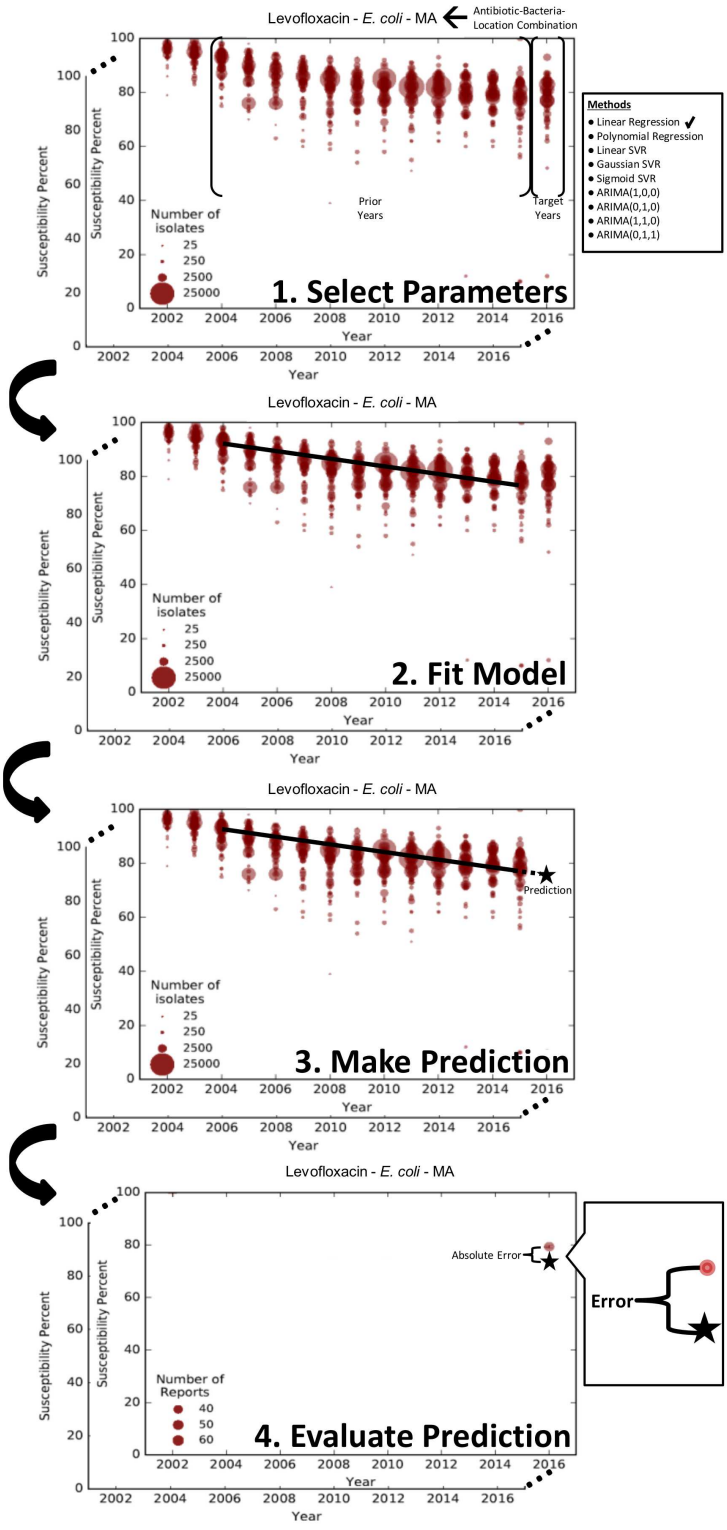

Figure 5. Global Methodology.

3.2. Global Models

We apply the global methodology with a variety of regression-based methods. Specifically, these include regression methods, support vector methods (SVR), and autoregressive integrated moving average models (ARIMA). As mentioned, the same method is applied uniformly to each A-B-L combination to create a set of predictive models. As these methods are modeling susceptibility percent, if the resulting predictions for target year Y are below 0 or above 100, the predictions are readjusted to be 0 or 100 respectively [26].

3.2.1. Regression Models. Regression is a method of analyzing the impact of an independent variable, year, on a dependent variable, susceptibility. Specifically, this is completed by minimizing the residual sum of squares (RSS), Equation 2, between the report susceptibility and the modeled susceptibility over time [12].

$$RSS = \sum_{i=1}^{n} (f_i - A_j)^2 \tag{2}$$

where n is the total number of reports, f_i is the forecasted mean susceptibility for the ith report, j is the year of the ith report, and A_j is the actual mean susceptibility in the jth year.

In this research we utilize two types of standard regressions: simple linear regression and second degree polynomial regression [24]. Linear regression assumes the relationship between susceptibility and year is linear while polynomial regression assumes the relationship between susceptibility and year is polynomial. We experiment with regression models where the reports are not weighted- every report has the same weight- and are weighted by the number of isolates [26].

3.2.2. Support Vector Regression Models. Support Vector Regression (SVR) is a regression technique that utilizes the support vector algorithm to model the data. Specifically, SVR fits a function where errors smaller than an identified margin of error are considered acceptable [23]. The benefit to using SVR for predictive analytics is the method's generalization ability and noise tolerance [29].

In particular, the support vector algorithm allows for the utilization of kernel functions to map the data to a different input space. This is useful if the data does not conform to a linear distribution. We implement SVR models with linear, Gaussian, and sigmoid kernels [24]. As with the standard regression models, we experiment with weighting and not weighting the reports by the number of isolates.

3.2.3. Autoregressive Integrated Moving Average Models. Autoregressive Integrated Moving Average (ARIMA) is a popular forecasting method. ARIMA's predictions are a sum of prior values; the number and weight of the prior values is dependent on the parameters p, d, and q. As such, ARIMA(p,d,q) models can be tailored to consider a different number of autoregressive terms p, the number of nonseasonal differences d, and the number of lagged forecast errors q [15]. We consider first-order autoregressive models ARIMA(1,0,0), random walk model ARIMA(0,1,0), differenced first-order autoregressive model ARIMA(1,1,0), and simple exponential smoothing model ARIMA(0,1,1) [26].

3.3. Customized Methodology

The global methodology assumed that every A-B-L combination should be modeled using the same method. However, model selectors offer the flexibility for different A-B-L combinations to be modeled using different methods. Thus, in addition to the global methodology, we propose a customized methodology which leverages model selectors to select among a set of predictive methods for each A-B-L combination [26].

The customized methodology is composed of seven steps for each subset of MA-SA: (1) choose model selector parameters, (2) select model parameters, (3) building models, (4) choose best method, (7) update model, (6) make prediction, (7) evaluate prediction. Step 1 involves choosing the set of methods $\{M_1,...,M_k\}$ that the model selector can select amongst and the strategy for selecting a model S. Step 2 involves selecting the parameters to build the models such as the A-B-L combination, the target year Y, and the number of years of historic data H. Step 3 involves using each of the methods in the method step $\{M_1,...,M_k\}$ to establish a model set $\{D_1,...,D_k\}$ for the data using the historic years of data H. Step 4 involves using the strategy S to select a model $D_d \in \{D_1,...,D_k\}$. Step 5, which is optional depending on the model selection strategy S, involves using the method $M_d \in \{M_1,...,M_k\}$ that built D_d to establish an updated model D. Step 6 involves using the updated model D to predict the mean susceptibility P for the target year Y. Lastly, Step 7 involves utilizing evaluation metrics to measure the effectiveness of the model D in predicting current susceptibility percent for the A-B-L combination by comparing the predicted mean susceptibility P to the actual mean susceptibility A in target year Y. This methodology is repeated for each A-B-L combination in the dataset [26].

As with the global methodology, we implement a sliding window to avoid overfitting. Specifically, we use 12 years of historic data to make predictions for 2014, 2015, and 2016.

3.4. Model Selectors

Tailoring the method with a model selector involves two unique challenges: deciding the methods to be included within the model selector and determining the best manner to select the model. In our previous work, we determined that choosing the model that minimized the mean squared error was not effective [24]. As such, we focus on the model selector we designed called Previous Year Prediction Error Reduction (PYPER) and the variant with error distinction (PYPERed) [24].

3.4.1. PYPER. PYPER aims to select the method that builds a model that will produce the smallest error in the subsequent year. This is accomplished by selecting the method that built the model that minimizes the absolute error in the prior year to make predictions for the target year Y [24].

Specifically, in step 3 of the customized methodology, PYPER utilizes susceptibility data in years $y_{-12}, ..., y_{-2}$ to build the model set $\{D_1, ..., D_k\}$. The strategy PYPER implements in step 4 to select the model is to determine which model produces the lowest absolute error when predicting the susceptibility in year y_{-1}; if there is a tie, the global method the lowest error when applied to all antibiotic-bacteria-location combinations is selected. As such, step 5 of the customized methodology is necessary to build an updated model with susceptibility data in years $y_{-12}, ..., y_{-1}$ [26].

3.4.2. PYPERed. PYPERed, a variant of PYPER model with error distinction, capitalizes on the fact that the for some A-B-L combinations the susceptibility changes minimally over time. As such, PYPERed employs the customized methodology when the actual mean susceptibility in year y_{-1} is different from the actual mean susceptibility in year y_{-2} by a predetermined threshold T. Otherwise, in this research, PYPERed assumes that the susceptibility in year Y will be similar to the susceptibility in year y_{-1} [26].

We have used two different thresholds T in our research. The first threshold T we used was based on the standard error formula (Equation 6 in Section 4.2.2 with an imposed population standard devation of 10. [24]. However, in this document we will be focusing on the second threshold T: the standard deviation (SD) of the reports in year y_{-1}. As such, PYPER is only employed if the change in susceptibility between year y_{-1} and y_{-2} is greater than the standard

deviation SD in year y_{-1}. Otherwise, the average susceptibility in year y_{-1} is used [26]. This is motivated by the fact that we observed that time series with small standard deviations have stable resistance trends. The standard deviation (SD) formula is shown in Equation 3 [12].

$$SD(y_{-1}) = \sqrt{\frac{1}{n_{y_{-1}}} \sum_{k=1}^{n_{y_{-1}}} (s_i - \bar{s})^2} \tag{3}$$

where $n_{y_{-1}}$ is the number of reports in year y_{-1} for the given A-B-L combination, s_i is the susceptibility percent of the kth report in year y_{-1}, and $\bar{s}$ is the average susceptibility of all reports in year y_{-1}. This average susceptibility is different than the actual mean susceptibility as it is not weighted by the number of isolates.

The standard deviation for report susceptibility of the AP dataset at the state granularity ranges from 0 to 43.3, the distribution of which is shown in Figure 6 [26].

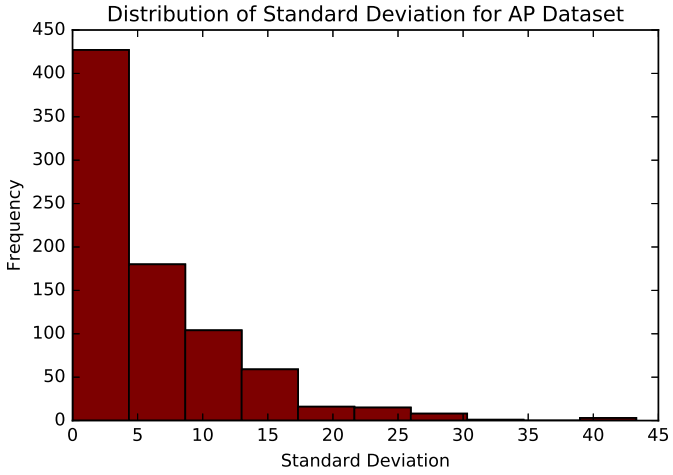

Figure 6. Distribution of the susceptibility standard deviation for AP state dataset.

3.5. Data Cleaning

As mentioned in Section 2.3, some of these hospitals do not submit reports to the MPDH with fewer than 20 to 30 isolates. Thus, we only consider reports that have at least 20 isolates [24]. In addition to matching the original guidelines, this requirement decreases the impact of an outlier bacterial infection [26]. For modeling purposes, each antibiotic-bacteria pair must have at least 6 years of data. Specifically, each of these antibiotic-bacteria pairs must have at least one report in

the target year Y and the two prior years y_{-1} and y_{-2}.

Under the aforementioned cleaning, the AP dataset, ACP dataset, and MCP dataset have just over 270, 180, and 40 antibiotic-bacteria pairs and just over 1650, 1250, and 290 antibiotic-bacteria-county pairs, respectively. In addition, for target years $2014-2016$, these data subsets have $2417-2823$, $2000-2301$, and $517-616$ antibiotic-bacteria-hospital pairs, respectively. The number of antibiotic-bacteria-hospital pairs increase from $2014-2016$ as a greater percent of state acute-care hospitals submit antibiograms to the MDPH in more recent years [26].

3.6. Mean Absolute Error

We evaluate the prediction ability of the models by comparing the actual mean susceptibility A to our forecasted mean susceptibility f for target year Y for each antibiotic-bacteria-location combination in the dataset [26]. We accomplish this using a common evaluation metric for assessing the quality of numeric predictions: mean absolute error (MAE) [14].

MAE, defined in Equation 4, will both be used to compare methods and interpret the usefulness of the forecast. It is useful for health professionals to know the average error when determining whether to use the prediction method [26].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|f_i - A_j| \tag{4}$$

where n is the total number of reports, f_i is the forecasted mean susceptibility for the ith report, j is the year of the ith report, and A_j is the actual mean susceptibility in the jth year.

3.7. Results

For each antibiotic-bacteria pair in the all antibiotic-bacteria pairs (AP) dataset, we use the global regression-based methods constructed with between 6 and 12 years of data to make predictions for the subsequent year. We evaluate the 813 predictions for 2014, 2015, and 2016 with MAE, as seen in table 1 [26].

All four of the ARIMA methods perform better than any of the other regression and SVR methods. Except for polynomial regression, the weighted version of these regression-based methods performed better than the unweighted version of the regression-based methods. Specifically, linear regression, linear SVR, and Gaussian SVR have the lowest MAE of these weighted regression-

Table 1. Global methods for AP state dataset.

Method	MAE
Unweighted Linear Regression	2.699
Unweighted Poly. Regression	2.679
Unweighted Linear SVR	2.747
Unweighted Gaussian SVR	3.028
Unweighted Sigmoid SVR	3.275
Weighted Linear Regression	2.524
Weighted Poly. Regression	3.505
Weighted Linear SVR	2.622
Weighted Gaussian SVR	2.585
Weighted Sigmoid SVR	3.131
ARIMA(1,0,0)	2.502
ARIMA(0,1,0)	2.143
ARIMA(1,1,0)	2.280
ARIMA(0,1,1)	2.494

based methods; the 2016 absolute errors of these methods for antibiotic-bacteria pairs containing *E. coli* in the MCP dataset are compared in Figure 7 [26].

From Figure 7, we can see that different methods perform better for different antibiotic-bacteria pairs, motivating the use of model selectors. As such, we apply PYPER and PYPERed on weighted regression-based methods and ARIMA variations, the results of which are in Table 2. The weighted regression-based methods only include linear regression (L-Reg), linear SVR (L-SVR), and Gaussian SVR (G-SVR); including polynomial regression and Sigmoid regression cause PYPER to perform worse [26].

Table 2. Customized methods for AP state dataset.

Method	MAE
PYPER(L-Reg,L-SVR,G-SVR)	2.099
PYPERed(L-Reg,L-SVR,G-SVR)	1.882
PYPER(ARIMA)	2.100
PYPERed(ARIMA)	2.043

PYPERed with weighted linear regression, linear SVR, and Gaussian SVR has the lowest

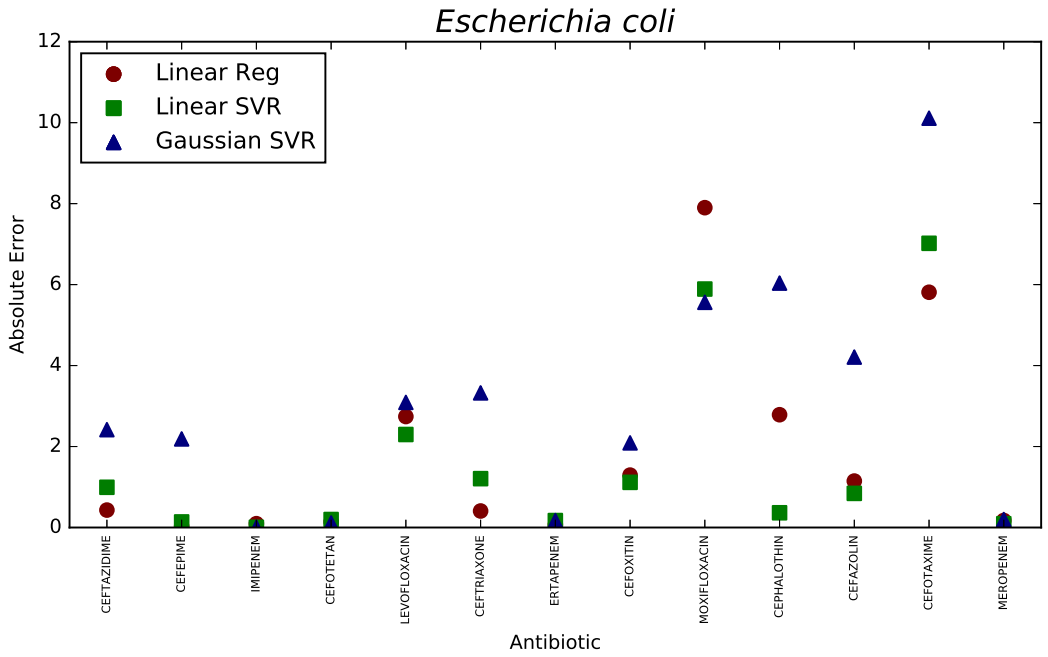

Figure 7. Comparison of 2016 **absolute errors for different weighted regression-based methods.**

MAE. PYPERed 2016 absolute errors for antibiotic-bacteria pairs containing *E. coli* in the MCP dataset are shown in Figure 8 [26].

Even though at the global level the ARIMA models had lower errors than the other regression models, PYPERed with ARIMA performs worse than PYPERed with the other regression models. As such, PYPER and PYPERed will further refer to the use of these model selectors with the methods linear regression, linear SVR, and Gaussian SVR. Using PYPERed, we are able to forecast the susceptibility in the target year Y on average within 1.882 susceptibility percent for the AP dataset at the state level [26].

3.8. Clinically Relevant Data Subsets

PYPERed remains an effective method at reducing susceptibility forecasting error at the state level regardless of the data subset selected. Table 3 compares PYPERed's error to linear regression, weighted linear SVR, and weighted Gaussian SVR for the all clinically revelent antibiotic-bacteria pairs (ACP) and most clinically relevant antibiotic-bacteria pairs (MCP) datasets. The state ACP dataset contains 543 predictions and the state MCP dataset contains 124 predictions for target years 2014, 2015, and 2016 [26].

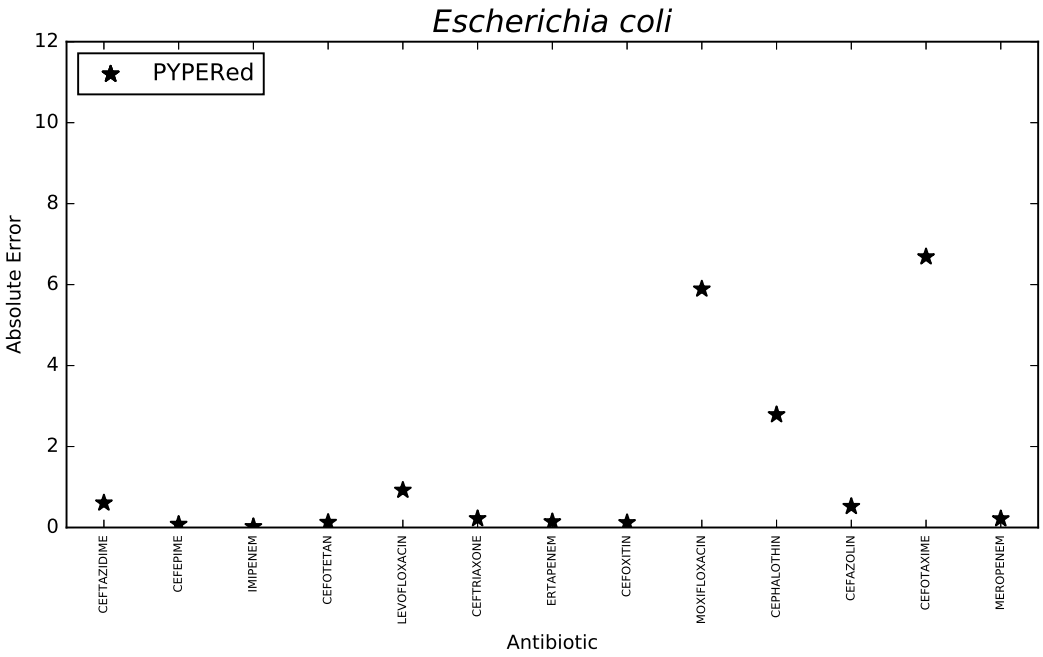

Figure 8. 2016 **absolute errors for PYPERed.**

Table 3. Model comparison for state level datasets.

Method	ACP	MCP
Linear Regression	1.844	1.384
Linear SVR	2.073	1.675
Gaussian SVR	2.102	2.135
PYPERed	1.453	1.293

Regardless of the clinical significance of the antibiotic-bacteria pairs in the three datasets, implementing PYPERed proves to produce better forecasts than any of the individual regression-based methods. PYPERed has a MAE of 1.822 (Table 2), 1.453 (Table 3), and 1.293 (Table 3) for the AP, ACP, and MCP state level datasets respectively. In other words, as the clinical significance of the antibiotic-bacteria pairs increases, the MAE of the subsequent year forecasts decreases [26].

3.9. Location Granularity

We can analyze the data at the state, county, and hospital levels for the AP, ACP, and MP data subsets. The PYPERed MAE of these data subsets at the different location granularities are displayed in table 4. In addition, 5 contains the number of predictions for the antibiotic-bacteria-

location (A-B-L) combinations for target years 2014, 2015, and 2016 [26].

Table 4. PYPERed comparison at different location granularities.

Granularity	AP	ACP	MCP
State	1.822	1.453	1.293
County	2.490	2.479	1.657
Hospital	2.934	2.911	2.396

Table 5. Number of A-B-L predictions at different location granularities.

Granularity	AP	ACP	MCP
State	813	543	124
County	4981	3772	875
Hospital	7779	6392	1677

For every combination of dataset and location granularity, PYPERed MAE is lower than any globally applied regression-based methods. Consistent with the conclusions of Section 3.8, as the clinical significance of the A-B-L combination increases, the MAE decreases for every location granularity. Additionally, for every dataset, as the location becomes more granular, the error increases. This makes sense given that the number of predictions for the A-B-L combinations increases greatly as the location becomes more granular. For instance, we can forecast within 1.822 susceptibility percent for 813 antibiotic-bacteria pairs and within 2.934 susceptibility percent for 7779 antibiotic-bacteria-hospital combinations [26].

3.10. Longitudinal Data Importance

We use the state datasets to explore the importance of longitudinal data in forecasting future antibiotic susceptibility. In particular the MCP dataset is ideal for this task as, since the antibiotic-bacteria pairs are most relevant, the number of years of data has little impact to the number of antibiotic-bacteria pairs. We explore the impact of using between $6-12$, $3-12$, $3-9$, and $6-9$ years of historic data, seen in Table 6 [26].

For both the AP and ACP datasets, using between $6-12$ years, as was done prior portions of the paper, has the lowest MAE of all of longitudinal restrictions. This is not true for the MCP dataset; not permitting more than 9 years of data reduces the error. However, the MCP dataset only contains a maximum of 45 antibiotic-bacteria pairs per year so this may be the influence of

Table 6. State level forecasting MAE for different length time series with PYPERed.

Prior Years	AP	ACP	MCP
$6-12$	1.822	1.453	1.293
$3-12$	1.940	1.462	1.294
$3-9$	1.965	1.469	1.124
$6-9$	1.920	1.473	1.110

a single antibiotic-bacteria pair: Cephalothin and *Klebsiella oxytoca*. As seen in Figure 9, with $6-12$ years of data and $3-9$ years of data, PYPERed predicted the 2015 with an absolute error of 24.667 and 11.008, respectively [26].

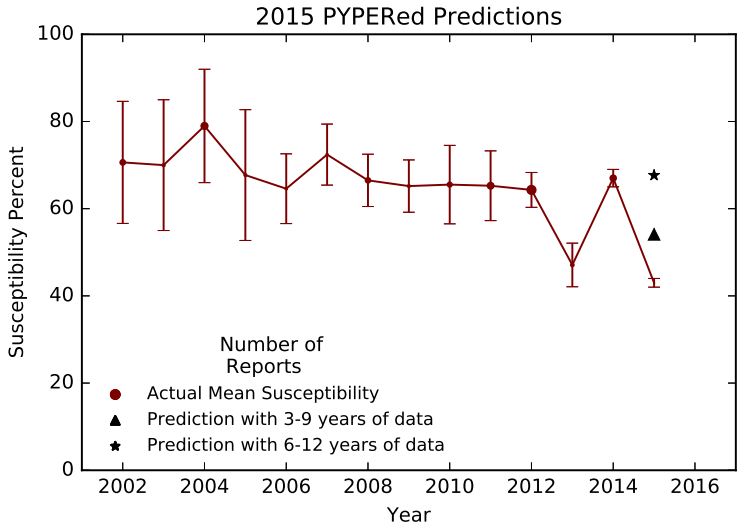

Figure 9. Cephalothin and *Klebsiella oxytoca* susceptibility.

We can attribute the results for the MCP data subset in Table 6 in part to the outlying mean susceptibilities in the resistance trend for Cephalothin and *Klebsiella oxytoca*. Thus, we conclude that it is advisable to use as much data as possible when making predictions for future antibiotic susceptibility. Though, we acknowledge that the importance of many years of data decreases as the subset of antibiotic-bacteria pairs become more clinically significant [26].

3.11. Forecasting Multiple Years into the Future

When forecasting multiple years into the future, we use between 6 and 12 years of historic data to forecast the three subsequent years at the state level for each data subset. Specifically, we

use prior years of data y_{-12} to y_{-1} to forecast target years Y_0, Y_{+1}, and Y_{+2}. The sliding window works such that forecasts Y_0, Y_{+1}, and Y_{+2} are still for years 2014, 2015, and 2016. Given that the MA-SA dataset only contains 15 years of data, this means that there may not be up to 12 years of data available with which to make these forecasts; as seen in Section 3.10, this would likely reduce the forecasting ability for AP and ACP datasets even without forecasting further into the future. These forecasts multiple years into the future are seen in Table 7 [26].

Table 7. State level forecasting multiple years into the future with PYPERed.

Forecast Year	AP	ACP	MCP
$Year_0$	1.822	1.453	1.293
$Year_1$	2.506	1.939	1.500
$Year_2$	3.134	2.459	2.047

For all three datasets, forecasting an additional year into the future consistently increases the error by around a quarter to a third of the previous error. These future forecasts indicate trends and would allow for health professionals to prepare for outbreaks[26].

3.12. $2017-2018$ Forecasting

We applied our forecasting methodology to years of data not within MA-SA. Specifically, we analyzed the trends of the antibiotic-bacteria pairs within the MCP data subset. These forecasts indicate that *E. coli, K. oxytoca, and K. pneumoniae* susceptibility to carbapenems will remain stable. *Klebsiella*'s susceptibility to cephalosporins and *E. coli*'s susceptibility to fluoquinolones are forecasted to become stable. The forecasts indicate that *Klebsiella* susceptibility to fluoro1uinolones will increase while *E. coli* susceptibility to Cephalosporins will decrease. This indicates that further actions need to be taken to prevent further *E. coli* resistance to cephalosporins [26].

3.13. Subpar Evaluations

The antibiotic-bacteria pairs with susceptibilities that are not predicted well at the state level can be sorted into three nonexclusive categories: (1) the bacteria are known to be resistant to that antibiotic, (2) there are not many data points for that antibiotic-bacteria pair, and (3) a change in CLSI guidelines caused a very sudden change in susceptibility [24].

The most common reason that antibiotic-bacteria pairs have subpar susceptibility predictions

is that the antibiotic is known not to be effective in treating the bacterial infection. For instance antibiotics in the fluoroquinolone and macrolides families were repeatedly parts of pairs that were not predicted correctly. While also part of other incorrectly predicted pairs, these antibiotics were frequently predicted badly when paired with any one of the three *Staphylococcus aureus* bacterial infections to which it is known to be frequently resistant. Ampicillin is also not used to treat *Staphylococcus aureus* infections because of high prevalence of resistance. Additionally, Nitrofurantoin, to which *Klebsiella ssp*, *Enterobacter* ssp, and *Pseudomonas* ssp are known to be resistant, is involved in multiple pairs with larger absolute errors between the predicted and actual mean susceptibilities. As these antibiotics are not being used to treat infections caused by these bacteria, it is not as important for final medical treatment if we can predict the future susceptibility of these antibiotic-bacteria pairs [24].

The second cause that antibiotic-bacteria pairs with susceptibilities that are not well predicted is a lack of reports for each year. In particular, this is an issue for the less common bacterial infections such as *Stenotrophomonas maltophilia* and *Acinetobacter baumannii*. This cause is understandable as, if there are only a small handful of reports each year, a single outlying report can greatly impact the model and prediction ability. This was the case for cephalothin and *Klebsiella oxytoca* in Section 3.10. This cause could be mitigated by implementing more rigorous cleaning strategies; however, this would also remove antibioitic-bacteria pairs that are well predicted.

Lastly, there are a few antibiotic-bacteria pairs with sudden changes in susceptibility percent due to CLSI guidelines changes that a model based on prior data could not anticipate. This is the reason that the susceptibilities for very clinically significant antibiotic-bacteria pairs, namely, *Enterobacter aerogenes* with carbapenems meropenem and imipenem, are not always well predicted. To demonstrate, Figure 10 contains the actual mean susceptibility percents a with standard deviation of the antibiotic-bacteria pair imipenem and *Enterobacter aerogenes* from 2002 to 2016. CLSI guidelines changed from 2010 to 2013, resulting in universal decreased susceptibility rates for carbapenems [20]. This change in CLSI guidelines explains the sudden observed decrease in susceptibility percent and the varied adherence to these new guidelines explains the sudden increase in standard deviation after 2010 for *Enterobacter aerogenes* and imipenem, as seen in Figure 10. Thus, by monitoring changes in CLSI guidelines, it is possible to anticipate certain antibiotic-bacteria pairs that may not be able to be predicted reliably [24].

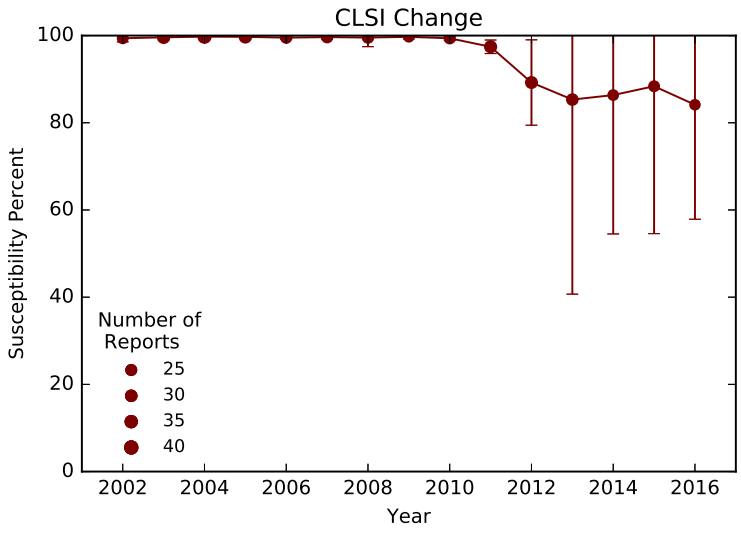

Figure 10. Imipenem and *Enterobacter aerogenes* susceptibility.

4. Clinical and Statistical Significance

In this section, we introduce **C**linically **A**nd **S**tatistically **S**ignificant **I**dentification **A**ssistant for decreases in susceptibility between two years, referred to as CASSIA [25]. Upon receiving a high volume of reports, CASSIA is designed to assist healthcare professionals by immediately identifying antibiotic-bacteria pairs demonstrating decreases in susceptibility that are both clinically and statistically significant. Thus, CASSIA outputs these antibiotic-bacteria pairs which identified for further monitoring.

4.1. Methodology

CASSIA has three stages: the input stage, the filter stage, and the output stage. During the input phase, CASSIA receives all susceptibility reports for two years. From Section 2, we know on average there are around 7000 reports per year. During the filter stage, CASSIA calculates the clinical and statistical significance of the decreases in antimicrobial susceptibility between the two years. Lastly, during the output stage, CASSIA output the antibiotic-bacteria pairs with observed decreases in susceptibility that need to be monitored, sorted by the magnitude of the decrease in susceptibility. In other words, CASSIA outputs the antibiotics that will not be as effective as a treatment to a particular bacterial infection as it was in prior years. These stages are seen in Figure 11.

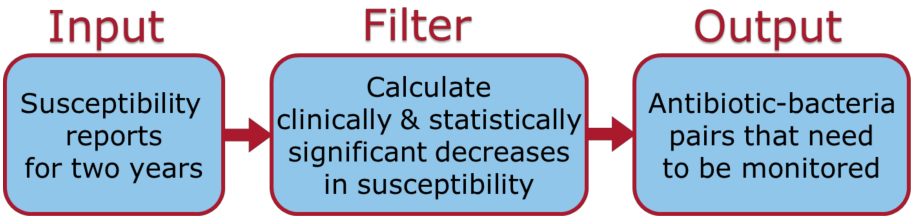

Figure 11. CASSIA pipeline.

The filter stage utilizes both the number of isolates and the number of reports in the first year to make these determinations. This is done in a 4-step methodology. Step 1 is to identify decreases in susceptibility. Step 2 is to identify which decreases demonstrate significant significance. Step 3 is to identify which decreases demonstrate clinically significant. Step 4 is to identify the antibiotic-bacteria pairs associated with these clinically and statistically significant decreases in susceptibility sorted by the magnitude of the decrease. This process is demonstrated in Figure 12 for the antibiotic-bacteria pairs that have reports in 2014 and 2015 in MA-SA [25].

4.2. Decrease in Susceptibility

First, CASSIA determines if there is a decrease in mean susceptibility between year y_1, and the subsequent year y_2 is defined. The difference d between susceptibilities in y_1 and y_2 is defined in Equation 5 where n is the total number of reports statewide in y_1, m is the total number of reports statewide in y_2, s_{1i} and t_{1i} are the susceptibility and number of isolates of the ith report in y_1, respectively, and $s_2 j$ and $t_2 j$ are the susceptibility and number of isolates of the jth report in y_2, respectively. There is a decrease in mean susceptibility between year y_1 and y_2 if $d < 0$ [25].

$$d = \frac{\sum_{i=1}^{n} s_{1i} t_{i1}}{\sum_{i=1}^{n} t_{1i}} - \frac{\sum_{j=1}^{m} s_{1j} t_{j1}}{\sum_{j=1}^{m} t_{1j}} < 0 \tag{5}$$

4.2.1. Statistical Significance. As is the current standard, CASSIA calculates the statistical significance of the decrease by using the chi-square test on the proportion of susceptible isolates. The number of isolates in y_2 is normalized to match the total number of isolates in y_1 so the test only detects the change in susceptibility. CASSIA uses the corresponding p-value with a 0.05 significance level to determine if the change in susceptibility is statistically significant [25]

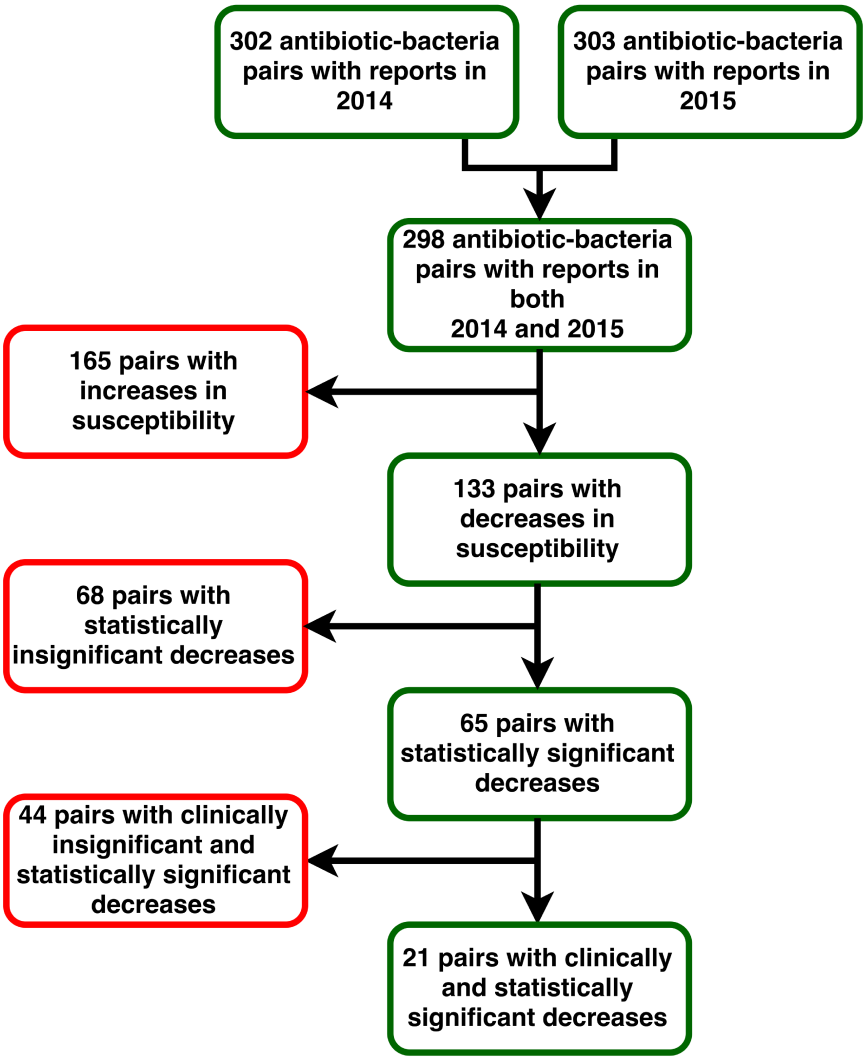

Figure 12. Amount of antibiotic-bacteria pairs with reports in 2014 and 2015 at each stage of the CASSIA methodology.

4.2.2. Clinical Significance. Domain experts indicated that for a single report in Massachusetts, an annual change in susceptibility greater than 10 susceptibility percent is considered an actionable change. This is the maximum potential change that could be attributed to errors. Thus, for this work, the imposed population standard deviation $\sigma = 10$. CASSIA uses the standard error SE formula [12], Equation 6, to propagate this clinically relevant change to n reports, which in this example is the number of reports statewide in year y_1. Thus, the calculated SE is the minimum change in susceptibility that must occur statewide during the next year for the change to be considered clinically relevant.

$$SE = \frac{\sigma}{\sqrt{n}} \tag{6}$$

Once the standard error SE is calculated, CASSIA uses Equation 7 to determine if the observed change in susceptibility demonstrates clinical significance. In other words, if the absolute difference in mean susceptibility between y_1 and y_2 (5) is greater than the calculated standard error SE, the change is considered clinically significant.

$$|d| - SE > 0 \tag{7}$$

4.3. Results

From $2006-2015$, MA-SA contains at least one report for 578 antibiotic-bacteria pairs with number of antibiotic-bacteria pairs tested in consecutive years varying from 283 to 383. Figure 13 shows the statistically significant changes in susceptibility for $2013-2014$ and $2014-2015$ [25]. Depending on the consecutive years, between $40.4-64.3$ percent of those pairs demonstrate decreases in susceptibility. Just over 50 percent of the observed decreases were statistically significant. Additionally, only 44.6 percent and 45.8 percent of the statistically significant decreases in susceptibility are greater than negative one when $y_1 = 2014$ and $y_1 = 2013$. Specifically, the statistically significant decreases ranged from 0.06 to 24.01 susceptibility percent when $y_1 = 2014$ and from 0.08 to 18.27 susceptibility percent when $y_1 = 2013$ [25]

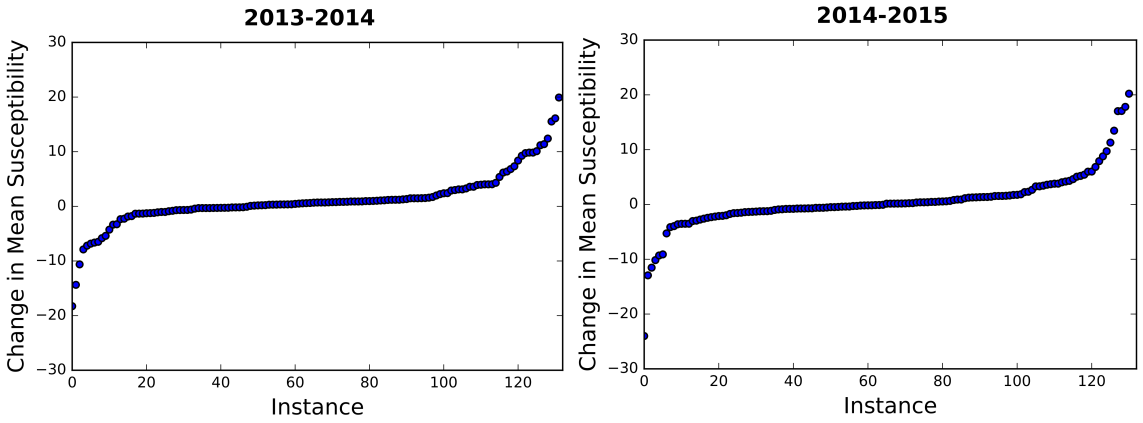

Figure 13. The statistically significant differences d between susceptibilities in y_1 and y_2 for each antibiotic-bacteria pair.

Small changes in susceptibility between years may be statistically significant based on the

Table 8. CASSIA output for $2013-2014$

Antibiotic	**Bacteria**	n	t_1	s_1	d
Nitrofurantoin	*P. aeruginosa*	11	2167	18.4	-18.3
Cefotetan	*E. cloacae*	4	257	27.9	-14.4
Cefuroxime	*P. aeruginosa*	4	762	11.3	-10.6
Moxifloxacin	*Staph. aureus*	7	5259	85.0	-7.2
Moxifloxacin	*E. coli*	10	6423	85.8	-6.8
Ticarcillin-Clav.	*P. aeruginosa*	6	557	86.9	-6.5
Cefuroxime	*E. cloacae*	15	879	40.5	-5.8
Ceftriaxone	*P. aeruginosa*	26	4610	14.5	-5.4
Ampicillin-Sulb.	*E. cloacae*	15	1109	26.1	-3.4
Moxifloxacin	MRSA	19	4805	61.8	-3.3
Levofloxacin	*S. marcescens*	40	2262	96.5	-2.3
Ampicillin-Sulb.	*E.coli*	58	104239	64.6	-1.9
Ciprofloxacin	*S. marcescens*	44	2434	94.1	-1.8

number of isolates in year y_1 but decreases of such magnitude are unlikely to be clinically significant. Thus, CASSIA also considers clinical significance based on the number of reports in year y_1. Depending on the year, only between 27 and 49 percent of the statistically significant decreases were also clinically significant. Additionally, the decreases identified as clinically significant ranged from 1.40 percent to 24.01 susceptibility percent when $y_1 = 2014$ and from 1.79 to 18.27 susceptibility percent when $y_1 = 2013$ [25]. Thus, the antibiotic-bacteria pairs with decreases in susceptibility that are clinically significant have decreases with a larger magnitude.

In the two most recent sets of consecutive years, CASSIA detected the fewest clinically and statistically significant decreases in susceptibility. CASSIA only identified 13 antibiotic-bacteria pairs in $2013-2014$ and only 21 antibiotic-bacteria pairs in $2014-2015$. The specific pairs for these years can be viewed in Tables 8 and 9, respectively. These tables contain the number of reports in the first year y_1, the number of isolates t_1 for year y_1, the susceptibility s_1 for y_1, and the susceptibility decrease d between years y_1 and y_2 [25].

It was confirmed that the antibiotic-bacteria pairs with decreases that have greatest impact on clinical practice and human health were identified by CASSIA. However, CASSIA outputs false positives as not all of the antibiotic-bacteria pairs identified by CASSIA are clinically significant.

Table 9. CASSIA output for $2014-2015$

Antibiotic	**Bacteria**	n	t_1	s_1	d
Cephalothin	*K. oxytoca*	2	87	67.0	−24.0
Cephalothin	*K. pneumonia*	2	493	89.1	−12.9
Ciprofloxacin	*S. malophilia*	2	353	40.5	−11.5
Cephalothin	*E. coli*	2	2489	43.4	−10.2
Cefazolin	*E. cloacae*	33	2916	10.3	−9.3
Amoxicillin-Clav.	*Staph. aureus*	4	2777	63.1	−9.1
Moxifloxacin	MRSA	22	5291	58.5	−5.3
Ampicillin	*E. aerogenes*	14	695	9.4	−4.1
Moxifloxacin	*E. cloacae*	8	182	96.0	−4.0
Ticarcillin-Clav.	*P. aeruginosa*	10	716	80.4	−3.5
Cefazolin	*K. oxytoca*	40	3165	63.0	−3.5
Moxifloxacin	*E. coli*	13	7748	79.0	−3.0
Ceftriaxone	*P. aeruginosa*	24	4393	9.1	−3.0
Ampicillin	*E. cloacae*	17	1610	6.9	−2.8
Tobramycin	*P. aeruginosa*	46	9675	90.5	−2.4
Cefuroxime	*E. coli*	22	25294	89.1	−2.3
Tobramycin	*E. aerogenes*	36	1248	99.0	−2.2
Ampicillin-Sulb	*K. oxytoca*	38	2595	69.0	−2.0
Ampicillin-Sulb.	*K. pneumonia*	48	14151	96.2	−1.6
Ciprofloxacin	*E. coli*	52	97056	79.7	−1.6
Ampicillin	*E. coli*	56	103981	55.9	−1.4

This can be addressed at any of the three stages: the input stage, the filter stage, and the output stage. At the input stage, only reports that contain clinically significant antibiotic-bacteria pairs can be entered into CASSIA. At the filter stage, an additional filter can be implemented to reduce antibiotic-bacteria pairs that are not clinically significant from being output by CASSIA. Examples of such filters would be to restrict antibiotic-bacteria pairs with susceptibility percent above a certain threshold and/or requires a minimum report or isolate membership in y_2 which would remove less tracked antibiotic-bacteria pairs. Lastly, users could simply disregard antibiotic-bacteria pairs that are not clinically significant at the output stage. As the definition of clinical significance, data quality, and expected number of reports/isolates can differ by place, different solutions may be

preferred by different users.

A major benefit of incorporating clinical significance is the drastic decrease in the number of antibiotic-bacteria pairs identified. Only 10.6 of the total antibiotic-bacteria pairs had clinically and statistically significant decreases in susceptibility. Figure 14 compares the number of antibiotic-bacteria pairs with decreases, statistically significant decreases, clinically significant decreases, and decreases that are both statistically and clinically significant. For consecutive years from 2006 to 2015, CASSIA returning between 13 and 66 antibiotic-bacteria pairs depending on the specific years [25].

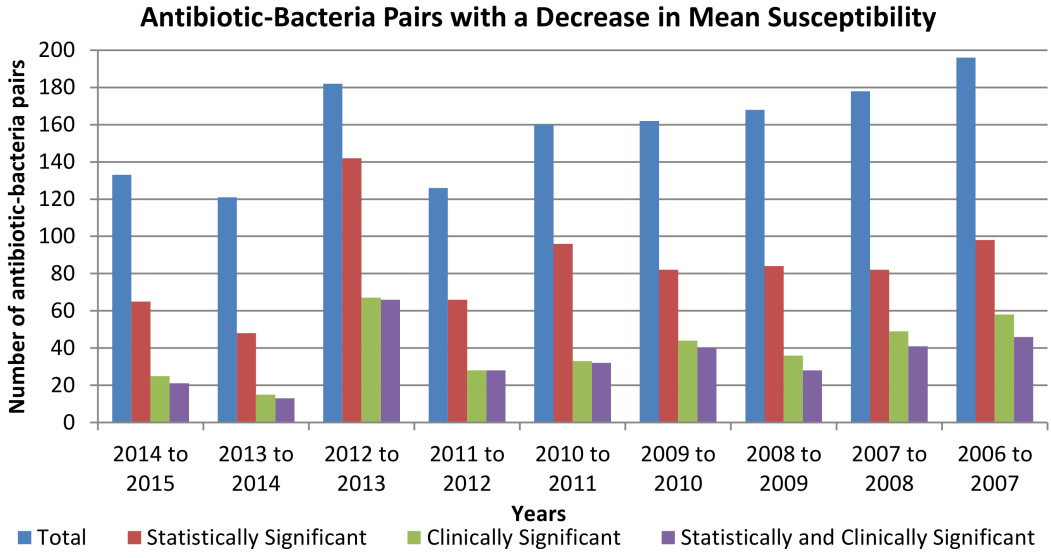

Figure 14. Number of antibiotic-bacteria pairs with a decrease in mean susceptibility by year.

CASSIA rapidly, consistently, and efficiently provides health professions with the antibiotic-bacteria pairs with clinically and statistically significantly decreases in susceptibility. A manual examination of the reports for every antibiotic-bacteria would be both more time consuming and subjective. CASSIA provides health professions with a small subset of the antibiotic-bacteria pairs immediately upon receiving the reports. The output of CASSIA is easily consumable and allows the user to quickly notice if there is an antibiotic that will no longer be effective.

It is important to note that CASSIA assists with prioritizing the analysis and monitoring of antibiotic-bacteria pairs by rapidly highlighting significant pairs within large volumes of antibiogram data. If there is an antibiotic or bacteria of particular concern, a healthcare professional

can still access the raw data and perform a manual analysis [25]. CASSIA is meant to assist, not replace, a trained healthcare professional.

5. Anomalous Trend Detection

In this section, we will introduce Anomalous Resistant Trend *I*dentification system, further referred to as ARTI. This system removes antibiotic-bacteria which have anomalous antibiotic susceptibility trends to increase the forecasting ability of the remaining antibiotic-bacteria pairs. As the MPDH currently assumes the prior year of resistance is reflective of the resistance in the current year, we will use ARTI to remove antibiotic-bacteria pairs where this assumption is erroneous. For this purpose, our objective is to tackle the challenging problem of optimizing two negatively correlated goals: minimizing the prediction error and minimizing the number of antibiotic-bacteria pairs removed.

5.1. ARTI Methodology

ARTI, like CASSIA in Section 4.1, has three basic steps: the input stage, the detect stage, and the output stage. During the input phase, ARTI receives all antibiogram reports as well as user preferences. Specifically, the user is able to specify the percent of antibiotic-bacteria pairs it is acceptable to exclude; this parameter is further referred to as $c\%$. Also, the user can choose to enter a prediction method M, but this research focuses on using the default which assumes the prior year of resistance. During the detect phase, for every set of parameters, ARTI will calculate the outlier score for all of the resistance trends; the $c\%$ with the highest outlier score in the parameter set with the lowest error are considered outliers. During the output stage, ARTI outputs the set of antibiotic-bacteria pairs that do not have anomalous antibiotic resistance trends. In other words, ARTI will output the antibiotic-bacteria pairs that can be predicted with method M. This process is summarized in Figure 15.

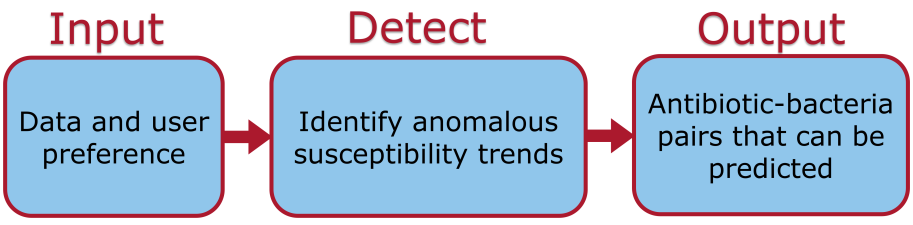

Figure 15. ARTI pipeline.

We originally apply ARTI to find anomalous trends in 12 years of data: $2002-2013$ and $2003-2014$. Each susceptibility time series must have at least 3 years of data. Then we identify the parameter set with the lowest mean absolute error (MAE), Section 3.6, between the predicted and actual mean susceptibilities for target years 2014 and 2015. We validate ARTI by applying that identified parameter set to data from $2004-2015$ to identify anomalous trends in 2016. This indicates that the results were not year specific.

ARTI detects the anomalous susceptibility trends in a 4 step methodology. Step 1 involves identifying the antibiotic-bacteria pairs with the highest outlier scores for each parameter set. We employ two different unsupervised discriminative approaches f to calculate these outlier scores [9]: Euclidean distance with k-means clustering and dynamic time warping distance between all time series. Step 2 involves using these approaches f with parameters $p_1 \dots p_q$ in the parameter set to compute the number of antibiotic-bacteria pairs (nAB) with outlying resistance trends and the MAE, as seen in Equation 8

$$(nAB, MAE) = f(p_1, \dots, p_q) \tag{8}$$

Step 3 involves adjusting the values of nAB and MAE based on user input. Given the differences in scale, adjusting these values allows for the user to determine the importance in minimizing each. Thus, nAB_{adj} is weighted with w, which is determined by user input $c\%$. The formulas for nAB_{adj} and MAE_{adj} for the ith parameter set are given in Equations 9 and 10, respectively. Step 4 involves selecting the parameter set that has the minimal distance between the Euclidean distance between (nAB_{adj}, MAE_{adj}) and $(0,0)$.

$$nAB_{adj} = w\frac{nAB_i}{max(nAB)} \tag{9}$$

$$MAE_{adj} = \frac{MAE_i - min(MAE)}{max(MAE) - min(MAE)} \tag{10}$$

5.1.1. Euclidean Clustering Approach. We identify the anomalous resistance trends by comparing these trends to the common shapes of resistance. This is accomplished through k-means clustering with Euclidean distance. Utilizing k-means clustering is a popular strategy to cluster temporal data [9]. For this research, Euclidean distance is used to determine the distance between time series at each time step and cluster the time series. Thus, the time series must be the same length. The

outlier score is calculated as the Euclidean distance between the time series and the centroid of the closest cluster [9].

5.1.2. Dynamic Time Warping Approach. A more traditional distance measure for comparing time series is Dynamic Time Warping (DTW) distance. Specifically, DTW compares time series at their most similar points, regardless of the time stamp [18]. This is an advantage as time series do not need to be the same length to compare nor do the trends need to occur at the same time throughout the time series. The DTW distance between each time series is calculated. The outlier score for each time series is either the maximum or average of its distance from the other time series.

5.1.3. The Parameter Sets. The parameters that are applicable to both Euclidean and DTW methodology are the maximum number of outliers (mOut), the normalization method (normM), the time-alignment (tAlign). As we are interested in minimizing the number of outliers, we experiment with mOut values of 1, 5, 10, and 15. The mOut value is not the same as user input $c\%$ since other parameters and the approach may influence the percent of pairs excluded.

We implement two different normalization techniques: (1) subtracting each value of the time series by the first value of the time series and (2) subtracting each value of the time series by the previous value in the time series. When no normalization technique is applied, k-means clustering in Section 5.1.1 yields a series of almost horizontal lines at different levels of susceptibility. As our goal is to identify resistance trends with abnormal shapes of resistance instead of resistance trends with abnormal levels of resistance, normalization of the time series is crucial. We also consider the time series with and without time-alignment; Figure 16 demonstrates this difference. A series without time-alignment considers only the ordering of the year when forming the timeseries. A series with time-alignment estimates the missing years of data as an average of the susceptibility directly before and after the missing year.

There are parameters that are unique to Euclidean clustering approach. These include the number of clusters (nk), the window length (lenWindow), the set of windows (sWindows), and the set type (sType). We consider between 1 and 5 clusters when k-means clustering (Section 5.1.1). Calculating the Euclidean distance between time series requires the time series to be the same length. As such, we consider overlapping windows of our time series [9]. We consider windows of length 3 to length 12. Also, we consider either all of the windows or only the last window of the time series.

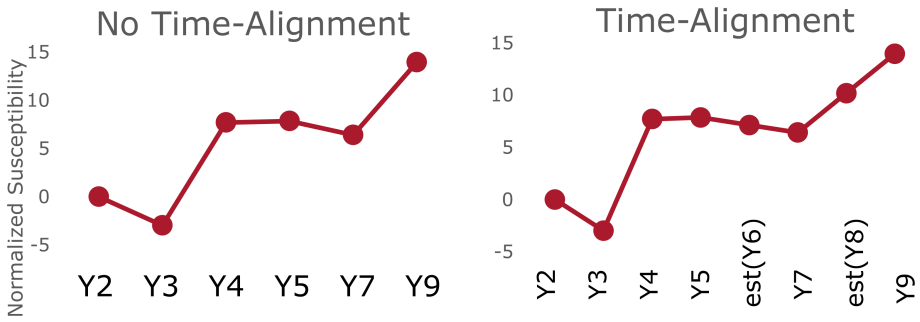

Figure 16. Comparison of a time series with and without time-alignment.

Lastly, we combine sets of outlying resistance in different window lengths. The sType can be individual, union, intersection, unionEnd, or intersectionEnd. Individual is if we simply take a single set of outlying resistance for a given window length. For the union, we include the time series if it falls within the outlier sets for any window length. For the intersection, we include the time series if it falls within every outlier set for any window length. For unionEnd and intersectionEnd, we only consider whether the time series is an outlier in the smaller window lengths. Thus, calculating the (nAB, MAE) becomes a function f_E of the parameters for the Euclidean clustering approach, Equation 11.

$$(nAB, MAE) = f_E(mOut, normM, tAlign, nk, lenWindow, sWindows, sType) \tag{11}$$

The parameters that are unique to the DTW approach are the maximum years of data (maxT) and the manner of calculating the outlier score (outScore). First, we consider the most recent 3 to the most recent 12 years of data so maxT ranges from 3 to 12. As mentioned in Section 5.1.2, the outlier score is calculated as either the maximum or average of its distance from the other time series. Thus, calculating the (nAB, MAE) becomes a function f_{DTW} of the parameters for the DTW approach, Equation 12.

$$(nAB, MAE) = f_{DTW}(mOut, normM, tAlign, maxT, outScore) \tag{12}$$

5.2. Results

When no outliers are removed, $(nAB, MAE) = (0, 1.778)$ for target years 2014 and 2015. There are a total of 568 susceptibility time series in these years. There are 2560 sets of parameters, the (nAB, MAE) of these sets are plotted in Figure 17.

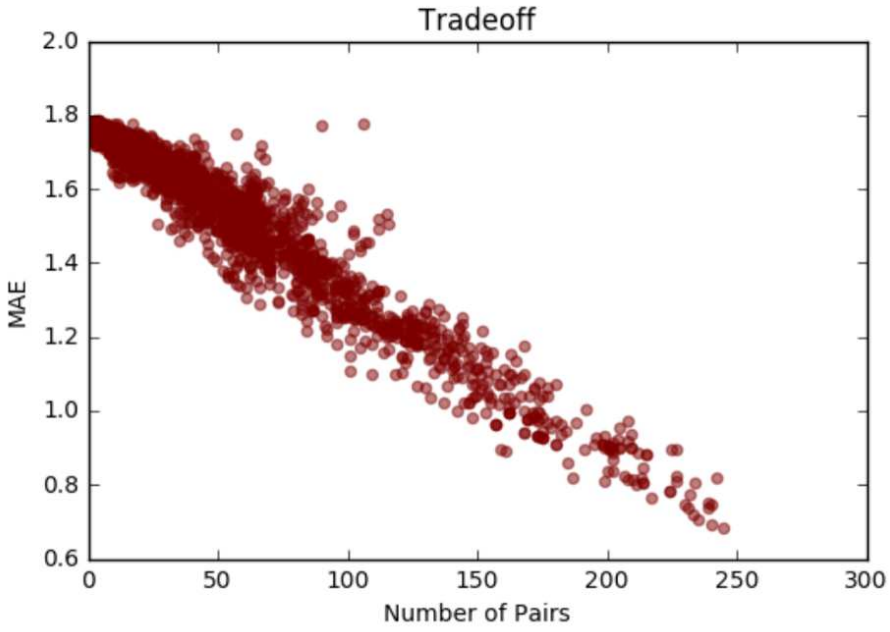

Figure 17. The number of antibiotic-bacteria pairs and MAE for all 2560 **parameter sets.**

From Figure 17, we can see that there is a negative linear relationship between the number of antibiotic-bacteria pairs excluded and the MAE once those pairs are removed. The most antibiotic-bacteria pairs removed is 245 and the smallest MAE is 0.682. However, given the domain, we wish to minimize both of these numbers. The weight given to each is determined by the user. The greater percent of antibiotic-bacteria pairs the user is willing to exclude $c\%$, the smaller the resulting MAE will be.

We demonstrate ARTI's effectiveness for $c\% = 20\%$, $c\% = 10\%$, and $c\% = 5\%$ in Table 10 for target years 2014 and 2015. These percents correspond to weights $w = 1$, $w = 2$, and $w = 3$ in Equations 9 and 10. This process means that while the number of antibiotic-bacteria pairs excluded may not be exactly $c\%$, the (nAB, MAE) corresponding to the selected set will be a nearby local minima.

Table 10. ARTI detection results for target years 2014 **and** 2015**.**

w	c%	nAB%	nAB	MAE	MAE Reduction
1	20%	19.2%	109	1.10	0.68
2	10%	10.0%	57	1.34	0.44
3	5%	5.3%	30	1.46	0.32

For $w = 1$ and $w = 3$, a Euclidean methodology best minimized both nAB and MAE. However, for $w = 2$, a DTW methodology minimized (nAB, MAE). There was no clear parameter set that was preferable. We apply the parameter sets that produced the results in Table 10 to $2004 - 2015$ data to validate these parameter sets for 2016 (Table 11).

Table 11. ARTI detection results verified on 2016**.**

w	c%	nAB%	nAB	MAE	MAE Reduction
1	20%	19.2%	63	1.66	1.1
2	10%	10.0%	57	1.87	0.89
3	5%	5.3%	20	1.99	0.77

The MAE in 2016 before outliers were removed was much higher than the MAE in 2014 and 2015. As such, it makes sense that the MAE reduction for 2016 after excluding anomalous time series is higher than for 2014 and 2015. ARTI is effective at reducing MAE which means it is excluding antibiotic-bacteria pairs where the prior resistance is not reflective of the current resistance. If an antibiotic-bacteria pair of interest is excluded, the user can re-run ARTI with a smaller $c\%$. The 2016 anomalous resistant trends that correspond to $c\% = 5\%$ are show in Figure 18.

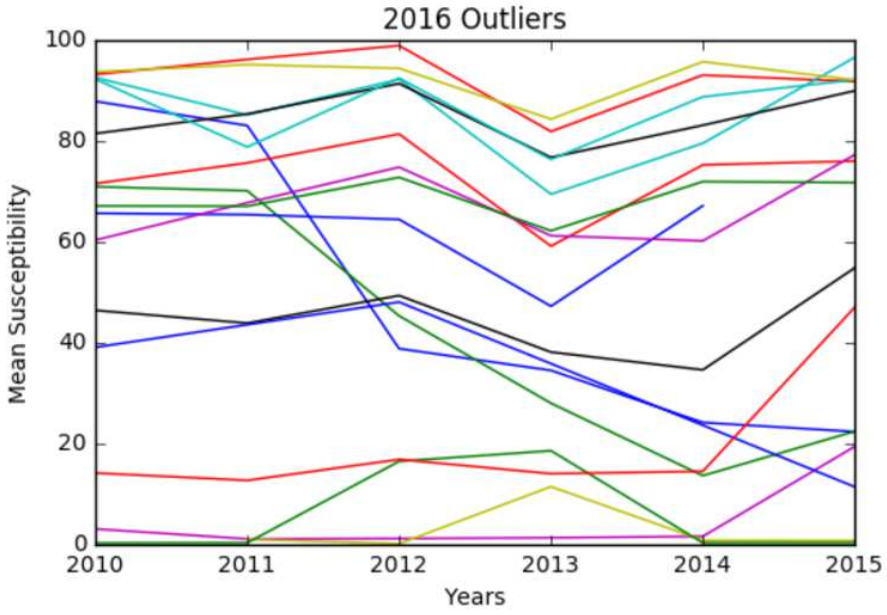

Figure 18. Identified anomalous resistance trends.

Overall, there is no set of parameters that performed the best. However, there individual parameters that clearly produced better results for this particular application. For instance, only the last window should be included when using the Euclidean methodology and the maximum distance should be used when calculating the DTW outlier score. In general, the shorter time series are more effective at identifying anomalous antibiotic resistant trends in this domain.

ARTI can be expanded to include additional or different unsupervised discriminative approaches to outlier detection. In addition, the maximum outliers in the current approaches can be increased to remove a greater percent of outliers if desired. This permits for flexibility when ARTI is applied to other domains.

6. Additional Research

During the completion of the aforementioned research, we also completed two additional tasks that did not fall into the above categories. These tasks involved isolating the impact of county on antimicrobial resistance and comparing the susceptibility in different counties.

6.1. Isolating Geospatial Impact

The second task was motivated by another study that isolated the impact of time on susceptibility resistance [13]. As location is the only other variable in our dataset, we explored the impact of county on susceptibility using multivariate regression. However, our results showed that the county did not statistically significantly influence antibiotic susceptibility. As such, we are not able to improve predictions by including county information.

6.2. Geospatial Susceptibility Comparison

The MA-SA dataset from $2002-2015$ contains 2654 antibiotic-bacteria-year combinations. We compared whether the susceptibilities of the counties for all of these antibiotic-bacteria-year combinations was different using ANOVA [12]. The susceptibilities were only different in 14.3 percent of antibiotic-bacteria-year combinations. From this we conclude that neighboring counties susceptibility should have no influence on prescribing practices for hospitals within a particular county the majority of the time. However, for around 25 antibiotic-bacteria pairs each year, neighboring counties will have different antibiotic susceptibility for certain bacterial infection. If a

nearby county has a lower antibiotic susceptibility, a different antibiotic may be preferred in case the resistant bacterial infection has spread to nearby counties. Such information could be displayed in a manner similar to Figure 19 overlaid on a map of Massachusetts counties; the size of the mean susceptibility for each county is scaled with the number of reports.

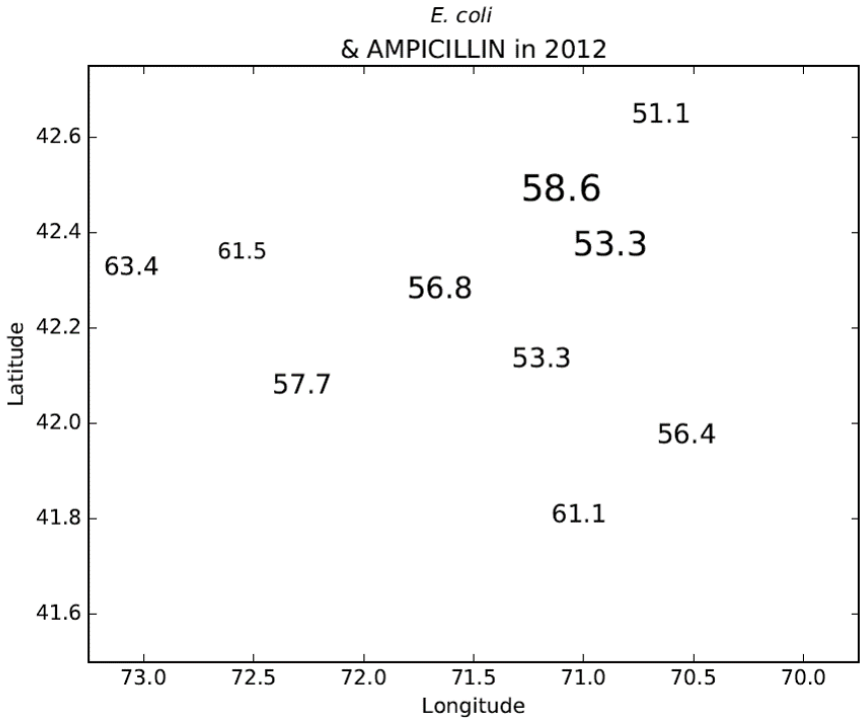

Figure 19. County susceptibilities for an antibiotic-bacteria-year combination.

7. Conclusions

Prescribing antibiotics based on the most current resistance levels possible is important to reducing further growth of resistance. Currently the Massachusetts Department of Health (MDPH) is prescribing antibiotics based on outdated resistance knowledge. We use historic data to make current susceptibility forecasts as well as develop tools to help improve the use of this outdated data. Our research helps fight antibiotic resistance by improving antibiotic prescribing.

First, we design a model selector called PYPERed which selects weighted regression-based methods to make predictions about the current year of resistance. This is important as not all antibiotic-bacteria pairs have a linear distribution over time. We were able to predict the current susceptibility of over 270 antibiotic-bacteria pairs on average within 1.882 susceptibility percent

of the actual mean susceptibility using PYPERed.

CASSIA is an assistant which immediately identifies antibiotic-bacteria pairs with clinically and statistically significant decreases in susceptibility once data is available. This research introduces a way to mathematically calculate clinical significance. CASSIA allows for rapid reaction to changing antibiotic susceptibilities.

ARTI is a system that identifies and removes antibiotic-bacteria pairs with anomalous resistance trends. It allows for the user to indicate the percent of data that is acceptable to be excluded. Removing these anomalous trends helps reduce the prediction error.

Our next step is, instead of predicting the future susceptibility percent, predicting if the change in antimicrobial susceptibility between years will be clinically and statistically significant [25]. Other future antibiogram work would involve forecasting with data is that collected either more or less frequently [26]. Instead of removing outliers at the time series level, removing reports with outlying susceptibilities is another future step. Lastly, PYPERed, CASSIA, and ARTI should be applied to other datasets.

References

[1] Alvarez-Uria, G., Gandra, S., Mandal, S., Laxminarayan, R. (2018). Global forecast of antimicrobial resistance in invasive isolates of Escherichia coli and Klebsiella pneumoniae. International Journal of Infectious Diseases, vol. 68, pp. 50-53.

[2] Anderson, D., Miller, B., Marfatia, R., and Drew, R. (2012). Ability of an antibiogram to predict pseudomonas aeruginosa susceptibility to targeted antimicrobials based on hospital day of isolation. Infection Control & Hospital Epidemiology, 33(6):589–593.

[3] Beaulac, K., Tlachac, M., Rundensteiner, E., Barton, K., Troppy, S., Doron, S. (2018) Utilization of Predictive Modeling to Identify Emerging Statewide Antibiotic Resistance Trends. The Society for Healthcare Epidemiology of America (SHEA). Conference Presentation.

[4] Bureau of Infectious Disease and Laboratory Sciences. (2016). 2015 statewide antibiogram report. Massachusetts State Public Health Laboratory. Accessed 24 Jan 2017.

[5] Centers for Disease Control and Prevention. (2015). Antibiotic resistance threats in the United States, 2013. U.S. Department of Health and Human Services. https: //www.cdc.gov/drugresistance/pdf/ar-threats-2013-508.pdf.

[6] Centers for Disease Control and Prevention. (2018). About Antimicrobial Resistance. Antibiotic/Antimicrobial Resistance. https://www.cdc.gov/drugresistance/about.html. Accessed May, 2018.

[7] Crnich, C., Safdar, N., Robinson, J., and Zimmerman, D. (2007). Longitudinal trends in antibiotic resistance in US nursing homes, 2000-2004. Infection Control and Hospital Epidemiology, 28(8):1006–1008.

[8] Food and Drug Administration (2016). Stewardship guidelines. The National Antimicrobial Resistance Monitoring System.

[9] Gupta, M., Gao, J., Aggarwal, C., Han, J. (2014). Outlier Detection for Temporal Data: A Survey. IEEE Transactions on Knowledge and Data Engineering, vol. 26(9).

[10] Hare, K., Singleton, R., Grimwood, K., Valery, P., Cheng, A., Morris, P., Leach, A., Smith-Vaughan, H., Chatfield, M., Redding, G., Reasonover, A., McCallum, G., Chikoyak, L., McDonal, M., Brown, M., Torzillo, P., Chang, A. (2013). Longitudinal nasopharyngeal carriage and antibiotic resistance of respiratory bacteria in indigenous Australian and Alaska children with bronchiectasis. PLOS ONE, 8(8).

[11] Hastey, C., Boyd, H., Schuetz, A., Anderson, K., Citron, D., Dzink-Fox, J., Hackel, M., Hecht, D., Jacobus, N., Jenkins, S., Karlsson, M., Knapp, C., Koeth, L., Wexler, H., and Roe-Carpenter, D. (2016). Changes in the antibiotic susceptibility of anaerobic bacteria from 2007-2009 to 2010-2012 based on CLSI methodology. Anaerobe, 42:27–30.

[12] James, G.,Witten, D., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning with Applications in R. Springer-Verlag, New York, 1 edition.

[13] Lagace-Wiens, P., Adam, H., Low, D., Blondeau, J., Baxter, M., Denisuik, A., Nichol, K., Walkty, A., Karlowsky, J., Mulvey, M., Hoban, D., and Zhanel, G. (2013). Trends in antibiotic resistance over time among pathogens from canadian hospitals: Results of the canward study 2007-11. Journal of Antimicrobial Chemotherapy, 6:i23–i29.

[14] Moore, D. (2007). The Basic Practice of Statistics. WH Freeman, New York, 4 edition. Rennie, R. and Jones, R. (2014). Effects of breakpoint changes on carbapenem susceptibility rates of enterobacteriaceae: Results from the sentry antimicrobial surveillance program, united states, 2008 to 2012. Canadian Journal of Infectious Diseases and Medical Microbiology, 25(5):285–287.

[15] Nau, R. (2018). Statistical forecasting: notes on regression and time series analysis. Fuqua School of Buisness, Duke University. https://people.duke.edu/rnau/411home.htm.AccessedMay,2018.

[16] Neumtu, R., Ahsan, R., Rundensteiner, E., Sarkozy, G. (2016). Interactive time series exploration powered by the marriage of similarity distances. Proceedings of the Very Large Data Bases Endowment, 10(3).

[17] O'Neill, J. (2016). Tackling Drug-Resistant Infections Globally: Final Report and Reccomendations. The Review on Antimicrobial Resistance, https://amr-review.org/sites/default/files/160525Final\%20paperwith\%20cover.pdf.

[18] Rakthanmanon, T., Campana, B., Mueen, A., Batista, G., Westover, B., Zhu, Q., Zakaria, J., Keogh, E. (2012). Searching and mining trillions of time series subsequences under dynamic time warping. Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining: 262-270.

[19] Raposo, L., Arruda, M. (2014). Probabilistic neural network for predicting resistance to HIV-protease inhibitor nelfinavir. In Proceedings of the International Conference on Bioinformatics Models, Methods and Algorithms: 17-23.

[20] Rennie, R. and Jones, R. (2014). Effects of breakpoint changes on carbapenem susceptibility rates of enterobacteriaceae: Results from the SENTRY antimicrobial surveillance program, United States, 2008 to 2012. Canadian Journal of Infectious Diseases and Medical Microbiology, 25(5):285–287.

[21] Sanders, D., Smidt, R. (2000). Statistics: a first course. McGraw-Hill Companies, Inc., 6th edition.

[22] Seidman, J., Johnson, L., Levens, J., Mkocha, H., Munoz, B., Silbergeld, E., West, S., Coles, C. (2016). Longitudinal comparison of antibiotic resistance in diarrheagenic and non-pathogenic Escherichia coli from young tanzian children. Frontiers in Microbiology: 7: 1420.

[23] Smola, A. and Scholkopf, B. (2004). A tutorial on support vector regression. Statistics and Computing, 14(3):199–222.

[24] Tlachac, M., Rundensteiner, E., Barton, K., Troppy, S., Beaulac, K., Doron, S. (2018). Predicting Future Antibiotic Susceptibility using Regression-based Methods on Longitudinal Massachusetts Antibiogram Data. In Proceedings of the 11th International Conference on Health Informatics.

[25] Tlachac, M., Rundensteiner, E., Barton, K., Troppy, S., Beaulac, K., Doron, S., Zou, J. (2018). CASSIA: An Assistant for Identifying Clinically and Statistically Significant Decreases in Antimicrobial Susceptibility. 2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI).

[26] Tlachac, M., Rundensteiner, E., Troppy, S., Beaulac, K., Doron, S., Zou, J., Barton, K. (2018). Predictive Modeling of Emerging Antibiotic Resistance Trends. Communications in Computer and Information Science, Springer. In preparation.

[27] Ventola, L. (2015). The antibiotic resistance crisis. Pharmacy and Therapeutics, 40(4):277–283.

[28] World Health Organization. (2014). Antimicrobial resistance global report on surveillance 2014. World Health Organization. http://apps.who.int/iris/bitstream/handle/10665/112642/9789241564748eng.pdf.

[29] Yang, H. and King, L. (2009). Localized support vector regression for time series prediction. Nuerocomputing, 72(10-12):2659–2669.