DEVELOPMENT AND EVALUATION OF A CLINICAL DECISION SUPPORT SYSTEM FOR THE PREDICTION OF METHICILLIN RESISTANT *STAPHYLOCOCCUS AUREAUS* SURGICAL SITE INFECTIONS IN PATIENTS UNDERGOING MAJOR SURGICAL PROCEDURES IN THE UNITED STATES.

By

Kevin A. Wilson

A Dissertation Submitted

In Partial Fulfilment of the Requirements for the Degree of

Doctor of Philosophy in Biomedical Informatics

Department of Health Informatics

Rutgers, The State University of New Jersey

School of Health Professions

March, 2019

RUTGERS
BIOMEDICAL AND
HEALTH SCIENCES

Final Dissertation Approval Form

DEVELOPMENT AND EVALUATION OF A CLINICAL DECISION SUPPORT SYSTEM FOR THE PREDICTION OF METHICILLIN RESISTANT *STAPHYLOCOCCUS AUREAUS* SURGICAL SITE INFECTIONS IN PATIENTS UNDERGOING MAJOR SURGICAL PROCEDURES IN THE UNITED STATES.

By

Kevin Anthony Wilson

Dissertation Committee:
Shankar Srinivasan, PhD, Committee Chair
Frederick Coffman, PhD, Committee Member
Suril Gohel, PhD, Committee Member

Approved by the Dissertation Committee:

_____ Date: _____

_____ Date: _____

_____ Date: _____

Abstract

DEVELOPMENT AND EVALUATION OF A CLINICAL DECISION SUPPORT SYSTEM FOR THE PREDICTION OF METHICILLIN RESISTANT *STAPHYLOCOCCUS AUREAUS* SURGICAL SITE INFECTIONS IN PATIENTS UNDERGOING MAJOR SURGICAL PROCEDURES IN THE UNITED STATES.

By

Kevin Anthony Wilson

Methicillin-resistant *Staphylococcus aureus* (MRSA) is the leading cause of antibiotic resistance related mortality in surgical patients. Effective prediction of MRSA and MRSA-related SSI would facilitate the prophylactic use of appropriate antibiotics or application of other prevention techniques, which have been shown to improve clinical outcomes. While there is a range of factors that have been shown to increase the risk of MRSA-related infections, research is less clear on the best approaches to developing predictive models for incorporation into a clinical decision support system. This study compared two common modeling approaches — logistic regression (LR) and artificial neural networks (ANN) — for the prediction of MRSA infection and MRSA-related SSI in patients undergoing major surgical procedures (MSPs) in the United States.

The data source for analysis is the National Inpatient Sample, which contains approximately 7 million discharges each year. A descriptive analysis was performed to identify potential predictors for each of three research hypotheses and ANN and LR models were developed and evaluated for the prediction of: (1) MRSA infection in patients undergoing MSPs; (2) MRSA-related SSI in patients undergoing MSPs; and (3) MRSA-related SSI in patients with *S. aureus* infection.

The ANN model performed best for Hypothesis 1, with an AUC of 0.87, sensitivity of 0.86 and specificity of 0.74; the LR model achieved an AUC of 0.85, sensitivity of 0.79 and specificity of 0.75. For Hypothesis 2, the ANN model achieved an AUC of 0.86, sensitivity of 0.73 and specificity of 0.87; the LR model achieved an AUC of 0.85, sensitivity of 0.77 and specificity of 0.76. For Hypothesis 3, the ANN model achieved an AUC of 0.67, sensitivity of 0.57 and specificity of 0.67; the LR model achieved an AUC of 0.68, sensitivity of 0.61 and specificity of 0.64.

This study assessed the feasibility of LR and ANN for the prediction of MRSA-related infections in surgical patients using a range of demographic, clinical, procedural, and hospital-related factors. The results showed that both algorithms are effective modeling approaches with reasonable sensitivity and specificity and suggest that a clinical decision support tool based on either model could be informative in clinical practice.

Acknowledgements

I would like to thank the faculty and staff of the Department of Health Informatics, particularly my advisor, Dr. Shankar Srinivasan and committee members, Dr. Frederick Coffman and Dr. Suril Goehl. Their guidance, support and advice through the dissertation process has been instrumental in the successful completion of this dissertation.

I appreciate the insight and guidance of Dr. Per Ashorn, Dr. Cindy Lawler, and Dr. Guy White, all of whom supported me through the process and helped me stay on the right track.

I would also like to thank Christopher Siege and Dr. Dennis Wallace from RTI International for their support, flexibility, and advice through the process. Completing such a long and complex work would not be possible without the flexibility and support of my employer and coworkers.

Finally, I would like to thank my wife, Julie, for her unending support, for taking care of me through this process, and taking care of everything that I haven't been able to do.

To

The people who have supported me the most

and without whom none of this would be possible:

My parents and my wife ❤

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Hospital-acquired infections (HAI) are a major cause of morbidity and mortality and lead to prolonged hospital stays, with approximately 15% of all hospitalized patients worldwide experiencing a HAI.[1-6] HAIs are defined as infections acquired during a hospital stay or within 30 days of discharge and are the most common complication in hospitalized patients.[1,4,5,7,8] At any given time, over 1.4 million people experience HAIs, representing approximately 10% of all hospital admissions in high-income countries.[1,5] HAIs acquired in an acute care setting, which are known as nosocomial infections, are particularly prevalent in intensive care units and when invasive surgical procedures are performed.[2,8] These infections are responsible for up to 100,000 deaths annually in the United Sates (US) and are the fifth leading cause of death in acute care hospitals.[1,8,9]

The Centers for Disease Control and Prevention (CDC) estimates that of the 722,000 hospital-acquired infections that occurred in US hospitals in 2011, around 75,000 resulted in death.[6,10-12] At any given time, approximately 1 in 25 admitted hospital patients experiences at least one HAI and, in addition to increased morbidity and mortality, HAIs result in significantly increased costs for patient care.[4,12] It is estimated that, when controlling for disease severity and ICU status, the cost of caring for a patient increases from $6767 to $15,275 when an HAI is present, and the overall cost of HAIs annually is $33 billion.[7,13] Thus, it is important to identify and prevent infections as early as possible

in the patient's care.[3] Prevention and treatment of HAIs is dependent on both the pathogen causing the infection and the nature of the infection.[11]

The majority of HAIs are caused by a small number of bacterial pathogens.[1,6,7,14] Common bacterial pathogens include *Staphylococcus aureus*, *Escherichia coli*, *Vancomycin-resistant enterococci*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, and *Clostridium difficile*.[14] In addition, approximately 5% of HAIs are caused by viruses, such as Hepatitis B and C, influenza, HIV, rotavirus, and herpes-simplex, and an even smaller number are caused by fungal parasites, such as *Aspergillus spp*.[1] In a study of pediatric HAIs, the most common pathogen was *S. aureus*, which accounted for 17% of all infections.[6] These pathogens are associated with a range of infection sites and types.[14]

The most common HAIs in the United States are urinary tract infections, surgical and soft tissue infections, bloodstream infections, pneumonia and gastrointestinal infections.[11,12,14] Invasive devices and procedures, such as central line associated blood stream infections, catheter-associated urinary tract infections, and ventilator associated pneumonia are particularly prevalent.[7,15] Bloodstream infections have been estimated to have a mortality rate of up to 25% and surgical site infections have been shown to result in prolonged hospitalization.[1,15] A national study of 183 hospitals estimated that device associated infections collectively accounted for 47% of all HAIs, gastro-intestinal infections accounted for 9.2%, and pneumonia and surgical site infections (SSIs) accounted for 21.8% each.[11] Surgical procedures were the single most common source of HAIs.

Worldwide over 234 million surgeries are performed annually.[16] In the US alone, it is estimated that between 20 and 27 million surgical procedures are performed each year.[16,17] Surgical episodes are the single most expensive hospital cost and approximately $400 billion per year is spent on surgical care.[18] According to the Health Care Utilization Project (HCUP), each year there are 7.2 million inpatient surgeries and 9.9 million ambulatory surgeries, with over 17 million hospital visits that include invasive, therapeutic surgery.[19] Patients undergoing surgical procedures are at an elevated risk of acquiring a HAI.[7,15,20]

Despite progress in their prevention, SSIs are among the most common HAIs in the US.[11,20-22] SSIs occur in approximately 5% of all surgical procedures and range in severity from easily management to serious and life threatening.[20] Although estimates vary, studies have found that SSIs account for between 11 and 26% of all healthcare associated infections.[21] The incidence of SSIs varies by operative procedure, with rates as high as 10.5% for cardiac surgery, 7% for vascular procedures, 2.4% for orthopedic procedures, and 4.8% for breast surgery.[22] As with other HAIs, SSIs are known to increase, morbidity, mortality, length of stay, and treatment cost.[23]

1.1 Background of the Problem

SSIs are a leading cause of morbidity and mortality in the US.[16,23,24] Formally, an SSI is defined as an infection that occurs within 30 days of surgery and SSIs are typically classified as deep incisional, superficial incisional, or organ space.[23,24] The CDC classifies surgical wounds as clean, clean-contaminated, contaminated and dirty with overall risk of SSI ranging from 3% in a clean wound to over 27% in a dirty wound.[23,25]

In the US, SSIs occur in approximately 150,000 patients each year and cost the healthcare system $10 billion. It is estimated that up to 60% of SSIs are preventable.[23] For prevention and treatment to be effective, it is important to identify the cause and type of the infection.

SSIs are generally caused by one of a small number of pathogens, including *Escherichia coli*, *Clostridium difficile* and *Staphylococcus aureus*.[1,6,11,14,15] SSI-causing pathogens differ based on the type of surgery, the location in the body, and whether a prosthesis is used.[21] The most common pathogen that causes SSIs is *S. aureus*, which accounts for up to 22% of all SSIs.[6] The *S. aureus* pathogen is endogenous to patients and resides on the skin and in nasal passages.[23] Approximately 20% of patients are persistently colonized with *S. aureus* with a further 30% intermittently colonized.[14] *S. aureus* is transmitted on the skin of hospital workers and patients, and enters the body through open wounds.[1,26] *S. aureus* can be difficult to treat due to the development of antibiotic resistance.

S. aureus is increasingly becoming resistant to standard antibiotic treatments.[20,27,28] Antibiotic resistance in *S. aureus* began almost immediately after the invention of Penicillin in the 1940s.[29,30] In recent years, the continued development of resistance has been fueled by the overuse of broad spectrum antibiotics.[30] Methicillin-resistant *Staphylococcus aureus* (MRSA) is the leading cause of antibiotic resistance related mortality in the US.[31] Approximately 60% of all *S. aureus* infections are now caused by MRSA, which results in the pathogen being resistant to all Beta-lactam antibiotics.[30,32] The large proportion of SSIs caused by *S. aureus* and the associated risk of Methicillin resistance presents major challenges to the prevention and treatment of SSIs.[33]

MRSA is endemic in many hospital settings and an increasing number of *S. aureus* infections are becoming resistant.[16,17,20,31] SSIs that are the result of a MRSA infection are not treatable by standard antibiotics, such as clarithromycin, cotrimoxazole, and gentamycin.[20] While estimates vary, some studies estimate that *S. aureus* is responsible for up to 30% of SSIs, and between 50 and 60% of *S. aureus* infections are caused by MRSA.[20,32] Other studies indicate that MRSA is responsible for between 26 and 31% of all SSIs.[6] Despite discrepancies in these estimates, it is clear that MRSA is a prevalent cause of SSIs and is a significant cause of morbidity and mortality.

MRSA-related SSI is associated with increased morbidity, mortality, length of hospitalization, antimicrobial resistance, and healthcare cost.[1,7,16,23] The mortality rate of patients infected with MRSA is known to be at least double that of patients infected with Methicillin Sensitive *Staphylococcus aureus* (MSSA).[28,34] One study estimated 90-day mortality to be 21% for MRSA SSI patients, in comparison to 7% for MSSA SSI patients.[20] MRSA is also known to increase post-operative length of stay.[8,22] There are also significant increases in treatment cost, with estimates ranging from a 10-20% increase to almost double.[21,34] Thus, the prevention or treatment of MRSA-related SSIs has the potential to significantly reduce morbidity, mortality, length of stay and cost.

There is limited evidence on the efficacy of standard treatments (e.g., Vancomycin and Linezolid) for MRSA-related SSIs, which makes the early identification of the infection critically important.[35,36] In particular, while Vancomycin is the first-line treatment for MRSA, there is increasing evidence both of its difficulty in penetrating infections and of the development of resistance, resulting in Vancomycin-resistant *Staphylococcus aureus* (VRSA) infections.[37] There is also conflicting evidence of the

efficacy of mupirocin ointment in the prevention of both MRSA and MSSA.[38,39] Several researchers have asserted that it is critical to identify the resistance profile of the *S. aureus* pathogen prior to the prescription of preventive measures or treatment.[22,40] In the case of an inappropriately treated MRSA infection, selective genomic pressure can result in subsequent severe MSSA infection. It is therefore important to identify the specific pathogen prior to the administration of preventive measures, which may include antibiotic prophylaxis.

Prediction of MRSA SSI allows for the prophylactic use of appropriate antibiotics or application of other prevention techniques, which have been shown to improve clinical outcomes.[3,21,41,42] Antibiotic prophylaxis is an important technique in the prevention of SSIs, however it is important to select an appropriate narrow-spectrum antibiotic based on the most likely pathogen.[21] While preoperative screening for MRSA and subsequent eradication therapy is effective in reducing SSIs, PCR-based tests are often cost prohibitive and it is often not possible to wait for the results of slower, microbiological tests.[28,42] In a study conducted in the UK, approximately 81% of MRSA infected patients received ineffective antimicrobial prophylaxis due to a delay in receiving screening results.[42] To address these issues, the risk of MRSA infection could be estimated through consideration of key risk factors.[3,28,41]

There is a range of patient, procedural and hospital level factors that increase the risk of a surgical patient contracting a MRSA infection.[4,16,23,28,41-46] Risk factors such as transfer within hospital, presence of comorbidities, length of stay and time in surgery are known predictors of both MRSA-related SSI and MRSA colonization in general.[41,42,45,47] Demographic factors such as older age, black or American Indian race, and insurance

status are also predictors of MRSA-related SSI.[16,47] In addition, hospital and procedural factors, such as time in surgery, adherence to safety procedures, hospital size, and hospital teaching status, have also been shown to increase risk.[23] Many of these risk factors are available in routinely collected hospital administrative data, and as such, could be used to assess risk of MRSA-related SSI with minimal burden.

1.2 Statement of the Problem

The specific problem is that the need for, and type of, antibiotic prophylaxis for patients at risk of MRSA-related surgical site infection is difficult to determine because routine testing of patients for MRSA infection is not performed.[20,21,47] While the perioperative administration of prophylactic antibiotics has been shown to significantly reduce the incidence of MRSA-related SSI, in practice administration is often sub-optimal.[39,47] Colonization at hospital admission is not a significant predictor of MRSA and many infections are acquired during or after surgery.[39,47] There is also disagreement between researchers about the circumstances under which antibiotic prophylaxis should be administered, in particular around the cleanliness of the wound.[20,21] Selection of the appropriate antibiotic should be based on the most likely organism, and as such it is necessary to identify effective predictors to inform this selection.[20]

There is a range of patient and hospital level risk factors of MRSA-related SSI, including age, smoking status, obesity, type of surgery, use of prophylaxis, and nursing home residency, number of comorbidities, and hospitalization duration.[16,17,21,27,41] Both hospital and individual level factors, such as the presence of an active MRSA surveillance process, are also predictive of risk.[16,21] Although factors such as prior

antimicrobial use and other clinical factors also inform risk of MRSA-related SSI, the majority of significant factors can be obtained from *routinely collected* demographic, comorbidity, procedural, and hospital-level variables.[16,21,41] For example, teaching hospitals and larger hospitals have been shown to be associated with reduced incidence of MRSA-related SSIs.[16] These types of variables are routinely collected as part of national health care quality initiatives.[16,19,34]

Studies to date disagree about whether MRSA is increasing or decreasing in the US and if MRSA colonization is a predictor of MRSA-related SSI; many studies to identify risk factors have been subject to small sample sizes.[16,17,22,40,42,48] While preoperative screening in general has been shown to reduce risk of SSI, patients with a history of MRSA are still at increased risk.[22,42,49] Although risk models for MRSA-related SSI exist, such models are rarely based on systematically-collected, national level data, due to in effective and inconsistent screening/reporting.[16,42,48] In a study of MRSA surveillance systems, lack of consistent reporting has illustrated significant discrepancies in prevalence across different research settings.[48] More reliable prevalence estimates and risk factors can be generated through the use of nationally representative data.[16]

Although there has been significant research into the risk factors of MRSA-related SSIs, a predictive model that incorporates both hospital and patient level variables, based on a large, national database has not been developed.[3,16,27] Limitations in the consistency and timeliness of screening for MRSA, can result in a significant number of patients receiving inappropriate prophylaxis or other preventive measures.[42] Development of a predictive model reliable enough to inform clinical practice requires a large sample size based on nationally-representative data.[16,27] While the majority of studies support the

feasibility of developing a model to predictive model to correctly distinguish between MRSA and MSSA infections, at least one study was unable to do so.[50] However, one nationally representative study identified 13 statistically significant predictors of MRSA-related SSI, including both hospital- and patient-level characteristics.[16] These risk factors were not incorporated into a model, which could serve to predict MRSA-related SSIs.

Therefore, the purpose of this quantitative research study is to develop and evaluate a model for the prediction of MRSA related SSIs in patients that underwent major surgical procedures in the United States between 2010 and 2014.[3,41,51] This study will use several databases generated by the Health Care Utilization Project, including the National Inpatient Sample (NIS). The NIS contains a range of routinely collected patient and hospital level characteristics, including ICD-9-CM diagnosis and procedure codes, patient demographic characteristics, hospital characteristics, financial information, discharge status, and a range of severity and comorbidity measures.[16,52] Statistical modeling methods will accommodate the potentially non-linear nature of the data and the model will be evaluated within a machine learning framework. The primary modeling approach will be regression-based to ensure that the effect of identified predictors can be explicitly quantified.[53]

1.3 Objectives of the Study

The overarching objective of this study is to develop and evaluate a model to predict MRSA-related SSI based on a range of demographic, clinical, procedural, and hospital factors. The specific research objectives are:

1. To estimate the prevalence of MRSA, SSI, MRSA-related SSI, and MSSA-related SSI in patients undergoing major surgical procedures in the US;
2. To identify and assess the significance of demographic, clinical, procedural and hospital-level risk factors for MRSA infection, SSI, MRSA-related SSI, and MSSA-related SSI;
3. To determine the optimal modeling approach for the prediction of MRSA and MRSA-related SSI in patients undergoing major surgical procedures (i.e., Artificial Neural Network or Logistic Regression);
4. To develop and evaluate a predictive model to identify patients at high risk of MRSA infection and MRSA-related SSI;
5. To incorporate the predictive models into a Clinical Decision Support System.

1.4 Significance of the Study

This study aims to develop and evaluate a clinical decision support system for the prediction of MRSA and MRSA-related SSI in patients undergoing major surgical procedures in the United States. Development of this system will facilitate the early prediction of MRSA-related SSI and allow for appropriate prophylaxis or other preventive measures, and in doing so, could reduce mortality, morbidity, cost, and length of hospital stay for patients undergoing surgical site infections. There are approximately 1.35 million SSIs in the US each year, of which 30% are caused by MRSA.[16,17,20,32] The 90-day mortality for MRSA-related SSI is 21%, which accounts for 81,000 deaths annually, compared to 66,100 for MSSA-related SSIs.[16,20] Some studies estimate that up to 81% of patients receive inadequate prophylaxis due to delays in receiving test results.[42]

If all MRSA-related infections were prevented or treated successfully, this would account for reduction of 81,000 deaths annually. Similarly, prevention of all SSIs would save the US $10 billion per year, reduce the cost of individual treatment and length of stay by up to 50%.[8,21,22,24,34] Although the model will not have perfect predictive accuracy and other factors will impact outcomes, given that 60% of SSIs are preventable, significant reductions in mortality, length of stay, and cost are likely.[23]

CHAPTER 2

LITERATURE REVIEW

2.1 Literature Search Strategy

The literature search strategy and associated search terms were generated directly from the domains identified in the purpose statement defined in Chapter 1. The purpose of this dissertation is to develop a clinical decision support system for the prediction of MRSA related SSIs in patients that underwent major surgical procedures in the United States. Key domains and search terms used were:

- "Surgical Site Infection" or "SSI"
- "MRSA" or "Methicillin Resistant Staphylococcus aureus"
- ("MRSA" or "Methicillin Resistant Staphylococcus aureus") AND ("Surgical Site Infection" or "SSI")
- "Clinical Decision Support System" or "CDSS"
- ("Clinical Decision Support System" or "CDSS" AND ("MRSA" or "Methicillin Resistant Staphylococcus aureus")
- ("Clinical Decision Support System" or "CDSS" AND ("MRSA" or "Methicillin Resistant Staphylococcus aureus" AND "Surgical Site Infection" or "SSI")

All searches were performed using the Rutgers Library Web Portal and were filtered to include peer reviewed journal articles, refereed conference proceedings, text books, doctoral dissertations, and official reports. The Rutgers Library has access to a range of resources from a wide range of publishers, including Elsevier, EBSCO, PubMed and Medline. The most relevant abstracts were reviewed and, if applicable to the dissertation,

were downloaded into an EndNote library. Papers were grouped into four categories, representing the four domains contained in the purpose statement: Surgical Site Infections (SSI), Methicillin Resistant Staphylococcus aureus (MRSA), MRSA-related SSI, and Clinical Decision Support Systems (CDSS). Final totals for each category were:

- SSI – 46 papers, book chapters, conference proceedings or dissertations
- MRSA – 50 papers, book chapters, conference proceedings or dissertations
- MRSA-Related SSI – 31 papers, book chapters, conference proceedings or dissertations
- CDSS – 46 papers, book chapters, conference proceedings or dissertations

Additional searches were performed during the writing of the literature review as any gaps were defined and in cases where additional evidence or definitions were needed. Each of the domains above forms a section of this chapter, with the goal of identifying the complementary and overlapping research gaps in each area and thus justifying the conduct of this study.

2.2 Surgical Site Infections

2.2.1 Overview of SSI

According to the Centers for Disease Control and Prevention (CDC), a surgical site infection (SSI) is defined as an infection that occurs as a result of a surgical incision within 30 days of an operative procedure.[54-57] SSIs are often severe infections and among the most commonly reported hospital acquired infection (HAI).[58,59] Many SSIs do not present until after the patient has been discharged from the hospital, which suggests a need for careful post-surgical surveillance.[60] Diagnosis of SSIs occurs through culturing

of wound drainage and consideration of supporting signs and symptoms.[56] SSIs can result in significantly increased morbidity and mortality.[61]

Although significant progress has been made in the prevention and treatment of SSIs, incidence continues to be high and SSIs remain a leading cause of nosocomial-related morbidity and mortality.[62,63] While advances in the use of antiseptics and antibiotics have revolutionized the prevention and treatment of SSIs, an increase in the frequency and invasiveness of surgery along with increasing general morbidity have resulted in increased incidence over time.[58] Increases in the proportion and number of surgeries performed in an ambulatory setting combined with shorter hospital stays have also contributed to increasing incidence.[64,65] It is estimated that approximately 1.4 million SSIs occur in the US each year, which accounts for between 2% and 5% of all surgeries and results in costs of over 1 billion dollars annually.[63] By some estimates, SSIs are the most common type of hospital acquired infection.[66]

SSIs are the most common hospital acquired infections among surgical patients.[61,67-69] Estimates of incidence vary based on the type of surgery but can be as high as 30% for invasive procedures such as lower gastrointestinal and colorectal surgeries.[70,71] SSIs are routinely identified as the most common post-operative infection, with some estimates as high as 38%.[67] Overall, SSIs are one of the most common nosocomial infections and although some researchers suggest they are the most prevalent HAI, estimates vary significantly.[61,66,69] There is consensus that SSIs result in increased morbidity, increased cost, longer hospital stays, and a higher probability of readmission.[68]

SSIs have routinely been shown to have a significant effect on morbidity and mortality.[65,72-74] In addition to prolonged hospital stays and increased therapeutic costs,

SSIs result in significantly higher health care costs.[60,62,73] SSIs are also associated with increased mortality and higher hospital readmission rates.[54,56,58,64] Researchers estimate that surgical site infections cost the health care system up to $10 billion per year, increase time in hospital by between 7 and 11 days, and increase the risk of mortality by between 2 and 11 times.[54,66] Despite improvements in prevention and treatment protocols, the incidence of SSIs and HAIs in general remains high.[43,72]All surgical wounds are subject to some degree of contamination, which can significantly increase a patient's risk of infection, depending on the type and location of the wound. [23,75]

Surgical wounds and SSIs are classified in various ways to facilitate prevention, treatment and analysis.[60,65,70,72,76,77] Surgical wounds are assigned to one of four categories based on the degree of contamination.[23,56,78] Standard surgical wound classification (SWC) categories are clean, clean-contaminated, contaminated, and dirty, with each category representing an increasing risk of SSI.[17,23,65,79] In addition, SSIs are post-operatively classified into different types based on the nature and location of the infection.[65,70] The CDC divides SSIs into three levels based on the anatomic level of the infection: superficial incisional, deep incisional, and organ space.[54,60,70,76,77] Superficial incisional SSIs generally occur in skin and subcutaneous tissue, whereas deep incisional SSIs occur in the fascial and muscular layers of soft tissue.[60,71,72] Organ space SSIs are the most invasive and serious infections and can have the most significant impact on outcomes.[80] While SSIs are typically defined as infection that occur within 30 days of surgery, some researchers also classify them as early (within 3 months of surgery), delayed (between 3 months and 2 years from surgery) and late (greater than 2 years from

surgery).[54-57,81] Regardless of the classification of the infection, most SSIs are caused by a small number of bacterial pathogens, most notably *Staphylococcus aureus*.[55]

Staphylococcus aureus is a virulent and common pathogen and is currently the most common cause of SSIs.[6,62,63,82] Treatment of SSIs is further complicated by the development of antimicrobial resistance to the Staphylococcus aureus bacterium, and in particular by the substantially increasing prevalence of Methicillin-resistant *Staphylococcus aureus* (MRSA).[62,63] MRSA bacteria have acquired genetic mutations that encode resistance to Methicillin and other penicillin and Beta-lactam antibiotics.[75] The increasing prevalence of antibiotic resistance contributes to the morbidity and mortality of SSIs and complicates treatment regimens.[55,75] By some estimates, MRSA now represents up to 63% of all *Staphylococcus aureus* infections, and thus treatment or prevention of SSIs requires knowledge about the resistance profile of the potential pathogen.[55,62] Specifically, it is important to identify risk factors that facilitate the differentiation of MRSA and Methicillin-susceptible *Staphylococcus aureus* (MSSA) SSIs so that appropriate preventive therapy may be administered.[3,21,41,42]

The identification of risk factors for SSI and the development of risk models that can accurately predict MRSA and MSSA SSI in surgical patients is critical in determining the appropriate prevention and treatment strategies.[58,74,83] The probability of a patient experiencing an SSI is affected by a range of demographic, clinical, procedural, and hospital-related factors, and a number of risk scores have been developed using these factors.[23,83,84] Under the auspices of the National Nosocomial Infections Surveillance (NNIS) System, the CDC developed an SSI risk index, which incorporates the American Society of Anesthesiologist (ASA) Score, wound classification (clean, clean-

contaminated, contaminated, dirty) and procedure duration.[85] Other researchers have developed methodology to screen patients for risk of SSI at admission.[58,86] While predictive models can be developed using encounter data collected for the purposes of quality assessment, significant, subtle variations between MRSA and MSSA require large sample sizes and careful classification of predictors.[64,73,86,87] Nevertheless, differentiating between MRSA and MSSA and determining risk of infection is critical in the prevention of SSIs.[57]

Prevention of SSIs is a multifaceted process that requires consideration of a range of patient and hospital factors and the administration of multiple preventive measures, including use of antiseptics, preoperative antibiotic therapy, and clean surgical processes.[71,88-90] Antibiotic prophylaxis is a key component of SSI prevention, however inconsistent implementation of screening protocols and lack of universal screening can hinder its success.[23,65,66,75,87] There is a significant body of evidence to support antibiotic prophylaxis, however almost universally, researchers indicate the need for the targeting of the antibiotic to the specific pathogen.[57,67,68] Specifically, the routine prophylactic use of vancomycin for the prevention of *Staphylococcus aureus* SSIs should be avoided unless the resistance profile of the pathogen is known.[61,67]

2.2.2 Epidemiology of SSI in the United States

SSIs are a significant cause of morbidity and mortality and place substantial burden on the healthcare system in the United States.[59,77,84,91] Annually, of the 30 million operative procedures and 1.7 million HAIs that occur in the United States, over 500,000 are SSIs.[57,77,92,93] These infections affect approximately 157,000 patients per year, cost $20,000 per infection, and cost the healthcare system a total of $3.3 million annually.[59,84]

SSIs represent approximately 22% of all HAIs, result in over 8000 deaths annually, and have been shown to be increasing as a proportion of HAIs.[89,92] Thus, SSIs occur in a significant proportion of surgical patients.

SSIs occur in a significant proportion of all surgical patients, with rates of infection varying by type of surgery.[23,66,74,82,84,85] Researchers generally agree that between 2% and 5% of all surgical patients experience an SSI, however some estimates are as high as 38%.[23,63,81,82,85,92] While estimates vary regarding the overall proportion of HAIs that are SSIs, most researchers asserting that SSIs are the most common nosocomial infection, representing between 30% and 40% of all HAIs.[23,55,60,64,74,76,92] Rates vary by surgical procedure with SSIs occurring in 2% to 5% of patients undergoing hip or knee arthroplasty, 15% of patients undergoing spinal surgery, 30% of patients undergoing colorectal surgery, 33% of patients undergoing abdominal surgery, and up to 45% in patients undergoing head or neck surgery.[56,77,81,86,88,91] The majority of SSIs – up to 79% – occur within 30 days of surgery, and because a substantial proportion of these occur after discharge, identifying key risk factors and taking preventative action is critically important.[81]

Key risk factors for SSI include age, type of surgery, sex, SWC, use of prophylactic antibiotic therapy (PAT) and NNIS risk index.[55,59,78,88,94] Incidence of SSI has been shown to be higher in females than males (78% vs 77.3%), and also higher in the 15-30 years age group.[55] In contrast, other researchers have found that incidence of SSI is higher in patients aged 70 or over.[59] PAT was generally found to be an effective prevention strategy.[87,94] For example, in a study of over 5000 patients undergoing surgery for lumbar disc problems, the risk of SSI for patients administered PAT was one third

that of those not administered PAT, with a number need to treat (NNT) of 43.[94] Type of surgery is not only a risk factor for SSI, but also has a significant impact on SSI sequelae.[84]

Patients with SSI experience a range of complications, both before and after diagnosis, which result in longer hospital stays, reduced quality of life, higher mortality, and abnormal wound symptoms.[54,70,84,90] Mortality rates and costs differ by type of surgery: neurological SSIs are generally the most expensive, costing around $23,755 per SSI and SSIs in spinal surgery associated with significantly increased mortality.[69] Outcomes are also affected by the pathogen causing the SSI, the most prominent of which is *Staphylococcus aureus*, which is causative in up to 57% of SSIs.[90]

Although *Staphylococcus aureus* is a major cause of SSI, estimates vary by type of surgery, with different studies reporting widely different results.[55,60,63,93] While some researchers estimate that as few as 20% of SSIs are caused by *Staphylococcus aureus*, most estimate that this pathogen is present in over 50% of SSIs.[55,60] Antibiotic resistance is a major issue with *Staphylococcus aureus* infections that complicates treatment and negatively impacts morbidity and mortality, due to the development of further resistance to limited treatment options, such as Vancomycin.[61] Reducing the incidence of SSIs and slowing the development of antibiotic resistance requires an epidemiological study based on a large-scale nationally-representative sample of major surgical procedures in the United States.[52]

For the purposes of epidemiological data analysis of SSIs, the National Inpatient Sample (NIS) provides a large, nationally-representative sample of hospital admissions,

collected over the course of several years.[16,52,95] The NIS collects data from over 1000 hospitals in the United states and contains data on over 8 million hospital discharges.[95] The NIS defines a major surgical procedure based on ICD-9 codes and includes both diagnostic and therapeutic procedures that occur in an operating room.[16,52] Analysis of the NIS will facilitate the identification and validation of predictors of SSI along with the generation of epidemiological measures such as estimates of incidence and prevalence.[16]

2.2.3 Modeling and Prediction of SSI

A range of statistical and machine learning models have been applied successfully to the prediction of SSI from patient demographic factors, clinical factors, microbial factors, and hospital factors.[72,84,89,96] Hierarchical multivariate logistic regression can be used to differentiate SSI patients from non-SSI patients by incorporating variables relating to the surgical procedure, patient factors, and hospital factors, with comorbidities and surgical procedures identified by ICD-9 codes.[77,80,96] Researchers determined that simpler models with fewer covariates were effective predictors of SSI, one group obtaining sensitivity of 72% and specificity of 64%.[43,84,97] Key predictors identified include RACHS-1 Score (for cardiac surgery), ASA score greater than or equal to 3, concurrent infection, wound class, blood transfusion, BMI, MRSA colonization, duration of surgery longer than 3 hours, and most notably preoperative clinical severity and comorbidities.[43,89,97] One study demonstrated a 9-fold increase in risk based on nasal colonization alone.[89] It is likely that predictors of SSI differ due to both the underlying population and the existing risk profile of that population (i.e., the presence of comorbidities).[80]

In addition to patient demographics, such as socioeconomic status, age, and BMI/obesity status, the presence and number of comorbidities is a significant predictor of SSI.[67,68,80,92] Comorbidities shown to increase risk include, diabetes, chronic obstructive pulmonary disease (COPD), obesity, pre-operative hyperglycemia, anemia, inflammatory bowel disease.[67,68,80,85,92] The American Association of Anesthesiologists (ASA) Score is another measure of the severity of illness of surgical patients.[98]

2.2.4 Predictors of Surgical Site Infection

General Risk Factors. Researchers have identified a broad set of risk factors that are useful in the prediction of SSI, however many studies are subject to small sample sizes and non-representative samples and thus are not powered to detect statistically significant differences.[71,77,88,90] Common preoperative clinical and demographic factors include age, surgery time, surgical duration, glucose level, length of stay, smoking status, and comorbid diabetes.[77,89] Procedural factors, such as operating room condition, surgical hand preparation, antibiotic prophylaxis, and screening for *Staphylococcus aureus*, while amenable to intervention may be problematic from a modeling perspective due to their lack of inclusion in standardized data collection protocols.[88,95] Other known risk factors include the skill of the surgeon, antibiotic administration, contamination status of the wound, temperature, results of microbiological culturing (e.g., results of biological testing for *Staphylococcus aureus*), and administration of blood transfusion.[68,77,86,93,96] In addition to general risk factors, there are a range of specific risk factors that are related to specific surgical procedures.[43] For example, risk factors in gastrointestinal surgery also include corticosteroid use and malnutrion.[71] Risk factors common to most surgery types are identified from the literature and are described in more detail below.[2]

Age. Researchers generally agree that extreme of age is a significant predictor of SSI, with the very young and very old experiencing higher risk.[55,59]

Sex. There is no clear consensus on whether a patient's sex increases risk of SSI.[55,59,83] While a number of researchers have identified female sex as increasing risk, further research needs to be performed to explore this relationship in more detail.[55,59]

Weight. SSI has been shown to be significantly associated with increasing weight.[69,70,73] Patients with higher than normal BMI has been shown to have significantly higher rates of superficial and deep/organ-space SSI.[70] Incidence of SSI has been shown to be as high as 30% for severely obese patients.[73] One recent study of patients undergoing cervical spine surgery identified BMI of greater than 35 kg/m^2 as an independent risk factor for SSI.[69]

ASA Score. The ASA score is a measure of a patient's health prior to a surgical operation, and as such, broadly captures the impact of comorbidities and the patients general health and is a useful predictor of SSI.[74,98] A number of researchers include the ASA score in the development of predictive models and risk indices.[74,80,97] The ASA score is often combined with other covariates for the purpose of risk adjustment, the most common of which is SWC.[74,80,97] For example, the National Nosocomial Infections Surveillance System (NNIS) risk index includes SWC, ASA score, and surgical duration.[97]

SWC. Surgical Wound Classification, in conjunction with ASA score and surgical duration, is used to identify patients at risk of SSI.[69,76,78] SWC describes the overall degree of wound contamination and assigns a wound to one of four categories:

clean, clean-contaminated, contaminated, and dirty.[74] For example, In colorectal surgery, rates of SSI in clean-contaminated wounds can exceed 20% and may be affected by the use of antibiotic prophylaxis.[76,87]

***Staphylococcus aureus* Colonization.** Colonization with *Staphylococcus aureus* or Methicillin Resistant *Staphylococcus aureus* (MRSA) has been shown to increase risk of SSI.[62,81,93] Nasal carriage of *Staphylococcus aureus* has been reported as a major risk factor for the development of SSI, and consistent with that finding, treatment with intranasal mupirocin and a chlorohexidine scrub have been shown to be protective factors.[75,90,93] MRSA has been shown to be the leading cause of SSI in several different types of surgery, including vascular, orthopedic, and cardiac procedures.[62] Predicting the causative agent of an SSI and understanding its antibiotic resistance profile is critical in applying appropriate preventive measures.[61,67]

One way of understanding the relative importance of each predictor is to consider the number of citations in the literature. Table 1 shows the number of citations for each of the top 20 predictors.

Table 1: Top 20 SSI Predictors

Predictor	Citation Count
BSI	11
ASA Score	8
MRSA - Colonization	7
Surgical Hand Preparation	6
Colon Surgery	5
Diabetes	5
Sex	4
Smoking	4
Surgical Wound Class	4
Blood Transfusion	3
Nutritional Status	3
Steroid Use	3
Wound Class	3
Anemia	2
COPD	2
Infection	2
Surgery Type	2
Absence of Splenic Function	1
Alcoholism	1
Antibiotic Administration/Use (Risk Factor)	1

2.3 Methicillin Resistant Staphylococcus Aureus

2.3.1 Background

Hospital acquired infections (HAI), also known as nosocomial infections, are responsible for significant morbidity and mortality, prolonged hospitalization, and increased healthcare expenditures.[5,6,15] Hospitalized patients, particularly those undergoing invasive procedures, are at a significantly higher risk of acquiring an infection, the most common of which are urinary tract infections (UTI), SSI, and bloodstream infections (BSI).[15] Nosocomial infections are defined as infections for which there was no evidence at the time of admission that are acquired within 48 hours of admission hospital, or appear within 30 days if discharge, and manifest themselves as a clinical disease.[5,15,99] The National Healthcare Safety Network (NHSN) classifies nosocomial infection based on 50 different body location and according to 13 types.[14]

Many nosocomial infections are caused by microorganisms (e.g., bacteria) that have developed resistance to common antimicrobial treatments.[6]

The development of antimicrobial resistance to a range of bacterial pathogens, including *Staphylococcus aureus*, is an increasing public health issue.[100-103] Use of antibiotics for the treatment of bacterial infections, and particularly the use of broad spectrum antibiotics, has led to increasing selective pressure in which susceptible pathogens are killed and resistant strains become more prevalent and thus, over all the pathogen develops further resistance.[5,47,101,102,104,105] Antimicrobial resistance has been recognized as an inevitable consequence of antibiotic use as far back as 1945, when Alexander Fleming observed that inappropriate use of penicillin could lead to the development of mutant forms of the *Staphylococcus aureus* pathogen.[29,32,105] Within 10 years of the introduction of Penicillin, virtually all strains of *Staphylococcus aureus* exhibited resistance due to the spread of the Beta-lactam gene through the species.[14,29,106] Resistance to Beta-lactam antibiotics led, in the 1950s, to the introduction of Methicillin – a synthetic derivative of penicillin, however resistance quickly developed due to the acquisition of *mecA* and *mecC* genes, and was prevalent by the 1980s.[107] By some estimates, global consumption of antibiotics has increased by up to 65% between 2000 and 2015, which has accelerated the development of Methicillin Resistant *Staphylococcus aureus* (MRSA), particularly in the hospital setting and in immunocompromised patients.[6,31]

MRSA is a bacterial pathogen that is resistant to Beta-lactam antibiotics, including methicillin, oxacillin, and cefoxitin and is a leading cause of HAIs in the United States and Europe.[30-32,34,103,108] MRSA typically resides on the skin and in nasal

passages and can cause infections if it enters the body through a cut or wound, and so surgical patients are at particularly high risk.[8,34,109] The first MRSA infections were reported in the UK as early as 1961, MRSA emerged as a major infection by the 1980s, and it became endemic in US and European hospitals in the last 20 years.[2,8,29,30,32,35,106,108] In contrast to Methicillin Susceptible *Staphylococcus aureus* (MSSA), MRSA infection is associated with greater severity of illness, poorer clinical outcomes and higher economic burden, and by some estimates is the leading cause of HAI-related mortality in the United States.[31,109,110] The development of newer, community-based strains of MRSA, combined with increasing resistance to Vancomycin – the primary treatment for MRSA – has led to a critical need for more effective screening and prediction of MRSA infections so that appropriate prophylaxis and other preventive measures can be applied to minimize further resistance.[32,50,111]

There is significant debate about the cost-effectiveness and efficacy of testing for MRSA at hospital admission.[35,39,41,104,112] While there are no official recommendations for the most effective screening/testing strategies, most researchers agree that screening/testing protocols should target patients at the highest risk of infection.[8,39,41,109,113] In addition, lack of cost-effective rapid tests that exhibit high sensitivity and specificity make universal screening infeasible and result in physicians administering antibiotics based on inconclusive diagnoses.[46] Therefore, in order to reduce the number of unnecessary screening tests, many institutions have developed clinical prediction tools to identify patients at high risk of infection.[112,114,115] These methods are important because, by some measures, lack of MRSA screening can result in approximately half of all carriers being missed.[114] Developing more effective methods for

predicting MRSA infection has the potential to reduce incidence, facilitate the more effective use of screening tests, and improve outcomes.[41,113]

Because of the need to target high-risk patients for more in-depth screening and testing protocols and to improve the timeliness of diagnosis, significant research has been conducted to develop predictive models for both MRSA colonization and subsequent risk of infection.[3,27,50,113,114,116,117] Researchers have developed predictive algorithms for MRSA carriage at hospital admission and to differentiate between resistant and non-resistant infections.[40,45-47,114] Data used to develop these models, and any indicated preventive measures, should be based on local epidemiological information that reflects the population of interest.[3,48,101,112] One successful example is a study of MRSA risk in pneumonia patients, in which patients were stratified into 3 risk groups based on a small number of clinical factors (age, prior IV antibiotics, cardiovascular disease, sex, diabetes, and nursing home or long term care residency).[118] New machine learning techniques, the increasing availability of data, such as the MIMIC-III dataset, and standardized (ICD-9) coding of MRSA in healthcare data, have supported the development of effective clinical decision support systems (CDSS), which have been shown to reduce the reliance on laboratory testing.[31,100,108,118,119] Machine learning techniques, because of their ability to model non-linear and other complex relationships, have the potential to differentiate between MRSA and MSSA, support the implementation of CDSS, and the rapid administration of appropriate therapy.[3,40,108,115] Advanced modeling techniques, such as logistic regression, support vector machines, and random forest classifiers have been shown to be effective predictors of MRSA infection and outcomes.[40,108] Effective

prediction methods facilitate the administration of appropriate prevention and treatment regimens, including the administration of prophylactic antibiotics.[5]

Increasing resistance to standard treatments for both MSSA and MRSA (e.g., Vancomycin) due to overuse of broad spectrum antibiotics require that the pathogen's resistance profile be determined prior to administration of prophylaxis.[30,101,109,110,120] Current prophylaxis protocols do not account for the resistance profile of the probably pathogen, is necessarily broad spectrum, and as a consequence does not explicitly target MRSA.[82,115] However, the development of new antibiotics with improved activity against MRSA (e.g., daptomycin, linezolid, etc.) combined with more effective screening procedures, could facilitate a more targeted approach.[1,48,107] The use of appropriate antibiotic prophylaxis is particularly important in the prevention of SSI.[1,5,116]

2.3.2 MRSA Epidemiology

MRSA infections have been shown to increase the duration of hospitalization, to significantly increase the risk of mortality, and to lead to long term disability.[1,26,31,46] Patients with MRSA infections are more likely to fail treatment within 1 week and stay an average of 6 times as long as patients admitted to an intensive care unit.[26,50] MRSA is the leading cause of death attributed to antimicrobial resistance in the United States, with some studies estimating a 33% mortality rate for MRSA patients in comparison to 22% for non-MRSA patients.[31,46] Patients infected with resistant organisms, such as MRSA, require substantially greater hospital care, which results in increased costs of up to $30,000 per infection due to increased medical complications.[34,121] The incidence and prevalence of MRSA are likely underestimated due to the lack of universal testing and high probability of colonization in the general population.[113]

MRSA is one of the most common nosocomial infections in the United States and Europe and accounts for between 40 and 50% of all *Staphylococcus aureus* infections.[1,27,103,122] It is a Gram-positive bacterium that resides on the skin of proportion of the population, these asymptomatic carriers acting as vectors, with estimates of prevalence ranging from 1% to 20% depending on region (infection rates are higher in the southern US) and other factors.[26,31,41]A significant proportion of MRSA infections are categorized as skin and soft tissue infections, and of these infections, approximately 50% are the result of community-acquired MRSA.[50,117] Community-acquired MRSA (CA-MRSA) is an increasing public health problem and in many settings incidence exceeds that of hospital-acquired MRSA. There is also evidence that CA-MRSA is a leading cause of hospitalizations in the United States, due primarily to the high incidence of skin and soft tissue infections.[31,111,117,123]

Although the number of MRSA-related hospitalizations has increased due to the presence of CA-MRSA, other types of MRSA infections, such as central line-associated blood stream infections have decreased.[123] In general, researchers disagree regarding the prevalence of MRSA infections and MRSA related hospitalizations, suggesting the presence of confounding in the studies or underlying differences in the population being studied. For example, whereas a Veterans' Administration study reported an 80% reduction in MRSA infections, over the same period, the NHSN reported no change.[31,48] It may also be the case that decreases in hospitalization rates for skin and soft tissue infections were offset by increases in invasive infections, such as sepsis and SSI. Thus, overall rates of MRSA remain high and the presence of multiple strains present significant challenges for prevention and treatment due to increasing resistance to

Vancomycin and other antibiotics and increasing reliance on broad spectrum antibiotics.[103,107,122,124] Despite these challenges, use of a Clinical Decision Support System (CDSS) based on the identification of discriminatory predictors has been shown to improve appropriate antibiotic prescription.[115]

2.3.3 Predictors of MRSA

General Risk Factors. Researchers have identified a broad range of clinical predictors of MRSA colonization, MRSA infection, and MRSA-associated outcomes.[99,113,114,121] Demographic predictors include sex, age, and prior contact with the healthcare system, in particular prior hospitalization.[113,114,117] Multiple studies have determined that vital signs and clinical measurements, including temperature, hematocrit value, white blood count, and c-reactive protein level are predictive of MRSA infection.[40,110] Other predictors include prior antibiotic use, previous hospitalization, infection acquired during hospitalization (as opposed to in the community or other health care setting) and underlying severity of illness.[26,113,117,121] No differences have been observed between males and females.[117]

Antibiotic Exposure. Exposure to broad spectrum antibiotics prior to hospital admission is a significant predictor of MRSA infection.[2,40,41,47,102,109,120] In one analysis, researchers found that 45% of patients with MRSA had documented antibiotic use at admission, versus 17% of patients with MSSA.[40] Consistent with this result, other researchers found that patients with recent antibiotic exposure, when controlling for other factors (prior hospitalization, age>70) were over 4 times more likely to have MRSA carriage at admission than patients without antibiotic exposure.[47] Other researchers obtained similar results in a range of multivariate analyses, suggesting that overuse and

misadministration of broad spectrum antibiotics is a major contributor to antimicrobial resistance generally, and to MRSA specifically.[41,46,100,109]

Prior Hospitalization. Prior hospitalization is a known risk factor for both MRSA colonization and MRSA infection.[8,112,117,120] In a univariate analysis, hospitalization of more than one week in the previous 6 months was identified as a significant predictor of MRSA infection, and another study found that recent hospitalization doubled the risk of infection.[41,117] Researchers also identified intra-hospital transfer and hospitalized patients with compromised immune systems as major contributors to risk.[14,112]

Age. While estimates vary across studies, older age has been shown to be correlated with MRSA colonization and infection.[11,26,99] Patients over the age of 75 are almost twice as likely as younger patients to be colonized with MRSA.[47] Other studies find similar results with slightly different age cut-offs, including 70 years and 79 years.[41,118] While much of the research is focused on older patients, very young patients are also at higher risk, and because of the prevalence of MRSA in older patients, colonization and infection in younger patients is likely to be missed.[27,112]

Indwelling devices. Indwelling devices, such as percutaneous endoscopic gastrostomy (PEG) tubes, are colonized with a range of organisms, including *Staphylococcus aureus*, and are associated with higher rates of MRSA infection.[2,104,122] By some estimates, over 90% of external feeding tubes are colonized with harmful bacteria.[104] A number of studies have identified mechanical ventilation (MV) tubes and central venous catheters (CVC) as strong risk factors for MRSA infection.[2]

Clinical predictors. A number of studies have identified a broad range of clinical predictors of MRSA infection, including fever, loss of function, multifocality, total white blood cell count, neutrophil count, c-reactive protein level, and temperature.[27,110,123] However, results differ by study and the risk factors considered with some researchers unable to identify clinical factors that could effectively distinguish between MRSA and MSSA.[41,50] In contrast, other studies were able to develop risk prediction models with strong performance characteristics.[27,114,118,121] In a study of MRSA prediction in bone and join infections, a c-reactive protein value of greater than 13.9 mg/L was found to predict MRSA infection with 93% sensitivity and 79% specificity, with an area under the curve of 89%.[27] Discrepancies between these results suggest a need for further research into demographic and clinical predictors that can effectively differentiate between MRSA and MSSA infections. Table 2 summarizes the 20 most cited MRSA predictors.

Table 2: Top 20 MRSA Predictors

Predictor	Citation Count
Hospitalization - Previous or Transfer	15
Antibiotic Administration/Use (Risk Factor)	12
Age	11
Decubitus Ulcers	6
Length of Hospitalization	5
Sex	5
Cerebrovascular Disease	4
Colon Surgery	4
Nursing Home	4
Temperature	4
Antibiotic - Broad Spectrum	3
Hematocrit	3
Urinary Catheter	3
WBC Count	3
Chronic Skin Disease	2
Diabetes	2
ICU	2
Indwelling Devices	2
Loss of Function	2
MRSA - Colonization	2

2.4 MRSA-related Surgical Site Infections

2.4.1 Overview of MRSA-related SSI

Despite significant medical advances in the last 20 years, including improvements in surgical techniques, infection control practices, and extensive use of antibiotic prophylaxis prior to surgery, surgical site infections (SSI) remain a major cause of morbidity, mortality, longer hospital stays, and higher healthcare costs.[7,36,63,99,125] SSIs account for up to 26% of all healthcare-associated infections (HAI) and can increase costs by more than $61,000 per infection.[21,126] Increasing antibiotics resistance due to the overuse of broad spectrum antibiotics has led to significant challenges in the administration of appropriate antibiotic prophylaxis, with a significant proportion of *Staphylococcus aureus* infections becoming resistant to common antibiotics.[7,127,128] As a result, Methicillin-Resistant *Staphylococcus aureus* has become the predominant pathogen in SSI.[63,128]

MRSA is now the leading pathogenic cause of SSI in the United States.[24,36,63,128] *Staphylococcus aureus* is the single most common bacterial cause of SSI, with MRSA accounting for over 50% of these cases, and 20% of SSIs overall.[36,128,129] In contrast, MSSA *Staphylococcus aureus* infections, MRSA-related infections also result in increased hospital costs, a higher number of 90-day readmissions, and poorer outcomes.[128,129] Lack of information about MRSA-related SSI predictors and risk factors, makes treatment decisions difficult and suggests a need for further research.[39,47,130] Early identification of MRSA or accurate prediction of MRSA-related SSI enables the administration of appropriate prophylactic antibiotics (i.e., vancomycin for MRSA infection or Beta-lactams for MSSA).[131] Thus, development of a clinical prediction tool

to accurately predict MRSA-related SSI could significantly reduce morbidity, mortality, duration of stay, and healthcare costs.[127]

Antimicrobial prophylaxis is the administration of a course of antibiotics shortly before surgery with the aim of boosting the host defenses' ability to fight any infection resulting from surgery.[17,36,63] Identifying the infective agent is critical, and failure to do so can have a significant effect on the efficacy of the prophylaxis.[63,129,132] Many healthcare systems are reluctant to administer Vancomycin prophylactically, even though it is the first line treatment for MRSA infections, because there is increasing evidence of resistance for Vancomycin due to inappropriate administration to patients with MSSA, administration of Vancomycin alone increases risk of MSSA-related infection, and Vancomycin is ineffective against MSSA.[24,36,49,129] Unfortunately, because of these challenges many patients that are later determined to have MRSA-related infections, do not receive appropriate prophylaxis.[39,42,133] Although there is significant debate about whether universal testing for MRSA should be performed, researchers agree that identifying the likeliest pathogen prior to inform appropriate prophylaxis is of critical importance.[20,22,49,63,132,133]

2.4.2 Epidemiology of MRSA-related SSI

MRSA-related SSI is a significant cause of morbidity and mortality in the United States.[17,125,134] SSI increases morality risk by between 2 and 11 times, increases hospitalization time by 1 week per infection, and increase healthcare costs between $12,000 an $35,000 per infection.[16,125,128,135,136] While estimates vary, the incidence of SSI is increasing and SSIs are generally considered as one of the top three nosocomial infections, accounting for 20% of all HAIs in the United States.[17,24] Each year

approximately 234 million major surgical procedures are performed and of these, between 2 and 5% result in SSI.[16,24,125,134] Limited epidemiological research has been performed with respect to MRSA-related SSI.[16] Approximately 1% of all major surgical procures performed in the United States each year result in a MRSA-related SSI.[16,22] Furthermore, MRSA infection is an independent risk factor that significantly increases the risk of mortality compared with non-MRSA infections, with some estimates suggesting 12 times increase in risk compared to control patients.[82,128,137] Increased mortality is particularly common in patients undergoing invasive surgeries, such as coronary artery bypass surgery and cardiac valve surgery.[35]

MRSA and MSSA are the most common pathogens involved in SSI due to high levels of nasal colonization within the US population.[20,125,126,128] Over 50% of *Staphylococcus aureus* infections are due to MRSA, which now represents the most common cause of SSI.[28,36,138] Patients with MRSA-related SSI are 30 times more likely than non-infected patients to be readmitted to hospital, 7 times more likely to die within 3 months of surgery, and on average, spend an extra 16 days in hospital.[135] Compared to MSSA-related SSI patients, MRSA-related SSI patients were almost 3 times more likely to die within 3 months of surgery , and on average spend 3 more days in hospital.[22,79,135] Incidence of MRSA-related SSI differs by facility and surgery type, but can be up to 33% for invasive surgeries.[35] Estimates of the proportion of SSIs attributed to MRSA differ and range from 15% to 35%.[35,133,139] It is also possible that MRSA-related SSIs remain under reported by a significant proportion, by some estimates up to 60%, are suspected to present after discharge.[20,21] Despite this limitation, researchers have demonstrated an association between MRSA, SSI and poor outcomes, however further research is

necessary to better understand the epidemiology of MRSA-related SSI and to identify clinical predictors that can be used to inform prevention and treatment regimens.[16,35,42,47]

2.4.3 MRSI-related SSI Predictors

MRSA-related SSI predictors can be categorized broadly into patient-related factors, hospital/healthcare-related factors, and surgical factors.[7,17,134] Patient-related factors include standard demographic characteristics, such as age, race, and sex, and also factors relating to previous hospitalization, severity of illness, and the presence of comorbidities.[17,137,138] Hospital/healthcare related factors include the use of invasive devices, infection prevention methods, and antibiotic use policy.[21,128,138] Surgical factors include type of surgery, duration of surgery, surgical wound classification, and blood loss/transfusion.[22,47,136] Several studies have attempted to differentiate between MRSA-related and MSSA-related SSI using multivariate logistic regression models, although the performance of the models varies widely.[125,127,129,132,133] Patient-related factors, which can be broken down into immutable demographic characteristics, and modifiable risk factors are generally easy to measure, and thus are useful covariates in predictive models.[17]

Age. As with SSI and MRSA in general, older age is found to be a predictor of MRSA-related SSI in comparison to MSSA-related SSI.[17,20,21,24,79,127,130-132,138] While the specific age cutoff varies between studies, most researchers suggest that patients over the age of 65 are more likely to experience a MRSA-related SSI in contrast to a MSSA-related SSI, however it must be noted that some researchers do not identify age as a predictor or explicitly indicate a lack of association.[7,125,136] It may be that age is confounded by other variables associated with MRSA-related SSI, such as an increased number of comorbidities or increased severity of illness.[24] Other studies found that risk

was elevated in the very young and also those in middle age, which suggests that further research is needed to fully understand the relationship between age and MRSA-related SSI, and its utility in differentiating between MRSA-related SSI and MSSA-related SSI.[17,20]

Sex. There is limited evidence to support an association between sex and MRSA-related SSI, however some researchers have found that males are between 35% and 70% more likely to experience a MRSA-related SSI compared with MSSA-related SSI.[20,130]

Race. While a literature search did not yield and studies that indicated race as is a predictor of MRSA-related SSI, one study did identify African American race as a predictor of MRSA-colonization.[127] In this respect, further study of race as a possible differentiator between MRSA-related SSI and MSSA-related SSI is necessary, using a large, nationally-representative dataset, which has sufficient statistical power to uncover a possibly small effect.[16,27]

Smoking Status. Two studies identified smoking as a risk factor for MRSA-related SSI.[17,136] One additional study identified Chronic Obstructive Pulmonary Disease (COPD) as a significant predictor of MRSA-related SSI, suggesting that, as with other predictors that may have been evaluated with small sample sizes, further research into smoking status using a larger, nationally-representative sample, may yield valid associations.[21]

MRSA Colonization. Colonization with MRSA upon admission to hospital has been shown to be a significant predictor of MRSA-related SSI.[28,42,49,126,132] Some researchers identify MRSA colonization as the most important predictor of MRSA-

related SSI, and suggest the need for screening of high-risk patients, however studies disagree on whether the administration of pre-operative decolonization is an effective preventive technique.[22,28,42,49,126] In one study, almost 7% of patients developed an SSI, despite the application of preoperative eradication therapy.[49] And while decolonization has been shown to reduce the incidence of SSI in orthopedic surgery, risk of SSI still remains high in comparison to patients with no history of MRSA colonization.[42] Other studies support this result in that they did not find significant correlation between MRSA carriage at admission and subsequent SSI.[39,47,114] In practice, screening for MRSA in a timely manner is difficult because surgical patients are received from multiple preoperative areas, which suggests a need for a predictive algorithm based on easily measured clinical signs.[128]

Prior Antibiotic Use. Prior antibiotic use has been consistently shown to be a predictor of MRSA-related SSI, due to the overuse of non-pathogen specific broad-spectrum antibiotics.[131,133,137,138] In contrast, a number of studies do not identify prior antibiotic use as a predictor of MRSA-related SSI, although some suggest that other concomitant medications, particularly immunosuppressants or steroids may play role.[7,17,130,132]

Prophylaxis. While the administration of Vancomycin has been shown to reduce the incidence of MRSA-related SSI, lack of screening leading to administration of inappropriate antibiotics has been shown to increase risk.[21,127,138] Administration of appropriate antibiotics between 1 and 2 hours prior to surgery has been shown to reduce the incidence of SSI.[128,130,138] In contrast, administration of inappropriate antibiotics (i.e., administration of antibiotics for which resistance is present) is associated with an

increased risk of SSI, prolonged hospital stay, and higher mortality.[127,137] By some estimates, approximately 32% of MRSA infections are inappropriately treated, reinforcing the need for better, more timely screening protocols.[137]

Hospitalization Factors. As with the prediction of SSI and MRSA in general, MRSA-related SSI is also significantly associated with prior or prolonged hospitalization.[7,132,134,137] In particular, the total duration of time in hospital, regardless of whether it is measured as total time, post-operative time, or duration of a previous hospitalization, is consistently associated with an increased risk of MRSA-related SSI.[21,137,138] Researchers differ in estimates of the number of days stay that indicates increased risk. Manian, Meyer, Setzer, Senkel [138] indicate that a post-operative stay of greater than 3 days significantly increases risk, whereas Sganga, Tascini, Sozio, Carlini, Chirletti, Cortese, Gattuso, Granone, Pempinello, Sartelli, Colizza [21] suggest that a current stay of more than 16 days increases risk, and Stevens et al.[137] find that a previous stay of greater than 8.4 days is a reliable predictor. Other researchers suggest that the number of previous visits is a predictor of MRSA-related SSI, with one study suggesting that patients with at least 5 previous hospital visits are at increased risk.[132]

Surgical Factors. There is a range of surgical factors that have been shown to increase risk of MRSA-related SSI, including timing of prophylaxis, type of surgery, and duration of surgery.[22,24,47,129,133,134,136,138] Multiple studies have demonstrated that longer surgical duration is significantly associated with increased risk of MRSA-related SSI.[47,129,133,134,136] Estimates differ based on the type of surgery, but surgical duration of longer than 137 minutes has been shown to increase risk by 3 times.[136] Surgery type has also been shown to increase risk with cardiothoracic and other cardiac procedures

exhibiting significantly higher risk than other types of surgery.[22,133] In addition, orthopedic surgeries that require placement of a prosthetic device also increased risk of SSI.[138] in general, elective procedures reduced risk by 62%.[16] Thus, severity of the condition being treated and the invasiveness of the procedure may be key predictors of MRSA-related SSI.[7,125,129]

Severity of Illness and Comorbidities. Severity of illness, number and type of comorbidities, and type of surgery are all strong predictors of MRSA-related SSI.[7,20,79,127,129-133] Security of illness, as measured by the American Society of Anesthesiologists (ASA) physical status classification, is correlated with MRSA-related SSI.[130,132] The ASA score is a qualitative assessment of a patient's perioperative fitness, where patients are assigned a score from 1 (normal healthy patient) to 6 (brain-dead patient), based on their physical health.[140] A one point increase of ASA score increases risk of SSI by a factor of two.[130] Similarly, the presence of comorbidities, as measured by the Charlson comorbidity index, results in increased risk of MRSA-related SSI.[21,24,129] The Charlson comorbidity index is predictor of 10-year mortality calculated by assigning different conditions a score from 1 to 6 and then computing a weighted average based on the severity of each comorbidity.[141] In one study researchers found that patients with a Charlson score of 1 or 2 were 4 times more likely to suffer a MRSA-related SSI compared to those with a Charlson score or 0 and patients with a Charlson score of 3 exhibited 6 times the risk.[129] In addition, the presence of specific comorbidities has been shown to increase risk of MRSA-related SSI. Examples include diabetes, cancer, and skin infections.[17,127,131,136,137]

Other Predictors. A range of other, less common MRSA-related SSI predictors has been identified, however many of these predictors reflect the specific health condition, surgery, and population being studies, and thus are less useful for a broadly applicable predictive model.[17,79,134] Examples include poor functional status, poor nutritional status, anemia, and lack of independence with activities of daily living (ADL).[17,79,129] Table 3 summarizes the 20 most cited MRSA-related SSI predictors.

Table 3: Top 20 MRSA-related SSI Predictors

Predictor	Citation Count
Age	12
MRSA - Colonization	10
Surgical Duration	9
Colon Surgery	8
ASA Score	7
Antibiotic Prophylaxis (Protective)	6
Sex	6
Wound Class	6
Antibiotic Administration/Use (Risk Factor)	4
Charlson Score	4
Immuno-compromised/Therapy	4
Hospitalization - Previous or Transfer	3
Laparoscopic Surgery - Protective	3
Surgery Type	3
Abdominal Surgery	2
Activities of Daily Living (ADL)	2
Antibiotic - Broad Spectrum	2
Blood Loss	2
BMI/Obesity	2
Diabetes	2

2.5 Clinical Decision Support Systems

2.5.1 Overview of Clinical Decision Support Systems

There has been a significant interest in the development of computational tools to support clinical diagnosis since the late 1950s.[142-145] It is estimated that almost 100,000 patients die each year in the United States as a result of preventable medical errors.[143] Despite the growing evidence base being generated by an ever increasing number of clinical trials, epidemiological studies and systematic reviews, non-adherence to guidelines in healthcare remains a significant problem in clinical practice.[143,146] Computational tools, such as clinical decision support systems (CDSS), are able to address deficiencies in clinical care and improve the timeliness and accuracy of diagnosis by providing specific recommendations based on patient-specific factors.[143,147]

Effective clinical decision making requires assessment of risk of possible treatment options. CDSS are able to provide evidence-based recommendations based on empirical data and individual patient characteristics and can easily be incorporated into a clinician's workflow.[143,148] CDSS incorporate computer-based rules or algorithms that provide clinical recommendations based on empirical knowledge gained from the scientific literature.[149,150] Patient-specific information, such as demographics, results of lab tests and other clinical characteristics can be obtained from a patient's electronic health record (EHR), is used to tailor the output of a CDSS to a specific patient's needs.[151-154]

It is important to recognize that the primary aim of a CDSS is to provide decision support, that is, to provide physicians with treatment recommendations based on the combination of information about the specific disease, known risk factors, and patient characteristics.[155,156] CDSS are advantageous because they can distill empirical data and provide guidance to physicians in a timely manner.[157]

Generally, CDSS comprise a knowledge base, a reasoning engine, and a user interface.[142,155,158] The knowledge base contains detailed information about a range of medical conditions, signs and symptoms, and clinical guidelines, augmented with patient-specific information from the EHR.[156,159] The reasoning, or inference, engine applies either a series of decision rules or a computational algorithm to provide recommendations, and it is this feature that separates CDSS from medical reference programs.[149,156,160,161] The user interface is of critical importance as it drives the collection of patient-specific information, such as current signs and symptoms – and in doing so ensures that the physician asks an appropriate and complete set of questions – and provides feedback to the physician in the form of context-sensitive alerts.[158,159,161]

While a CDSS can employ multiple algorithms for performing inference, it is important that a decision/recommendation is presented with an explanation of why the recommendation is optimal.[162] Thus, usability and other factors relating to the user interface, such as the delivery of reminders and automated prompts, are critical factors in both system uptake and performance.[155,157,160]

CDSS have been developed to address a range of clinical conditions and can be categorized into multiple types based on the level of information provided to the user.[155,156,160] In the simplest case, CDSS can present relevant information to the user to enable them to make a decision, whereas more complex systems can analyze patient trends, generate recommendations, and even learn over time based on the collection of new data.[155,156] It is the latter types of CDSS that provide the most utility to physicians by providing recommendations and alerts, such as the identification of drug-drug interactions, warnings of potential antimicrobial resistance, flagging of abnormal lab values, and recommendations for disease management.[152,153,163] CDSS are currently used for multiple medical conditions in both general and specialty practice, including intensive care, pediatrics, and cancer.[151] In many cases, CDSS are deployed as independent systems, outside of the EHR and in this way provide a tailored user interface that minimizes cognitive burden on the user that is ultimately more effective than embedding the CDSS within the EHR.[153,163]

While research has shown that deployment of CDSS can reduce medical errors, improve adherence to clinical guidelines, improve prescribing practices and enhance antimicrobial stewardship, there exist a number of challenges to their successful implementation and adoption.[143,148,149,160,164-166] Uptake of CDSS can be limited if

systems are not intuitive and easy to use and if the systems do not present clear recommendations in a way that is understandable to the user.[149,163,166] Acceptance of CDSS by physicians varies considerably, with some physicians reporting increased patient confidence with others reluctant to use a CDSS in front of a patient.[155] Other challenges include ensuring the validity of the underlying knowledge base and inference models, and incorporation of patient preferences into the guidance generated.[148,149] It must also be noted that the financial benefit of implementing CDSS has not been clearly and consistently demonstrated because many studies focus only on effectiveness.[164]

CDSS have been shown to improve the performance of physicians and other medical providers, improve patient outcomes, and minimize errors.[142,146,151,157,164] Some studies estimate that the implementation of CDSS can have up to a 20% magnitude of effect on patient morbidity and may also reduce mortality.[167,168] Other researchers, in meta-analyses and systematic reviews of CDSS effectiveness, find that 70% of studies assessing CDSS report statistically significant improvements to patient outcomes.[150,164] Additional benefits of CDSS include the reduction in medical errors, and thus there is the potential for CDSS to reduce the occurrence of medical malpractice lawsuits.[156,157] To maximize the utility of CDSS and improve the quality of care, researchers suggest a need to identify and monitor key performance indicators.[54]

Additional benefits of CDSS include their ability to enforce treatment and diagnosis guidelines, standardize approaches to patient care, and improve the consistency of clinical interventions.[167-169] CDSS that generate alerts relating to patient safety have been shown to be particularly useful, especially when the content of the alerts is outside of the physician's area of expertise.[165,170-172] Alerts that identify potential medication

issues, such as the recommendation for administration of appropriate antibiotics, can reduce inappropriate medication user and potentially limit the inappropriate prescribing of broad spectrum antibiotics.[161,173]

The long term success of CDSS requires that the underlying algorithms are up-to-date, complete, and reflect current clinical guidelines.[156,165,174] While integration with EHR systems can be advantageous in maintaining the knowledgebase, portable, stand-alone systems have also been shown to be effective.[161] Other challenges to successful implementation include system complexity, which can occur if too many data elements are required to be entered into the system and results in poor compliance.[150,168,175]

Despite the potential advantages of CDSS, utilization for diagnostic purposes remains low.[157] In addition to the system complexity and knowledgebase maintenance issues described above, a proportion of physicians question the ability of CDSS to accurately capture the complexity inherent in the medical diagnosis process, and further warn that over-reliance on an automated system may, in fact, lead to medical mistakes and poorer outcomes.[176] However, despite these concerns, CDSS have potential to improve outcomes.[142,146,164]

2.5.2 Applications of Clinical Decision Support Systems

Overuse and inappropriate administration of antibiotics in the clinical setting is one major public health issue that can be improved through the use of CDSS.[53,54,100,175] Antimicrobial resistance is a growing problem with over 2 million people becoming infected with resistant pathogens each year in the United States.[53,100] Overuse of broad spectrum antibiotics and lack of uniformity in testing has resulted in increasing rates of antibiotic resistance, with some estimates indicating that half of all antibiotic prescribing

is inappropriate.[54,146,154] One of the main reasons for overuse of broad spectrum antibiotics is the inability of clinicians to identify the infection causing pathogen in a timely manner, due to the time required to obtain lab test results.[100,146] Up to 40% of all hospitalized patients diagnosed with infections are treated with antibiotics, despite the fact that the resistance profile of the pathogen is not considered and even that some of the infections are not bacterial.[161,173,177,178]

In addition to being a major clinical issue that contributes to morbidity and mortality, inappropriate prescribing of antibiotics can have significant financial implications for a health care system.[53,100,177] By some estimates, annual costs associated with antimicrobial resistance are up to $35 billion, and for these reasons, developing tools for enhanced antimicrobial stewardship that can ensure the administration of appropriate antibiotics when needed, could have a significant impact on morbidity, mortality, and healthcare costs.[154]

Use of CDSS in antimicrobial stewardship applications has been limited and where used has met with limited success.[113,153,163] Lack of rapid screening tests for common pathogens, such as MRSA, in combination with the poorer outcomes observed when treatment is delayed, leads to clinicians prescribing inappropriate antibiotics.[46,51] One study found that guidelines for Vancomycin use – the first line treatment for MRSA infection – were not followed 68% of the time.[175] Thus, there is a clear need for CDSS in the prediction of potential AMR infections and subsequent recommendation of appropriate treatment.[54,163]

Several studies have attempted to use machine learning methods to identify resistant pathogens and subsequently recommend appropriate treatment.[40,46,53,113]

Researchers have adopted different approaches to providing clinical decision support in the context of AMR pathogens, such as MRSA.[45,46,51,175] Successful CDSS implementation for the enforcement of Vancomycin guidelines requires the timely alerting of the physician at the time of potential antibiotic prescription.[115,154] In addition, to address issues of timeliness, the CDSS knowledgebase should incorporate robust epidemiological data so that recommendations can be based on observable and measurable clinical signs and symptoms.[40,46,47,113] In the case of MRSA infection, differentiating between resistant and non-resistant infection is a significant challenge, and thus developing an effective CDSS for MRSA infection remains an open problem.[40,108]

Despite challenges in the development and adoption of CDSS for AMR applications, CDSS are used successfully in a broad range of medical domains.[115,152,168,179] Examples include treatment of skin and soft tissue infections, IV drug administration, treatment and management of sepsis, diagnosis and treatment of chronic obstructive pulmonary disease, and treatment of acute respiratory distress syndrome.[147,150,165,168,169] CDSS can result in increased adherence to clinical guidelines, to triage patients upon presentation to an emergency room, support early detection of disease, and inform treatment decisions.[115,179]

Benefits of CDSS are emphasized when the condition being treated is complex and requires a detailed review of the patient's medical chart, such as with the screening of cervical cancer – in this case the CDSS can more effectively, efficiently and consistently distill the necessary information and provide a recommendation consistent with guidelines.[174] This benefit is also realized in the oncology domain, which requires integration of multiple types of data from the patient's clinical record, including imaging,

pathology and clinical data.[152] CDSS have also been shown to reduce prescribing errors and improve standardization of drug administration, including intravenous and oral rehydration solution administration.[152,169,176] Thus, it can be seen that CDSS are most successful when applied to guideline adherence and application of standardized treatment protocols in situations where there is a significant cognitive burden.[152,162,167,176]

2.5.3 Computational Approaches to Clinical Decision Support Systems

Architecturally, a CDSS consists of three components: a knowledgebase, an inference engine, and a user interface.[142,155,158] The knowledgebase is the core of a CDSS system and contains key data that the inference engine uses to make diagnosis of treatment recommendations.[177] The user interface (UI) facilitates the collection of a variety of patient data, which could include medical history and quality of life factors.[147] The UI also provides the primary output mechanism and presents diagnosis and treatment recommendations to the physician.[142,147] In addition, when incorporated effectively into the physician's workflow the UI can present medication and other alerts, which prevent medical errors and medication interactions, and ensure compliance with necessary guidelines.[115,147,160,166] The HELP system was an early and successful example of a CDSS, and key to its success was the generation of automated alerts based on discrepancies in the patient's medical record.[142,180] CDSS differ in how the inference engine interacts with the knowledge base and generally use rule-based or machine-learning approaches, although with the advent of data-driven approaches and integration with EHRs, the lines between these two approaches are blurred.[142]

Traditionally, CDSS contained an inference engine that applied computational logic to patient input based on a series of system-defined rules.[142,181] The MYCIN system

is an early example of an expert-based CDSS, which uses predicate logic and the chaining of rules to generate recommendations.[181] Expert-based approaches are able to request additional input to narrow down potential diagnoses. The DXPlain system is a good example of an expert based system that generates a list of potential diagnoses based on input of clinical signs and symptoms.[182] Modern approaches are typically data driven, with the rules automatically extracted from clinical trial data and a matching algorithm used to determine the applicability of different rules based on the input of patient characteristics.[148,162,178] In some senses, these systems are almost full machine-learning systems, where the knowledgebase is essentially a broad dataset that incorporates a range of patient and hospital level data, and the inference engine is a statistical model trained on these data.[174,178] While rule-based systems are effective, they can require significant maintenance and systems that automatically incorporate data based on an EHR or clinical data warehouse and generate recommendations based on machine-learning have become the preferred approach.[142,158,174,178]

Machine learning comprises a series of computational techniques for learning from data.[53,147,178] A significant increase in the amount of data that is captured both during medical encounters and as part of hospital quality assurance procedures has led to the development of a range of machine learning algorithms that aim to learn models from data.[53,178] These models can then be used to predict outcomes based on a set of predictors obtained from the patient and hospital, and have been shown to accommodate increasingly large datasets, such as those generated by large-scale genome wide association studies.[100,147,176,179] The most prevalent machine learning approach used for CDSS is known as supervised learning or classification.[147,183] In this learning approach,

data are provided about two or more classes of patients – for example, patients with MRSA and patients without MRSA – and the algorithm builds a model that is able to predict the outcome when new data are supplied. Thus, the learning process requires labeled training data in which the outcome for each patient is known.[184]

A wide range of supervised learning algorithms have been used to develop CDSS, including logistic regression, artificial neural networks, decision trees, support vector machines, classification and regression trees, random forests, and Bayesian networks.[51,158,183,185] While these algorithms differ significantly in their expressiveness, it is unclear which is better in a given situation.[183,184] Although artificial neural networks and support vector machines have been popular in the development of CDSS due to their expressive modeling power they are considered "black box" algorithms because it is not clear to the user how the algorithm obtained its prediction.[53] In contrast, there is a general preference for algorithms such as logistic regression and decision trees where the contribution of each predictor is easily assessed by the user.[53] Thus, it is important to understand the strengths and limitations of each approach prior to developing a CDSS.[184]

Multiple logistic regression is a statistical modeling technique that estimates the probability that a certain set of predictors result in a given binary outcome.[183,184,186] The dependent variable in a logistic regression model is estimated using a linear combination of predictors, with the result indicating the relative contribution of each predictor, presented as an odds ratio with p-value.[184] Predictors can be added or removed from logistic regression model based on the significance of their contribution to the outcome, using automated learning techniques, including forward and backward stepwise variable selection.[51,184] Enhancements to the typical algorithm include use of shrinkage

techniques, which allow for a greater number of predictors, and conditional logistic regression, in which positive and negative cases are matched based on a set of predictors.[114,184]

A number of researchers have successfully used logistic regression methods for the prediction of MRSA-related illnesses.[46,108,114,148] In a case-control study, Harbarth, Sax, Fankhauser-Rodriguez, Schrenzel, Agostinho, Pittet [114] used conditional logistic regression with 1:4 matching of cases to controls to predict MRSA carriage at hospital admission; In a study of MRSA-related pneumonia, Jung, Kang, Park, Park, Leem, Kim, Chung, Kim, Kim, Chang, Jung [46] found that patients with a previous history of MRSA were six times more likely to have pneumonia than those without such history; and, Goodman, Lessler, Cosgrove, Harris, Lautenbach, Han, Milstone, Massey, Tamma, Antibacterial Resistance Leadership [51] used multivariate logistic regression to predict extended-spectrum β-lactamase (ESBL) bacteremia. One additional advantage of logistic regression is the ability to determine the probability of class assignment, which can be used to develop risk scores.[148]

Training of logistic regression models is generally computationally tractable and the results of the model are easily interpretable to clinicians, making the technique an ideal choice for CDSS.[184,186] Potential disadvantages include an inability to model non-linear relationships without explicitly including interaction and higher-order terms in the model specification, which is difficult to accomplish and can significantly increase complexity.[184,186] Alternative approaches, such as artificial neural networks (ANNs), can address these issues through their inherent ability to model non-linear relationships.[186]

ANNs are non-parametric models that contain multiple layers of neurons that attempt to simulate the learning processes of the human brain.[186-188] The combination of multiple hidden layers of neurons and the weighting of inputs as they propagate the network, enable the ANN to model complex, non-linear relationships.[158,179] As with logistic regression models, ANNs are able to estimate a probability for class assignment and so can also be used as risk models, however because of the complex way the result is calculated, it is difficult to determine how the ANN generates a result and the impact of individual predictors on the result.[158] It is also difficult to determine both the optimal number of hidden layers and the number of neurons in each layer, although it is possible to approach this systematically if a sufficient amount of training data is available.[162] Despite these limitations, ANNs are very powerful machine learning methods, and are commonly used in the development of CDSS.[186]

ANNs have been used extensively in the development of CDSS.[108,147,158,162,179] In a study of 5-year survival after surgery, Buzaev, Plechev, Nikolaeva, Galimova [162] used an ANN to predict survival based on whether patients underwent percutaneous coronary intervention or coronary artery bypass surgery; Anakal, Sandhya [147] used an ANN to predict disease severity in patients with chronic obstructive pulmonary disease; Pombo, Araujo, Viana [179], in a systematic review of knowledge discovery for pain management models, identified a large number of studies that used ANNs as the primary modeling approach; and, Hsu, Lin, Chen, Liu, Muder [189] used an ANN to predict MRSA colonization with 90% accuracy.

Training of ANNs can be computationally challenging, however improvements in computing power have largely ameliorated this concern when clinical data are being used

to train the algorithm.[184,186,187] The main advantage of ANNs is their expressiveness, i.e. their ability to model non-linear relationships between multiple predictors.[184] Disadvantages include the "black box" nature of the algorithm, in which decisions are hidden from the user, although this limitation may be offset by improved predictive ability.[186] Although use of ANNs in MRSA-related models is limited, their expressiveness may be well suited to elucidate the subtle differences between MRSA-related and MSSA-related SSI.[16]

Tree-based methods, such as decision trees, random forests, and classification and regression trees are alternative machine learning methods based on the recursive partitioning of data based on cut-off values of predictors.[51,108,190,191] Decision tree methods can be considered an extension of the risk prediction tools that many physicians use, and as such are an intuitive representation of quantitative data that supports clinical decision making.[53,184] Challenges with decision trees are that they are sensitive to the order of variables and the defined cut-points and are prone to overfitting.[184,190] These challenges are largely addressed through random forests, which consider multiple trees with each tree containing a random combination of predictors.[108,179,184]

A number of researchers have successfully applied decision trees to clinical decision support problems.[51,53,183] In study to determine whether patients were infected with ESBL-producing bacteria, Goodman, Lessler, Cosgrove, Harris, Lautenbach, Han, Milstone, Massey, Tamma, Antibacterial Resistance Leadership [51] developed a decision tree that contained five predictors, including history of colonization, presence of indwelling hardware, age, recent hospitalization, and antibiotic exposure; Gerald, Tang, Bruce, Redden, Kimerling, Brook, Dunlap, Bailey [192] used a decision tree to predict

whether contacts of tuberculosis patients were likely to have positive TB test results; and Rundensteiner, Kong, Teeple, Brownell, Sen, Hartvigsen [108] achieved excellent results (AUC > 0.95) using random forests to predict MRSA infection based on known risk factors.

Other machine learning methods used in CDSS include support vector machines (SVM) and Bayesian networks (BN).[108,179,185] SVMs aim to separate data into two distinct classes using a hyperplane that maximizes the distance between classes in high-dimensional space.[179,193,194] It is possible to use the "kernel trick" to map data into a higher dimensional space to enable data to be linearly separable, thus SVM is a good method to use when dealing with high dimensional or noisy data.[184,193,194] As with ANNs, disadvantages of SVM include the computational complexity of the training process and the fact that it is a "black box" method.[184] BNs are probabilistic graphical models that model the many-to-many relationships between variables using a conditional factorization of the joint probability distribution over all the variables.[185,195,196] Thus, BNs can be used to model multiple outcomes and determine the impact of perturbations to variables on these outcomes.[195]

Although less popular than other methods, a number of researchers have used SVM and BN for predictive modeling.[185] In a study of sepsis patients, Gultepe, Green, Nguyen, Adams, Albertson, Tagkopoulos [185] used both SVM and BN to predict mortality; and Rundensteiner, Kong, Teeple, Brownell, Sen, Hartvigsen [108] used SVM for the early prediction of MRSA infection using data from EHR, although it must be noted that performance of SVM on this dataset was lower than that of logistic regression and random forests. Thus, while these results demonstrate the applicability of machine

learning algorithms to CDSS, it is important to choose an appropriate modeling technique and carefully evaluate the results.[108,183]

Evaluation of a machine learning model usually occurs through the application of the trained algorithm to a held out test data set using a range of metrics, including the receiver operating characteristic (ROC) curve, sensitivity, specificity, positive and negative predictive value.[158,184,197-199] A ROC curve is a plot of the true positive rate against the true negative rate for different cut-off values, both measured between 0 and 1, in which the area under the curve (also known as the c-statistic) is a measure of predictive performance.[183,198] An area of 0.5 represents a chance result and an area of 1 represents a perfect predictor.[158] Sensitivity, measured between 0 and 1 is the true positive rate, i.e. the number of positive cases that are correctly identified by the algorithm, whereas specificity is the opposite – the number of negative cases correctly identified by the algorithm.[200] Positive predictive value and negative predictive value are the proportion of positive and negative cases identified by the algorithm respectively.[158]

CDSS can achieve excellent diagnostic performance with ROC values of around 0.9 and high levels of sensitivity and specificity.[40,177,185] In a model that predicted the inappropriate administration of antibiotics, Beaudoin, Kabanza, Nault, Valiquette [178] achieved a positive predictive value of 0.74 and sensitivity of 0.96; Ju, Zurakowski, Kocher [40], in a study of MRSA infection in children, achieved a c-statistic of 0.94; and, Rundensteiner, Kong, Teeple, Brownell, Sen, Hartvigsen [108] achieved a c-statistic of over 0.9 when using several different algorithms for early prediction of MRSA infection. While CDSS have had a significant impact on the delivery of health care and contributed

to improved morbidity and mortality, further research is needed to identify the optimal algorithm that should be used in a given situation.[143,146,163,173]

2.6 Summary of the Research Gap

Although a significant amount of research has been conducted looking independently at the causes of SSIs and MRSA, limited epidemiological research has looked at the causes of MRSA-related SSI. Estimates of MRSA incidence differ and researchers disagree about whether MRSA-related SSI is increasing or decreasing. Further, identifying clinical and hospital-level predictors that can effectively differentiate between MRSA-related SSI and MSSA-related SSI, requires a large-scale nationally representative database and the application of models that can adequately capture the subtle differences between these groups of patients. It is also unclear which models, whether artificial neural networks, logistic regression, or other modeling approaches, will perform the best within a CDSS. Thus, this research aims to build and evaluate a CDSS based on a large-scale, nationally-representative dataset of clinical and hospital-level predictors. In doing so, the work will also summarize the epidemiology and risk factors for MRSA-related SSI and identify and appropriate modeling approach.

CHAPTER 3

METHODOLOGY

3.1 Overview

This study uses a range of statistical analysis and machine learning methods to build and evaluate a series of predictive models to identify surgical patients at high-risk of Methicillin Resistant *Staphylococcus aureus* (MRSA) infections, specifically MRSA-related Surgical Site Infections (SSIs). A descriptive analysis of prevalence and risk factors was performed using SAS 9.4 with a large, multi-year nationally-representative healthcare dataset. Artificial Neural Network (ANN) and Logistic Regression (LR) models were built and evaluated using R, and the models were used to develop a clinical decision support system (CDSS) using R Shiny.[201] The sections that follow describe the characteristics of data source, data elements used, data processing and cleaning methods, configuration of the learning algorithms, evaluation methods, and design for the CDSS.

3.2 Data Source

The primary data source for then analyses specified in this chapter is the National Inpatient Sample (NIS). The NIS is the largest all-payer (i.e., it includes Medicaid, Medicare, privately insured and uninsured patients) hospital discharge database in the United States. Collected annually as part of the Health Care Utilization Project (HCUP) and sponsored by the Agency for Healthcare Research Quality (AHRQ), the NIS is a

nationally-representative sample of discharges from participating community hospitals in the United States. The sample does not include rehabilitation hospitals or long term care hospitals. In 2012, the NIS sampling strategy was updated to select 20% of all discharges across participating hospitals across 46 participating states, which represents approximately 97% of the United States population. The NIS database contains data on approximately 7 million discharges for each year of data collection.

The NIS comprises approximately 100 variables, which include a range of demographic, clinical, procedural, and hospital-related factors. Patient factors collected include age, sex, race, type of admission (urgent or elective), reason for hospitalization, ICD-9-CM codes for up to 24 secondary diagnoses, ICD-9-CM codes for up to 15 procedures, final disposition of the patient, indicator variables for 29 specific comorbidities, length of hospital stay, hospital charges for care, and mortality. Hospital characteristics include the teaching status of the hospital, a categorization of hospital size (small, medium or large) based on number of beds, urban or rural location, state, and ownership of hospital.

Analyses will use data from the 2010 - 2014 data collection years. The number and proportion of Major Surgical Procedures (MSP), SSI, MRSA infections, MRSA-related SSI, and total number of discharges sampled is shown in Table 4 below.

Table 4: Summary of NIS Sample for Years 2010-2014

Year	MSP	SSI	MRSA	MRSA SSI
2010	2,265,467 (31.04%)	24,761 (1.09%)	24,156 (1.07%)	2,483 (0.1%)
2011	2,323,627 (31.83%)	26,163 (1.13%)	25,802 (1.11%)	2,615 (0.1%)
2012	2,090,896 (28.65%)	21,613 (1.03%)	22,416 (1.07%)	2,111 (0.1%)
2013	2,048,380 (28.77%)	20,735 (1.01%)	21,550 (1.05%)	1,960 (0.1%)
2014	2,023,168 (28.61%)	20,130 (1.00%)	21,105 (1.04%)	1,888 (0.1%)

3.3 Data Elements

The table below lists the NIS data elements considered for inclusion in the model. The data element list includes a range of patient-level demographic factors, a range of ICD-9 codes representing diagnoses, procedures, chronic conditions and comorbidities, in addition to a set of hospital level variables. The variable *ORPROC* is used to define the study population — patients undergoing a major surgical procedure. In addition, a small number of analysis variables will be derived, including case/control, MRSA, MSSA, and SSI indicators.

Table 5: NIS Variables Considered for Inclusion

Variable	Description
AGE	Age in years at admission
APRDRG_Risk_Mortality	All Patient Refined DRG: Risk of Mortality Subclass
APRDRG_Severity	All Patient Refined DRG: Severity of Illness Subclass
CM_AIDS	AHRQ comorbidity measure for ICD-9-CM codes: acquired immune deficiency syndrome
CM_ALCOHOL	AHRQ comorbidity measure for ICD-9-CM codes: alcohol abuse
CM_ANEMDEF	AHRQ comorbidity measure for ICD-9-CM codes: deficiency anemias
CM_ARTH	AHRQ comorbidity measure for ICD-9-CM codes: rheumatoid arthritis/collagen vascular diseases
CM_BLDLOSS	AHRQ comorbidity measure for ICD-9-CM codes: chronic blood loss anemia
CM_CHF	AHRQ comorbidity measure for ICD-9-CM codes: congestive heart failure
CM_CHRNLUNG	AHRQ comorbidity measure for ICD-9-CM codes: chronic pulmonary disease
CM_COAG	AHRQ comorbidity measure for ICD-9-CM codes: coagulopathy
CM_DEPRESS	AHRQ comorbidity measure for ICD-9-CM codes: depression
CM_DM	AHRQ comorbidity measure for ICD-9-CM codes: diabetes, uncomplicated
CM_DMCX	AHRQ comorbidity measure for ICD-9-CM codes: diabetes with chronic complications
CM_DRUG	AHRQ comorbidity measure for ICD-9-CM codes: drug abuse
CM_HTN_C	AHRQ comorbidity measure for ICD-9-CM codes: hypertension (combine uncomplicated and complicated)
CM_HYPOTHY	AHRQ comorbidity measure for ICD-9-CM codes: hypothyroidism
CM_LIVER	AHRQ comorbidity measure for ICD-9-CM codes: liver disease
CM_LYMPH	AHRQ comorbidity measure for ICD-9-CM codes: lymphoma
CM_LYTES	AHRQ comorbidity measure for ICD-9-CM codes: fluid and electrolyte disorders
CM_METS	AHRQ comorbidity measure for ICD-9-CM codes: metastatic cancer
CM_NEURO	AHRQ comorbidity measure for ICD-9-CM codes: other neurological disorders
CM_OBESE	AHRQ comorbidity measure for ICD-9-CM codes: obesity
CM_PARA	AHRQ comorbidity measure for ICD-9-CM codes: paralysis

CM_PERIVASC	AHRQ comorbidity measure for ICD-9-CM codes: peripheral vascular disorders
CM_PSYCH	AHRQ comorbidity measure for ICD-9-CM codes: psychoses
CM_PULMCIRC	AHRQ comorbidity measure for ICD-9-CM codes: pulmonary circulation disorders
CM_RENLFAIL	AHRQ comorbidity measure for ICD-9-CM codes: renal failure
CM_TUMOR	AHRQ comorbidity measure for ICD-9-CM codes: solid tumor without metastasis
CM_ULCER	AHRQ comorbidity measure for ICD-9-CM codes: peptic ulcer disease excluding bleeding
CM_VALVE	AHRQ comorbidity measure for ICD-9-CM codes: valvular disease
CM_WGHTLOSS	AHRQ comorbidity measure for ICD-9-CM codes: weight loss
DXn	ICD-9-CM Diagnosis
ELECTIVE	Elective versus non-elective admission
HOSP_BEDSIZE	Bed size of hospital
HOSP_LOCTEACH	Location/teaching status of hospital
HOSP_REGION	Region of hospital
HOSP_RNFTEAPD	RN FTEs per 1000 adjusted inpatient days
HOSP_RNPCT	Percentage of RNs among all nurses (RNs and LPNs)
HOSP_TEACH	Teaching status of hospital
LOS	Length of stay, cleaned
NCHRONIC	ICD-9-CM Number of chronic conditions
NDX	Number of ICD-9-CM diagnoses on this discharge
NPR	Number of ICD-9-CM procedures on this discharge
ORPROC	Major operating room ICD-9-CM procedure indicator
PAY1	Expected primary payer, uniform
RACE	Race
SEX	Sex of the patient
TRAN_IN	Indicator of a transfer into the hospital
TRAN_OUT	Transfer out of the hospital
YEAR	Calendar year
ZIPINC	Median household income for patient's ZIP Code (based on 1999 demographics)

Derived variables are listed in Table 6.

Variable	**Description and Derivation**
ANTIBX	Long-term (current) antibiotic use: *IF DX1 = 'V5862' OR DX2 = 'V5862' OR DX3 = 'V5862' OR DX4 = 'V5862' OR DX5 = 'V5862' OR DX6 = 'V5862' OR DX7 = 'V5862' OR DX8 = 'V5862' OR DX9 = 'V5862' OR DX10 = 'V5862' OR DX11 = 'V5862' OR DX12 = 'V5862' OR DX13 = 'V5862' OR DX14 = 'V5862' OR DX15 = 'V5862' OR DX16 = 'V5862' OR DX17 = 'V5862' OR DX18 = 'V5862' OR DX19 = 'V5862' OR DX20 = 'V5862' OR DX21 = 'V5862' OR DX22 = 'V5862' OR DX23 = 'V5862' OR DX24 = 'V5862' OR DX25 = 'V5862' OR DX26 = 'V5862' OR DX27 = 'V5862' OR DX28 = 'V5862' OR DX29 = 'V5862' OR DX30 = 'V5862' THEN ANTIBX = 1; ELSE ANTIBX = 0;*
COMORBIDITIES	Number of comorbidities: *Calculated as the total number of comorbidities present.*
MORTAL_SCORE	Elixhauser comorbidity index — mortality score: *Calculated using Elixhauser comorbidity software.*
MRSA	*Methicillin Resistant Staphylococcus aureus infection:*

	IF DX1 = '04112' OR DX2 = '04112' OR DX3 = '04112' OR DX4 = '04112' OR DX5 = '04112' OR DX6 = '04112' OR DX7 = '04112' OR DX8 = '04112' OR DX9 = '04112' OR DX10 = '04112' OR DX11 = '04112' OR DX12 = '04112' OR DX13 = '04112' OR DX14 = '04112' OR DX15 = '04112' OR DX16 = '04112' OR DX17 = '04112' OR DX18 = '04112' OR DX19 = '04112' OR DX20 = '04112' OR DX21 = '04112' OR DX22 = '04112' OR DX23 = '04112' OR DX24 = '04112' OR DX25 = '04112' OR DX26 = '04112' OR DX27 = '04112' OR DX28 = '04112' OR DX29 = '04112' OR DX30 = '04112' OR DX1 = '03812' OR DX2 = '03812' OR DX3 = '03812' OR DX4 = '03812' OR DX5 = '03812' OR DX6 = '03812' OR DX7 = '03812' OR DX8 = '03812' OR DX9 = '03812' OR DX10 = '03812' OR DX11 = '03812' OR DX12 = '03812' OR DX13 = '03812' OR DX14 = '03812' OR DX15 = '03812' OR DX16 = '03812' OR DX17 = '03812' OR DX18 = '03812' OR DX19 = '03812' OR DX20 = '03812' OR DX21 = '03812' OR DX22 = '03812' OR DX23 = '03812' OR DX24 = '03812' OR DX25 = '03812' OR DX26 = '03812' OR DX27 = '03812' OR DX28 = '03812' OR DX29 = '03812' OR DX30 = '03812' OR DX1 = '48242' OR DX2 = '48242' OR DX3 = '48242' OR DX4 = '48242' OR DX5 = '48242' OR DX6 = '48242' OR DX7 = '48242' OR DX8 = '48242' OR DX9 = '48242' OR DX10 = '48242' OR DX11 = '48242' OR DX12 = '48242' OR DX13 = '48242' OR DX14 = '48242' OR DX15 = '48242' OR DX16 = '48242' OR DX17 = '48242' OR DX18 = '48242' OR DX19 = '48242' OR DX20 = '48242' OR DX21 = '48242' OR DX22 = '48242' OR DX23 = '48242' OR DX24 = '48242' OR DX25 = '48242' OR DX26 = '48242' OR DX27 = '48242' OR DX28 = '48242' OR DX29 = '48242' OR DX30 = '48242' THEN MRSA = 1; ELSE MRSA = 0;
MRSA_SSI	MRSA-related SSI: *IF MRSA = '1' AND SSI = '1' THEN MRSA_SSI = '1' ELSE MRSA_SSI = '0';*
MRSAHX	History of MRSA infection: *IF DX1 = 'V1204' OR DX2 = 'V1204' OR DX3 = 'V1204' OR DX4 = 'V1204' OR DX5 = 'V1204' OR DX6 = 'V1204' OR DX7 = 'V1204' OR DX8 = 'V1204' OR DX9 = 'V1204' OR DX10 = 'V1204' OR DX11 = 'V1204' OR DX12 = 'V1204' OR DX13 = 'V1204' OR DX14 = 'V1204' OR DX15 = 'V1204' OR DX16 = 'V1204' OR DX17 = 'V1204' OR DX18 = 'V1204' OR DX19 = 'V1204' OR DX20 = 'V1204' OR DX21 = 'V1204' OR DX22 = 'V1204' OR DX23 = 'V1204' OR DX24 = 'V1204' OR DX25 = 'V1204' OR DX26 = 'V1204' OR DX27 = 'V1204' OR DX28 = 'V1204' OR DX29 = 'V1204' OR DX30 = 'V1204' THEN MRSAHX = 1; ELSE MRSAHX = 0;*
MSSA	Methicillin-Susceptible Staphylococcus aureus: *IF DX1 = '04111' OR DX2 = '04111' OR DX3 = '04111' OR DX4 = '04111' OR DX5 = '04111' OR DX6 = '04111' OR DX7 = '04111' OR DX8 = '04111' OR DX9 = '04111' OR DX10 = '04111' OR DX11 = '04111' OR DX12 = '04111' OR DX13 = '04111' OR DX14 = '04111' OR DX15 = '04111' OR DX16 = '04111' OR DX17 = '04111' OR DX18 = '04111' OR DX19 = '04111' OR DX20 = '04111' OR DX21 = '04111' OR DX22 = '04111' OR DX23 = '04111' OR DX24 = '04111' OR DX25 = '04111' OR DX26 = '04111' OR DX27 = '04111' OR DX28 = '04111' OR DX29 = '04111' OR DX30 = '04111' THEN MSSA = 1; ELSE MSSA = 0;*
MSSA_SSI	MSSA-related SSI: *IF MSSA = '1' AND SSI = '1' THEN MSSA_SSI = '1' ELSE MSSA_SSI = '0';*
READMIT_SCORE	Elixhauser comorbidity index — readmission score: *Calculated using Elixhauser comorbidity software.*

SSI	Surgical site infection: *IF DX1 = '99859' OR DX2 = '99859' OR DX3 = '99859' OR DX4 = '99859' OR DX5 = '99859' OR DX6 = '99859' OR DX7 = '99859' OR DX8 = '99859' OR DX9 = '99859' OR DX10 = '99859' OR DX11 = '99859' OR DX12 = '99859' OR DX13 = '99859' OR DX14 = '99859' OR DX15 = '99859' OR DX16 = '99859' OR DX17 = '99859' OR DX18 = '99859' OR DX19 = '99859' OR DX20 = '99859' OR DX21 = '99859' OR DX22 = '99859' OR DX23 = '99859' OR DX24 = '99859' OR DX25 = '99859' OR DX26 = '99859' OR DX27 = '99859' OR DX28 = '99859' OR DX29 = '99859' OR DX30 = '99859' THEN SSI = 1; ELSE SSI = 0;*
SUSMRSA	Suspected MRSA infection: *IF DX1 = 'V0254' OR DX2 = 'V0254' OR DX3 = 'V0254' OR DX4 = 'V0254' OR DX5 = 'V0254' OR DX6 = 'V0254' OR DX7 = 'V0254' OR DX8 = 'V0254' OR DX9 = 'V0254' OR DX10 = 'V0254' OR DX11 = 'V0254' OR DX12 = 'V0254' OR DX13 = 'V0254' OR DX14 = 'V0254' OR DX15 = 'V0254' OR DX16 = 'V0254' OR DX17 = 'V0254' OR DX18 = 'V0254' OR DX19 = 'V0254' OR DX20 = 'V0254' OR DX21 = 'V0254' OR DX22 = 'V0254' OR DX23 = 'V0254' OR DX24 = 'V0254' OR DX25 = 'V0254' OR DX26 = 'V0254' OR DX27 = 'V0254' OR DX28 = 'V0254' OR DX29 = 'V0254' OR DX30 = 'V0254' THEN SUSMRSA = 1; ELSE SUSMRSA = 0;*

Table 6: Derived Variables

3.4 Study Design, Sample and Hypotheses

The study will apply the case-control methodology to test three hypotheses relating to MRSA infection in surgical patients. The underlying population being studied is patients undergoing a Major Surgical Procedure (MSP) in the United States between 2010 and 2014. These years were selected due to changes in the NIS database after 2014. Prior to 2015, ICD-9-CM codes were used and during 2015 ICD-10 CM codes were introduced. In all cases, MSP is defined using the NIS *ORPROC* variable. Specific hypotheses and case/control definitions are described below.

3.4.1 Hypothesis 1 — Prediction of MRSA in Surgical Patients

It is possible to develop a predictive model of MRSA infection in patients undergoing a Major Surgical Procedure in the United States. For this hypothesis, MRSA

infection is defined consistently with Allareddy, Das, Lee, Nalliah, Rampa, Allareddy, Rotta [16] as the union of three ICD-9-CM codes: Occurrence of MRSA infection (041.12), MRSA septicemia (038.12) and MRSA pneumonia (482.42). Cases are patients with MRSA infection per this definition (041.12 ∪ 038.12 ∪ 482.42) and controls are all other patients who underwent a MSP.

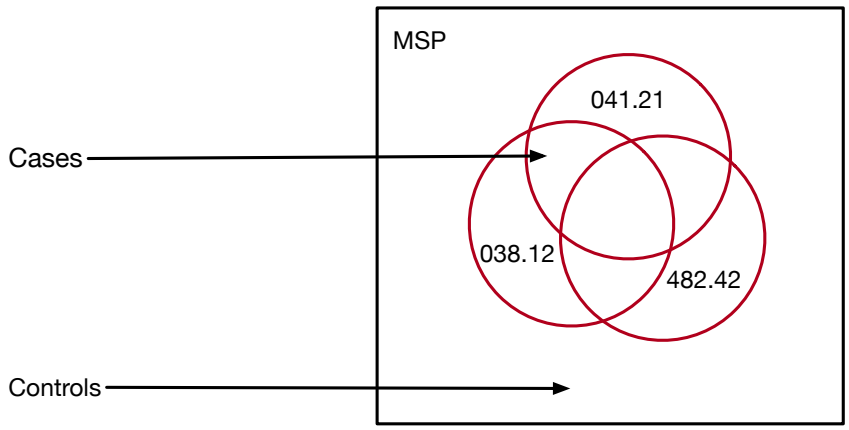

Figure 1: Case and Control Definition for Hypothesis 1

3.4.2 Hypothesis 2 — Prediction of MRSA-related SSI in Surgical Patients

It is possible to develop a predictive model of MRSA infection in patients undergoing a MSP who also experience an SSI in the United States. For this hypothesis, MRSA is defined as the union of occurrence of MRSA infection (041.12), MRSA septicemia (038.12) and MRSA pneumonia (482.42). SSI is defined using the ICD-9-CM code, post-operative wound infection (998.59). Cases are patients with MRSA infection per this definition [(041.12 ∪ 038.12 ∪ 482.42) ∩ 998.59] and controls are all other patients who underwent a MSP.

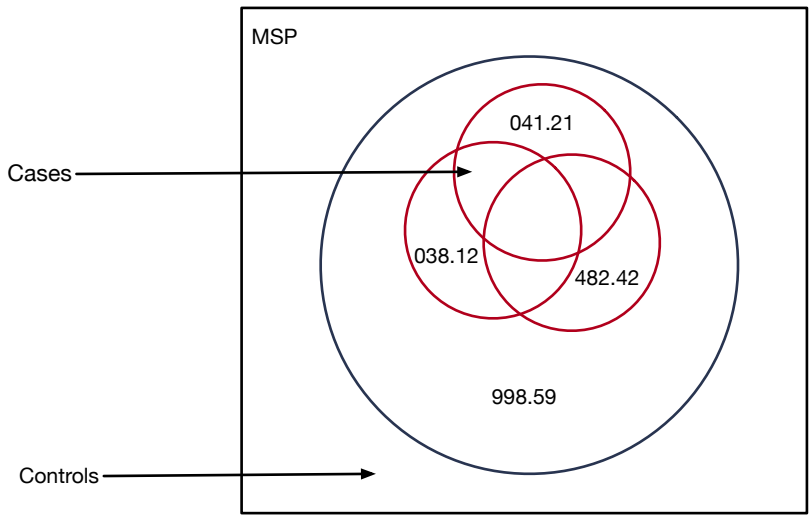

Figure 2: Case and Control Definition for Hypothesis 2

3.4.3 Hypothesis 3 — Prediction of MRSA Infection in S. aureus Patients

It is possible to develop a predictive model of MRSA-related SSI vs MSSA-related SSI in patients undergoing a MSP in the United States. For this hypothesis, MRSA infection is defined as the union of occurrence of MRSA infection (041.12), MRSA septicemia (038.12) and MRSA pneumonia (482.42). SSI is defined using the ICD-9-CM code, post-operative wound infection (998.59). Cases are patients with MRSA infection and SSI [(041.12 $\cup$ 038.12 $\cup$ 482.42) $\cap$ 998.59] and controls are all patients with MSSA-infection and SSI (041.11 $\cap$ 998.59) who underwent a MSP.

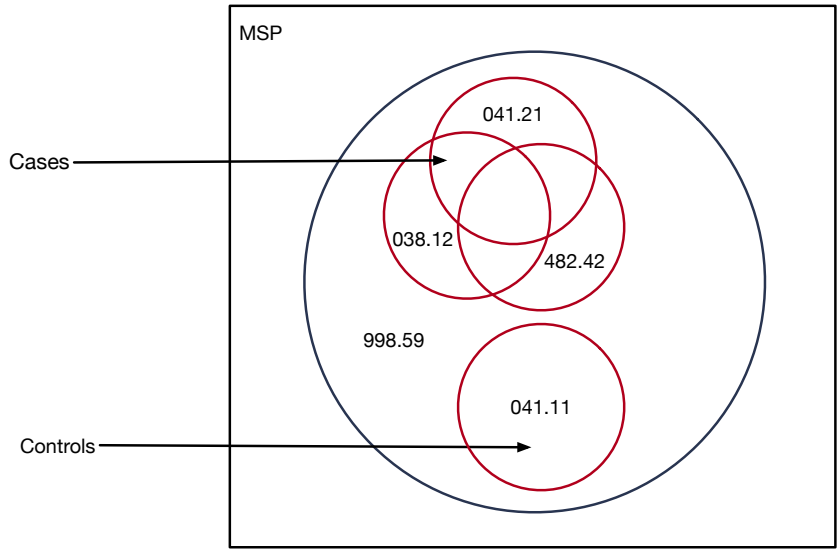

Figure 3: Case and Control Definition for Hypothesis 3

3.5 Descriptive Analysis

A number of descriptive analyses, aligned with the above hypotheses, will be performed to describe the study population:

1. Summary of basic characteristics of the NIS dataset, including frequency of MRSA, SSI, MRSA-related SSI, MSSA and MSSA-related SSI;
2. Nationally-representative prevalence estimates of MRSA, SSI, MRSA-related SSI, MSSA and MSSA-related SSI;
3. Univariate association of demographic, clinical and hospital-related factors for each research hypothesis.

For all hypothesis the statistics will be calculated for each variable and *p* values generated for the difference between case and control groups. For normally distributed continuous variables, a t-test will be performed. For non-normally distributed continuous variables

the Mann Whitney test will be used. For categorical variables, a chi-square test will be used. The results of these tests, in addition to variables identified in the literature review, will inform the selection of variables for the logistic regression and artificial neural network models.

3.6 Logistic Regression Model

Logistic regression is a statistical technique in which a linear combination of variables are used to predict a dichotomous outcome. [184,202-205] In contrast to linear regression, which generates a continuous outcome, logistic regression uses a sigmoid function, known as the *logistic function*, to bound its output between 0 and 1.

Logistic regression is based on the calculation of odds, which can be defined as the probability of an event occurring divided by the probability of the event not occurring:

$$odds = \frac{p}{1-p}$$

The logit function is defines the log of the odds ratio in terms of a linear combination of predictors:

$$logit\ (odds) = \beta_0 + \beta_1 X_1 + \ \beta_2 X_2 + \cdots + \ \beta_n X_n$$

where X_1, X_2, …. X_n are independent variables in the dataset,

β_0 is the intercept, and β_1 to β_n are the computed coefficients of each variable and represent the contribution each variable makes to the probability of the outcome. The probability of being a case is defined by the odds ratio:

$$\hat{p} = \frac{odds}{1 + \ odds}$$

Thus, the probability can be derived directly from the predictors of the logistic regression model:

$$\hat{p} = \frac{e^{\beta_0+\beta_1 X_1+ \beta_2 X_2+\cdots+ \beta_n X_n}}{1 + e^{\beta_0+\beta_1 X_1+ \beta_2 X_2+\cdots+ \beta_n X_n}}$$

This analysis will model the probability of MRSA infection in patients undergoing an MSP. The relationship between odds ratios and probabilities is also useful in determining the relative contribution of each predictor to the outcome. The statistical significance of each predictor can be assessed to determine which predictors should be included in the final model. The coefficient of each predictor represents the contribution each predictor makes to the outcome while holding all the other predictors constant. Thus, a logistic regression model with multiple predictors is an effective approach to controlling confounding.

There are multiple methods for determining which independent variables should be included in the model. With forward stepwise selection, all possible single predictor models are generated and the predictor with the lowest p value is kept. In each subsequent step, another predictor is added. If the predictor's *p* value is below a certain threshold value, it is kept in the model, otherwise it is dropped. The process continues until all covariates with p values below the threshold are included in the model. During the process, it is possible for a covariate's p-value to change, and in this case, the covariate would be dropped from the model. In contrast, backwards stepwise selection beings with a model that contains all the predictors, and eliminates one by one those that do not meet the p-value threshold. In addition, both forwards and backwards selection methods can be used with global model scoring metrics, such as Akaike Information

Criterion (AIC), which has been shown to be particularly effective in reducing the likelihood of overfitting.[206]

Although efficient, care needs to be taken with forward and backward selection methods due to the use of multiple statistical tests (multiple comparisons) and the possibility of type 1 errors. Thus, it is important that selection of variables for the model be principled, i.e. care must be taken to ensure the model makes sense from a biomedical perspective. In addition, it is important to evaluate the performance of the model use a validation data set and a range of objective measures. Model evaluation is discussed in detail below.

3.7 Artificial Neural Network Model

An artificial neural network (ANN) is biologically-inspired predictive modeling technique. Structurally, the model comprises a series of neurons that are connected together to form a network. Because the ANN can support multiple layers, typically an input layer, followed by one or more hidden layers, and an output layer, and because the number of neurons in each layer is arbitrary, the ANN is an effective technique for modeling non-linear relationships between the input variables and the output. In some respects, ANN can be seen as a generalization of regression models, and in fact, ANNs are built from individual binary classifiers, known as *perceptrons*.

A perceptron is a binary linear classifier that accepts a vector of values as input and learns a set of weights for each input that in combination map a set of input values to an output class. The goal of the training process is to learn the weight vector that maximizes performance of the classifier. Given an input vector, x containing d values,

and a weight vector, *w* also containing *d* values, the output, *y*, of a perceptron is defined as:

$$y = \sum_{j=1}^{d} w_j x_j + w_0$$

Where w_0 is known as the bias unit. Thus, the perceptron is an example of a linear model, in which the input space is divided into two classes based on identification of the optimal hyperplane. Although perceptrons can, as a result, only model linear relationships, they can be combined together to form a *multilayer perceptron*, which facilitates the more complex modeling of non-linear relationships.

The classic ANN is best described as a feed-forward multi-layer perceptron. In this type of network, there is an input layer that contains one neuron for each of the input values, one or more hidden layers each containing and arbitrary number of neurons, and in the case of a binary classifier, two output neurons. The network is known as a feed-forward network because the weights for each connection are fed forward into the next layer, but not backwards. Figure 4 is an example of a feed-forward multi-layer perceptron used for binary classification. In this example, there are n input variables, 1 hidden layer with m neurons, and an output layer with 2 neurons. The ANN is also able to model non-linear relationships between the input variables and the output classes.

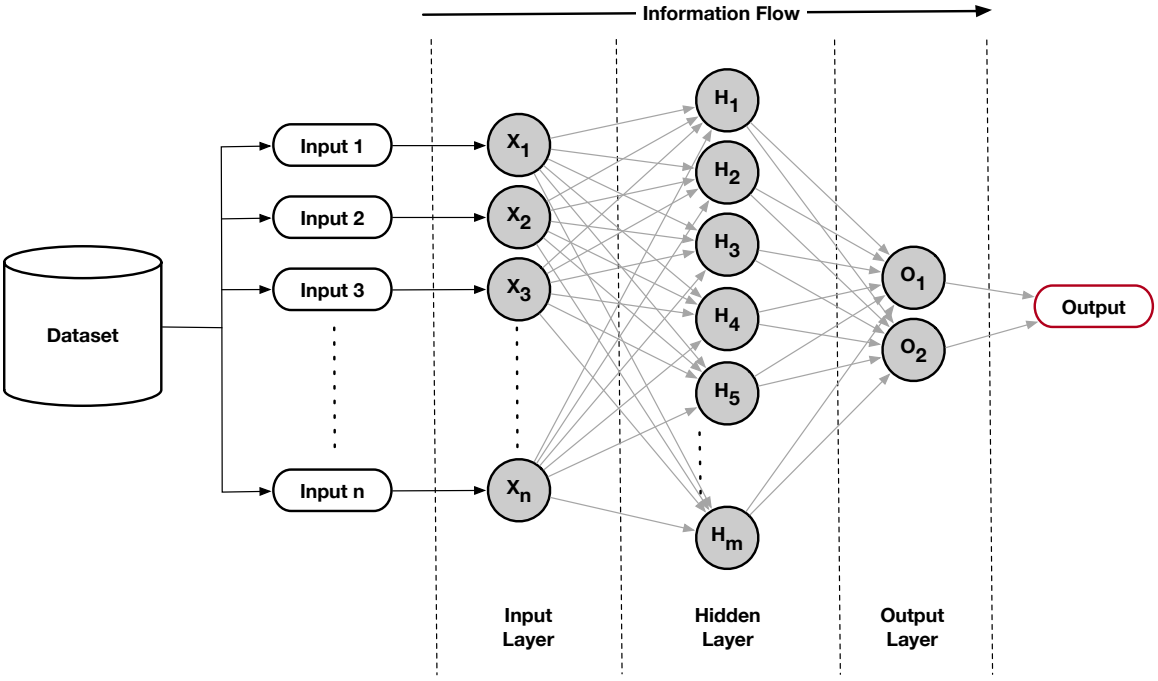

Figure 4: Artificial Neural Network Example

There is no principled way to determine the optimal number of neurons within the hidden layer, thus several models will be developed using cross-validation, and the final model will be validated using a held-out test data set. The algorithm for determining the optimal number of neurons in the hidden layer is illustrated in Figure 5.

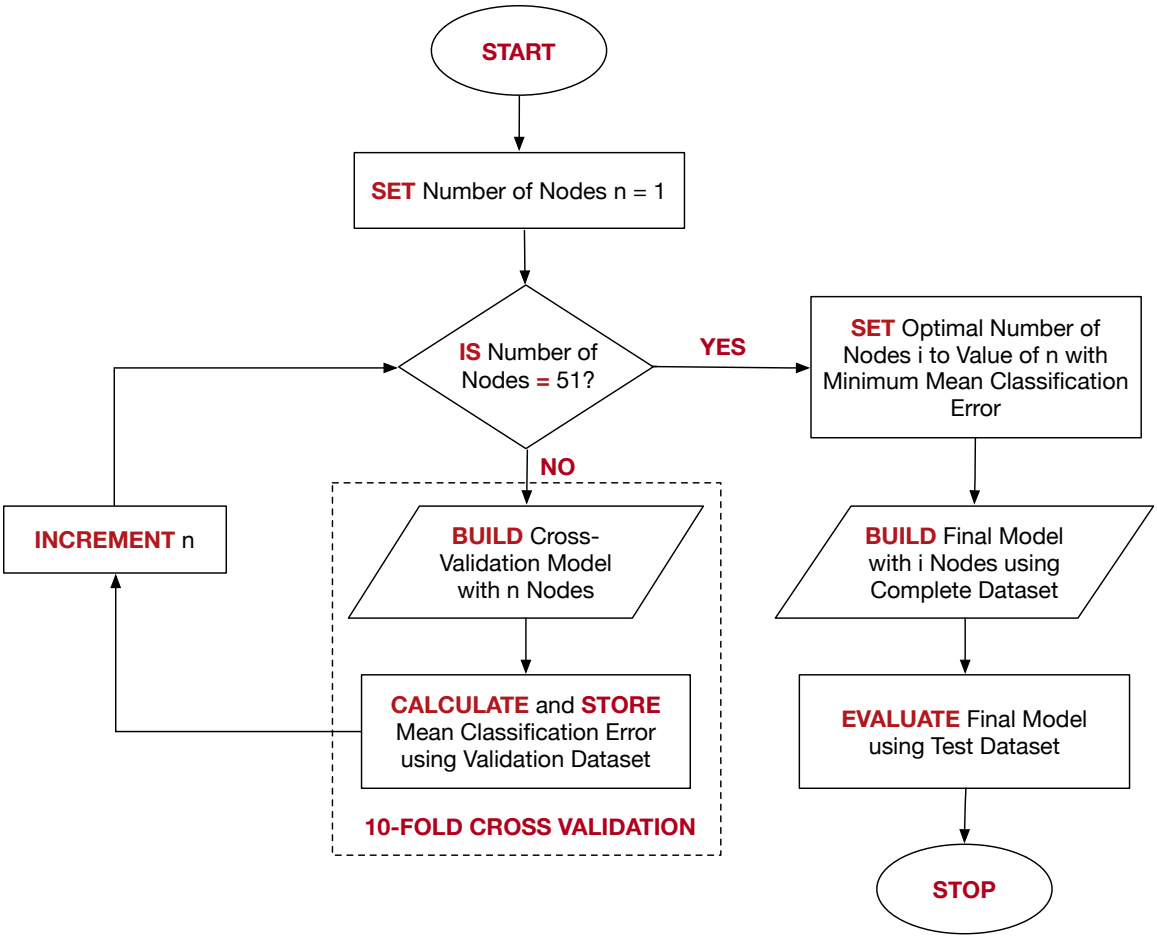

Figure 5: Algorithm for Determining the Optimal Number of Neurons in the Hidden Layer of the ANN

The algorithm iterates through every possible number of nodes, beginning with 1 and ending with 50. During each iteration, a model is trained and evaluated using the cross-validation method illustrated in Figure 6. After each sequence of cross-validation training, an error rate is calculated, and the configuration with the minimum error rate is used for the final mode. Once the final model configuration has been determined, the model is retrained with all of the training data and its performance is assessed using the held-out test set.

For prediction of a binary outcome, several different neuron activation functions could be used. In this case, a *sigmoid* function will be used:

$$y(k) = S\left(\sum_{j=1}^{d} w_i(k) \cdot x_i(k) + w_0\right)$$

where the *sigmoid* function is:

$$S(t) = \frac{1}{1 + e^t}$$

The configuration of the ANN will differ for each hypothesis, but in all cases will contain only one hidden layer, as ANNs with only one layer have been shown to be sufficient in modeling non-linear relationships.[184]

3.8 Model Training

Data were preprocessed to create an analysis dataset that was used to train and evaluate the neural network. In order to ensure that the algorithms could be trained effectively, controls were under sampled due to the large difference between the numbers of cases and controls.[207,208] The neural network was configured with 1 input layer, 1 hidden layer, and 1 output layer. The input layer contained one node for each variable in the dataset, and the output layer contained two nodes, one for the Case variable and one for the Control variable. The number of hidden layers and the number of nodes within each layer was determined using cross-validation, as described in Figure 5.

The neural network was trained using 10-fold cross-validation. Using this approach, the dataset was randomly split into 10 equally sized parts and the network was trained with 9 of these parts and evaluated with the 10th. This process was repeated 10 times ensuring that all the data were used for training and evaluation without biasing the results. The output class (case or control) was determined using the higher value of the

two variables in the output layer. All analyses were performed using R version 3.1.3. The *Caret* package was used to assess the performance of the neural network, *ROCR* was used to generate the ROC curve, and *NeuralNet* was used to train the neural network.[209-211] The cross-validation process is illustrated in Figure 6.

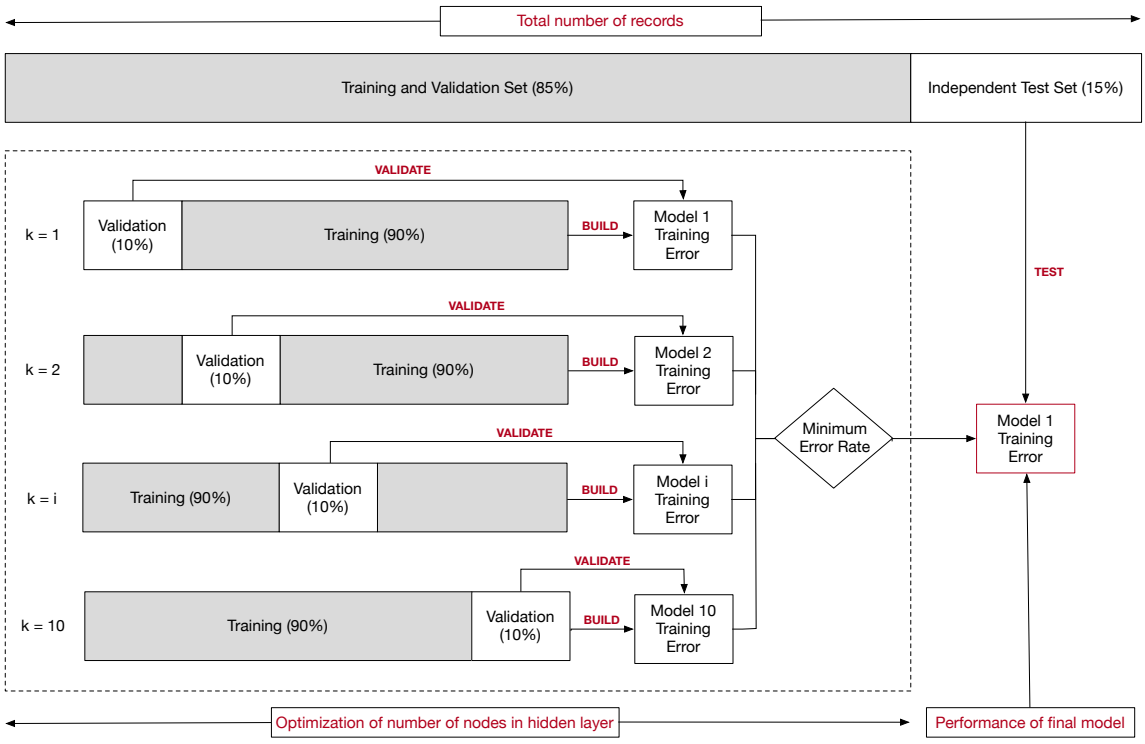

Figure 6: Cross-validation of the ANN Models

3.9 Model Evaluation

The performance of the neural network was assessed with a range of metrics including sensitivity, specificity, positive predictive value, and negative predictive value.

The aim of a binary classifier is to predict a binary outcome with 100% accuracy, however, it is usually not possible for a test to be 100% accurate, and thus there are multiple possible outcomes:

- A *true positive* occurs when the classifier correctly identifies a person who has the outcome (e.g., MRSA-related SSI).
- A *false positive* occurs when the test identifies a person as having the disease but they do not.
- A *true negative* occurs when a test identifies a person as not having the disease and they do not have the outcome.
- A *false negative* occurs when the test identifies the person has not having the disease but they do have the outcome.

An effective test will maximize true positives and true negatives and minimize false positives and false negatives.

Both logistic regression and ANN models are algorithms that aim to assign patients to one of two classes (in this case diseased or not diseased) and they are typically assessed using a range of metrics, including *sensitivity* and *specificity*. Sensitivity is the *true positive rate (TPR)*: the proportion of people with the disease that are correctly classified as having the disease. Specificity is the *true negative rate (TNR)*: the proportion of people without the disease that are correctly identified as not having the disease. The *false positive rate (FPR)* is calculated as 1 – specificity and the sum of the true positive rate and the false positive rate is 1. Thus, a classifier with a higher true positive rate will necessarily have a lower false positive rate and vice versa. These rates are controlled by a threshold value, above which it is assumed the test result is *positive* and below which it is assumed the classifier result is *negative*. This value can be adjusted to maximize the true positive rate and minimize the false positive rate.

Additional metrics include positive predictive value (PPV) and negative predictive value (NPV). These metrics are useful in helping the clinician and patient understand the implications of a positive or negative test result. PPV is the probability that the patient has the disease, given a positive test result and NPV is the probability that the patient does not have the disease, given a negative test results. Both of these metrics are affected by the underlying prevalence of the disease in the population which, in essence, is the pre-test probability when viewed within a Bayesian framework. In contrast to sensitivity and specificity, which are characteristics of the test (or model), PPV and NPV are related to the patient.[212] Possible outcomes of a predictive model are illustrated in Table 7.

Table 7: Possible Outcomes of a Predictive Model

		Actual	
		Positive	**Negative**
Predicted	**Positive**	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Using these definitions, it is possible to derive sensitivity, specificity, PPV, and NPV:

$$Sensitivit = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

$$PPV = \frac{TP}{TP + FP}$$

$$NPV = \frac{TN}{TN + FN}$$

PPV and NPV can also be defined in terms of sensitivity, specificity and prevalence of the disease in the population:

$$PPV = \frac{Sensitivity \times Prevalence}{Sensitivity \times Prevalance + (1 - Specificity) \times (1 - Prevalence)}$$

$$NPV = \frac{Specificity \times (1 - Prevalance)}{(1 - Sensitivity) \times Prevalance + Specificity \times (1 - Prevalence)}$$

Performance was also visualized using the receiver operating characteristic (ROC) curve method. A ROC curve is a plot of the true positive rate on the y-axis against the false positive rate on the x-axis. An example of a ROC curve for a neural network classifier of pediatric brain injury is shown below.

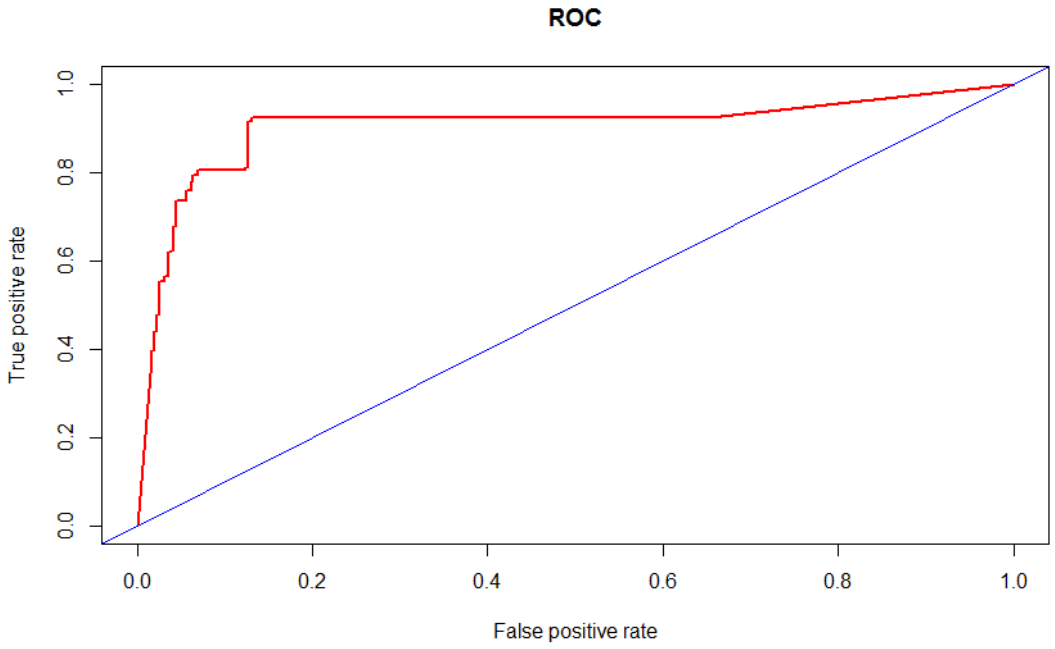

Figure 7: Receiver Operating Characteristic (ROC) Curve Example

The red line on the ROC curve is a plot of the true positive rate vs false positive rate for each value of the classifier threshold. The ROC curve can be used to identify the optimal threshold value, i.e. the value to use as a cutoff to assign a positive and negative result such that the TPR is maximized and the FPR is minimized.

The area under the curve (AUC) in a ROC curve is a measure of a model's ability to correctly discriminate between patients with and without a disease. The better the model's performance, the higher the curve will be above the diagonal line, and thus the greater the area under the curve. A perfect model would have an AUC of 1, indicating a 100% TPR and a 0% FPR. The blue line represents an equal TPR and FPR, which results in an AUC of 0.5 and signifies that the model classifies people with the disease correctly 50% of the time. This means that the performance of the model is no better than would be obtained by tossing a coin. Thus, this line effectively represents a model with no discriminatory power. In this example, the AUC is approximately 0.85 and the optimum sensitivity (TPR) is 0.7 meaning the model correct identifies patients with the disease 70% of the time, which is significantly higher than chance.

As discussed above, changing the threshold used to determine a positive result can impact the sensitivity and specificity of the model. In addition to their use in optimizing the threshold, these metrics also have significant clinical importance. A model with higher sensitivity will minimize the number of false negatives while increasing the number of false positives. This is advantageous when the disease is a serious one for which treatment is available: it is better to err on the side of identifying "too many" people who don't have the disease rather than missing those who do have it. In contrast,

a model with higher specificity would have the reverse effect – we would increase the number of false negatives and decrease the number of false positives. This is advantageous when the disease is not serious and treatment is expensive. Thus, sensitivity and specificity can be used to optimize the model in the clinical setting based on the seriousness of the disease and the cost of treatment.

3.10 Clinical Decision Support System

The clinical decision support system will be based on the logistic regression and artificial neural network models described above. Because the underlying models will be developed using R, the CDSS will be developed in R Shiny – a web application platform based on the R language.[201] In this way, the CDSS can access the information contained within the models. The application will be developed as an online application and users will be able to access it through a web browser on a laptop, tablet or phone. The application will employ responsive design principles in that it will adapt and present a user-friendly interface regardless of the device being uses. Users will be prompted to enter the key predictors needed to inform the model and a recommendation will be generated in real time.

CHAPTER 4

RESULTS

4.1 Introduction

This chapter presents the results of the three hypotheses defined in Chapter 3 along with a descriptive analysis of the NIS dataset. Five years of data (2010-2014) are used for the descriptive analyses, with 15% of data held out as a testing dataset for the artificial neural network and logistic regression models.

The chapter begins with an overview of the demographic characteristics of the NIS dataset and characterizes patients undergoing surgical procedures in the United States between 2010 and 2014. Frequencies of outcome variables required to support the study hypotheses are described, including SSI, confirmed MRSA infection, confirmed MSSA infection, MRSA-related SSI, and MSSA-related SSI. Additional characteristics described include mean age, length of stay, number of chronic conditions, number of diagnoses, number of procedures performed, number of days to first procedure, number of comorbidities, gender, primary expected payer, race, surgery type, hospital region, hospital bed size, and hospital teaching status.

The chapter continues with one section for each hypothesis, with each section containing a detailed descriptive analysis, logistic regression model, artificial neural network model, model evaluation, and discussion. The descriptive analysis, which itself is informed by the literature review, is used to further refine the set of predictors used in

the models. Finally, the chapter concludes with a description of the design and implementation of the clinical decision support system, along with a discussion of its potential application in clinical practice.

4.2 Demographic Characteristics of NIS Data

This section describes the characteristics of the NIS dataset for patients that have undergone a major surgical procedure in the years 2010 to 2014. The total number of admissions meeting these criteria is 10,751,538. Although this is a significant number, the proportion of patients meeting the addition criteria of MRSA infection, SSI, MSSA infection, and MRSA/MSSA-related infections, is significantly smaller. These proportion of patients meeting these criteria is summarized in Table 8.

Table 8: Basic Characteristics of the NIS Data (Outcome Variables)

Variable	Proportion (%)	Frequency	% Missing
Confirmed MRSA infection	1.07	115,029	0.00
Surgical Site Infection	1.05	113,402	0.00
Confirmed MSSA infection	0.66	71,099	0.00
MRSA-related SSI	0.10	11,057	0.00
MSSA-related SSI	0.08	9,032	0.00

Of the 10,751,538 admissions sampled in years 2010 to 2014, 115,029 or 1.07% had confirmed MRSA infection as defined by ICD-9 codes 041.12, 038.12 and 482.42; 113,402, or 1.05%, had a confirmed SSI as defined by ICD-9 code 998.59; and 71,099, or 0.66%, had a confirmed MSSA infection as defined by ICD-9 code 041.11. The proportion of admissions with both MRSA and SSI was much smaller at 11,057, or 0.10%; and the proportion of admissions with both MSSA and SSI was similar at 9,032,

or 0.08%. Because these categories are assigned based on ICD-9 codes from the diagnosis fields, there are no missing data.

Characteristics of patients undergoing major surgical procedures in the United States between the years 2010 to 2014 are described in Table 9 (continuous variables) and Table 10 (categorical variables).

Table 9: Basic Characteristics of the NIS Data (Continuous Variables)

Variable	Mean (Standard Deviation)	N	% Missing
Age in years at admission (years)	48.20 (25.29)	10,744,926	0.06
Length of stay (days)	5.07 (8.04)	10,751,278	0.00
Number of chronic conditions	3.76 (3.38)	10,751,538	0.00
Number of diagnoses	8.23 (5.76)	10,751,538	0.00
Number of procedures	3.07 (2.58)	10,751,438	0.00
Number of days to first procedure	1.16 (3.63)	10,152,388	5.57
Number of comorbidities	1.74 (1.83)	10,751,538	0.00

Table 10: Basic Characteristics of the NIS Data (Categorical Variables)

Variable	Proportion (%)	Frequency	% Missing
Gender			
Male	46.61	5,004,131	0.13
Female	53.39	5,732,628	
Primary expected payer			
Medicare	33.98	3,645,555	0.22
Medicaid	17.37	1,863,893	
Private insurance	40.42	4,336,556	
Self-pay	3.95	423,368	
No charge	0.40	42,532	
Other	3.88	415,984	
Race			
White	69.82	6,869,218	8.50
Black	12.59	1,238,673	
Hispanic	10.89	1,710,074	
Asian/Pacific Islander	2.59	254,906	
Native American	0.61	59,964	
Other	3.50	344,298	
Surgery type			
Non-elective	52.36	5,610,245	0.34
Elective	47.64	5,104,382	
Hospital region			
Northeast	18.36	1,973,788	0.00
Midwest	23.21	2,495,201	
South	38.73	4,163,933	
West	19.71	2,118,616	
Hospital bed size			
Small	13.12	1,402,358	0.60
Medium	25.03	2,674,446	
Large	61.85	6,610,127	
Hospital teaching status			

Rural	8.65	924,536	0.60
Urban non-teaching	35.00	3,739,944	
Urban teaching	56.35	6,022,451	

The average age of patients at admission was 48.20 years (SD 23.29), with the average length of stay being 5.07 days (SD 8.04). Patients, on average, had 3.76 chronic conditions (SD 3.38), 8.23 diagnoses (SD 5.76), and underwent 3.07 procedures (2.58). The number of days from admission until the first procedure was performed was 1.16 (SD 3.63) and the average number of comorbidities was 1.74 (SD 1.83). The proportion of missing data is low for all variables, with the highest amount of missing data 5.57% for number of days to first procedure.

Approximately 46.51% of the sampled admissions were male and 53.30% were female. With respect to race, the majority of patients were white (69.82%), and 12.29% were black or African American, 10.89% were Hispanic, 2.59% were Asian/Pacific islander, and 3.88% were another race. The primary expected payer was Medicare in 33.98% of cases, suggesting that surgical patients tend to be older, and of the remaining patients 17.37% were paid by Medicaid, 40.42% by private insurance, 3.92% by the patient themselves, and 3.88% by other means. In a very small percentage (0.40%) of admissions, no charge was levied for the surgical procedure. The majority of surgeries were non-elective (52.36%) and 47.64% were elective.

The majority of hospitals, representing 38.73% of admissions were in the South, with 18.36% in the Northeast, 23.21% in the Midwest, and 19.71% in the West. Approximately 25.03% of hospitals were of medium size, 13.12% were classified as small, and 61.85% were large. Over half of the hospitals were teaching hospitals located

in large urban centers (56.35%), 35% were non-teaching urban hospitals, and 8.65% were rurally located.

The proportion of missing data is low for most variables, with the exception of Race, for which approximately 8.5% is missing.

4.3 National Prevalence Estimates

Because the NIS data is collected using a nationally-representative survey sampling scheme, national level estimates of prevalence can be extrapolated from the above sample-based estimates using SAS survey procedures. Table 11 summarizes 5-year and average annual prevalence for the key outcome variables being studied.

Table 11: National-level 5-year Prevalence Estimates

Variable	Prevalence	Average Annual Prevalence
Surgical procedure	53,302,100	10,660,420
Confirmed MRSA infection	570,907	114,181
Surgical Site Infection	562,670	112,534
Confirmed MSSA infection	353,901	70,780
MRSA-related SSI	54,913	10,983
MSSA-related SSI	44,811	8,962

Between 2010 and 2014, nationally 53,302,100 underwent a major surgical procedures, of which 570,907 had a confirmed MRSA infection, 562,670 had a SSI, 353,901 had a confirmed MSSA infection, 54,913 had MRSA-related SSI and 44,811 had an MSSA-related infection. Average annual prevalence of surgical procedures was 10,660,420, of which 114,181 experienced a MRSA infection, 112,534 had a SSI, 70,780 had a confirmed MSSA infection, 10,983 had a MRSA-related SSI, and 8,962 had MSSA-related SSI.

These results emphasis the scope and scale of the problem and support the need for models to predict MRSA, MSSA and SSI.

4.4 Hypothesis 1 — Prediction of MRSA Infection in Surgical Patients

The first hypothesis aims to predict the incidence of MRSA infection in surgical patients using a range of known demographic, clinical, and hospital-related variables. To facilitate and initial selection of variables for incorporation in the models, a descriptive analysis was performed. Descriptive statistics and p-values were calculated for MRSA and non-MRSA patients. Case and control definitions are illustrated in Figure 1.

For continuous variables, the Wilcoxon Rank Sum test was used. The Wilcoxon Rank Sum test is a non-parametric hypothesis test that is appropriate for both normally-distributed and non-normally-distributed data. The test compares the mean of data between the two groups and provides p-values that quantify the significance of the difference in means between the group. In all cases, a cut-off of $p < 0.05$ was used to assess statistical significance.

For categorical variables, the Chi-Square test of association was used to ascertain whether the difference in proportion in levels of a variable between MRSA and non-MRSA patients are statistically significant. The chi-square test requires the two variables to be categorical, the groups defined by the variables to be independent, and the sample size to be large (i.e., 80 percent of expected frequencies must be greater than 5 and all expected frequencies must be greater than 1).

Differences between the two groups are described using means (and standard deviations) for continuous variables and proportions (percentages) for categorical variables. Key demographic variables, for MRSA patients, are presented visually.

4.4.1 Descriptive Analysis

Table 12 summarizes the means, standard deviations, and p-values for continuous demographic variables. The null hypothesis in each case is that there is no difference in means between MRSA and non-MRSA patients. Variables analyzed were age in years at admission, length of stay in days, number of chronic conditions, number of diagnoses, number of procedures performed, number of days after admission until the first procedure was performed, and number of comorbidities.

Table 12: Demographic characteristics of NIS data stratified by MRSA status (continuous variables)

Variable	MRSA – Mean (SD)	No MRSA – Mean (SD)	P-Value
Age in years at admission (years)	56.59 (20.14)	48.11 (25.63)	< .0001
Length of stay (days)	13.59 (17.16)	4.98 (7.83)	< .0001
Number of chronic conditions	6.34 (3.74)	3.73 (3.36)	< .0001
Number of diagnoses	14.93 (6.84)	8.16 (5.70)	< .0001
Number of procedures	4.49 (3.62)	3.06 (2.56)	< .0001
Number of days to first procedure	3.89 (7.51)	1.13 (3.56)	< .0001
Number of comorbidities	3.44 (2.21)	1.72 (1.82)	< .0001

Patients with MRSA infection were generally older than those without MRSA, with a mean age of 56.59 (SD 20.14) compared to 48.11 (SD 25.63). Figure 8 illustrates the age distribution for MRSA and non-MRSA patients. Each bar represents the proportion of the given age group that has a MRSA infection. Thus, it can be seen that older patients tend to have a higher incidence of MRSA.

The average length of stay for MRSA patients was 13.59 days (SD 17.16) and 4.98 days (SD 7.83) for non-MRSA patients. Similarly, MRSA patients had a higher

number of chronic conditions, 6.34 (SD 3.74) compared with 3.73 (3.36); a larger number of diagnoses, 14.93 (SD 6.84) compared to 8.16 (SD 5.70); a larger number of procedures, 4.49 (SD 6.84) compared to 2.06 (SD 2.56); a longer time from admission until performance of the first procedure, 2.89 days (SD 7.51) compared to 1.13 (SD 3.56); and, a higher number of comorbidities, 3.44 (SD 2.21) compared to 1.72 (SD 1.82). In all cases, the differences between MRSA and non-MRSA patients were statistically significant, suggesting that these variables could be useful predictors of MRSA infection.

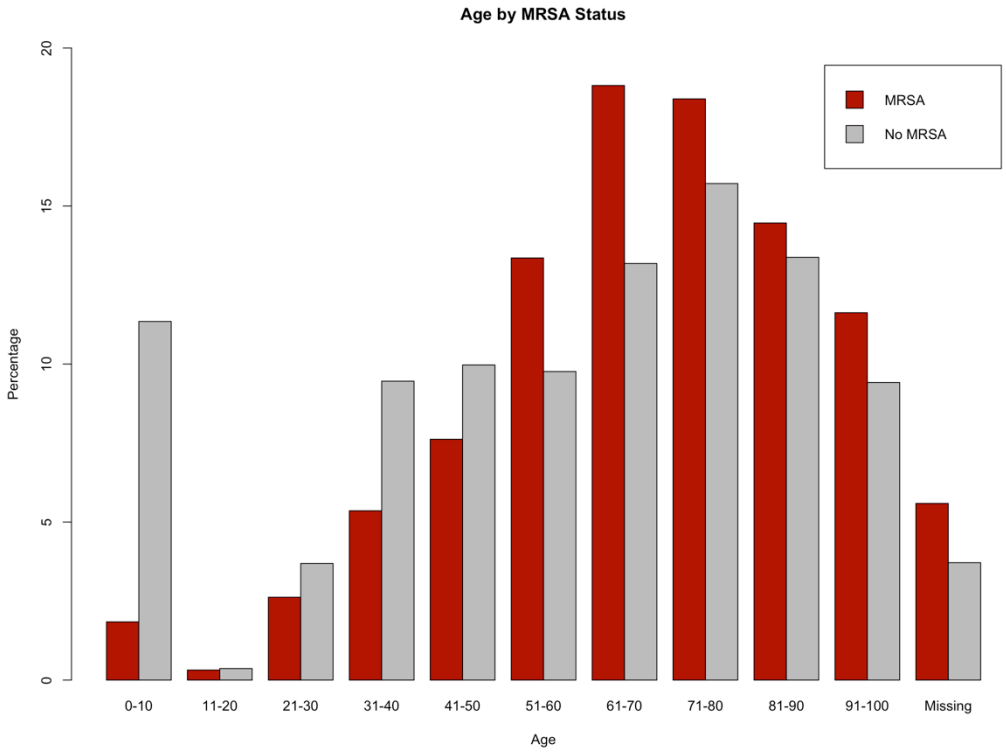

Figure 8: Age Distribution of MRSA Patients Compared with non-MRSA Patients

Categorical variables analyzed were Gender, Primary Expected Payer, Race, Median income of patient zip code (quartile), Surgery Type, Hospital Region, Hospital

Bed Size, and Hospital Teaching Status. Frequencies and percentages of levels of these variables are presented in Table 13 for MRSA and non-MRSA patients.

Table 13: Demographic Characteristics of the NIS Data Stratified by MRSA Status (Categorical Variables)

Variable	MRSA Frequency (%)	No MRSA Frequency (%)	P-Value
Gender			
Male	65,560 (57.01)	4,938,571 (46.49)	< .0001
Female	49,446 (42.99)	5,683,182 (53.51)	
Primary expected payer			
Medicare	57,502 (50.11)	3,588,053 (33.81)	< .0001
Medicaid	19,482 (16.98)	1,844,411 (17.38)	
Private insurance	25,297 (22.05)	4,311,259 (40.62)	
Self-pay	7,408 (6.46)	415,960 (3.92)	
No charge	895 (0.78)	41,637 (0.39)	
Other	41,460 (3.63)	411,824 (3.88)	
Race			
White	75,101 (69.95)	6,794,117 (69.82)	< .0001
Black	17,620 (16.41)	1,221,052 (12.55)	
Hispanic	9,752 (9.08)	1,061,322 (10.91)	
Asian/Pacific Islander	1,350 (1.26)	253,556 (2.61)	
Native American	981 (0.91)	58,983 (0.61)	
Other	2,554 (2.38)	341,744 (3.51)	
Median income of national quartile or patient ZIP code			
Quartile 1	38,769 (34.53)	2,746,107 (26.33)	< .0001
Quartile 2	29,528 (26.30)	2,670,338 (25.61)	
Quartile 3	25,483 (22.70)	2,630,590 (25,23)	
Quartile 4	18,502 (16.48)	2,380,564 (22.83)	
Surgery type			
Non-elective	90,676 (79.09)	5,519,569 (52.07)	< .0001
Elective	23,968 (20.91)	5,080,414 (47.93)	
Hospital region			
Northeast	18,095 (15.73)	1,955,693 (18.39)	< .0001
Midwest	24,416 (21.23)	2,470,785 (23.23)	
South	52,002 (45.21)	4,111,931 (38.66)	
West	20,516 (17.84)	2,098,100 (19.73)	
Hospital bed size			
Small	14,316 (12.52)	1,388,042 (13.13)	< .0001
Medium	28.313 (24.77)	2,646,133 (25.03)	
Large	71,688 (62.71)	6.528,439 (61.84)	
Hospital teaching status			
Rural	10,411 (9.11)	914,125 (8.65)	< .0001
Urban non-teaching	41,397 (36.21)	3,698,547 (34.98)	
Urban teaching	62,509 (54.68)	5,959,942 (56.37)	

As with the continuous variables, there were statistically significant differences between MRSA and non-MRSA patients. Figure 9 (below) illustrates the breakdown of gender for MRSA patients. A greater proportion of MRSA patients were male, 57.01%

compared to 46.49% of non-MRSA patients, and, as expected for an older population, a greater proportion utilized Medicare, 50.11% compared to 33.81%.

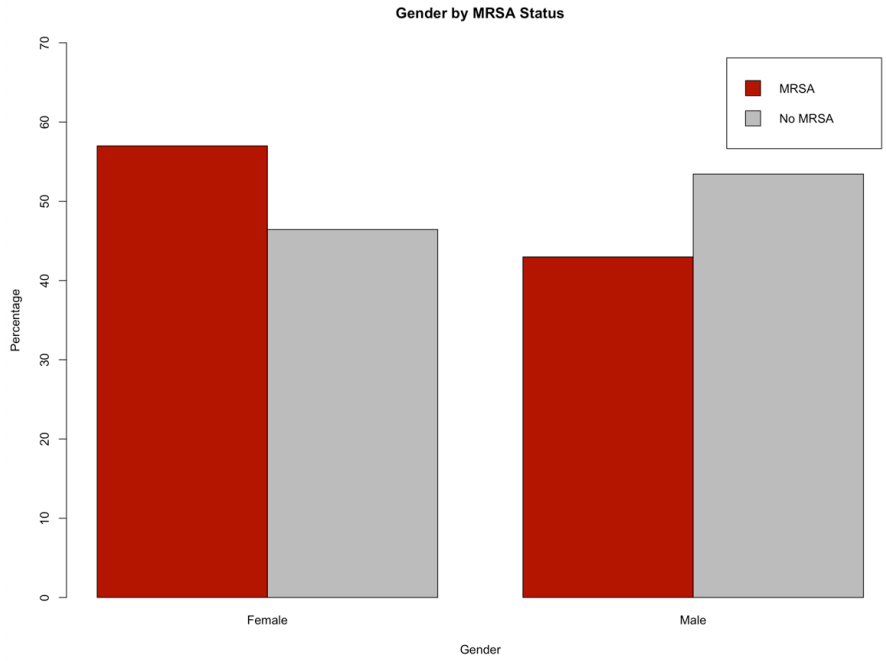

Figure 9: Gender of MRSA Patients Compared with non-MRSA Patients.

The racial breakdown across both groups is broadly similar, however MRSA patients are more likely to be black than non-MRSA patients, 16.1% compared to 12.55%. While there is a similar proportion of white patients across both groups, MRSA patients are also less likely to be Asian/Pacific Islander and Hispanic. Figure 10 summarizes the distribution of race for MRSA patients.

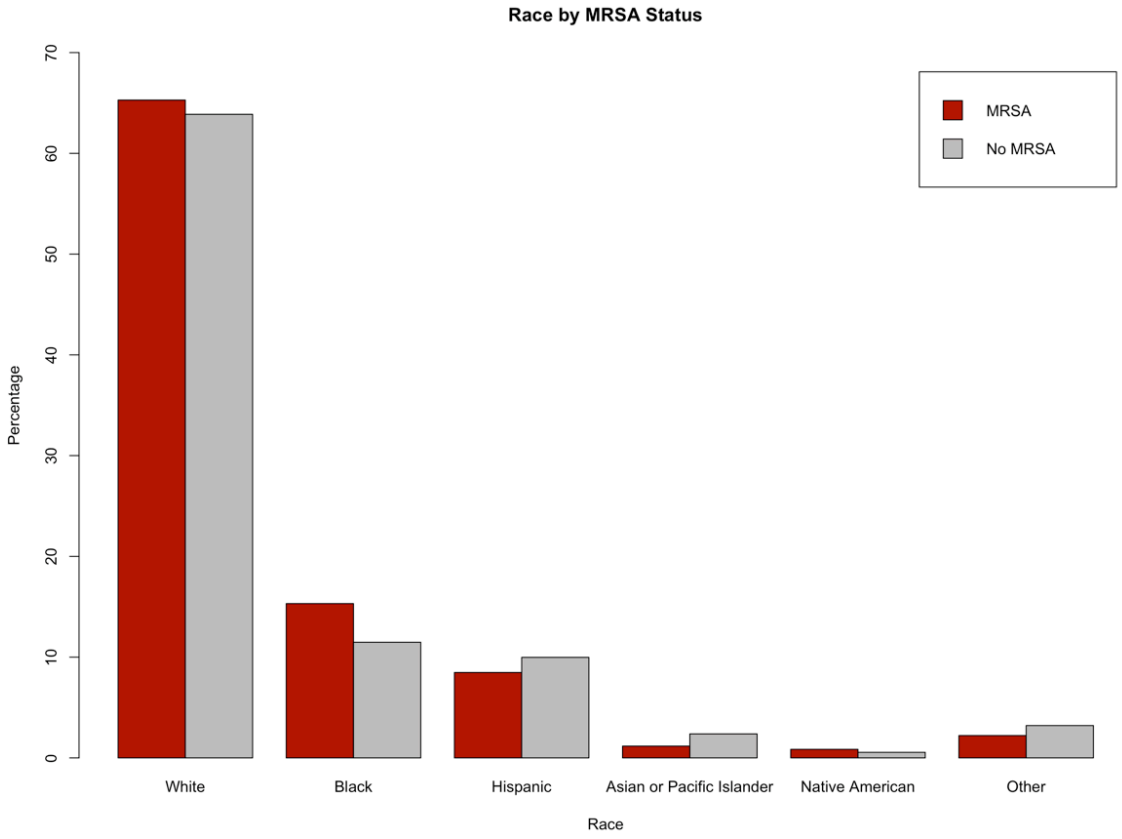

Figure 10: Distribution of Race for MRSA Patients

MRSA patients were more likely to be poorer than non-MRSA patients: 34.35% of MRSA patients were in the lowest quartile of income for their zip code compared with 26.33% of non-MRSA patients. The proportion of patients whose income was in the third or fourth quartile for their zip code was commensurately lower: 22.70% of MRSA-patients had income in the third quartile compared with 25.23% of non-MRSA patients, and 16.48% of MRSA-patients had income in the fourth quartile compared with 22.83% of non-MRSA patients.

A number of hospital and surgery related factors were also compared. For MRSA patients, 79.09% of surgeries were non-elective compared to 52.07% for non-MRSA patients. MRSA was more prevalent in hospitals in the South: 45.21% of MRSA patients

were treated in hospitals in the South compared with 28.66% of non-MRSA patients. Proportions of MRSA and non-MRSA patients were broadly similar across different hospital sizes, however despite these similarities, the differences remain statistically significant. Finally, a greater proportion of MRSA patients, 9.11%, were more likely to be treated in rural hospitals, compared with 8.65% of non-MRSA patients.

As with the continuous variables described above, these categorical factors all exhibit statistically significant differences between MRSA and non-MRSA patients, and thus, should be considered for inclusion into the logistic regression and artificial neural network models.

Finally, a number of additional risk factors were considered, including long-term antibiotic use, risk of mortality (APRDRG calculated score), severity (APRDRG calculated score), a broad range of comorbidities, history of MRSA infection, and transfer in or out of the hospital. These risk factors are described in Table 14.

Table 14: Additional Risk Factors Stratified by MRSA Status

Variable	MRSA (%)	No MRSA (%)	P-Value
Long-term (current antibiotic use)			
Yes	670 (0.58)	10,755 (0.10)	< .0001
No	114,359 (99.42)	10,625,754 (99.90)	
Risk of mortality (APRDRG)			
No class specified	15 (0.00)	6,379 (0.06)	< .0001
Minor likelihood of dying	34,271 (29.79)	7,351,856 (69.12)	
Moderate likelihood of dying	31,021 (26.97)	1,830,972 (17.21)	
Major likelihood of dying	28,657 (24.91)	961,362 (9.04)	
Extreme likelihood of dying	21,065 (18.31)	485,940 (4.57)	
Severity (APRDRG)			
No class specified	15 (0.01)	79 (0.06)	< .0001
Minor loss of function	10,857 (9.44)	4,611,076 (43.35)	
Moderate loss of function	30,022 (26.10)	3,668,364 (34.49)	
Major loss of function	42,139 (36.63)	1,708,781 (16.07)	
Extreme loss of function	31,997 (16.07)	31,966 (27.82)	
Comorbidity: AIDS			
Yes	500 (0.43)	11,157 (0.10)	< .0001
No	114,529 (99.57)	10,625,352 (99.90)	
Comorbidity: Alcohol			
Yes	4,832 (4.20)	208,610 (1.96)	< .0001
No	110,197 (95.80)	10,427,899 (98.04)	

Comorbidity: Deficiency anemias			
Yes	35,299 (30.69)	1,271,368 (11.95)	< .0001
No	79,730 (69.31)	9,365,141 (88.05)	
Comorbidity: Rheumatoid arthritis/collagen vascular diseases			
Yes	4,334 (3.77)	224,567 (2.11)	
No	110,695 (96.23)	10,411,942 (97.89)	< .0001
Comorbidity: Chronic blood loss anemia			
Yes	1,993 (1.73)	327,616 (3.08)	< .0001
No	113,036 (98.27)	10.308,893 (96.92)	
Comorbidity: Congestive heart failure			
Yes	15,093 (13.12)	389,519 (3.66)	< .0001
No	99,936 (86.88)	10,246,990 (96.34)	
Comorbidity: Chronic pulmonary disease			
Yes	23.806 (20.70)	1,379,809 (12.97)	< .0001
No	91,223 (79.30)	9,256,700 (87.03)	
Comorbidity: Coagulopathy			
Yes	8,077 (7.02)	380,542 (3.58)	< .0001
No	106,952 (92.98)	10,255,967 (96.42)	
Comorbidity: Depression			
Yes	15,069 (13.10)	856,355 (8.05)	< .0001
No	99,960 (86.90)	9,780,154 (91.95)	
Comorbidity: Diabetes (uncomplicated)			
Yes	21,824 (18.97)	1,475,743 (13.87)	< .0001
No	93,205 (81.03)	9,160,766 (86.13)	
Comorbidity: Diabetes with chronic complications	20,835 (18.11)	346,378 (3.26)	< .0001
Yes	94,194 (81.89)	10,290,131 (96.74)	
No			
Comorbidity: Drug abuse			
Yes	7,556 (6.57)	167,735 (1.58)	< .0001
No	107,473 (93.43)	10,468,774 (98.42)	
Comorbidity: Hypertension			
Yes	64,637 (56.19)	4,347,782 (40.88)	< .0001
No	50,392 (43.41)	6,288,727 (59.12)	
Comorbidity: Hypothyroidism			
Yes	12,043 (10.47)	916,788 (8.62)	< .0001
No	102,986 (89.53)	9,719,721 (91.38)	
Comorbidity: Liver disease			
Yes	4,770 (4.15)	184,130 (1.73)	< .0001
No	110,259 (95.85)	10,452,379 (98.27)	
Comorbidity: Lymphoma			
Yes	942 (0.82)	42,213 (0.40)	< .0001
No	114,087 (99.18)	10,594,296 (99.60)	
Comorbidity: Fluid and electrolyte disorders			
Yes	39,342 (34.20)	1,370,310 (12.88)	< .0001
No	75,687 (65.80)	9,266,199 (87.12)	
Comorbidity: Metastatic cancer			
Yes	2,071 (1,80)	217,682 (2.05)	< .0001
No	112,958 (98.20)	10,418,827 (97.95)	
Comorbidity: Other neurological disorders	10,285 (8.94)	410,663 (3.86)	< .0001
Yes	104,744 (91.06)	10,225,846 (96.14)	
No			
Comorbidity: Obesity			
Yes	19,177 (16.67)	1,229,468 (11.56)	< .0001

No	95,852 (83.33)	9,407.041 (88.44)	
Comorbidity: Paralysis			
Yes	7,532 (6.55)	165,218 (1.55)	< .0001
No	107,497 (93.45)	10,471,291 (98.45)	
Comorbidity: Peripheral vascular disorders			
Yes	17,158 (14.92)	561,508 (5.28)	< .0001
No	97,871 (85.08)	10,075,001 (94.72)	
Comorbidity: Psychoses			
Yes	6,817 (5.93)	239,687 (2.25)	< .0001
No	108,212 (94.07)	10,396,822 (97.75)	
Comorbidity: Pulmonary circulation disorders			
Yes	4,881 (4.24)	136,194 (1.28)	< .0001
No	110,148 (95.76)	10,500,315 (98.72)	
Comorbidity: Renal failure			
Yes	24,456 (21.26)	738,224 (6.94)	< .0001
No	90,573 (78.74)	9,898,285 (93.06)	
Comorbidity: Solid tumor without metastasis			
Yes	2,030 (1.76)	136,170 (1.28)	< .0001
No	112,999 (98.24)	10,500,339 (98.72)	
Comorbidity: Peptic ulcer disease excluding bleeding			
Yes	39 (0.03)	2,626 (0.02)	< .0001
No	114,990 (99.97)	10,633,883 (99.98)	
Comorbidity: Valvular disease			
Yes	5,051 (4.39)	249,967 (2.35)	< .0001
No	109,978 (95.61)	10,386,542 (97.65)	
Comorbidity: Weight loss			
Yes	15,679 (13.63)	329,088 (3.09)	< .0001
No	99,350 (86.37)	10,307,421 (96.91)	
History of MRSA infection			
Yes	2,862 (2.49)	44,031 (0.41)	< .0001
No	112,167 (97.51)	10,592,478 (99.59)	
Transfer in to hospital			
Not transferred in	4,896 (4.28)	146,11 (1.38)	< .0001
Transferred in from a different acute care hospital	100,073 (87.44)	10,024,890 (94.66)	
Transferred in from another type of health facility			
Transfer out of hospital			
Not transferred out	41.642 (36.23)	1,464,186 (13.77)	< .0001
Transferred out to a different acute care hospital	70,621 (61.45)	9,086,318 (85.47)	
Transferred out to another type of health facility			
Elixhauser Comorbidity Score (Readmission)	7,970,294.49	5,347,710.89	< .0001
Elixhauser Comorbidity Score (Mortality)	0.68	0.50	< .0001

All the listed comorbidities exhibited statistically-significant differences between MRSA and non-MRSA patients, suggesting that, in general, sicker patients are more likely to contract MRSA infection while in hospital. There may be some overlap or

collinearity between these predictors and it is likely that some of these will be selected out of the final models through the stepwise variable selection process. This hypothesis is supported by the Elixhauser comorbidity scores: both the readmission score and the mortality score are significantly higher for MRSA patients. This is also true for the APRDRG risk of mortality and severity measures. In essence all of these variables attempt to capture the same concept: the severity of illness and comorbidity burden. In the predictive models it will be important to include the best and most efficient measure of severity, and minimize the number of individual comorbidities. Similarly, patients transferring in to the hospital and patients that transferred out of the hospital are also at increased risk of MRSA infection. Both of these variables also address a component of severity of illness, although the transfer out variable cannot be used to predict the infection as, by definition, it occurs at the end of the hospitalization after the infection has been diagnosed. Finally, although previous history of antibiotic use is not included in the dataset, there is an ICD-9 code for long-term, current antibiotic use, which may be an appropriate proxy. Patients with long-term antibiotic use are over 5 times more likely to contract a MRSA infection, suggesting that this variable may also be a useful predictor.

The above descriptive analysis and comparisons between MRSA and non-MRSA patients clearly indicate that there are significant differences across these groups of patients with respect to a range of demographic, clinical and hospital related variables. The goal of the models presented in the following sections is to develop the best performing model using the most parsimonious set of predictors that are known to the clinician close to the time of admission. Thus, variables that are not available until later

in the hospital visit, or those that require complex calculations based on information from the medical record will be excluded.

4.4.2 Logistic Regression Model

As described in Chapter 3, a logistic regression model was developed to predict MRSA infection. Initially, all variables that were univariately significant according to Table 14 were included in the model and forwards-backwards stepwise variable selection was used to identify the optimal model. The model was optimized by minimizing Akaike Information Criterion (AIC) – a global comparative metric of model performance. The final logistic regression model for hypothesis 1 is shown in Figure 11.

```
MRSA ~ AGE + ELECTIVE + FEMALE + LOS + NCHRONIC + NDX + NPR +
   PAY1 + RACE + TRAN_IN + ZIPINC_QRTL + MRSAHX + ANTIBX + SUSMRSA +
   HOSP_BEDSIZE + HOSP_LOCTEACH + HOSP_REGION + APRDRG_Risk_Mortality +
   APRDRG_Severity + CM_AIDS + CM_ALCOHOL + CM_ANEMDEF + CM_ARTH +
   CM_BLDLOSS + CM_CHF + CM_CHRNLUNG + CM_COAG + CM_DEPRESS +
   CM_DM + CM_DMCX + CM_DRUG + CM_HTN_C + CM_LIVER + CM_LYMPH +
   CM_LYTES + CM_METS + CM_NEURO + CM_OBESE + CM_PARA + CM_PERIVASC +  CM_PSYCH
   CM_PULMCIRC + CM_RENLFAIL + CM_TUMOR + CM_VALVE +
   CM_WGHTLOSS
```

Figure 11: Logistic Regression Model Specification for Hypothesis 1

The final model includes a substantial number of predictors and achieves very strong predictive performance. The relative contribution of each predictor, while adjusting for all others, is presented below in Table 15. Odds ratios quantify the effect of each predictor on the outcome, 95% confidence intervals provide an estimate of variability, and p-values indicate statistical significance, with $p = 0.05$ used as the threshold value for significance.

Table 15: Odds ratios, Confidence Intervals, and p-values for Logistic Regression Model (Hypothesis 1)

Predictor	OR	2.5% CI	97.5% CI	P-Value
(Intercept)	0.08	0.04	0.17	< .0001
Age	1.01	1.01	1.01	< .0001
Surgery type (ref: Non-elective)				
Elective	0.45	0.44	0.46	< .0001
Gender (ref: Male)				
Female	0.65	0.63	0.67	< .0001
Length of stay	1.00	1.00	1.00	< .0001
Number of chronic conditions	0.80	0.79	0.80	< .0001
Number of diagnoses	1.18	1.18	1.19	< .0001
Number of procedures performed	0.97	0.97	0.98	< .0001
Primary expected payer (ref: Medicare)				
Medicaid	1.31	1.25	1.36	< .0001
Private insurance	0.95	0.92	0.99	0.007
Self-pay	2.36	2.22	2.50	< .0001
No charge	2.30	1.97	2.68	< .0001
Other	1.34	1.25	1.43	< .0001
Race (ref: White)				
Black	1.02	0.99	1.06	0.217
Hispanic	0.82	0.78	0.85	< .0001
Asian/Pacific Islander	0.55	0.50	0.61	< .0001
Native American	1.23	1.06	1.42	0.006
Other	0.71	0.66	0.76	< .0001
Transfer in to hospital (ref: Not transferred in)				
Transferred in from a different acute care hospital	1.11	1.05	1.16	< .0001
Transferred in from another type of health facility	1.69	1.57	1.83	< .0001
Median income of quartile of patient ZIP code (ref: Quartile 1)				
Quartile 2	0.86	0.84	0.89	< .0001
Quartile 3	0.83	0.80	0.86	< .0001
Quartile 4	0.71	0.68	0.73	< .0001
History of MRSA infection (ref: No history)	3.31	2.92	3.74	< .0001
Long-term (current) antibiotic use (ref: No)	5.02	3.80	6.62	< .0001
Suspected MRSA infection (ref: No)	9.31	7.82	11.09	< .0001
Hospital bed size(ref: Small)				
Medium	0.88	0.84	0.92	< .0001
Large	0.82	0.79	0.85	< .0001
Hospital teaching status (ref: Rural)				
Urban non-teaching	0.86	0.82	0.90	< .0001
Urban teaching	0.71	0.68	0.74	< .0001
Hospital region (ref: Northeast)				
Midwest	0.86	0.83	0.90	< .0001
South	1.11	1.07	1.15	< .0001
West	0.84	0.81	0.87	< .0001
Risk of mortality (APRDRG) (ref: No class specified)				
Minor likelihood of dying	11.98	5.99	23.99	< .0001
Moderate likelihood of dying	13.02	6.51	26.06	< .0001
Major likelihood of dying	9.76	4.88	19.53	< .0001
Extreme likelihood of dying	7.61	3.81	15.22	< .0001
Severity (APRDRG) (ref: No class specified)				
Minor loss of function	0.16	0.15	0.17	< .0001
Moderate loss of function	0.34	0.31	0.36	< .0001
Major loss of function	0.55	0.52	0.59	< .0001
Comorbidity: AIDS (ref: No)	1.88	1.45	2.45	< .0001
Comorbidity: Alcohol (ref: No)	0.92	0.85	0.99	0.019
Comorbidity: Deficiency anemias (ref: No)	1.33	1.29	1.38	< .0001

Comorbidity: Rheumatoid arthritis/collagen vascular diseases (ref: No)	1.56	1.45	1.67	< .0001
Comorbidity: Chronic blood loss anemia (ref: No)	0.41	0.38	0.44	< .0001
Comorbidity: Congestive heart failure (ref: No)	1.30	1.24	1.37	< .0001
Comorbidity: Chronic pulmonary disease (ref: No)	1.14	1.10	1.18	< .0001
Comorbidity: Coagulopathy (ref: No)	0.71	0.67	0.75	< .0001
Comorbidity: Depression (ref: No)	1.48	1.42	1.55	< .0001
Comorbidity: Diabetes (uncomplicated) (ref: No)	1.26	1.21	1.30	< .0001
Comorbidity: Diabetes with chronic complications (ref: No)	3.12	2.96	3.29	< .0001
Comorbidity: Drug abuse (ref: No)	2.70	2.52	2.89	< .0001
Comorbidity: Hypertension (ref: No)	1.04	1.01	1.08	0.007
Comorbidity: Liver disease (ref: No)	1.16	1.08	1.25	< .0001
Comorbidity: Lymphoma (ref: No)	1.17	1.01	1.37	0.041
Comorbidity: Fluid and electrolyte disorders (ref: No)	0.81	0.79	0.84	< .0001
Comorbidity: Metastatic cancer (ref: No)	0.50	0.46	0.54	< .0001
Comorbidity: Other neurological disorders (ref: No)	1.21	1.15	1.28	< .0001
Comorbidity: Obesity (ref: No)	1.09	1.06	1.13	< .0001
Comorbidity: Paralysis (ref: No)	2.18	2.03	2.34	< .0001
Comorbidity: Peripheral vascular disorders (ref: No)	1.32	1.26	1.38	< .0001
Comorbidity: Psychoses (ref: No)	1.91	1.79	2.04	< .0001
Comorbidity: Pulmonary circulation disorders (ref: No)	1.07	0.99	1.16	0.106
Comorbidity: Renal failure (ref: No)	1.19	1.14	1.24	< .0001
Comorbidity: Solid tumor without metastasis (ref: No)	0.89	0.81	0.98	0.014
Comorbidity: Valvular disease (ref: No)	0.95	0.89	1.02	0.156
Comorbidity: Weight loss (ref: No)	1.40	1.33	1.48	< .0001

Demographic variables most associated with MRSA infection include gender, primary payer, race, and income. Females were much less likely to suffer from MRSA infection than males (OR: 0.65; 95% CI: 0.63 to 0.67). In comparison to Medicare patients, Medicaid patients are more likely to be diagnosed with MRSA (OR: 1.31; 95% CI: 1.25 to 1.36), as are self-pay patients (OR: 2.36; 95% CI: 2.22 to 2.50) and those who were not charged for services (OR: 2.30; 95% CI: 1.97 to 2.68). In contrast, patients with private insurance were slightly less likely to be diagnosed with MRSA infection (OR: 0.95; 95% CI: 9.92 to 0.99). There was no meaningful difference in MRSA infection between white and black patients, however Hispanics (OR:0.82; 95% CI: 0.78 to 0.85), Asian/Pacific Islanders (OR: 0.55; 95% CI: 0.50 to 0.61), and Others (OR: 0.71; 95% CI: 0.66 to 0.76) were significantly less likely to be diagnosed with MRSA infection. In contrast, Native Americans exhibited higher risk (OR: 1.23; 95% CI: 1.06 to 1.42).

Income is also shown to affect the odds of MRSA infection, with patients in the lowest quartile being most susceptible. Patients in quartile 2 (OR: 0.86; 95% CI: 0.84 to 0.89), quartile 3 (OR: 0.83; 95% CI: 0.80 to 0.86), and quartile 4 (OR: 0.71; 95% CI: 0.68 to 0.73) exhibited progressively lower risk.

Hospital-related predictors included bed size, teaching status and region. Patients in medium (OR: 0.88; 95% CI: 0.84 to 0.92) and large (OR: 0.82; 0.79 to 0.95) hospitals were less likely to contract a MRSA infection compared to patients in small hospitals. In contrast to patients in rural hospitals, patients in urban, non-teaching hospitals (OR: 0.86; 95% CI: 0.83 to 0.90) and patients in urban teaching hospitals (OR: 0.71; 95% CI: 0.68 to 0.74) exhibited reduced risk. Patients in the South (OR: 1.11; 95% CI: 1.07 to 1.11) were more likely to be infected than those in the Northeast, in contrast to patients in the Midwest (OR: 0.86; 95% CI: 0.83 to 0.90) and West (OR: 0.84; 95% CI: 0.81 to 0.87) who were less likely to be infected. Finally, patients who transferred in from either a different acute care hospital (OR: 1.11; 95% CI: 1.05 to 1.16) or another type of health facility (OR: 1.69; 95% CI: 1.57 to 1.83) were at higher risk.

Procedure-related variables most predictive of MRSA infection include surgery type, with patients undergoing elective surgery substantially less likely to suffer MRSA infection compared with those undergoing urgent surgery (OR: 0.45; 95% CI: 0.44 to 0.46). In addition, both of the APRDRG severity measures were significant predictors of infection with the risk of mortality levels Minor likelihood of dying (OR: 11.98; 95% CI: 5.99 to 23.99) and Moderate likelihood of dying (OR: 13.02; 95% CI: 6.51 to 26.06) having the strongest impact, with Major likelihood of dying (OR: 9.76; 95% CI: 4.88 to 19.53) and Extreme likelihood of dying (OR: 7.61; 95% CI: 3.81 to 15.22) still having a

major impact. The Severity index, in contrast, is protective, with patients experiencing reduced risk. Patients with Minor loss of function (OR: 0.16; 95% CI: 0.15 to 0.17), Moderate loss of function (OR: 0.34; 95% CI: 0.31 to 0.36), and Major loss of function (OR: 0.55; 95% CI: 0.52 to 0.59) were all much less likely to contract MRSA. This contrast is likely explained through the amelioration of effect due to confounding as other, more specific comorbidities were added to the model.

Patients with history of MRSA infection (OR:3.31; 95% CI: 2.92 to 3.74), current antibiotic use (OR: 5.02; 95% CI: 3.80 to 6.62), and suspected MRSA infection on admission (OR: 9.31; 95% CI: 7.82 to 11.09) were all more likely to have the infection. In addition, several comorbidities were associated with increased risk of MRSA infection. These were AIDS (OR: 1.88; 95% CI: 1.45 to 2.45), deficiency anemias (OR: 1.33; 95% CI: 1.29 to 1.38), rheumatoid arthritis (OR: 1.56; 95% CI: 1.45 to 1.67), congestive heart failure (OR: 1.30; 95% CI: 1.24 to 1.37), chronic pulmonary disease (OR: 1.14; 95% CI: 1.10 to 1.18), depression (OR: 1.48; 95% CI; 1.42 to 1.55), uncomplicated diabetes (OR: 1.26; 95% CI: 1.21 to 1.30), diabetes with chronic complications (OR: 3.12; 95% CI: 2.96 to 3.29), drug abuse (OR: 2.70; 95% CI: 2.52 to 2.89), hypertension (OR: 1.04; 95% CI: 1.01 to 1.08), liver disease (OR: 1.16; 95% CI: 1.08 to 1.25), lymphoma (OR: 1.17; 95% CI: 1.01 to 1.37), other neurological disorders (OR: 1.21; 95% CI: 1.15 to 1.28), obesity (OR: 1.09; 95% CI: 1.06 to 1.13), paralysis (OR: 2.18; 95% CI: 2.03 to 2.34), peripheral vascular disorders (OR: 1.32; 95% CI: 1.26 to 1.38), psychoses (OR: 1.91; 95% CI: 1.79 to 2.04), renal failure (OR: 1.19; 95% CI: 1.14 to 1.24), and weight loss (OR: 1.40; 95% CI: 1.33 to 1.48). In contrast, alcohol (OR: 0.92; 95% CI: 0.85 to 0.99), chronic blood loss anemia (OR: 0.41; 95% CI: 0.38 to 044),

coagulopathy (OR: 0.71; 95% CI: 0.67 to 0.75), fluid and electrolyte disorders (OR: 0.81; 95% CI: 0.79 to 0.84), metastatic cancer (OR: 0.50; 95% CI: 0.46 to 0.54), solid tumor without metastasis (OR: 0.89; 95% CI: 0.81 to 0.98), and valvular disease (OR: 95; 95% CI: 0.81 to 0.98) were all significantly associated with reduced risk. While the model contains a relatively large number of predictors and the contribution of some predictors is small, the model performs very well.

As described in Chapter 3, the dataset was split into training and test sets, with 85% of the data used for training and 15% of the data used to evaluate the final model. The total sample size was 204,312, which was broken down into training/test and MRSA/Non-MRSA cases as shown in Table 16.

Table 16: Training Size for Hypothesis 1 Logistic Regression Model

		Training size	Test size
Class	Non-MRSA	86,258	15,105
	MRSA	87,407	15,542

Performance on the logistic regression model is summarized in Table 17 and Table 18 below. There were 13,325 true positives, 11,409 true negatives, 3,696 false positives, and 3,217 false negatives.

Table 17: Confusion Matrix for Hypothesis 1 Logistic Regression Model

		Actual	
		MRSA	Non-MRSA
Predicted	MRSA	12,325	3,696
	Non-MRSA	3,217	11,409

Assuming a cutoff of 0.5, sensitivity was calculated as 0.79 and specificity as 0.75. Figure 12 illustrates the overall performance of the model at different cutoff values.

Table 18: Performance of Hypothesis 1 Logistic Regression Model

		Performance
Statistic	Sensitivity	0.79
	Specificity	0.75
	AUC	0.85

The area under the curve was 0.85 which represents strong predictive performance.

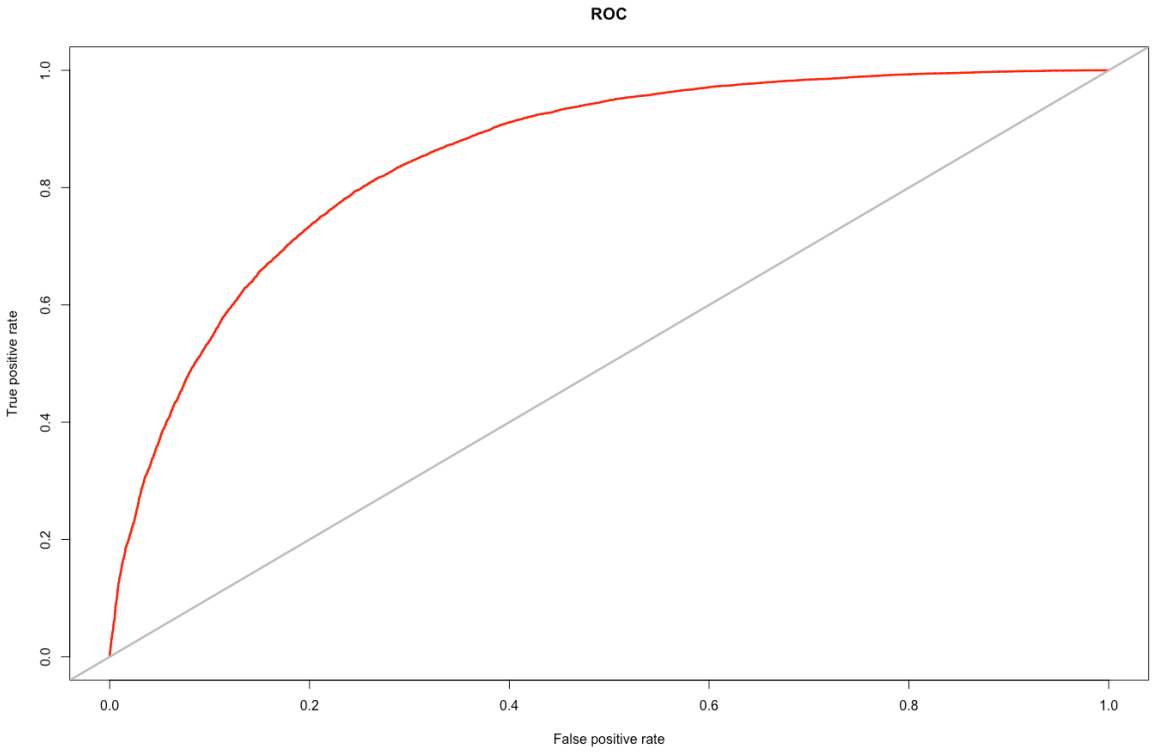

Figure 12: ROC Curve for Hypothesis 1 Logistic Regression Model

In order to assess the overall fit of the model and whether it is over or under fit, a learning curve was plotted. The learning curve is a plot of training and test error for increasing sample sizes. The goal of the learning curve is to qualitatively assess the bias and variance of the model. Bias is a measure of the difference between the model's outcome and the true value. A model with high bias would exhibit high training error. Variance is a measure of generalizability. Thus, a model with low variance would exhibit low training error but high test error. Thus by plotting both of these measures, it is possible to assess the bias/variance tradeoff.

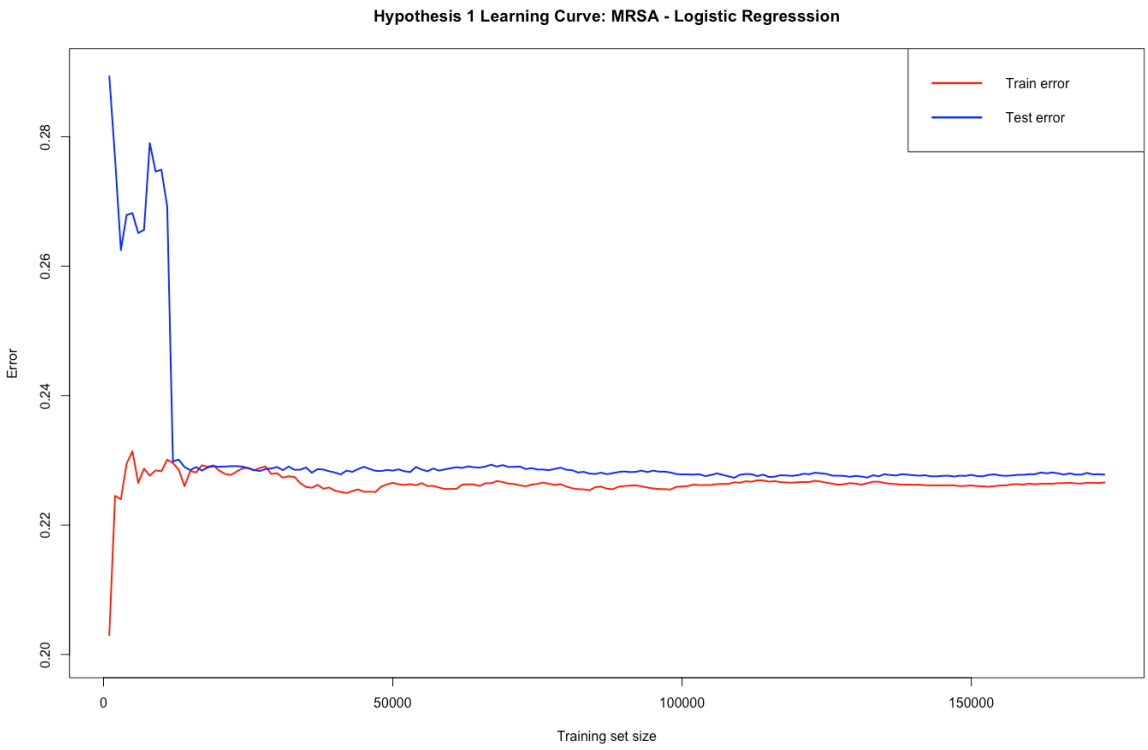

Figure 13: Learning Curve for Hypothesis 1 Logistic Regression Model

For this logistic regression model, training error and test error converge at a relatively small sample size and stay constant as the sample size increases. Adding more

training data neither has a positive nor negative effect as both training and test error remain low. Thus, it can be seen that the model has low bias (i.e., it reflects the true model) and high variance meaning it generalizes well to new data.

4.4.3 Artificial Neural Network Model

Predictors for the neural network were the same as those used in the logistic regression model. Similarly, the dataset was split into training and test sets, with 85% of the data used for training and 15% of the data used to evaluate the final model. The total sample size was 204,312, which was broken down into training/test and MRSA/Non-MRSA cases as shown in Table 16. Within the training set, 10-fold cross validation was used to determine the optimal number of nodes in the neural network's hidden layer. The neural network model specification is shown in Figure 14.

```
MRSA ~ AGE + ELECTIVE + FEMALE + LOS + NCHRONIC + NDX + NPR +
   PAY1 + RACE + TRAN_IN + ZIPINC_QRTL + MRSAHX + ANTIBX + SUSMRSA +
   HOSP_BEDSIZE + HOSP_LOCTEACH + HOSP_REGION + APRDRG_Risk_Mortality +
   APRDRG_Severity + CM_AIDS + CM_ALCOHOL + CM_ANEMDEF + CM_ARTH +
   CM_BLDLOSS + CM_CHF + CM_CHRNLUNG + CM_COAG + CM_DEPRESS +
   CM_DM + CM_DMCX + CM_DRUG + CM_HTN_C + CM_LIVER + CM_LYMPH +
   CM_LYTES + CM_METS + CM_NEURO + CM_OBESE + CM_PARA + CM_PERIVASC +
   CM_PSYCH  +  CM_PULMCIRC + CM_RENLFAIL + CM_TUMOR + CM_VALVE +
   CM_WGHTLOSS
```

Figure 14: Neural Network Model Specification for Hypothesis 1

The overall sample size and breakdown of training/test data and MRSA/non-MRSA cases is shown in Table 19.

Table 19: Training Size for Hypothesis 1 Neural Network Model

		Training size	Test size
Class	Non-MRSA	86,258	15,105
	MRSA	87,407	15,542

Performance on the neural network model is summarized in Table 20. There were 13,368 true positives, 11,173 true negatives, 3,932 false positives, and 2,174 false negatives.

Table 20: Confusion Matrix for Hypothesis 1 Neural Network Model

		Actual	
		MRSA	Non-MRSA
Predicted	MRSA	13,368	3,932
	Non-MRSA	2,174	11,173

Assuming a cutoff of 0.5, sensitivity was calculated as 0.86 and specificity as 0.74. Figure 12 illustrates the overall performance of the model at different cutoff values. The area under the curve was 0.87 which represents strong predictive performance. The optimal number of hidden nodes is 46 and the overall error rate is 0.20.

Table 21: Performance of Hypothesis 1 Neural Network Model

		Performance
Statistic	Sensitivity	0.86
	Specificity	0.74
	AUC	0.87
	Hidden Nodes	46
	Error Rate	0.20

As with the logistic regression model, a ROC curve was plotted to measure performance of the neural network model and a learning curve was plotted to understand the bias/variance tradeoff. These graphs are shown in Figure 15 and respectively.

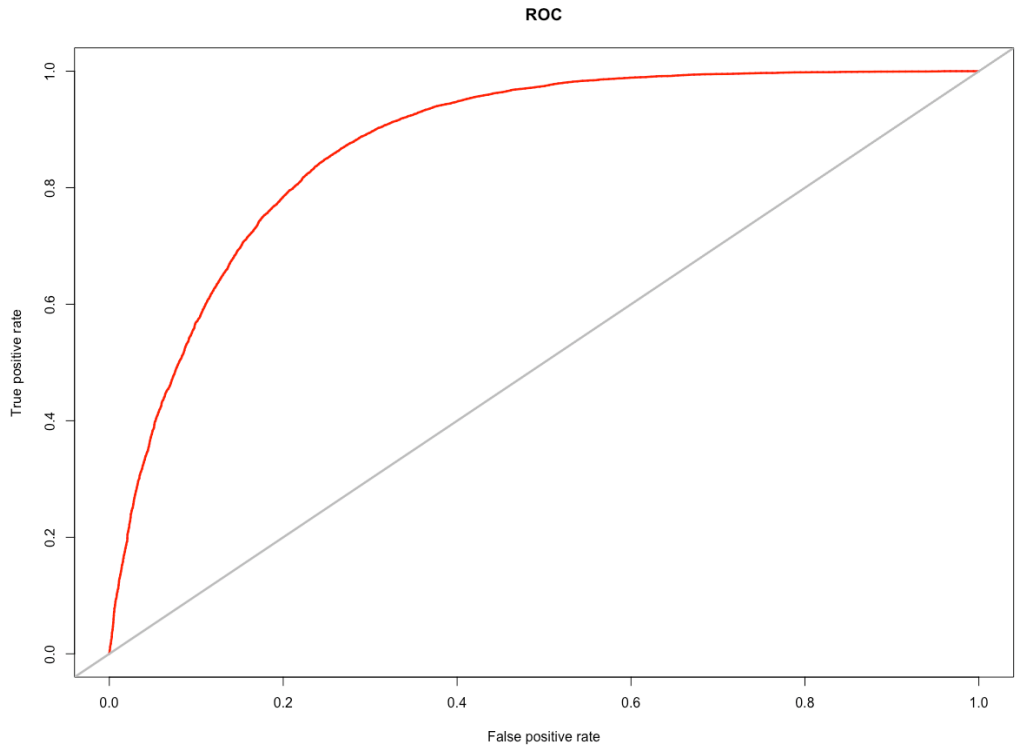

Figure 15: ROC Curve for Hypothesis 1 Neural Network Model

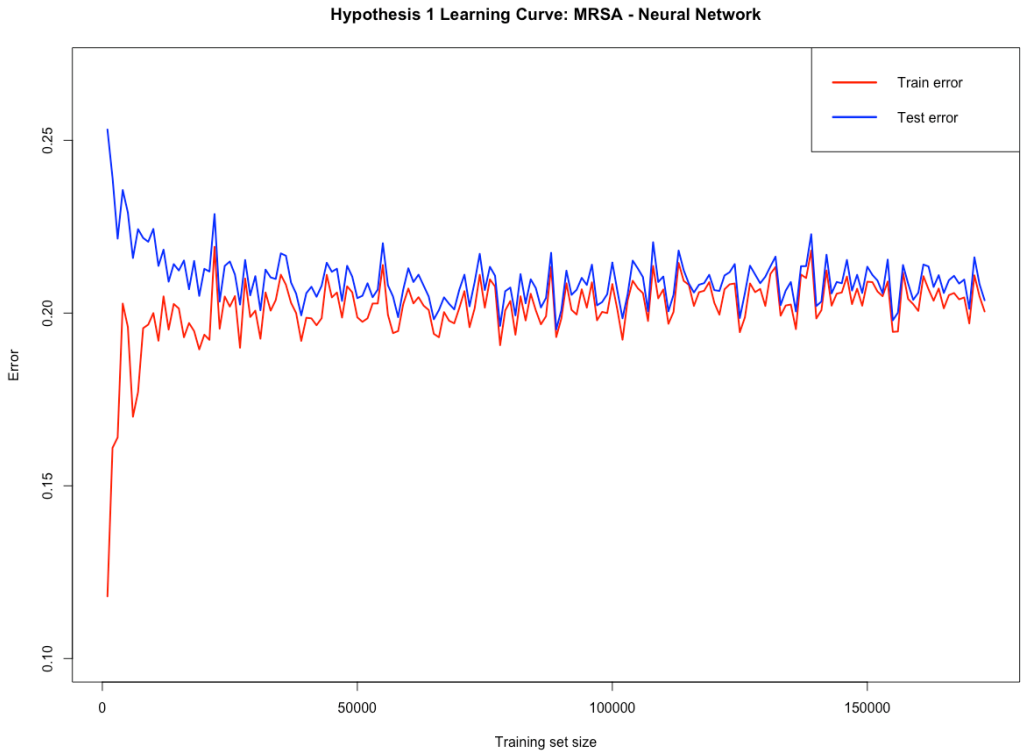

Figure 16: Learning Curve for Hypothesis 1 Neural Network Model

For this neural network model, training error and test error converge at a relatively small sample size and stay constant as the sample size increases. Adding more training data neither has a positive nor negative effect as both training and test error remain low. Thus, it can be seen that the model has low bias (i.e., it reflects the true model) and high variance, meaning it generalizes well to new data.

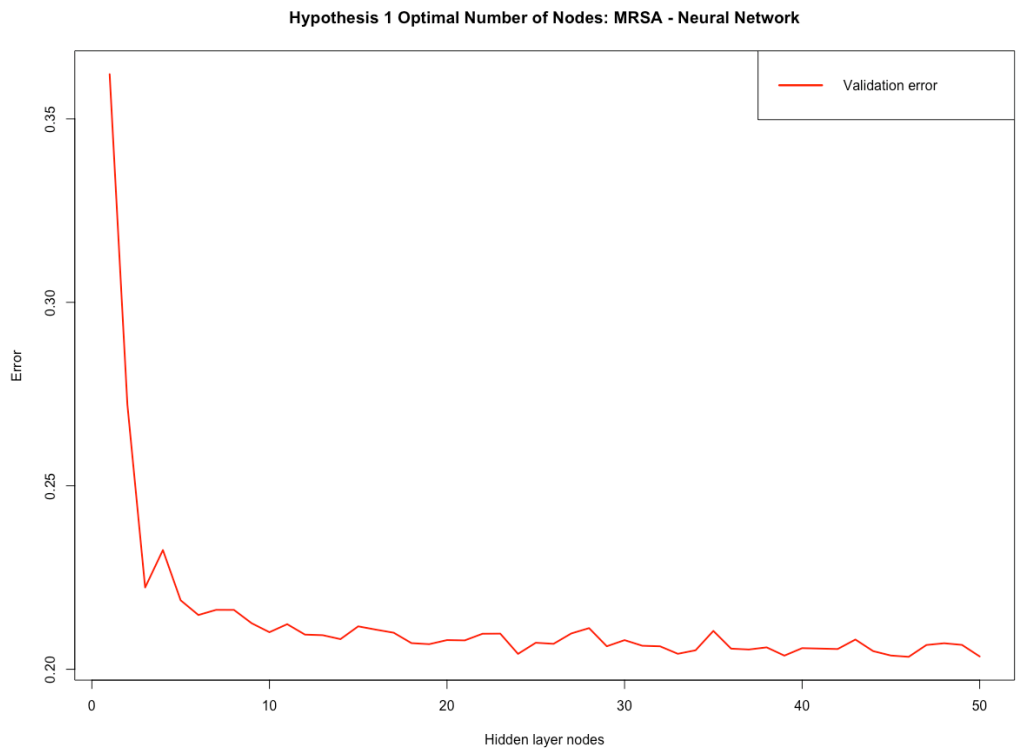

Figure 17: Optimal Number of Hidden Nodes for Hypothesis 1 Neural Network Model

One final tuning parameter for the neural network is the number of nodes in the hidden layer. As indicated above, the optimal number for the hypothesis 1 model is 46. Figure 27 illustrates the error rate for an increasing number of hidden layer nodes. As can be seen, after about 30 nodes the impact of additional nodes is minimal. Although computational power is not an issue and there is no reason to restrict the number of nodes in the hidden layer, this option could be considered if it were needed.

4.4.4 Comparative Performance of the Logistic Regression and Neural Network Models for Hypothesis 1

Table 22 summarizes the relative performance of the logistic regression and neural network models for the prediction of MRSA infection in surgical patients.

Table 22: Comparative Performance of Logistic Regression and Neural Network Models for Hypothesis 1

		Model	
		Logistic Regression	Neural Network
Statistic	Sensitivity	0.79	0.86
	Specificity	0.75	0.74
	AUC	0.85	0.87

Both models perform very well with an AUC of 0.85 or higher and high levels of sensitivity and specificity. On balance, the neural network model exhibits superior performance.

4.5 Hypothesis 2 — Prediction of MRSA-related SSI in Surgical Patients

The second hypothesis aims to predict the incidence of MRSA-related SSI in surgical patients using a range of known demographic, clinical, and hospital-related variables. To facilitate and initial selection of variables for incorporation in the models, a descriptive analysis was performed. Descriptive statistics and p-values were calculated for MRSA and non-MRSA patients. Case and control definitions are illustrated in Figure 2.

For continuous variables, the Wilcoxon Rank Sum test was used. The Wilcoxon Rank Sum test is a non-parametric hypothesis test that is appropriate for both normally-distributed and non-normally-distributed data. The test compares the mean of data between the two groups and provides p-values that quantify the significance of the difference in means between the group. In all cases, a cut-off of $p < 0.05$ was used to assess statistical significance.

For categorical variables, the Chi-Square test of association was used to ascertain whether the difference in proportion in levels of a variable between MRSA and non-MRSA patients are statistically significant. The chi-square test requires the two variables to be categorical, the groups defined by the variables to be independent, and the sample size to be large (i.e., 80 percent of expected frequencies must be greater than 5 and all expected frequencies must be greater than 1.)

Differences between the two groups are described using means (and standard deviations) for continuous variables and proportions (percentages) for categorical variables. Key demographic variables, for MRSA-related SSI patients, are presented visually.

4.5.1 Descriptive Analysis

Table 23 summarizes the means, standard deviations, and p-values for continuous demographic variables. The null hypothesis in each case is that there is no difference in means between MRSA and non-MRSA-related SSI patients. Variables analyzed were age in years at admission, length of stay in days, number of chronic conditions, number of

diagnoses, number of procedures performed, number of days after admission until the first procedure was performed, and number of comorbidities.

Table 23: Demographic Characteristics of the NIS Data Stratified by MRSA-related SSI Status (Continuous Variables)

Variable	MRSA-related SSI Mean (SD)	No MRSA-related SSI Mean (SD)	P-Value
Age in years at admission (years)	58.56 (17.85)	48.19 (25.60)	< .0001
Length of stay (days)	14.75 (18.17)	5.06 (8.02)	< .0001
Number of chronic conditions	5.70 (3.32)	3.76 (3.38)	< .0001
Number of diagnoses	14.74 (6.61)	8.22 (5.76)	< .0001
Number of procedures	4.77 (3.83)	3.07 (2.57)	< .0001
Number of days to first procedure	3.11 (6.30)	1.16 (3.63)	< .0001
Number of comorbidities	3.25 (2.08)	1.74 (1.83)	< .0001

Patients with a MRSA-related SSI were generally older than those without a MRSA-related SSI, with a mean age of 58.56 (SD 17.85) compared to 48.19 (SD 25.60). Figure 18 illustrates the age distribution for MRSA-related SSI in comparison to non MRSA-related SSI patients.

The average length of stay for MRSA patients was 15.75 days (SD 18.17) and 5.06 days (8.02) for non-MRSA patients. Similarly, MRSA-related SSI patients had a higher number of chronic conditions, 5.70 (SD 3.32) compared with 3.76 (SD 3.38); a larger number of diagnoses, 14.74 (SD 6.61) compared to 8.22 (SD 5.76); a larger number of procedures, 4.77 (SD 3.83) compared to 3.07 (SD 2.57); a longer time from admission until performance of the first procedure, 3.11 days (SD 6.30) compared to 1.16 (SD 3.63); and, a higher number of comorbidities, 3.25 (SD 2.08) compared to 1.74 (SD 1.83). In all cases, the differences between MRSA-related SSI and non MRSA-related SSI patients were statistically significant, suggesting that these variables could be useful predictors of MRSA-related SSI infection.

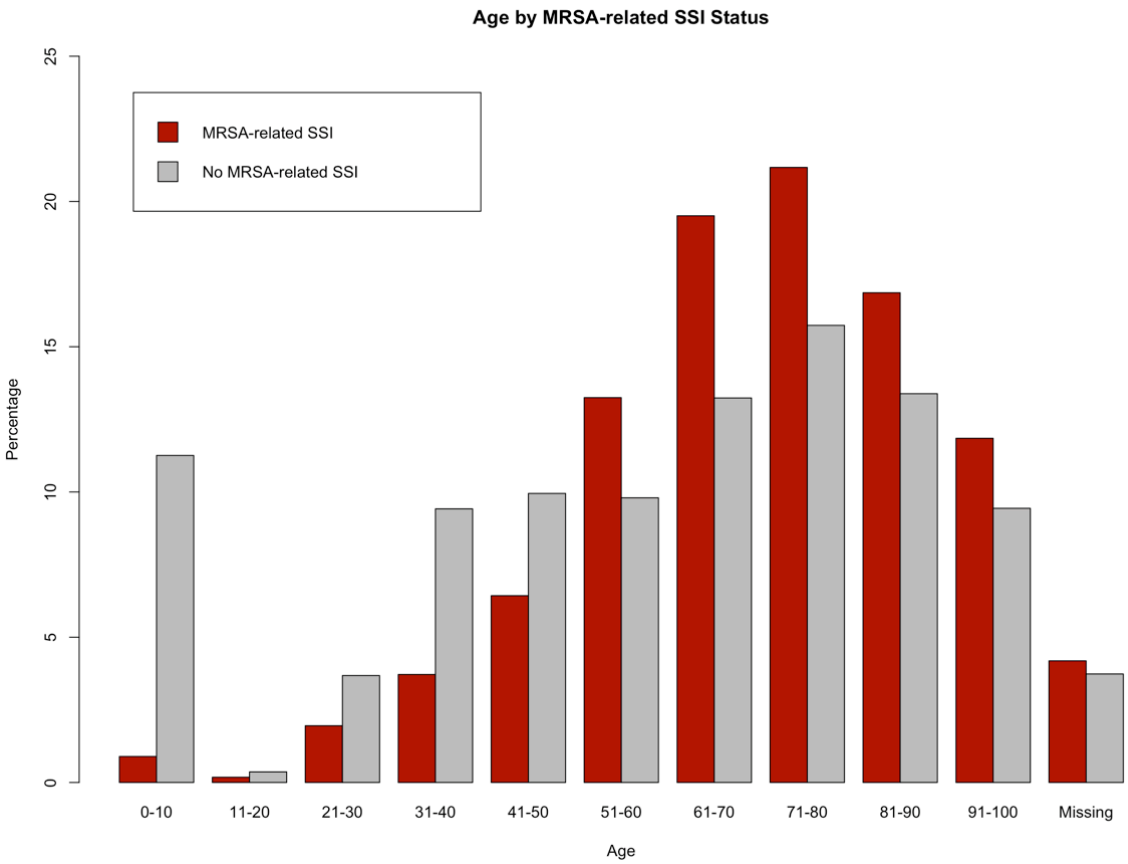

Figure 18: Age Distribution of MRSA-related SSI Compared with Non MRSA-related SSI Patients

Categorical variables analyzed were Gender, Primary Expected Payer, Race, Median income of patient zip code (quartile), Surgery Type, Hospital Region, Hospital Bed Size, and Hospital Teaching Status. Frequencies and percentages of levels of these variables are presented in Table 24 for MRSA and non-MRSA patients.

Table 24: Demographic Characteristics of the NIS data Stratified by MRSA-related SSI Status (Categorical Variables)

Variable	MRSA-related SSI Frequency (%)	No MRSA-related SSI Frequency (%)	P-Value
Gender			
Male	5,431 (49.13)	4,998,700 (46.60)	< .0001
Female	5,623 (50.87)	5,727,005 (53.40)	
Primary expected payer			
Medicare	5,579 (50.54)	3,639,976 (33.96)	< .0001

Medicaid	1,569 (14.21)	1,862,324 (17.38)	
Private insurance	2,936 (26.60)	4,333,620 (40.44)	
Self-pay	389 (3.52)	422,979 (3.95)	
No charge	51 (0.46)	42,481 (0.40)	
Other	515 (4.67)	415,469 (3.88)	
Race			
White	7,859 (76.34)	6,861,359 (69.82)	< .0001
Black	1,299 (12.62)	1,237,374 (12.59)	
Hispanic	751 (7.29)	1,070,323 (10.89)	
Asian/Pacific Islander	101 (0.98)	254,805 (2.59)	
Native American	67 (0.65)	59,897 (0.61)	
Other	218 (2.12)	344,080 (3.50)	
Median income of national quartile or patient ZIP code			
Quartile 1	2,288 (31.33)	2,781,488 (26.42)	< .0001
Quartile 2	2,882 (26.65)	2,696,984 (25.61)	
Quartile 3	2,542 (23.50)	2,653,531 (25.20)	
Quartile 4	2,003 (18.52)	2,397,063 (22.77)	
Surgery type			
Non-elective	7,675 (69.66)	5,602,570 (52.34)	< .0001
Elective	3,343 (30.34)	5,101,039 (47.66)	
Hospital region			
Northeast	1,765 (15.96)	1,972,023 (18.36)	< .0001
Midwest	2,325 (21.03)	2,492,876 (23.21)	
South	5,000 (45.22)	4,158,933 (38.72)	
West	1,967 (17.79)	2,116,649 (19.71)	
Hospital bed size			
Small	1,283 (11.68)	1,401,075 (13.12)	< .0001
Medium	2,622 (23.87)	2,671,824 (25.03)	
Large	7,078 (64.45)	6,603,049 (61.85)	
Hospital teaching status			
Rural	710 (6.46)	923,826 (8.65)	< .0001
Urban non-teaching	3,584 (32.63)	3,736,360 (35.00)	
Urban teaching	6,689 (60.90)	6,015,762 (56.35)	

As with the continuous variables, there were statistically significant differences between MRSA-related SSI and non MRS-related SSI patients. Figure 19illustrates the breakdown of gender for MRSA patients. A greater proportion of MRSA patients were male, 49.13% compared to 46.60%, and, as expected for an older population, a greater proportion utilized Medicare, 50.54% compared to 33.96%.

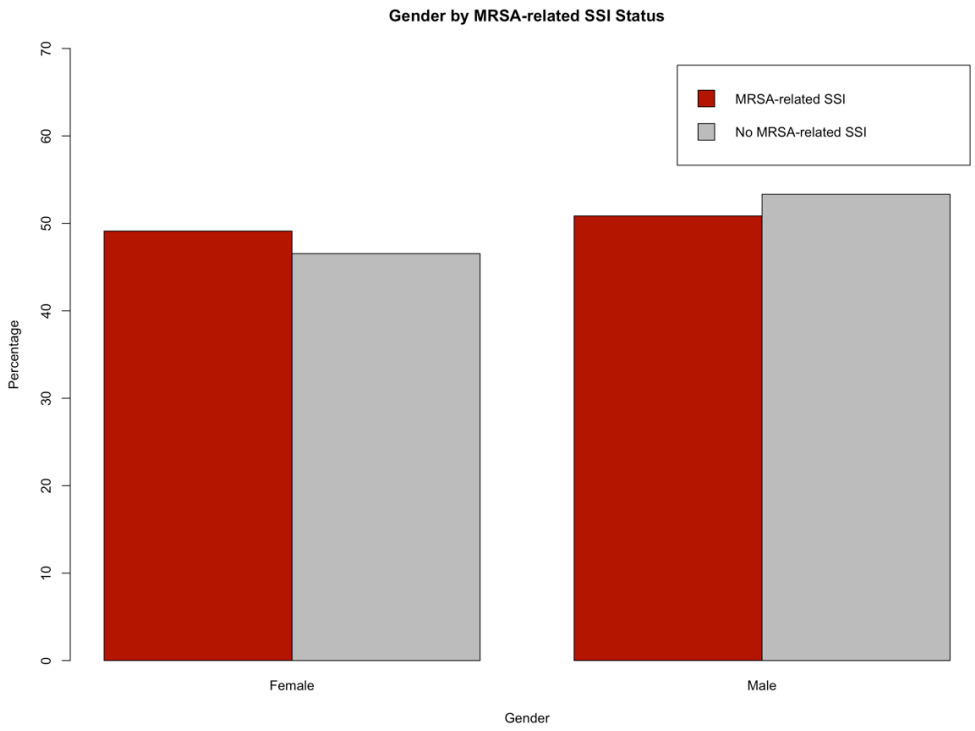

Figure 19: Gender of MRSA-related SSI Patients

The racial breakdown across both groups is broadly similar, however MRSA-related SSI patients are more likely to be white than non MRSA-related SSI patients, 76.34% compared to 69.82%. While there is a similar proportion of black patients across both groups, MRSA patients are also less likely to be Asian/Pacific Islander and Hispanic. Figure 20 summarizes the distribution of race for MRSA-related SSI patients.

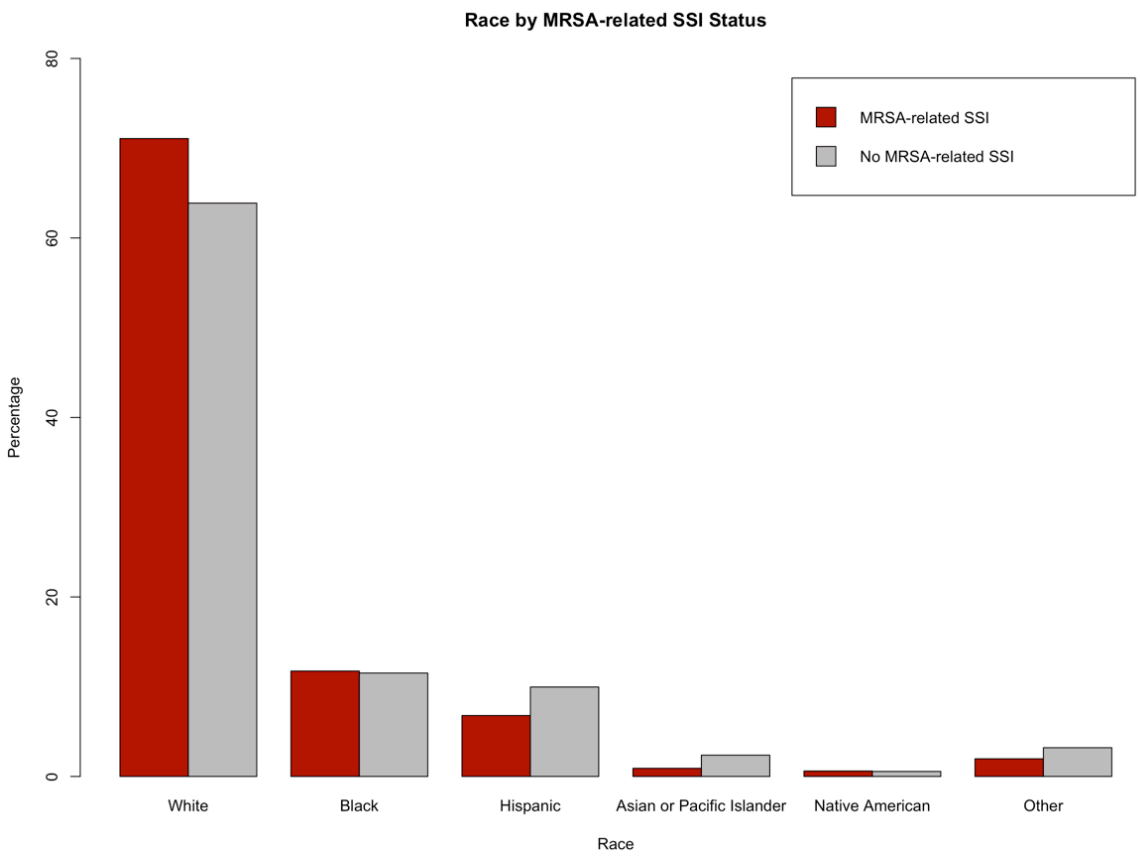

Figure 20: Distribution of Race for MRSA-related SSI Patients and Non MRSA-related SSI Patients

MRSA-related SSI patients were more likely to be poorer than non MRSA-related SSI patients: 31.33 % of MRSA-related SSI patients were in the lowest quartile of income for their zip code compared with 26.42% of non-MRSA patients. The proportion of patients whose income was in the third or fourth quartile for their zip code was commensurately lower: 23.50% of MRSA-related SSI patients had income in the third quartile compared with 25.20% of non MRSA-related SSI patients, and 18.52% of MRSA-related SSI patients had income in the fourth quartile compared with 22.77% of non MRSA-related SSI patients.

A number of hospital and surgery related factors were also compared. For MRSA-related SSI patients, 69.66% of surgeries were non-elective compared to 52.34% for non MRSA-related SSI patients. MRSA-related SSI was more prevalent in hospitals in the South: 45.22% of MRSA-related SSI patients were treated in hospitals in the South compared with 38.72% of non MRSA-related SSI patients. Proportions of MRSA-related SSI and non MRSA-related SSI patients were broadly similar across different hospital sizes, with a slightly higher proportion of surgeries taking place in large hospitals (64.45%) for MRSA-related SSI patients. Finally, a smaller proportion of MRSA-related SSI patients, 6.46%, were more likely to be treated in rural hospitals, compared with 8.65% of non MRSA-related SSI patients.

As with the continuous variables described above, these categorical factors all exhibit statistically significant differences between MRSA-related SSI and non MRSA-related SSI patients, and thus, should be considered for inclusion into the logistic regression and artificial neural network models.

Finally, a number of additional risk factors were considered, including long-term antibiotic use, risk of mortality (APRDRG calculated score), severity (APRDRG calculated score), a broad range of comorbidities, history of MRSA infection, and transfer in or out of the hospital. These risk factors are described in Table 25 (below).

Table 25: Additional Risk Factors Stratified by MRSA-related SSI Status

Variable	MRSA-related SSI Frequency (%)	No MRSA-related SSI Frequency (%)	P-Value
Long-term (current antibiotic use)			
Yes	70 (0.63)	11,355 (0.11)	< .0001
No	10,987 (99.37)	10,729,126 (99.89)	
Risk of mortality (APRDRG)			
No class specified	0 (0.00)	6,394 (0.06)	< .0001
Minor likelihood of dying	3,627 (32.80)	7,382,500 (68.74)	
Moderate likelihood of dying	3,282 (29.68)	1,858,711 (17.31)	

Major likelihood of dying	2,472 (22.36)	987,547 (9.19)	
Extreme likelihood of dying	1,667 (15.16)	505,329 (4.70)	
Severity (APRDRG)			
No class specified	0 (0.00)	6,394 (0.06)	< .0001
Minor loss of function	899 (8.13)	4,621,034 (43.02)	
Moderate loss of function	3,497 (31.63)	3,694,889 (34.40)	
Major loss of function	4,075 (0.23)	1,746,845 (16.26)	
Extreme loss of function	2,586 (23.39)	671,319 (6.25)	
Comorbidity: AIDS			
Yes	30 (0.27)	11,627 (0.11)	< .0001
No	11,027 (99.73)	10,728,854 (99.89)	
Comorbidity: Alcohol			
Yes	408 (3.69)	213,034 (1.98)	< .0001
No	10,649 (96.31)	10,527,447 (98.02)	
Comorbidity: Deficiency anemias			
Yes	3,040 (27.49)	1,303,627 (12.14)	< .0001
No	8,017 (72.51)	9,436,854 (87.86)	
Comorbidity: Rheumatoid arthritis/collagen vascular diseases			
Yes			
No	468 (4.23)	228,433 (2.13)	< .0001
	10,589 (95.77)	10,512,048 (97.87)	
Comorbidity: Chronic blood loss anemia			
Yes			
No	221 (2.00)	329,388 (3.07)	< .0001
	10,836 (98.00)	10,411,093 (96.93)	
Comorbidity: Congestive heart failure			
Yes	1,205 (10.90)	403,407 (3.76)	
No	9,852 (89.10)	10,337,074 (96.24)	< .0001
Comorbidity: Chronic pulmonary disease			
Yes	2,468 (22.32)	1,401,147 (13.05)	< .0001
No	8,589 (77.68)	9,339,334 (86.95)	
Comorbidity: Coagulopathy			
Yes	591 (5.35)	388,028 (3.61)	< .0001
No	10,466 (94.65)	10,352,453 (96.39)	
Comorbidity: Depression			
Yes	1,746 (15.79)	869,678 (8.10)	< .0001
No	9.311 (64.21)	9,870,803 (91.90)	
Comorbidity: Diabetes (uncomplicated)			
Yes	2,483 (22.46)	1,495,084 (13.92)	< .0001
No	8.574 (77.54)	9,245,397 (86.08)	
Comorbidity: Diabetes with chronic complications			
Yes	1,081 (9.78)	366,132 (3.41)	< .0001
No	9,976 (90.22)	10,374,349 (96.59)	
Comorbidity: Drug abuse			
Yes	415 (3.75)	174,876 (1.63)	< .0001
No	10,642 (96.25)	10,565,605 (98.37)	
Comorbidity: Hypertension			
Yes	6,509 (58.87)	4,405,910 (41.02)	< .0001
No	4,548 (41.13)	6,334,571 (58.98)	
Comorbidity: Hypothyroidism			
Yes	1,361 (12.31)	927,470 (8.64)	< .0001
No	9,696 (87.69)	9,813,011 (91.36)	
Comorbidity: Liver disease			
Yes	385 (3.48)	188,515 (1.76)	< .0001

No	10,672 (96.52)	10,551,966 (98.24)	
Comorbidity: Lymphoma			
Yes	94 (0.85)	43,061 (0.40)	< .0001
No	10,963 (99.15)	10,697,420 (99.60)	
Comorbidity: Fluid and electrolyte disorders			
Yes	3,322 (30.04)	1,406,330 (13.09)	< .0001
No	7,735 (69.96)	9,334,151 (86.91)	
Comorbidity: Metastatic cancer			
Yes	335 (3.03)	219,418 (2.04)	< .0001
No	10,722 (96.97)	10,521,063 (97.96)	
Comorbidity: Other neurological disorders			< .0001
Yes	966 (8.74)	419,982 (3.91)	
No	10,091 (91.26)	10,320,499 (96.09)	
Comorbidity: Obesity			
Yes	2,209 (19.98)	1,246,436 (11.61)	< .0001
No	8,848 (80.02)	9,494,045 (88.39)	
Comorbidity: Paralysis			
Yes	477 (4.31)	172,273 (1.60)	< .0001
No	10,580 (95.69)	10,568,208 (98.40)	
Comorbidity: Peripheral vascular disorders			
Yes	1,206 (10.91)	577,460 (5.38)	< .0001
No	9,851 (89.09)	10,163,021 (94.62)	
Comorbidity: Psychoses			
Yes	706 (6.39)	245,798 (2.29)	< .0001
No	10,351 (93.61)	10,494,683 (97.71)	
Comorbidity: Pulmonary circulation disorders			
Yes	341 (3.08)	140,734 (1.31)	< .0001
No	10,716 (96.92)	10,599,747 (98.69)	
Comorbidity: Renal failure			
Yes	1,486 (13.44)	761,194 (7.09)	< .0001
No	9,571 (86.56)	9,979,287 (92.91)	
Comorbidity: Solid tumor without metastasis			
Yes	348 (3.15)	137,852 (1.28)	< .0001
No	10,709 (96.85)	10,602,629 (98.72)	
Comorbidity: Peptic ulcer disease excluding bleeding			
Yes	3 (0.03)	2,662 (0.02)	0.8755
No	11,054 (99.97)	10,737,819 (99.98)	
Comorbidity: Valvular disease			
Yes	516 (4.67)	254,502 (2.37)	< .0001
No	10,541 (95.33)	10,485,979 (97.63)	
Comorbidity: Weight loss			
Yes	1,540 (13.93)	343,227 (3.20)	< .0001
No	9,517 (86.07)	10,397,254 (96.80)	
History of MRSA infection			
Yes	234 (2.12)	46,659 (0.43)	< .0001
No	10,823 (97.88)	10,693,822 (99.57)	
Transfer in to hospital			
Not transferred in	9,607 (87.35)	10,115,356 (94.59)	< .0001
Transferred in from a different acute care hospital	906 (8.24)	428,177 (4.00)	
Transferred in from another type of health facility	485 (4.41)	150,522 (1.41)	

Transfer out of hospital			
Not transferred out	6,780 (61.38)	9,150,159 (85.24)	< .0001
Transferred out to a different acute care hospital	267 (2.42)	83,199 (0.78)	
Transferred out to another type of health facility	3,999 (36.20)	1,501,829 (13.99)	
Elixhauser Comorbidity Score (Readmission)	7,582,286.40	5,373,497.96	< .0001
Elixhauser Comorbidity Score (Mortality)	6,380,866.09	5,374,734.78	< .0001

All the listed comorbidities, with the exception of Peptic Ulcer Disease, exhibited statistically-significant differences between MRSA-related SSI and non MRSA-related SSI patients, suggesting that, in general, sicker patients are more likely to contract MRSA-related SSI infection while in hospital. There may be some overlap or collinearity between these predictors and it is likely that some of these will be selected out of the final models through the stepwise variable selection process. This hypothesis is supported by the Elixhauser comorbidity scores: both the readmission score and the mortality score are significantly higher for MRSA-related SSI patients. This is also true for the APRDRG risk of mortality and severity measures. In essence all of these variables attempt to capture the same concept: the severity of illness and comorbidity burden. In the predictive models it will be important to include the best and most efficient measure of severity, and minimize the number if individual comorbidities included. Similarly, patients transferring in to the hospital and patients that transferred out of the hospital are also at increased risk of MRSA-related SSI infection. Both of these variables also address a component of severity of illness, although the transfer out variable cannot be used to predict the infection as, by definition, it occurs at the end of the hospitalization after the infection has been diagnosed. Finally, although previous history of antibiotic use is not included in the dataset, there is an ICD-9 code for long-term, current antibiotic use,

which may be an appropriate proxy. Patients with long-term antibiotic use are over 5 times more likely to contract a MRSA-related SSI infection, suggesting that this variable may also be a useful predictor.

The above descriptive analysis and comparisons between MRSA-related SSI and non MRSA-related SSI patients clearly indicate that there are significant differences across these groups of patients with respect to a range of demographic, clinical and hospital related variables. The goal of the models presented in the following sections is to develop the best performing model using the most parsimonious set of predictors that are known to the clinician close to the time of admission. Thus, variables that are not available until later in the hospital visit, or those that require complex calculations based on information from the medical record will be excluded.

4.5.2 Logistic Regression Model

A logistic regression model was developed to predict MRSA-SSI. Initially, all variables that were univariately significant according to Table 5 were included in the model and forwards-backwards stepwise variable selection was used to identify the optimal model. The model was optimized by minimizing Akaike Information Criterion (AIC) – a global comparative metric of model performance. The final logistic regression model for hypothesis 2 is shown in Figure 21.

```
MRSA_SSI ~ AGE + ELECTIVE + FEMALE + LOS + NCHRONIC + NDX + PAY1 +
    RACE + TRAN_IN + ZIPINC_QRTL + MRSAHX + ANTIBX + SUSMRSA +
    HOSP_LOCTEACH + HOSP_REGION + CM_AIDS + CM_ALCOHOL + CM_ANEMDEF + CM_ARTH +
    CM_BLDLOSS +  CM_CHF + CM_CHRNLUNG + CM_COAG + CM_DEPRESS + CM_DM + CM_DMCX
    CM_DRUG + CM_HTN_C + CM_HYPOTHY + CM_LIVER + CM_LYTES + CM_NEURO +
    CM_OBESE + CM_PARA + CM_PERIVASC + CM_PSYCH + CM_TUMOR + CM_VALVE +
    CM_WGHTLOSS
```

Figure 21: Logistic Regression Model Specification for Hypothesis 2

The final model includes a substantial number of predictors and achieves very strong predictive performance. The relative contribution of each predictor, while adjusting for all others, is presented below in Table 26. Odds ratios quantify the effect of each predictor on the outcome, 95% confidence intervals provide an estimate of variability, and p-values indicate statistical significance, with p = 0.05 used as the threshold value for significance.

Table 26: Odds Ratios, Confidence Intervals, and p-values for Logistic Regression Model (Hypothesis 2)

Predictor	OR	2.5% CI	97.5% CI	P-Value
(Intercept)	0.06	0.04	0.07	< .0001
Age	1.02	1.01	1.02	< .0001
Surgery type (ref: Non-elective)				
Elective	0.66	0.61	0.71	< .0001
Gender (ref: Male)				
Female	0.75	0.69	0.81	< .0001
Length of stay	1.00	1.00	1.00	< .0001
Number of chronic conditions	0.61	0.69	0.81	< .0001
Number of diagnoses	1.34	1.32	1.36	< .0001
Primary expected payer (ref: Medicare)				
Medicaid	1.07	0.93	1.23	0.321
Private insurance	1.04	0.94	1.15	0.469
Self-pay	1.21	0.98	1.50	0.075
No charge	1.09	0.60	1.98	0.777
Other	1.69	1.39	2.07	< .0001
Race (ref: White)				
Black	0.77	0.68	0.86	< .0001
Hispanic	0.63	0.55	0.73	< .0001
Asian/Pacific Islander	0.42	0.30	0.58	< .0001
Native American	1.04	0.60	1.79	0.889
Other	0.49	0.39	0.63	< .0001
Transfer in to hospital (ref: Not transferred in)				
Transferred in from a different acute care hospital	1.23	1.04	1.46	0.018
Transferred in from another type of health facility	2.40	1.85	3.11	< .0001
Median income of quartile of patient ZIP code (ref: Quartile 1)				
Quartile 2	0.89	0.80	0.98	0.020
Quartile 3	0.78	0.70	0.87	< .0001
Quartile 4	0.74	0.66	0.83	< .0001
History of MRSA infection (ref: No history)	1.80	1.23	2.64	0.003
Long-term (current) antibiotic use (ref: No)	1.99	0.99	3.98	0.053
Suspected MRSA infection (ref: No)	5.10	3.10	8.41	< .0001
Hospital teaching status (ref: Rural)				
Urban non-teaching	1.17	1.00	1.37	0.053
Urban teaching	1.35	1.16	1.58	< .0001
Hospital region (ref: Northeast)				

Midwest	0.83	0.73	0.94	0.004
South	1.14	1.02	1.27	0.017
West	0.85	0.75	0.97	0.014
Comorbidity: AIDS (ref: No)	2.62	1.06	6.45	0.037
Comorbidity: Alcohol (ref: No)	1.21	0.96	1.53	0.113
Comorbidity: Deficiency anemias (ref: No)	1.31	1.18	1.45	< .0001
Comorbidity: Rheumatoid arthritis/collagen vascular diseases (ref: No)	1.90	1.52	2.36	< .0001
Comorbidity: Chronic blood loss anemia (ref: No)	0.43	0.34	0.55	< .0001
Comorbidity: Congestive heart failure (ref: No)	1.26	1.08	1.49	0.005
Comorbidity: Chronic pulmonary disease (ref: No)	1.61	1.45	1.79	< .0001
Comorbidity: Coagulopathy (ref: No)	0.49	0.41	0.59	< .0001
Comorbidity: Depression (ref: No)	1.96	1.73	2.22	< .0001
Comorbidity: Diabetes (uncomplicated) (ref: No)	1.63	1.47	1.81	< .0001
Comorbidity: Diabetes with chronic complications (ref: No)	2.37	1.99	2.82	< .0001
Comorbidity: Drug abuse (ref: No)	1.82	1.41	2.36	< .0001
Comorbidity: Hypertension (ref: No)	1.45	1.31	1.59	< .0001
Comorbidity: Liver disease (ref: No)	1.36	1.07	1.74	0.012
Comorbidity: Fluid and electrolyte disorders (ref: No)	0.62	0.56	0.69	< .0001
Comorbidity: Other neurological disorders (ref: No)	1.45	1.23	1.71	< .0001
Comorbidity: Obesity (ref: No)	1.50	1.34	1.68	< .0001
Comorbidity: Paralysis (ref: No)	1.93	1.52	2.45	< .0001
Comorbidity: Peripheral vascular disorders (ref: No)	1.25	1.07	1.45	0.004
Comorbidity: Psychoses (ref: No)	3.58	2.88	4.47	< .0001
Comorbidity: Solid tumor without metastasis (ref: No)	1.90	1.46	2.46	< .0001
Comorbidity: Valvular disease (ref: No)	1.67	1.34	2.09	< .0001
Comorbidity: Weight loss (ref: No)	1.76	1.49	2.08	< .0001

Demographic variables most associated with MRSA infection include age, gender, primary payer, race, and income. Females were much less likely to suffer from MRSA-SSI than males (OR: 0.785; 95% CI: 0.69 to 0.81). In comparison to Medicare patients, other patients (OR: 1.69; 95% CI: 1.39 to 2.07) were the only group that reached statistical significance. There was a statistically significant difference in MRSA-SSI between White and Black patients (OR: 0.77; 95% CI: 0.68 to 0.86), Hispanics (OR:0.63; 95% CI: 0.55 to 0.73), and Asian/Pacific Islanders (OR: 0.42; 95% CI: 0.30 to 0.58), all of whom were less likely to experience MRSA-SSI. In contrast, Others (OR: 1.69; 95% CI: 1.39 to 2.07) exhibited higher risk. Income is also shown to affect the odds of MRSA infection, with patients in the lowest quartile being most susceptible. Patients in Quartile 2 (OR: 0.89; 95% CI: 0.80 to 0.98), Quartile 3 (OR: 0.78; 95% CI: 0.70 to 0.87), and

Quartile 4 (OR: 0.74; 95% CI: 0.66 to 0.83) exhibited progressively lower risk. Although the effect is small, age is also a significant predictor (OR: 1.02; 95% CI: 1.01 to 1.02).

Hospital-related predictors included teaching status and region. In contrast to patients in rural hospitals, patients in urban, non-teaching hospitals (OR: 1.17; 95% CI: 1.00 to 1.37) and patients in urban teaching hospitals (OR: 1.35; 95% CI: 1.16 to 1.58) exhibited higher risk. Patients in the South (OR: 1.14; 95% CI: 1.02 to 1.27) were more likely to be infected than those in the Northeast, in contrast to patients in the Midwest (OR: 0.83; 95% CI: 0.73 to 0.94) and West (OR: 0.85; 95% CI: 0.75 to 0.97) who were less likely to be infected. Finally, patients who transferred in from either a different acute care hospital (OR: 1.23; 95% CI: 1.04 to 1.46) or another type of health facility (OR: 2.40; 95% CI: 1.85 to 3.11) were at higher risk.

Procedure-related variables most predictive of MRSA-SS include surgery type, with patients undergoing elective surgery substantially less likely to suffer MRSA-SSI compared with those undergoing urgent surgery (OR: 0.66; 95% CI: 0.61 to 0.71). Number of diagnoses was associated with increased risk (OR: 1.34; 95% CI: 1.32 to 1.36) whereas number of chronic conditions was associated with reduced risk (OR: 0.61; 95% CI: 0.69 to 0.81).

Patients with history of MRSA infection (OR: 1.80; 95% CI: 1.23 to 1.64) and suspected MRSA infection on admission (OR: 5.10; 95% CI: 3.10 to 8.41) were all more likely to have the infection. In addition, several comorbidities were associated with increased risk of MRSA-SSI. These were AIDS (OR: 2.62; 95% CI: 1.06 to 6.45), deficiency anemias (OR: 1.31; 95% CI: 1.18 to 1.45), rheumatoid arthritis (OR: 1.90; 95% CI: 1.52 to 2.36), congestive heart failure (OR: 1.26; 95% CI: 1.08 to 1.49), chronic

pulmonary disease (OR: 1.61; 95% CI: 1.45 to 1.79), depression (OR: 1.96; 95% CI; 1.73 to 2.22), uncomplicated diabetes (OR: 1.63; 95% CI: 1.47 to 1.81), diabetes with chronic complications (OR: 2.37; 95% CI: 1.99 to 2.82), drug abuse (OR: 1.82; 95% CI: 1.41 to 2.36), hypertension (OR: 1.41; 95% CI: 1.31 to 1.59), liver disease (OR: 1.36; 95% CI: 1.07 to 1.74), other neurological disorders (OR: 1.45; 95% CI: 1.23 to 1.71), obesity (OR: 1.50; 95% CI: 1.34 to 1.68), paralysis (OR: 1.93; 95% CI: 1.52 to 2.45), peripheral vascular disorders (OR: 1.25; 95% CI: 1.07 to 1.45), psychoses (OR: 3.58; 95% CI: 2.88 to 4.47), valvular disease (OR: 1.67; 95% CI: 1.34 to 2.09), and weight loss (OR: 1.76; 95% CI: 1.49 to 2.08).

In contrast, chronic blood loss anemia (OR: 0.43; 95% CI: 0.34 to 0.55), coagulopathy (OR: 0.49; 95% CI: 0.41 to 0.59) and fluid and electrolyte disorders (OR: 0.62; 95% CI: 0.56 to 0.69) were significantly associated with reduced risk. While the model contains a relatively large number of predictors and the contribution of some predictors is small, the model performs very well.

As described in Chapter 3, the dataset was split into training and test sets, with 85% of the data used for training and 15% of the data used to evaluate the final model. The total sample size was 204,312, which was broken down into training/test and MRSA/Non-MRSA cases as shown in Table 27.

Table 27: Training Size for Hypothesis 2 Logistic Regression Model

		Training size	Test size
Class	Non MRSA-SSI	8,228	1,507
	MRSA-SSI	8,454	1,438

Performance on the logistic regression model is summarized in Table 17 and Table 18 below. There were 1,110 true positives, 1,152 true negatives, 355 false positives, and 328 false negatives.

Table 28: Confusion Matrix for Hypothesis 2 Logistic Regression Model

		Actual	
		MRSA-SSI	Non MRSA-SSI
Predicted	MRSA-SSI	1,110	355
	Non MRSA-SSI	328	1,152

Assuming a cutoff of 0.5, sensitivity was calculated as 0.77 and specificity as 0.76. Figure 22 illustrates the overall performance of the model at different cutoff values.

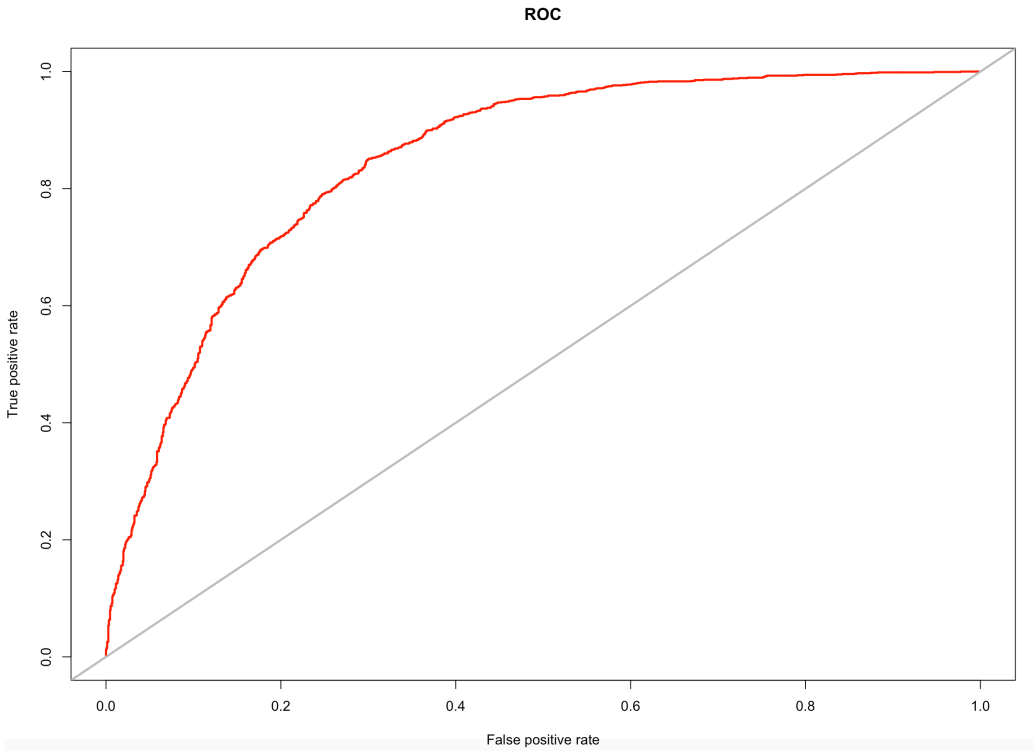

Figure 22: ROC Curve for Hypothesis 2 Logistic Regression Model

The area under the curve was 0.85 which represents strong predictive performance.

Table 29: Performance of Hypothesis 2 Logistic Regression Model

		Performance
Statistic	Sensitivity	0.77
	Specificity	0.76
	AUC	0.85

As with Hypothesis 1, a learning curve was plotted to determine the bias/variance tradeoff.

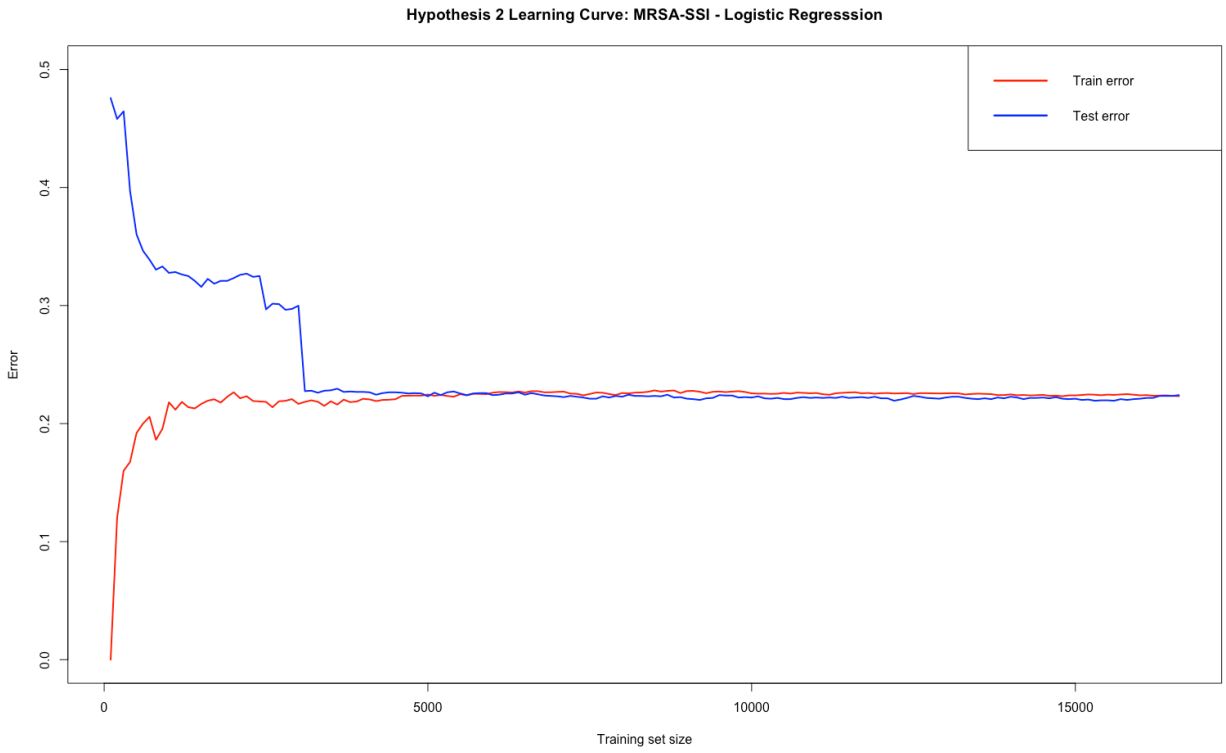

Figure 23: Learning Curve for Hypothesis 2 Logistic Regression Model

For this logistic regression model, training error and test error converge at a relatively small sample size and stay constant as the sample size increases. Adding more training data neither has a positive nor negative effect as both training and test error remain low. Thus, it can be seen that the model has low bias (i.e., it reflects the true model) and high variance meaning it generalizes well to new data.

4.5.3 Artificial Neural Network Model

Predictors for the neural network were the same as those used in the logistic regression model. Similarly, the dataset was split into training and test sets, with 85% of the data used for training and 15% of the data used to evaluate the final model. The total sample size was 19,627, which was broken down into training/test and MRSA/Non-MRSA cases as shown in Table 27.

Table 30: Training Size for Hypothesis 2 Neural Network Model

		Training size	Test size
Class	Non MRSA-SSI	8,228	1,507
	MRSA-SSI	8,454	1,438

Within the training set, 10-fold cross validation was used to determine the optimal number of nodes in the neural network's hidden layer. The neural network model specification is shown in Figure 24 and the overall sample size and breakdown of training/test data and MRSA/non-MRSA cases is shown in Table 30.

```
MRSA_SSI ~      AGE + ELECTIVE + FEMALE + LOS + NCHRONIC + NDX + PAY1 +
    RACE + TRAN_IN + ZIPINC_QRTL + MRSAHX + ANTIBX + SUSMRSA +
    HOSP_LOCTEACH + HOSP_REGION + CM_AIDS + CM_ALCOHOL + CM_ANEMDEF + CM_ARTH +
    CM_BLDLOSS +  CM_CHF + CM_CHRNLUNG + CM_COAG + CM_DEPRESS + CM_DM +
    CM_DMCX +   CM_DRUG + CM_HTN_C + CM_HYPOTHY + CM_LIVER + CM_LYTES +
    CM_NEURO  + CM_OBESE + CM_PARA + CM_PERIVASC + CM_PSYCH + CM_TUMOR +
    CM_VALVE + CM_WGHTLOSS
```

Figure 24: Neural Network Model Specification for Hypothesis 2

Performance on the neural network model is summarized in Table 31. There were 1,122 true positives, 1,236 true negatives, 385 false positives, and 202 false negatives.

Table 31: Confusion Matrix for Hypothesis 2 Neural Network Model

		Actual	
		MRSA-SSI	Non MRSA-SSI
Predicted	MRSA-SSI	1,236	385
	Non MRSA-SSI	202	1,122

Assuming a cutoff of 0.5, sensitivity was calculated as 0.73 and specificity as 0.87. Figure 12 illustrates the overall performance of the model at different cutoff values. The optimal number of hidden nodes is 24 and the overall error rate is 0.20.

Table 32: Performance of Hypothesis 2 Neural Network Model

		Performance
Statistic	Sensitivity	0.73
	Specificity	0.87
	AUC	0.86
	Hidden Nodes	24
	Error Rate	0.20

The area under the curve was 0.86 which represents strong predictive performance.

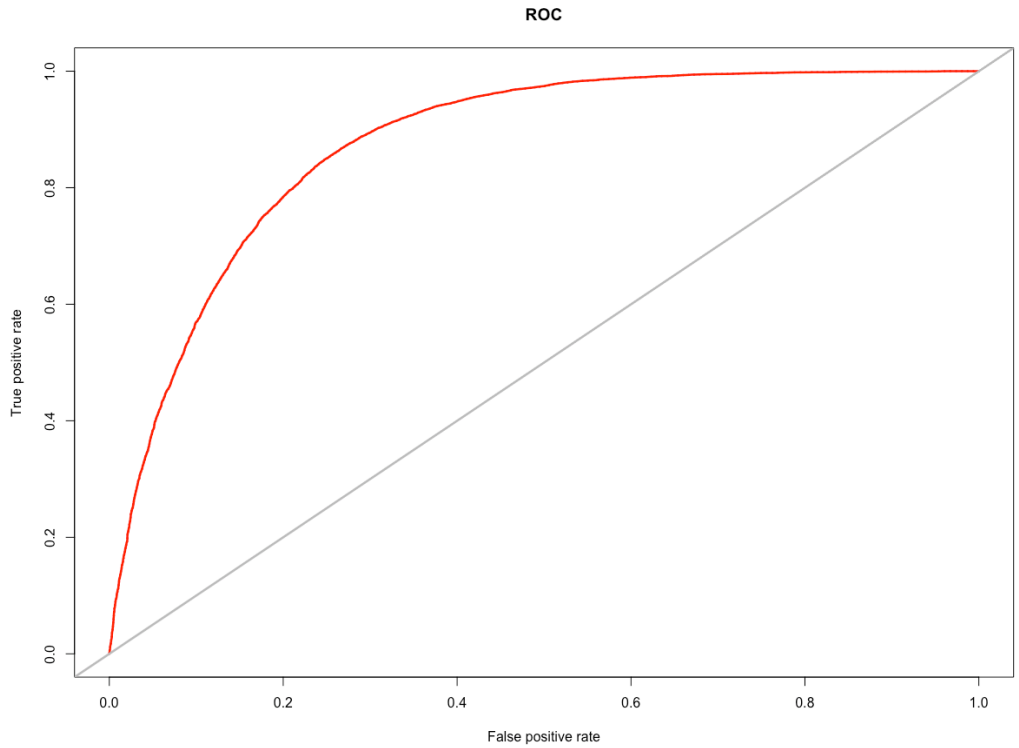

Figure 25: ROC Curve for Hypothesis 2 Neural Network Model

As with the logistic regression model, a learning curve was plotted to determine the bias/variance tradeoff.

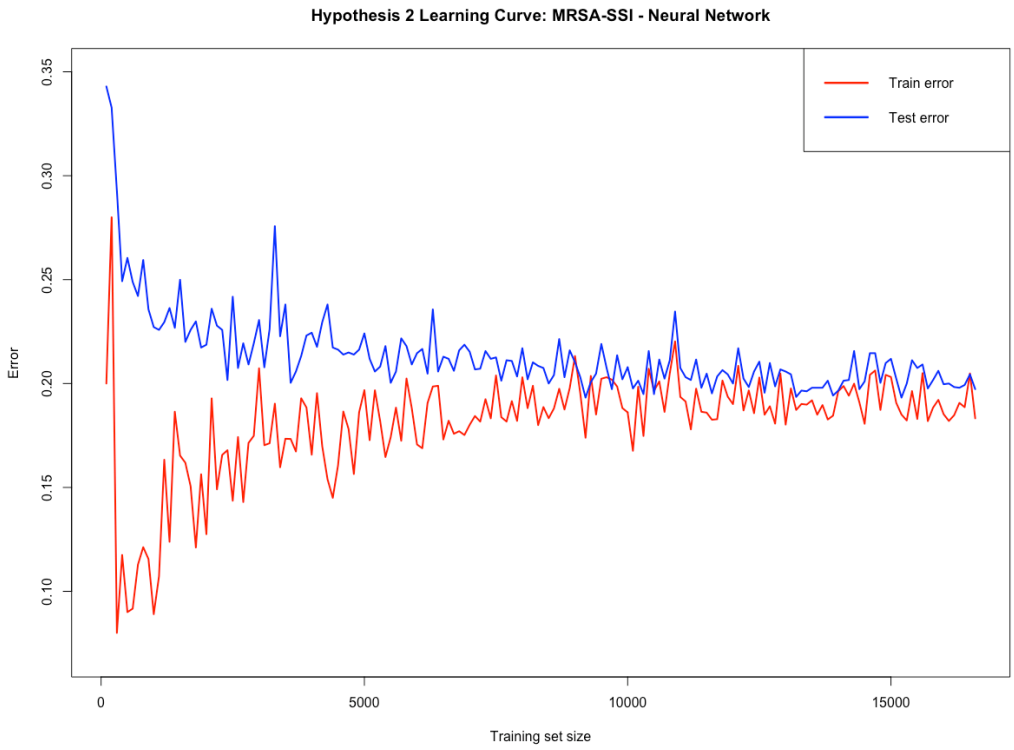

Figure 26: Learning Curve for Hypothesis 2 Neural Network Model

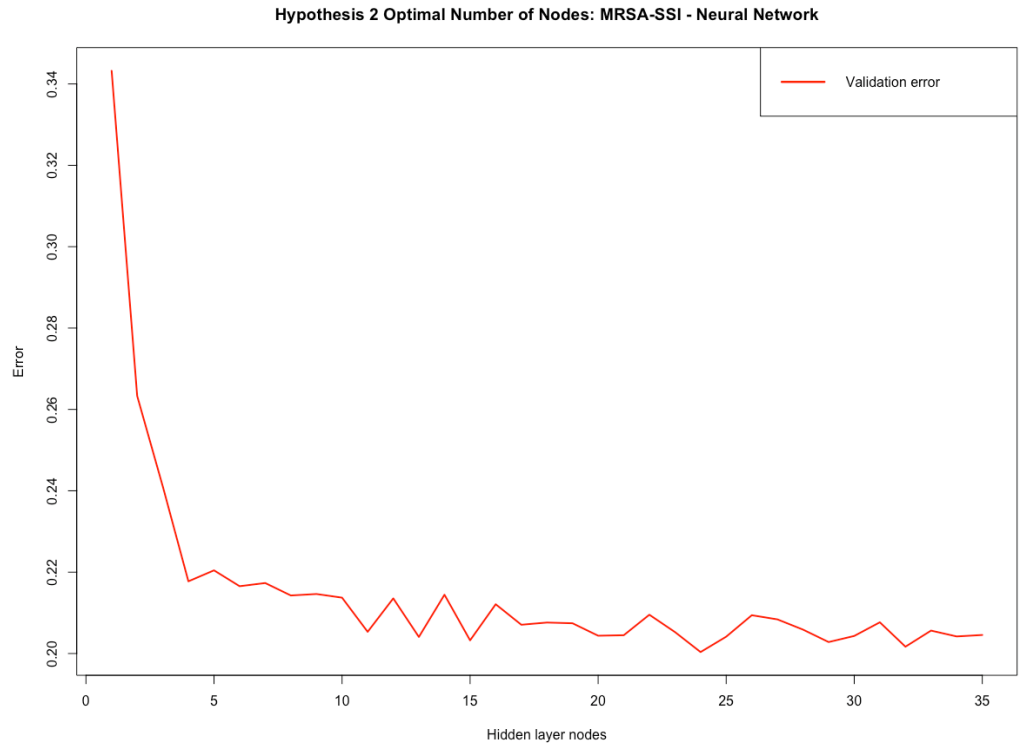

Figure 27: Optimal Number of Hidden Nodes for Hypothesis 2 Neural Network Model

One final tuning parameter for the neural network is the number of nodes in the hidden layer. As indicated above, the optimal number for the hypothesis 2 model is 24. Figure 27 illustrates the error rate for an increasing number of hidden layer nodes. As can be seen, after approximately 24 nodes the impact of additional nodes is minimal. Although computational power is not an issue and there is no reason to restrict the number of nodes in the hidden layer, this option could be considered if it were needed.

4.5.4 Comparative Performance of the Logistic Regression and Neural Network Models for Hypothesis 2

Table 33 summarizes the relative performance of the logistic regression and neural network models for the prediction of MRSA-SSI infection in surgical patients.

Table 33: Comparative Performance of Logistic Regression and Neural Network Models for Hypothesis 2

		Model	
		Logistic Regression	Neural Network
Statistic	Sensitivity	0.77	0.73
	Specificity	0.76	0.87
	AUC	0.85	0.86

Both models perform very well with an AUC of 0.85 or higher and good levels of sensitivity and specificity. On balance, the neural network model exhibits superior performance.

4.6 Hypothesis 3 — Prediction of MRSA Infection in *S. aureus* Patients

The first hypothesis aims to predict the incidence of MRSA infection in surgical patients using a range of known demographic, clinical, and hospital-related variables. To facilitate and initial selection of variables for incorporation in the models, a descriptive analysis was performed. Descriptive statistics and p-values were calculated for MRSA and non-MRSA patients. Case and control definitions are illustrated in Figure 3.

For continuous variables, the Wilcoxon Rank Sum test was used. The Wilcoxon Rank Sum test is a non-parametric hypothesis test that is appropriate for both normally-

distributed and non-normally-distributed data. The test compares the mean of data between the two groups and provides p-values that quantify the significance of the difference in means between the group. In all cases, a cut-off of $p < 0.05$ was used to assess statistical significance.

For categorical variables, the Chi-Square test of association was used to ascertain whether the difference in proportion in levels of a variable between MRSA and non-MRSA patients are statistically significant. The chi-square test requires the two variables to be categorical, the groups defined by the variables to be independent, and the sample size to be large (i.e., 80 percent of expected frequencies must be greater than 5 and all expected frequencies must be greater than 1.)

Differences between the two groups are described using means (and standard deviations) for continuous variables and proportions (percentages) for categorical variables. Key demographic variables, for MRSA patients, are presented visually.

4.6.1 Descriptive Analysis

Table 34 summarizes the means, standard deviations, and p-values for continuous demographic variables. The null hypothesis in each case is that there is no difference in means between MRSA and non-MRSA patients. Variables analyzed were age in years at admission, length of stay in days, number of chronic conditions, number of diagnoses, number of procedures performed, number of days after admission until the first procedure was performed, and number of comorbidities.

Table 34: Demographic Characteristics of the NIS Data Stratified by MRSA Status (Continuous Variables)

Variable	MRSA-SSI Mean (SD)	MSSA-SSI Mean (SD)	P-Value
Age in years at admission (years)	58.56 (17.85)	52.66 (19.03)	< .0001
Length of stay (days)	14.75 (18.17)	10.02 (12.86)	< .0001
Number of chronic conditions	5.70 (3.33)	4.54 (3.17)	< .0001
Number of diagnoses	14.74 (6.61)	11.91 (6.13)	< .0001
Number of procedures	4.77 (3.83)	3.83 (2.90)	< .0001
Number of days to first procedure	3.11 (6.30)	2.13 (5.05)	< .0001
Number of comorbidities	3.25 (2.08)	2.43 (1.93)	< .0001

Patients with MRSA-SSI were generally older than those with MSSA-SSI, with a mean age of 58.56 (SD 17.85) compared to 52.66 (SD 19.03). Figure 28 illustrates the difference in age distribution for MRSA-SSI and MSSA-SSI patients.

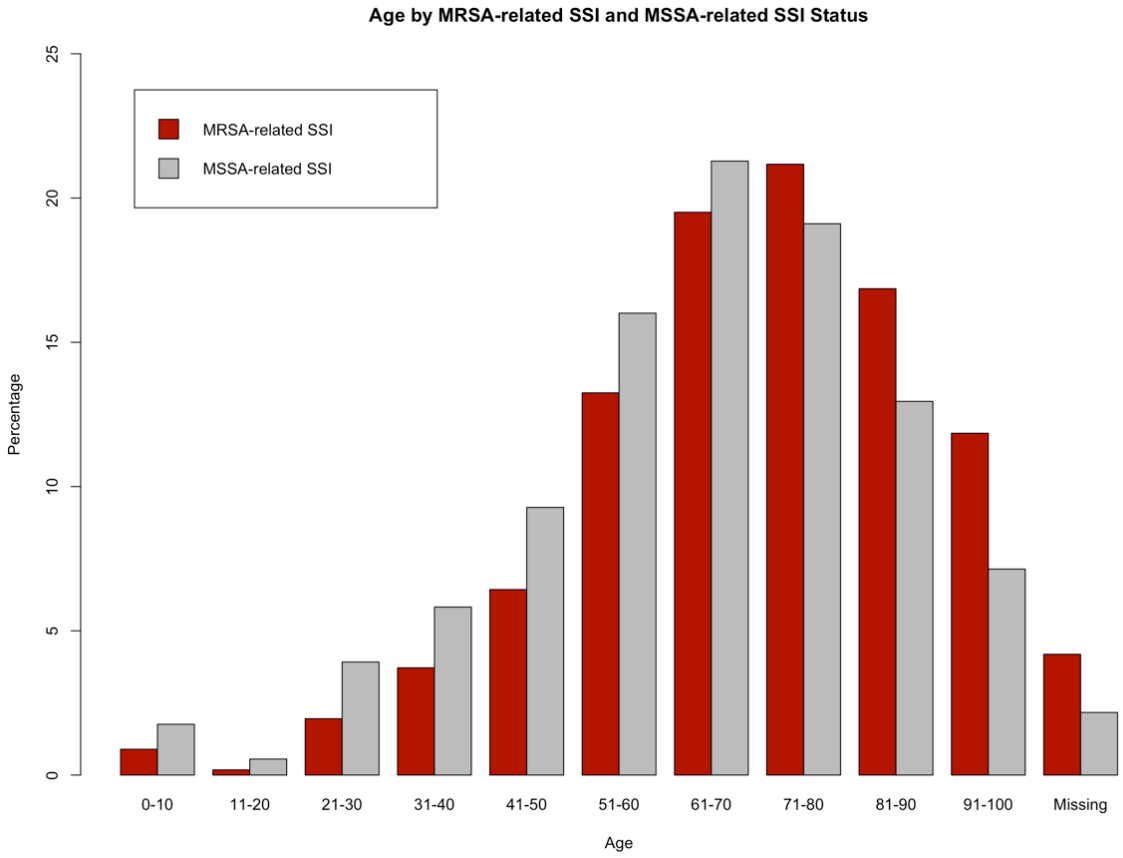

Figure 28: Age Distribution of MRSA-SSI and MSSA-SSI Patients

The average length of stay for MRSA-SSI patients was 14.75 days (SD 18.17) and 10.02 days (SD 12.86) for MSSA-SSI patients. Similarly, MRSA-SSI patients had a higher number of chronic conditions, 5.70 (SD 3.33) compared with 4.54 (3.17); a larger number of diagnoses, 14.74 (SD 6.61) compared to 11.91 (SD 6.13); a larger number of procedures, 4.77 (SD 3.83) compared to 3.83 (SD 2.90); a longer time from admission until performance of the first procedure, 3.11 days (SD 6.30) compared to 2.13 (SD 5.05); and, a higher number of comorbidities, 3.25 (SD 2.08) compared to 2.43 (SD 1.93). In all cases, the differences between MRSA-SSI and MSSA-SSI patients were statistically significant, suggesting that these variables could be useful in differentiating these infections.

Categorical variables analyzed were Gender, Primary Expected Payer, Race, Median income of patient zip code (quartile), Surgery Type, Hospital Region, Hospital Bed Size, and Hospital Teaching Status. Frequencies and percentages of levels of these variables are presented in Table 35 for MRSA and non-MRSA patients.

Table 35: Demographic Characteristics of the NIS Data Stratified by MRSA Status (Categorical Variables)

Variable	MRSA-SSI Frequency (%)	MSSA-SSI Frequency (%)	P-Value
Gender			
Male	5,431 (49.13)	4,918 (55.07)	< .0001
Female	5,623 (50.87)	4,012 (44.93)	
Primary expected payer			
Medicare	5,579 (50.54)	3,156 (35.57)	< .0001
Medicaid	1,569 (14.21)	1,330 (14.91)	
Private insurance	2,936 (26.60)	3,452 (38.69)	
Self-pay	389 (3.52)	364 (4.08)	
No charge	51 (0.46)	45 (0.50)	
Other	515 (4.67)	575 (6.44)	
Race			
White	7,859 (76.34)	6,414 (78.85)	< .0001
Black	1,299 (12.62)	729 (8.96)	
Hispanic	751 (7.29)	609 (7.49)	
Asian/Pacific Islander	101 (0.98)	112 (1.38)	
Native American	67 (0.65)	56 (0.69)	

Other	218 (2.12)	214 (2.63)	
Median income of national quartile or patient ZIP code			
Quartile 1	3,388 (31.33)	2,172 (24.87)	< .0001
Quartile 2	2,882 (26.65)	2,376 (27.21)	
Quartile 3	2,542 (23.50)	2,226 (25.49)	
Quartile 4	2,003 (18.52)	1,959 (22.43)	
Surgery type			
Non-elective	7,675 (69.66)	6,307 (70.86)	0.0660
Elective	3,343 (30.34)	2,594 (29.14)	
Hospital region			
Northeast	1,765 (15.96)	1,689 (18.91)	< .0001
Midwest	2,325 (21.03)	2,212 (24.76)	
South	5,000 (45.22)	3,212 (35.95)	
West	1,967 (17.79)	1,821 (20.38)	
Hospital bed size			
Small	1,283 (11.68)	1,077 (12.14)	0.6008
Medium	2,622 (23.87)	2,098 (23.65)	
Large	7,078 (64.45)	5,695 (64.21)	
Hospital teaching status			
Rural	710 (6.46)	505 (5.69)	< .0001
Urban non-teaching	3,584 (32.63)	2,546 (28.70)	
Urban teaching	6,689 (60.90)	5,819 (65.60)	

As with the continuous variables, there were statistically significant differences between MRSA-SSI and MSSA-SSI patients. Figure 29 illustrates the breakdown of gender for MRSA patients. In contrast to hypotheses 1 and 2, a greater proportion of MRSA-SSI patients were female, 50.87% compared to 44.93% of MSSA-SSI patients, and, as expected for an older population, a greater proportion utilized Medicare, 50.87% compared to 35.57%.

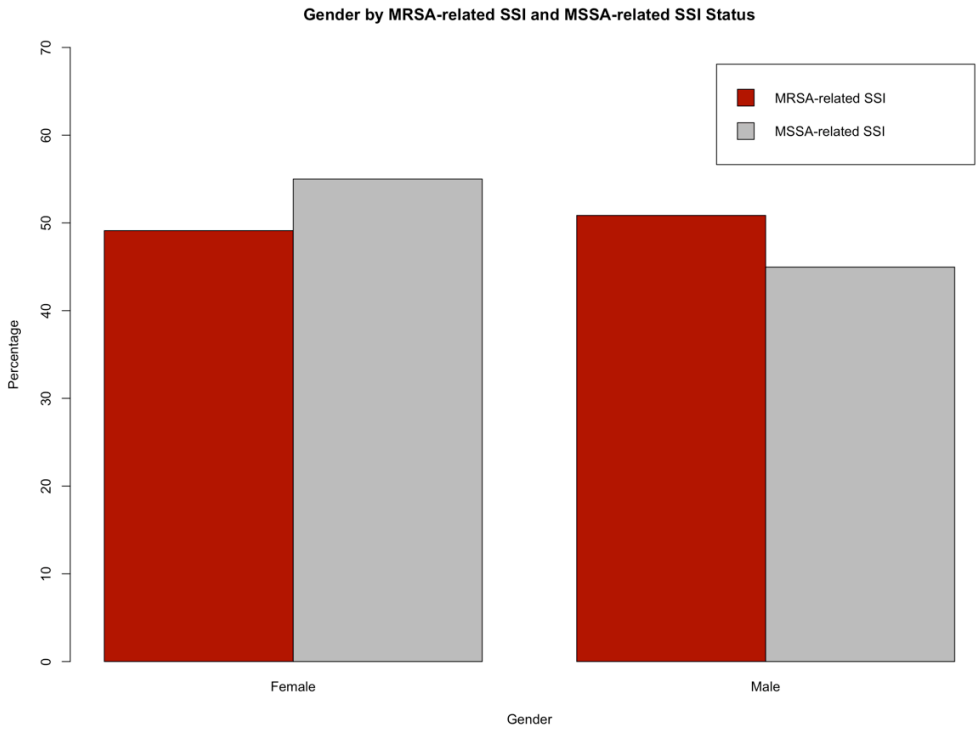

Figure 29: Gender of MRSA-SSI and MSSA-SSI Patients

The racial breakdown across both groups is broadly similar, however MRSA-SSI patients are more likely to be black than MSSA-SSI patients, 12.62% compared to 8.96%. Figure 30 summarizes the distribution of race for MRSA-SSI and MSSA-SSI patients.

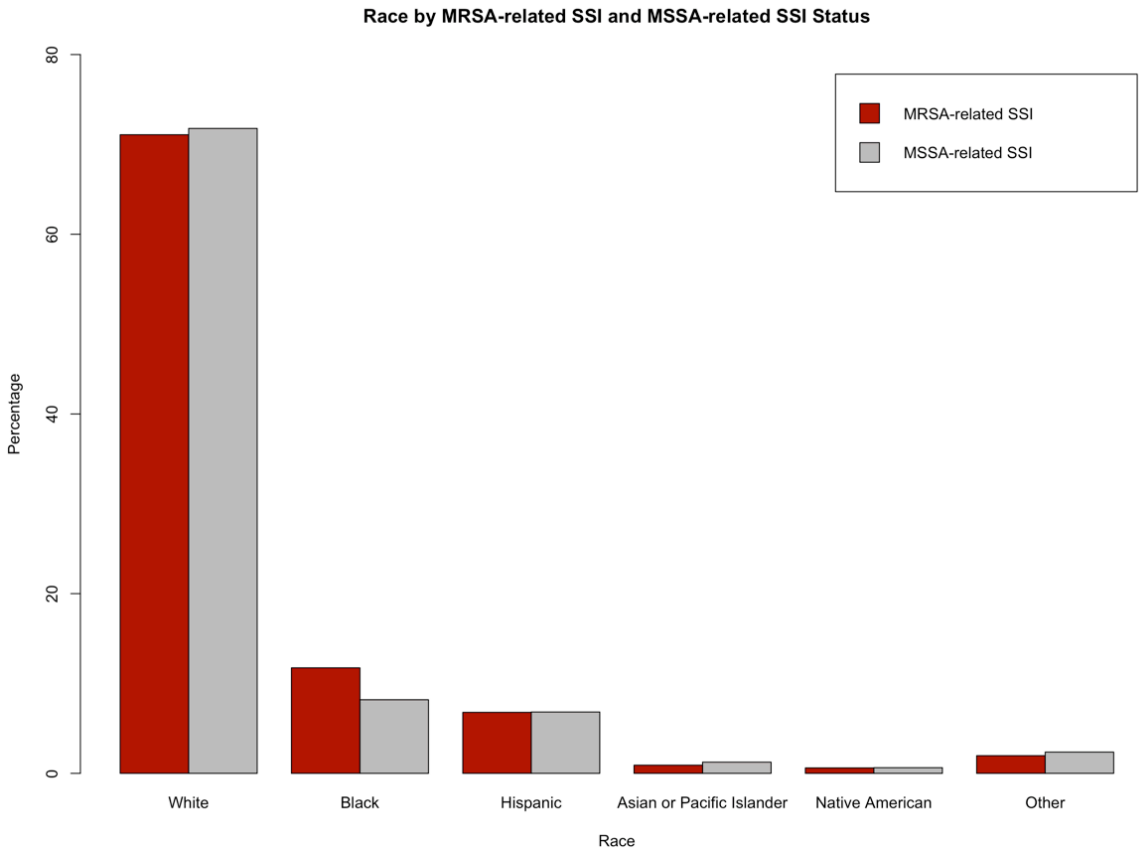

Figure 30: Distribution of Race for MRSA-SSI and MSSA-SSI Patients

MRSA-SSI patients were more likely to be poorer than non MSSA-SSI patients: 31.33% of MRSA-SSI patients were in the lowest quartile of income for their zip code compared with 24.87% of MSSA-SSI patients. The proportion of patients whose income was in the third or fourth quartile for their zip code was commensurately lower: 23.50% of MRSA-SSI patients had income in the third quartile compared with 25.49 of MSSA-SSI patients, and 18.52% of MSSA-SSI patients had income in the fourth quartile compared with 22.43% of non-MRSA patients.

A number of hospital and surgery related factors were also compared. There were no meaningful differences between MRSA-SSI and MSSA-SSI patients based on whether

surgery was elective or non-elective. As with previous hypotheses, MRSA-SSI was more prevalent in hospitals in the South: 45.22% of MRSA-SSI patients were treated in hospitals in the South compared with 35.95% of MSSA-SSI patients. Proportions of MRSA-SSI and MSSA-SSI patients were broadly similar across different hospital sizes and these differences were not statistically significant. Finally, a slightly greater proportion of MRSA-SSI patients, 6.46%, were more likely to be treated in rural hospitals, compared with 85.69% of MSSA-SSI patients.

As with the continuous variables described above, most categorical factors exhibit statistically significant differences between MRSA-SSI and MSSA-SSI patients, and thus, should be considered for inclusion into the logistic regression and artificial neural network models.

Finally, a number of additional risk factors were considered, including long-term antibiotic use, risk of mortality (APRDRG calculated score), severity (APRDRG calculated score), a broad range of comorbidities, history of MRSA infection, and transfer in or out of the hospital. These risk factors are described in Table 36.

Table 36: Additional Risk Factors Stratified by MRSA Status

Variable	MRSA-SSI Frequency (%)	MSSA-SSI Frequency (%)	P-Value
Long-term (current antibiotic use)			
Yes	70 (0.63)	65 (0.73)	0.4174
No	10,987 (99.37)	8,869 (99.27)	
Risk of mortality (APRDRG)			
No class specified	0 (0.00)	1 (0.01)	< .0001
Minor likelihood of dying	3,627 (32.80)	4,588 (51.35)	
Moderate likelihood of dying	3,282 (29.68)	2,592 (29.01)	
Major likelihood of dying	2,472 (22.36)	1,278 (14.30)	
Extreme likelihood of dying	1,676 (15.16)	475 (5.32)	
Severity (APRDRG)			
No class specified	0 (0.00)	1 (0.01)	< .0001
Minor loss of function	899 (8.13)	1,347 (15.08)	
Moderate loss of function	3,497 (31.63)	3,872 (43.34)	
Major loss of function	4,075 (36.85)	2,832 (31.70)	

Extreme loss of function	2,586 (23.39)	882 (9.87)	
Comorbidity: AIDS			
Yes	30 (0.27)	23 (0.26)	0.8495
No	11,027 (99.73)	8,911 (99.74)	
Comorbidity: Alcohol			
Yes	408 (3.69)	294 (3.29)	0.1274
No	10,649 (96.31)	8,640 (96.71)	
Comorbidity: Deficiency anemias			
Yes	3,040 (27.49)	1,760 (19.70)	< .0001
No	8,017 (52.77)	7,174 (80.30)	
Comorbidity: Rheumatoid arthritis/collagen vascular diseases			
Yes	468 (4.23)	379 (4.24)	0.9733
No	10,589 (95.77)	8,555 (95.76)	
Comorbidity: Chronic blood loss anemia			
Yes			
No	221 (2.00)	122 (1.37)	0.0006
	10,836 (98.00)	8,812 (98.63)	
Comorbidity: Congestive heart failure			
Yes	1,205 (10.90)	453 (5.07)	< .0001
No	9,852 (89.10)	8,481 (94.93)	
Comorbidity: Chronic pulmonary disease			
Yes	2,468 (22.32)	1,483 (16.60)	< .0001
No	8,589 (77.68)	7,451 (83.40)	
Comorbidity: Coagulopathy			
Yes	591 (5.35)	283 (3.17)	< .0001
No	10,466 (94.65)	8,651 (45.25)	
Comorbidity: Depression			
Yes	1,746 (15.79)	1,318 (14.75)	0.0428
No	9,311 (84.21)	7,616 (85.25)	
Comorbidity: Diabetes (uncomplicated)			
Yes			
No	2,283 (22.46)	1,646 (18.42)	< .0001
	8,574 (77.54)	7,288 (81.58)	
Comorbidity: Diabetes with chronic complications			
Yes	1,081 (9.78)	587 (6.57)	< .0001
No	9,976 (90.22)	8,347 (93.43)	
Comorbidity: Drug abuse			
Yes	415 (3.75)	326 (3.65)	0.6980
No	10,642 (96.25)	8,608 (96.35)	
Comorbidity: Hypertension			
Yes	6,509 (58.97)	4,463 (49.96)	< .0001
No	4,548 (41.13)	4,471 (50.04)	
Comorbidity: Hypothyroidism			
Yes	1,361 (12.31)	909 (10.17)	< .0001
No	9,696 (87.69)	8,025 (89.83)	
Comorbidity: Liver disease			
Yes	385 (3.48)	221 (2.47)	< .0001
No	10,672 (96.52)	8,713 (97.53)	
Comorbidity: Lymphoma			
Yes	94 (0.85)	64 (0.72)	0.2882
No	10,963 (99.15)	8,870 (99.28)	
Comorbidity: Fluid and electrolyte disorders			
Yes	3,322 (30.04)	1,724 (19.30)	< .0001
No	7,735 (69.96)	7,210 (80.70)	

Comorbidity: Metastatic cancer			
Yes	335 (3.03)	236 (2.64)	0.1014
No	10,722 (96.97)	8,698 (97.36)	
Comorbidity: Other neurological disorders			
Yes	966 (8.74)	638 (7.14)	< .0001
No	10,091 (91.26)	8,296 (92.86)	
Comorbidity: Obesity			
Yes	2,209 (19.98)	1,543 (17.27)	< .0001
No	8,848 (80.02)	7,391 (82.73)	
Comorbidity: Paralysis			
Yes	477 (4.31)	277 (3.10)	< .0001
No	10,580 (95.69)	8,657 (96.90)	
Comorbidity: Peripheral vascular disorders			
Yes	1,206 (10.91)	637 (7.13)	< .0001
No	9,851 (89.09)	8,297 (92.87)	
Comorbidity: Psychoses			
Yes	706 (6.39)	419 (4.69)	< .0001
No	10,351 (93.61)	8,515 (95.31)	
Comorbidity: Pulmonary circulation disorders			
Yes	341 (3.08)	158 (1.77)	< .0001
No	10,716 (96.92)	8,776 (98.23)	
Comorbidity: Renal failure			
Yes	1,486 (13.44)	646 (7.23)	< .0001
No	9,571 (86.56)	8,288 (92.77)	
Comorbidity: Solid tumor without metastasis			
Yes	348 (3.15)	282 (3.16)	0.9706
No	10,709 (96.85)	8,652 (96.84)	
Comorbidity: Peptic ulcer disease excluding bleeding			
Yes	3 (0.02)	1 (0.01)	0.4283
No	11,054 (99.97)	8,933 (99.99)	
Comorbidity: Valvular disease			
Yes	516 (4.67)	322 (3.60)	0.0002
No	10,541 (95.33)	8,612 (96.40)	
Comorbidity: Weight loss			
Yes	1,540 (13.93)	554 (6.20)	< .0001
No	9,517 (86.07)	8,380 (93.80)	
History of MRSA infection			
Yes	234 (2.12)	135 (1.51)	0.0016
No	10,823 (97.88)	8,799 (98.49)	
Transfer in to hospital			
Not transferred in	9,607 (87.35)	8,026 (90.34)	< .0001
Transferred in from a different acute care hospital	906 (8.24)	628 (7.07)	
Transferred in from another type of health facility	485 (4.41)	230 (2.59)	
Transfer out of hospital			
Not transferred out	6,780 (61.38)	6,918 (77.46)	< .0001
Transferred out to a different acute care hospital	267 (2.42)	143 (1.60)	
Transferred out to another type of health facility	3,999 (36.20)	1,870 (20.94)	
Elixhauser Comorbidity Score (Readmission)	110009.5956	8741.5422	< .0001

Elixhauser Comorbidity Score (Mortality)	10696.1256	9129.5025	< .0001

In contrast to hypotheses 1 and 2, not all the listed comorbidities exhibited statistically-significant differences between MSRA-SSI and MSSA-SSI patients. In particular, AIDS, alcohol abuse, rheumatoid arthritis, drug abuse, lymphoma, metastatic cancer, solid tumor, and peptic ulcer disease do not exhibit statistically-significant differences between the groups. There may be some overlap or collinearity between these predictors and it is likely that some of these will be selected out of the final models through the stepwise variable selection process. This hypothesis is supported by the Elixhauser comorbidity scores: both the readmission score and the mortality score are significantly higher for MRSA patients. This is also true for the APRDRG risk of mortality and severity measures. In essence all of these variables attempt to capture the same concept: the severity of illness and comorbidity burden. In the predictive models it will be important to include the best and most efficient measure of severity, and minimize the number if individual comorbidities included. Similarly, patients transferring in to the hospital and patients that transferred out of the hospital are also at increased risk of MRSA infection. Both of these variables also address a component of severity of illness, although the transfer out variable cannot be used to predict the infection as, by definition, it occurs at the end of the hospitalization after the infection has been diagnosed. Finally, current antibiotic use is not statistically-significant between the groups. Patients with long-term antibiotic use are over 5 times more likely to contract a MRSA infection, suggesting that this variable may also be a useful predictor.

The above descriptive analysis and comparisons between MRSA-SSI and MSSA-SSI patients clearly indicate that there are significant differences across these groups of

patients with respect to a range of demographic, clinical and hospital related variables. The goal of the models presented in the following sections is to develop the best performing model using the most parsimonious set of predictors that are known to the clinician close to the time of admission. Thus, variables that are not available until later in the hospital visit, or those that require complex calculations based on information from the medical record will be excluded.

4.6.2 Logistic Regression Model

A logistic regression model was developed to predict MRSA-SSI for patients with *S. aureus* infections. Initially, all variables that were univariately significant according to Table 5 were included in the model and forwards-backwards stepwise variable selection was used to identify the optimal model. The model was optimized by minimizing Akaike Information Criterion (AIC) – a global comparative metric of model performance. The final logistic regression model for hypothesis 3 is shown in Figure 31.

```
MRSA_SSI ~ AGE + ELECTIVE + FEMALE + LOS + NCHRONIC + NDX + NPR +
   PAY1 + RACE + TRAN_IN + ZIPINC_QRTL + MRSAHX + SUSMRSA +
   HOSP_BEDSIZE + HOSP_LOCTEACH + HOSP_REGION + APRDRG_Risk_Mortality +
   APRDRG_Severity + CM_ANEMDEF + CM_CHF + CM_CHRNLUNG + CM_DEPRESS +
   CM_DMCX + CM_LIVER + CM_METS + CM_NEURO + CM_PARA + CM_PSYCH +
   CM_RENLFAIL + CM_VALVE + CM_WGHTLOSS
```

Figure 31: Logistic Regression Model Specification for Hypothesis 3

The final model includes a substantial number of predictors and achieves very strong predictive performance. The relative contribution of each predictor, while adjusting for all others, is presented below in Table 37. Odds ratios quantify the effect of

each predictor on the outcome, 95% confidence intervals provide an estimate of variability, and p-values indicate statistical significance, with p = 0.05 used as the threshold value for significance.

Table 37: Odds Ratios, Confidence Intervals, and p-values for Logistic Regression Model (Hypothesis 3)

Predictor	OR	2.5% CI	97.5% CI	P-Value
(Intercept)	0.37	0.28	0.49	< .0001
Age	1.01	1.01	1.01	< .0001
Surgery type (ref: Non-elective)				
Elective	1.16	1.08	1.25	< .0001
Gender (ref: Male)				
Female	1.23	1.15	1.31	< .0001
Length of stay	1.00	1.00	1.00	< .0001
Number of chronic conditions	0.97	0.95	0.99	0.004
Number of diagnoses	1.02	1.00	1.03	0.005
Number of procedures	1.03	1.02	1.04	< .0001
Primary expected payer (ref: Medicare)				
Medicaid	0.84	0.74	0.94	0.002
Private insurance	0.69	0.64	0.76	< .0001
Self-pay	0.75	0.62	0.90	0.002
No charge	0.97	0.61	1.53	0.894
Other	0.73	0.62	0.85	< .0001
Race (ref: White)				
Black	1.44	1.29	1.61	< .0001
Hispanic	1.01	0.89	1.15	0.875
Asian/Pacific Islander	0.79	0.58	1.07	0.130
Native American	1.33	0.86	2.05	0.204
Other	0.85	0.68	1.05	0.124
Transfer in to hospital (ref: Not transferred in)				
Transferred in from a different acute care hospital	0.90	0.79	1.01	0.083
Transferred in from another type of health facility	1.28	1.07	1.53	0.008
Median income of quartile of patient ZIP code (ref: Quartile 1)				
Quartile 2	0.85	0.78	0.93	< .0001
Quartile 3	0.81	0.74	0.89	< .0001
Quartile 4	0.73	0.66	0.81	< .0001
History of MRSA infection (ref: No history)	1.42	1.11	1.80	0.005
Suspected MRSA infection (ref: No)	6.14	4.18	9.01	< .0001
Hospital bed size(ref: Small)				
Medium	1.15	1.02	1.20	0.-18
Large	1.10	0.99	1.22	0.074
Hospital teaching status (ref: Rural)				
Urban non-teaching	1.00	0.86	1.16	0.992
Urban teaching	0.84	0.73	0.97	0.019
Hospital region (ref: Northeast)				
Midwest	0.86	0.77	0.96	0.009
South	1.29	1.18	1.42	< .0001
West	0.98	0.88	1.09	0.709
Risk of mortality (APRDRG) (ref: No class specified)				
Minor likelihood of dying	1.00	0.91	1.10	0.936

Moderate likelihood of dying	1.17	1.03	1.34	0.019
Major likelihood of dying	1.39	1.14	1.70	0.001
Severity (APRDRG) (ref: No class specified)				
Minor loss of function	1.14	1.01	1.28	0.001
Moderate loss of function	1.34	1.16	1.54	0.030
Major loss of function	1.18	1.47	2.22	< .0001
Comorbidity: Deficiency anemias (ref: No)	1.15	1.07	1.25	< .0001
Comorbidity: Congestive heart failure (ref: No)	1.21	1.06	1.38	0.004
Comorbidity: Chronic pulmonary disease (ref: No)	1.16	1.07	1.25	< .0001
Comorbidity: Depression (ref: No)	1.08	0.98	1.19	0.108
Comorbidity: Diabetes with chronic complications (ref: No)	1.17	1.03	1.33	0.018
Comorbidity: Liver disease (ref: No)	1.30	1.07	1.57	0.007
Comorbidity: Metastatic cancer (ref: No)	0.80	0.66	0.98	0.027
Comorbidity: Other neurological disorders (ref: No)	1.10	0.98	1.25	0.112
Comorbidity: Paralysis (ref: No)	1.14	0.95	1.36	0.157
Comorbidity: Psychoses (ref: No)	1.36	1.18	1.58	< .0001
Comorbidity: Renal failure (ref: No)	1.19	1.05	1.34	0.005
Comorbidity: Valvular disease (ref: No)	0.88	0.75	1.04	0.139
Comorbidity: Weight loss (ref: No)	1.18	1.04	1.33	0.009

Demographic variables most associated with MRSA infection include age, gender, primary payer, race, and income. Females were much more likely to suffer from MRSA-SSI than males (OR: 1.23; 95% CI: 1.15 to 1.31). In comparison to Medicare patients, Medicaid patients (OR: 0.84; 95% CI: 0.74 to 0.94), Private insurance patients (OR: 0.69; 95% CI: 0.64 to 0.76), self-pay patients (OR: 0.75; 95% CI: 0.62 to 0.90), and other patients (OR: 0.73; 95% CI: 0.62 to 0.85) all exhibited reduced risk of MRSA-SSI infection. There was a statistically significant difference in MRSA-SSI between White and Black patients (OR: 1.44; 95% CI: 0.1.29 to 1.61) but no differences between other groups. Income is also shown to affect the odds of MRSA-SSI, with patients in the lowest quartile being most susceptible. Patients in Quartile 2 (OR: 0.85; 95% CI: 0.78 to 0.93), Quartile 3 (OR: 0.81; 95% CI: 0.74 to 0.89), and Quartile 4 (OR: 0.73; 95% CI: 0.66 to 0.81) exhibited progressively lower risk. Although the effect is small, age is also a significant predictor (OR: 1.01; 95% CI: 1.01 to 1.01).

Hospital-related predictors included bed size, teaching status and region. Patients in Medium sized hospitals (OR: 1.15; 95% CI: 1.02 to 1.20) and Large hospitals (OR:

1.10; 95% CI: 0.99 to 1.22) were more likely to contract MRSA-SSI than patients in Small hospitals. In contrast to patients in rural hospitals, patients in urban teaching hospitals (OR: 0.84; 95% CI: 0.73 to 0.97) exhibited reduced risk. Patients in the South (OR: 1.29; 95% CI: 1.18 to 1.42) were more likely to be infected than those in the Northeast, in contrast to patients in the Midwest (OR: 0.86; 95% CI: 0.77 to 0.96) who were less likely to be infected. Patients who transferred in from another type of health facility (OR: 1.28; 95% CI: 1.07 to 1.53) were at higher risk.

Procedure-related variables most predictive of MRSA-SSI include surgery type, with patients undergoing elective surgery substantially more likely to suffer MRSA-SSI compared with those undergoing urgent surgery (OR: 1.66; 95% CI: 1.08 to 1.25). Number of diagnoses was associated with slightly increased risk (OR: 1.02; 95% CI: 1.00 to 1.03) whereas number of chronic conditions was associated with reduced risk (OR: 0.97; 95% CI: 0.95 to 0.99). Number of procedures was also associated with increased risk (OR: 1.03; 95% CI: 1.02 to 1.04). In addition, both of the APRDRG severity measures were significant predictors of infection with the risk of mortality class Moderate likelihood of dying (OR: 1.17; 95% CI: 1.03 to 1.43) and Major likelihood of dying (OR: 1.39; 95% CI: 1.14 to 1.70) having a significant impact. The Severity index, similarly, indicates increased risk. Patients with Minor loss of function (OR: 1.14; 95% CI: 1.01 to 1.28), Moderate loss of function (OR: 1.34; 95% CI: 1.16 to 1.54), and Major loss of function (OR: 1.18; 95% CI: 1.47 to 2.22) were all more likely to contract MRSA-SSI.

Patients with history of MRSA infection (OR: 1.42; 95% CI: 1.11 to 1.80) and suspected MRSA infection on admission (OR: 6.14; 95% CI: 4.18 to 9.01) were all more

likely to have the infection. In addition, several comorbidities were associated with increased risk of MRSA-SSI. These were: deficiency anemias (OR: 1.15; 95% CI: 1.07 to 1.25), congestive heart failure (OR: 1.21; 95% CI: 1.06 to 1.38), chronic pulmonary disease (OR: 1.16; 95% CI: 107 to 1.25), diabetes with chronic complications (OR: 1.17; 95% CI: 1.03 to 1.33), liver disease (OR: 1.30; 95% CI: 1.07 to 1.57), psychoses (OR:1.36; 95% CI: 1.18 to 1.58), renal failure (OR: 1.19; 95% CI: 1.05 to 1.34), and weight loss (OR: 1.18; 95% CI: 1.04 to 1.33).

In contrast, metastatic cancer (OR: 0.80; 95% CI: 0.66 to 0.98) was significantly associated with reduced risk. While the model contains a relatively large number of predictors and the contribution of some predictors is small, the model performs very well.

As described in Chapter 3, the dataset was split into training and test sets, with 85% of the data used for training and 15% of the data used to evaluate the final model. The total sample size was 19,615, which was broken down into training/test and MRSA-SSI/MSSA-SSI cases as shown in Table 38.

Table 38: Training Size for Hypothesis 3 Logistic Regression Model

		Training size	Test size
Class	MSSA-SSI	8,161	1,467
	MRSA-SSI	8,511	1,476

Performance on the logistic regression model is summarized in

Table 39. There were 902 true positives, 952 true negatives, 515 false positives, and 574 false negatives.

Table 39: Confusion Matrix for Hypothesis 3 Logistic Regression Model

		Actual	
		MRSA-SSI	MSSA-SSI
Predicted	MRSA-SSI	902	515
	MSSA-SSI	574	952

Assuming a cutoff of 0.5, sensitivity was calculated as 0.61 and specificity as 0.64. Figure 22 illustrates the overall performance of the model at different cutoff values.

Table 40: Performance of Hypothesis 3 Logistic Regression Model

		Performance
Statistic	Sensitivity	0.61
	Specificity	0.64
	AUC	0.68

The area under the curve was 0.68 which represents reasonable predictive performance.

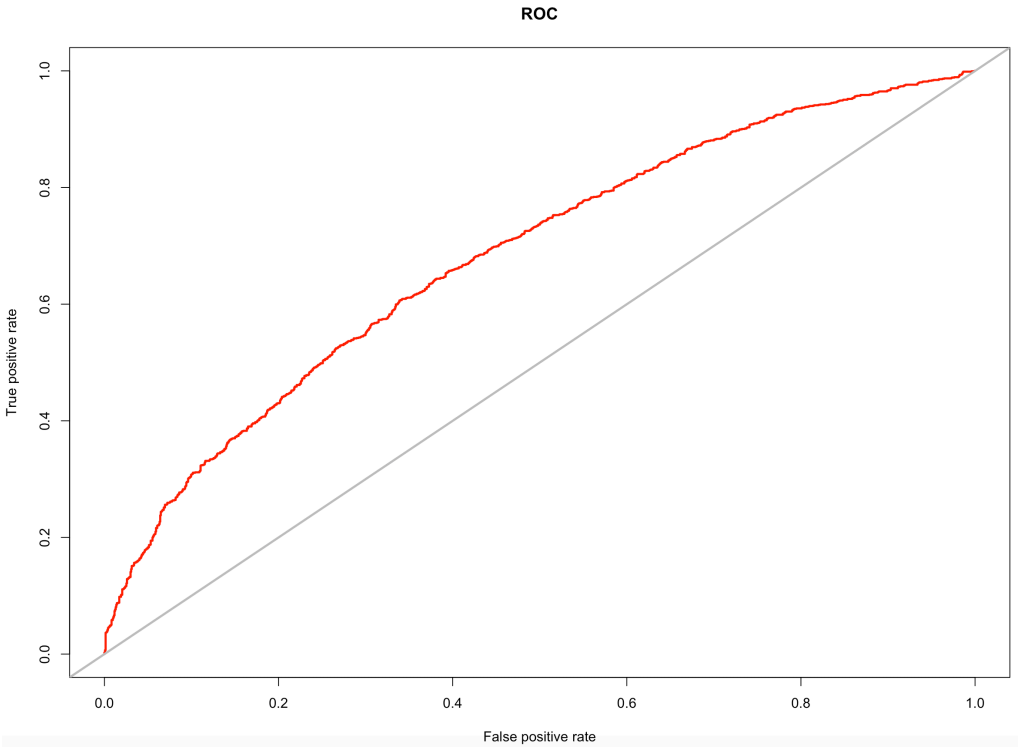

Figure 32: ROC Curve for Hypothesis 3 Logistic Regression Model

In order to assess the overall fit of the model and whether it is over or under fit, a learning curve was plotted.

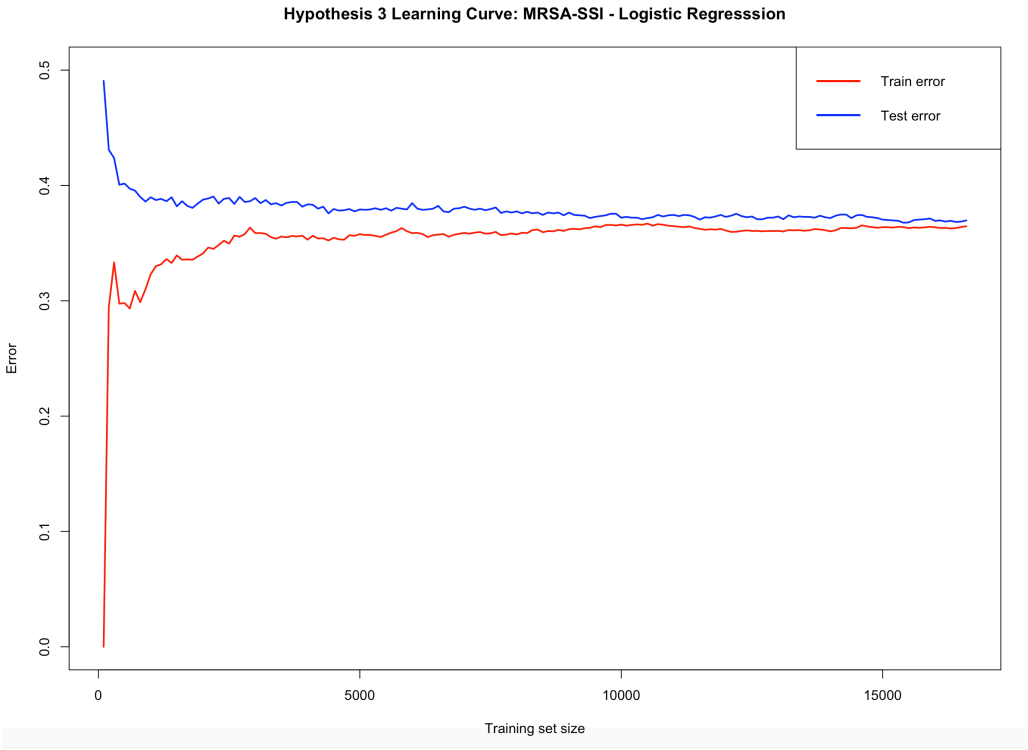

Figure 33: Learning Curve for Hypothesis 3 Logistic Regression Model

For this logistic regression model, training error and test error converge at a relatively small sample size and stay constant as the sample size increases. Adding more training data neither has a positive nor negative effect as both training and test error remain low. Thus, it can be seen that the model has low bias (i.e., it reflects the true model) and high variance meaning it generalizes well to new data.

4.6.3 Artificial Neural Network Model

Predictors for the neural network were the same as those used in the logistic regression model. Similarly, the dataset was split into training and test sets, with 85% of the data used for training and 15% of the data used to evaluate the final model. The

neural network model specification is shown in Figure 34 and the overall sample size and breakdown of training/test data and MRSA/non-MRSA cases is shown in Table 41.

```
MRSA_SSI ~ AGE + ELECTIVE + FEMALE + LOS + NCHRONIC + NDX + NPR +
   PAY1 + RACE + TRAN_IN + ZIPINC_QRTL + MRSAHX + SUSMRSA +
   HOSP_BEDSIZE + HOSP_LOCTEACH + HOSP_REGION + APRDRG_Risk_Mortality +
   APRDRG_Severity + CM_ANEMDEF + CM_CHF + CM_CHRNLUNG + CM_DEPRESS +
   CM_DMCX + CM_LIVER + CM_METS + CM_NEURO + CM_PARA + CM_PSYCH +
   CM_RENLFAIL + CM_VALVE + CM_WGHTLOSS
```

Figure 34: Neural Network Model Specification for Hypothesis 3

The total sample size was 19,615, which was broken down into training/test and MRSA-SSI/MSSA-SSI cases as shown in Table 41. Within the training set, 10-fold cross validation was used to determine the optimal number of nodes in the neural network's hidden layer.

Table 41: Training Size for Hypothesis 3 Neural Network Model

		Training size	Test size
Class	MSSA-SSI	8,161	1,467
	MRSA-SSI	8,511	1,476

Performance on the neural network model is summarized in Table 42. There were 847 true positives, 977 true negatives, 490 false positives, and 629 false negatives.

Table 42: Confusion Matrix for Hypothesis 3 Neural Network Model

		Actual	
		MRSA-SSI	MSSA-SSI
Predicted	MRSA-SSI	847	490
	MSSA-SSI	629	977

Assuming a cutoff of 0.5, sensitivity was calculated as 0.66 and specificity as 0.57.

Table 43: Performance of Hypothesis 3 Neural Network Model

		Performance
Statistic	Sensitivity	0.57
	Specificity	0.66
	AUC	0.67
	Hidden Nodes	26
	Error Rate	0.37

Figure 35 illustrates the overall performance of the model at different cutoff values.

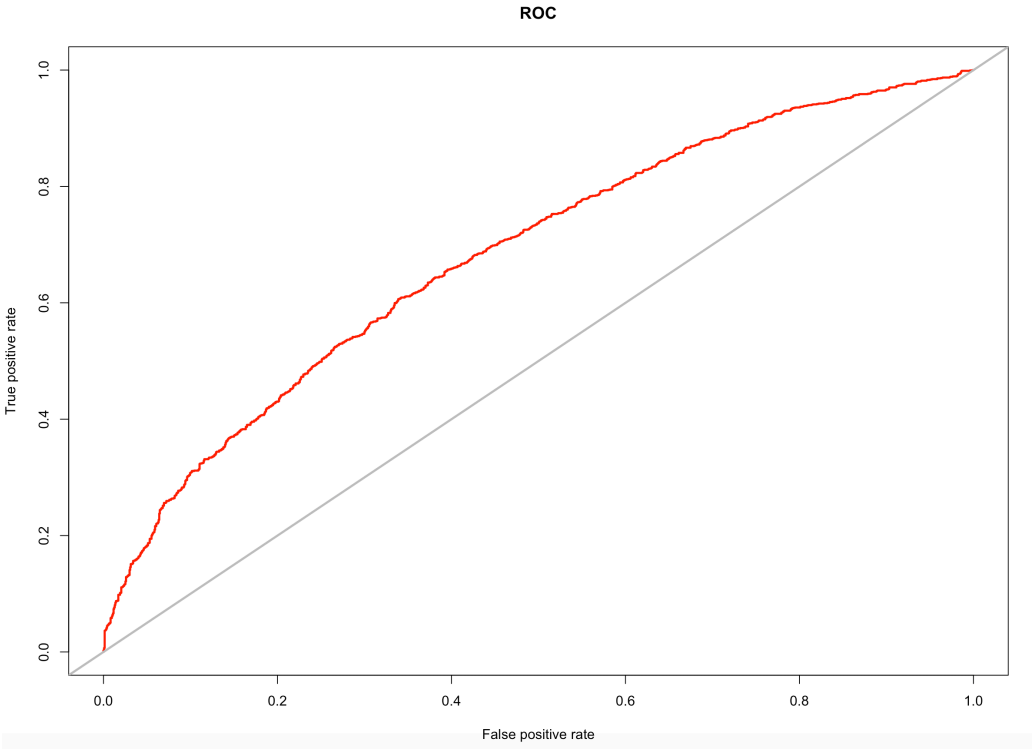

Figure 35: ROC Curve for Hypothesis 3 Neural Network Model

The area under the curve was 0.86 which represents strong predictive performance. The optimal number of hidden nodes is 24 and the overall error rate is 0.20

In order to assess the overall fit of the model and whether it is over or under fit, a learning curve was plotted.

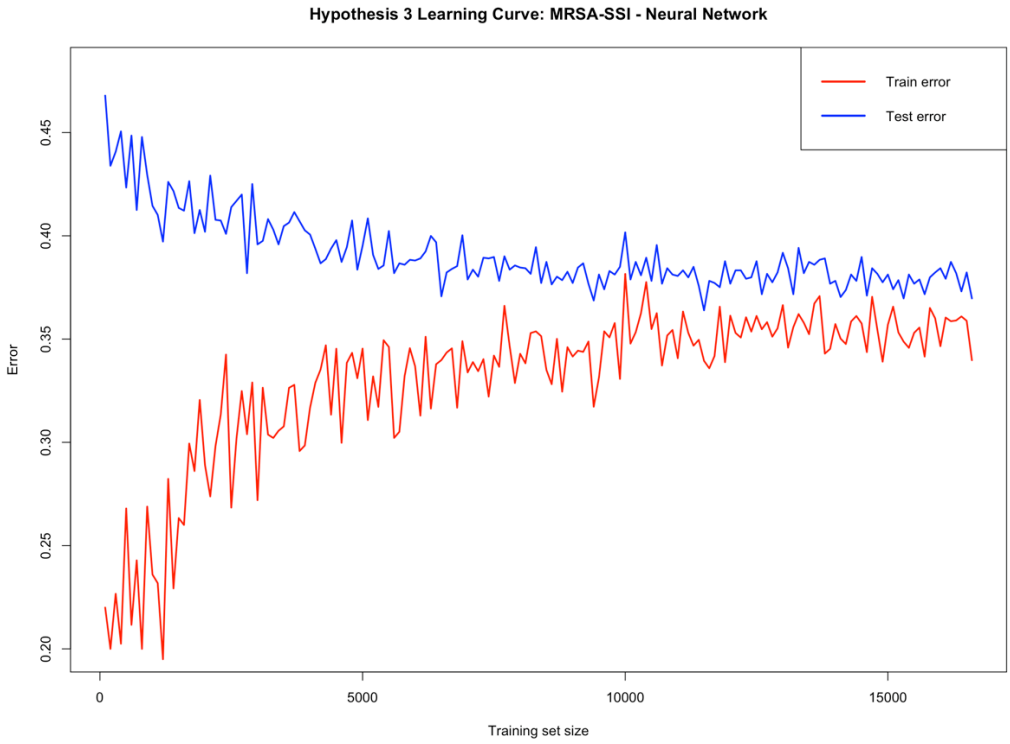

Figure 36: Learning Curve for Hypothesis 3 Neural Network Model

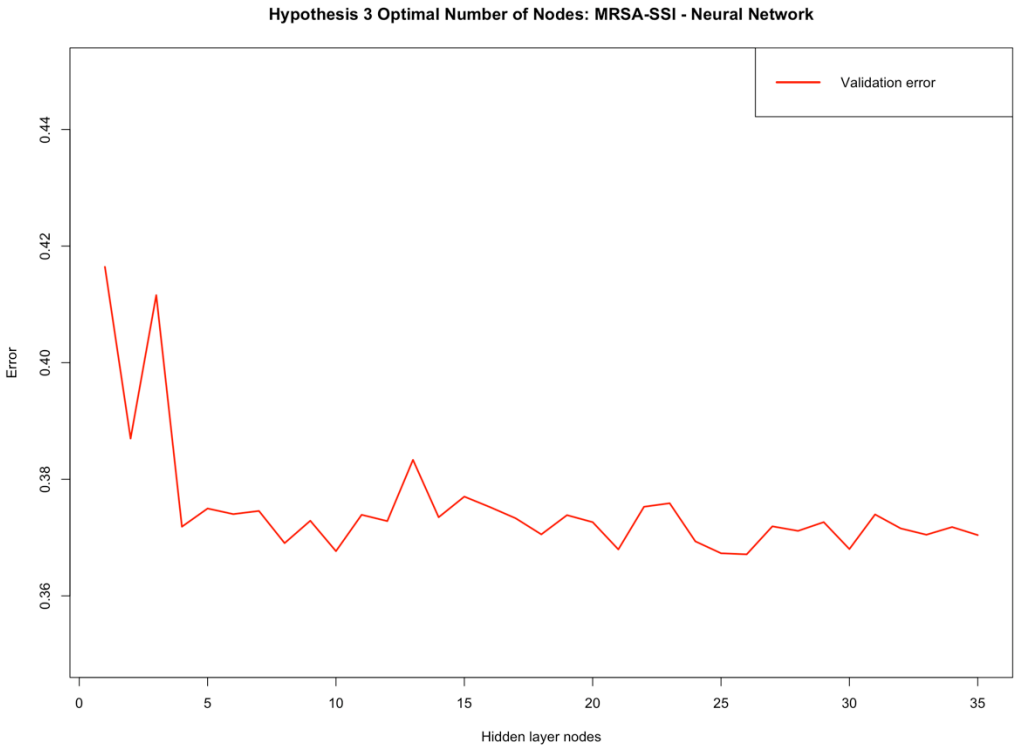

Figure 37: Optimal Number of Hidden Nodes for Hypothesis 3 Neural Network Model

One final tuning parameter for the neural network is the number of nodes in the hidden layer. As indicated above, the optimal number for the hypothesis 3 model is 26. Figure 27 illustrates the error rate for an increasing number of hidden layer nodes. As can be seen, after about 26 nodes the impact of additional nodes is minimal. Although computational power is not an issue and there is no reason to restrict the number of nodes in the hidden layer, this option could be considered if it were needed.

4.6.4 Comparative Performance of the Logistic Regression and Neural Network Models for Hypothesis 3

Table 44 summarizes the relative performance of the logistic regression and neural network models for the prediction of MRSA infection in surgical patients.

Table 44: Comparative Performance of Logistic Regression and Neural Network Models for Hypothesis 3

		Model	
		Logistic Regression	Neural Network
Statistic	Sensitivity	0.61	0.57
	Specificity	0.64	0.66
	AUC	0.68	0.67

Both models perform very well with an AUC of 0.67 or over and good levels of sensitivity and specificity. The logistic regression model exhibits superior performance.

4.7 Clinical Decision Support System

As described in Chapter 3, R Shiny was used to develop a clinical decision support system based on the logistic regression and artificial neural network models developed and evaluated for Hypothesis 1, 2 and 3. The CDSS provides a web-based interface for clinicians to enter the appropriate values for the required risk factors. The CDSS then calculates the probability of infection and makes a treatment recommendation. The application, which works equally well on mobile devices, is illustrated in Figure 38 and is available online.[213]

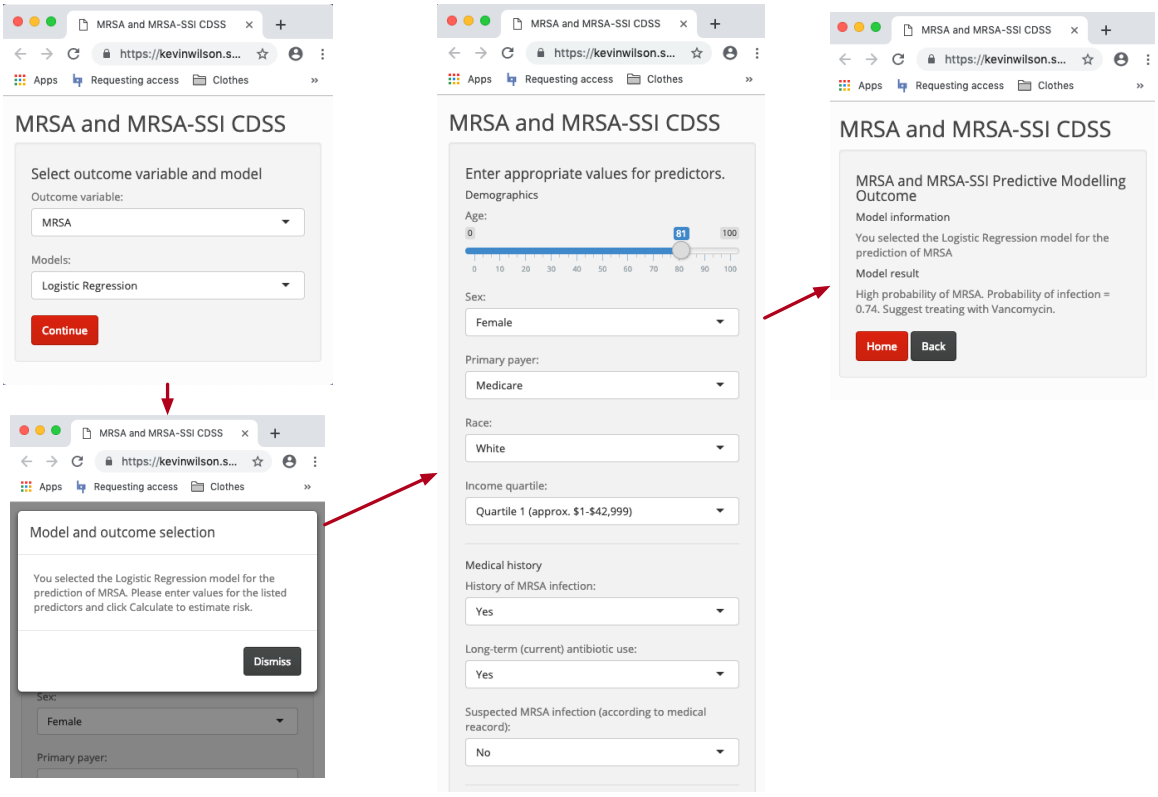

Figure 38: Clinical Decision Support System

CHAPTER 5

DISCUSSION

5.1 Overview

The overarching goal of this research was to develop and evaluate models for the prediction of MRSA and MRSA-related SSI in patients undergoing major surgical procedures in the United States. Additionally, the study aimed to develop a predictive model of MRSA-related SSI that is effectively able to differentiate between resistant and non-resistant strains of *Staphylococcus aureus*. In all cases, the goal was to identify high-risk patients a priori based on easily available demographic, clinical and hospital-level predictors so that appropriate prophylaxis may be given with the hope of preventing infection.

This goal was addressed through a series of research questions, which aimed to: (1) estimate the prevalence of MRSA, SSI, MRSA-related SSI, and MSSA-related SSI in patients undergoing a major surgical procedure in the United States; (2) identify demographic, clinical, and hospital-level risk factors for these infections; (3) develop and evaluate predictive models for these infections; (4) determine whether neural network or logistic regression approaches are more effective; and (5) incorporate the models into a CDSS that could be used by clinicians.

This chapter discusses the results obtained in comparison to the reviewed literature, summarizes key predictors, and provides recommendations for the development and use of CDSSs for the prediction of MRSA-related infections.

5.2 Prevalence

Prevalence was calculated from the NIS dataset using SAS survey procedures. These procedures correctly accounted for sample weights and thus produced nationally-representative estimates.

An average of 10,660,420 major surgical procedures (MSPs) took place annually between 2010 and 2014, and of these procedures, 112,534 resulted in an SSI, representing approximately 1%. Estimates of SSI in the literature are typically based on all surgical procedures, rather than only MSPs, and occur in between 1% and 5% of surgeries.[63,82,84,94] More recent estimates suggest that approximately 157,000 patients acquire an SSI annually in the United States across all surgery types, suggesting that the 112,534 infections estimated may disproportionately reflect major procedures.[59] Thus, patients undergoing major surgical procedures may be at higher risk of SSI than the general surgical population.

With respect to MRSA infection, estimates based on the NIS data suggest that an average of 114,181 patients undergoing a major surgical procedure contract MRSA, which represents approximately 1% and is consistent with previous studies.[16] A subset of these patients, approximately 10,983 annually, also suffer from an SSI and a further 8,962 experience an MSSA-related SSI. Although small, these numbers are consistent

with the literature which suggests that MRSA accounts for over 50% of S. aureus infections.[27]

Because the number of cases of MRSA- and MSSA-related SSI are relatively small, there may not be enough statistical power to differentiate between these infections, based on the available broad clinical predictors, however prediction of MRSA infection in general is likely to be more reliable. Prophylactic treatment of MRSA and MRSA-related SSI would be the same, thus prediction of MRSA infection alone is sufficient to potentially improve surgical outcomes.[3,108]

5.3 Risk Factors for MRSA, MRSA-related SSI, and MRSA-related SSI in S. aureus Patients

As described above, a descriptive analysis was performed prior to the development of predictive models for each of the three research hypotheses. Risk factors identified as part of these descriptive analyses were used as a starting point for determining, which predictors should be included in each of the models.

For Hypothesis 1, the study aimed to develop predictive models of MRSA infection. According to the descriptive analysis, there are several significant risk factors associated with MRSA infection in surgical patients. Consistent with the literature, the descriptive analysis found that patients who contract MRSA infection are typically older[26,41,47,118], spend more time in hospital[38,79,134], and are generally sicker, as measured by the number of chronic health conditions, number of diagnoses, number of procedures performed, and number of comorbidities.[26,113,117,121] Other risk factors identified include gender, with males significantly more likely to be infected than females; race, with black

patients at significantly higher risk than white patients; and socioeconomic-related factors, such as primary payer and income. There is conflicting information about the relationship between gender and MRSA infection.[113,114,117] There is also limited data available on the association between race and MRSA, with only one study having suggested a potential relationship between race and MRSA-related SSI.[16,27]Patients in the lowest income quartile are significantly more likely to be infected with MRSA, as are patients on Medicare. It is unclear from the univariate analysis whether primary payer is an independent predictor or if it is confounded by age.

Other identified risk factors include surgery type, with patients undergoing non-elective procedures being subject to significantly higher risk; region, with patients in the south of the United States experiencing higher risk; and hospital teaching status, with patients being treated in rural hospitals experiencing higher risk. As with age and primary payer, it is unclear from the univariate analysis whether the relationship between hospital region and MRSA infection is confounded by other sociodemographic variables. History of MRSA infection, transfer in to the hospital, and long-term antibiotic use are also associated with MRSA infection. As discussed above, patients that get a MRSA infection are invariably sicker than those that do not. This relationship is demonstrated by the significant association between the NIS-measured comorbidities and MRSA infection. These comorbidities are listed in Table 14 and in all cases, patients with the comorbidity are more likely to contract a MRSA infection than those who do not. These relationships are reinforced by significant differences in the Elixhauser comorbidity score and APRDRG severity scores. Previous studies support these associations, including exposure to antibiotics[2,40,41,47,102,109,120] and prior hospitalization.[8,112,117,120]

Identified risk factors for Hypothesis 2 were similar to those for Hypothesis 1. Patients with MRSA-related SSI were again found to be older, stayed in hospital longer, had a higher number of comorbidities and chronic conditions, had a higher number of procedures, and importantly were in hospital longer prior to their first procedure. Additional risk factors associated with SSI include smoking status[17,136], MRSA colonization,[28,42,49,126,132] and prior antibiotic use.[131,133,137,138] Although these variables are not represented in the NIS dataset, long-term antibiotic use is used as a proxy for prior antibiotic use, and history of MRSA and suspected MRSA infection are used as proxies for MRSA colonization.

As might be expected, for Hypothesis 3, differences between MRSA-related SSI patients and MSSA-related SSI patients were much less pronounced. Although these differences were similar to those for the first two hypotheses, the magnitude of effect was significantly reduced, although most predictors remained statistically significant. Thus, age, length of stay, number of chronic conditions and comorbidities, and number of days to first procedure, all remained risk factors for MRSA related-SSI. This funding supports the existing research, which asserts that MRSA and MSSA are difficult to distinguish clinically.[40]

Based on the results of the descriptive analysis, it seems likely that predictive models will be able to differentiate effectively between MRSA and non-MRSA patients, and indeed, this was the case. In addition, the availability of further risk factors suggests that prediction of MRSA-related SSI is also viable along with the differentiation of MRSA and MSSA in SSI patients. The final selection of variables for each model is discussed explicitly below.

In addition to demographic, hospital and comorbidity factors, a number of clinical and surgical factors have been shown to be associated with MRSA infection. Examples include the use of indwelling devices[2,104,122], surgical wound classification[69,76,78] and clinical factors, such as fever, white blood cell count, and other lab-based test results.[27,110,123] In general these factors are all measures of severity and risk and thus, NIS variables such as the APRDRG risk scores may serve as useful proxies for these factors.

5.4 Comparative Performance of Modeling Approaches

Despite there being a low prevalence of MRSA in the population and therefore proportionally lower prevalence of MRSA-related SSI and MSSA-related SSI, both the Logistic Regression (LR) models and the Artificial Neural Network (ANN) models produced reliable results. In addition, there is no evidence of overfitting, even with the Hypothesis 1 model which contains a relatively large number of predictors. This result is due, in large part, to the large sample size available in the NIS dataset and the use of 5 years' data to train the models.

Results of the three LR models and three ANN models along with their respective sample sizes is shown in Table 45. The performance of the models ranges from good to very good, with the Hypothesis 1 (MRSA) and Hypothesis 2 (MRSA-SSI) models performing very well. The Hypothesis 3 model, which aims to differentiate between MRSA-SSI and MRSA-SS is the weakest model but performance remains well above 0.50 AUC, which suggests that the model may still be useful in practice. The reason for the lower performance of this model is likely a combination of lower sample size and the need to differentiate conditions that are much closer clinically than with Hypothesis 1 or

2. It is reasonable to assume that the general predictors used to differentiate between the two types of *S. aureus* infections would not have as much discriminatory power as those used to differentiate between patients with no condition and MRSA or MRSA-related SSI as in Hypotheses 1 and 2.

Table 45: Comparative Performance of Models

		MRSA		MRSA-SSI		MRSA-SSI vs MSSA-SSI	
		Logistic Regression	Neural Network	Logistic Regression	Neural Network	Logistic Regression	Neural Network
Statistic	Sensitivity	0.79	0.86	0.77	0.73	0.61	0.57
	Specificity	0.75	0.74	0.76	0.87	0.64	0.66
	AUC	0.85	0.87	0.85	0.86	0.68	0.67
	Sample size	173,665		16,682		16,672	
	Prevalence	0.010		0.001		0.551	
	PPV	0.031	0.032	0.003	0.006	0.675	0.673
	NPV	0.997	0.998	1.000	1.000	0.572	0.556
	LR	3.16	3.31	3.17	5.62	1.69	1.68

The AUC scores across all models are very good to excellent, as are sensitivity and specificity scores. Within each hypothesis, relative scores for the logistic regression and artificial neural network approaches are very similar. For Hypothesis 1 — MRSA prediction — both models achieve almost identical performance with the neural network achieving slightly lower specificity and significantly higher sensitivity than the logistic regression model. For Hypothesis 2 — MRSA-SSI prediction — the logistic regression achieves significantly higher sensitivity and significantly lower specificity than the neural network model. For Hypothesis 3 — MRSA-SSI vs MSSA-SSI — sensitivity is slightly

higher with the logistic regression model and specificity is slightly lower than the neural network model. Given the importance of not missing positive cases for each hypothesis, it is therefore important to optimize for sensitivity, and in this respect, the LR models offer stronger performance for Hypothesis 1 and the ANN offers stronger performance for Hypotheses 2 and 3.

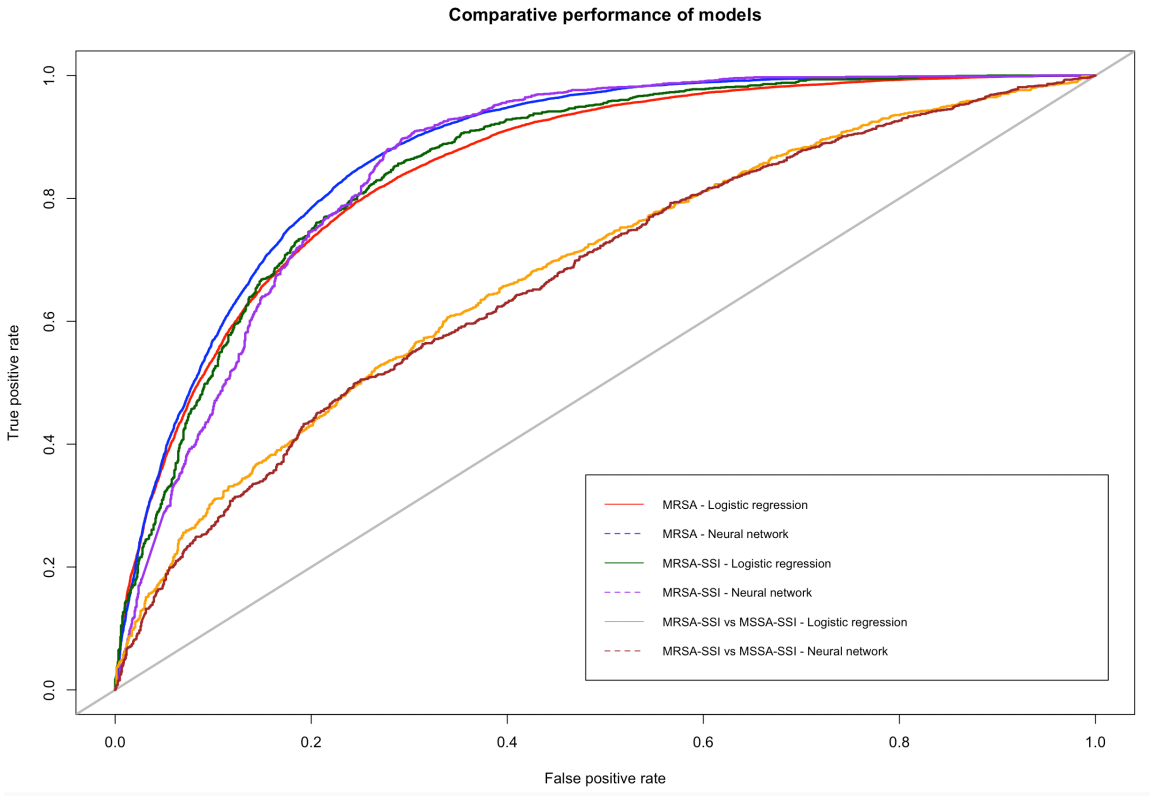

Figure 39: Comparative Performance of Models (AUC)

Overall performance of the models, as measured by AUC, is displayed in Figure 39. As can be seen from the superimposed ROC curves, Hypotheses 1 and 2 perform very similarly with only minor differences in sensitivity and specificity across the range of the ROC curve. With each Hypothesis, the comparative performance of the LR models and the ANN models is similar. As discussed above, differentiation between MRSA-SSI and

MSSA-SSI is more challenging and thus, the models do not perform as well, although the expressive power of logistic regression and the artificial neural network technique remain similar in this case.

5.5 Model Complexity and Predictors

For Hypothesis 1, the forwards-backwards stepwise selection method utilized in development of the LR model yielded a final model with 46 predictors. However, despite the large number of predictors, there was no evidence of overfitting when assessed using a bias/variance learning curve.

A number of predictors were protective, as assessed by odds ratios. Patients undergoing elective surgery, those with private insurance, those who were of Hispanic, Asian, or other race, those who were treated in urban, medium or large size hospitals, and those in the mid-west or west were all less likely to have MRSA. In contrast to the literature, females were also less likely to have MRSA infection.[117]

A range of predictors also increased risk, including age, number of diagnoses, patients with Medicaid, self-pay or those for whom services were not charged, patients of Black or Native American race, and most critically, patients with a history of MRSA, long-term antibiotic use, suspected MRSA infection, and those transferred into the hospital for treatment all exhibited higher risk.

Risk was also affected by a range of comorbidities, the most impactful of which were AIDS, rheumatoid arthritis, diabetes with complications, drug abuse and paralysis. As with the demographic and clinical variables, comorbidities ranged in effect with some being protective.

Variables included in the final LR model were also used for the ANN model and measure several key concepts, such as patient socioeconomics (demographic and financial information), hospital facilities (hospital factors, such as teaching status and bed size), and general severity of disease (comorbidities). Within the set of variables selected by the algorithm were a small number of contradictory findings. For example, number of chronic conditions and number of procedures performed were both found to be protective whereas number of diagnoses was found to increase risk. The most likely explanation for this seemingly discrepant finding is the amelioration of effect due to confounding. The use of AIC to optimize variable selection is known to not necessarily select the most parsimonious set of predictors.[214] Thus, it may be possible to reduce the number of predictors in the model while maintaining or improving performance. The exploration of other model optimization metrics and the development of more parsimonious models is a useful area for further research.

The final models included 8 of the top 20 cited predictors of MRSA infection (refer to Table 2), including prior antibiotic use, loss of function, and MRSA colonization (proxied by suspected MRSA infection/history of MRSA). As expected, some known clinical and surgical predictors could not be included in the model due to their absence from the NIS database. In spite of this the model yielded good performance, thus supporting the hypothesis that MRSA infection can be effectively predicted by demographic, limited clinical (including comorbidity) and hospital-related factors.

For Hypothesis 2, the variable selection algorithm yielded a slightly less complex model with only 34 predictors. As with Hypothesis 1, a number of these factors were protective, such as elective surgery, being female, number of chronic conditions, income

and location. Others, such as age, number of diagnoses, transfer into the hospital from another facility, history of MRSA, long-term (current) antibiotics use and suspected MRSA infection, were shown to increase risk.

In contrast to Hypothesis 1, a smaller number of comorbidities were shown to impact risk of MRSA-SSI. The strongest predictors were AIDS, rheumatoid arthritis, diabetes (both complicated and uncomplicated), presence of a solid tumor, and paralysis. As with Hypothesis 1, the overall performance of the model supports the hypothesis.

Hypothesis 3 yielded the least complex model with only 31 predictors. In contrast to Hypotheses 1 and 2, there was a different set of protective factors. For example, urgent surgery and being male were both protective. In addition, patients who transferred in from another acute care hospital were at reduced risk. Taken together, these results suggest that patients with a *S. aureus* infection are, in general, a sicker, higher-risk population, and one in which urgent hospital treatment is more critical. Demographically, this population is older, and risk of MRSA-SSI increases with higher disease severity and mortality risk.

In contrast to Hypotheses 1 and 2, only 13 comorbidities were included in the final models for hypothesis 3, suggesting that there is a different disease/comorbidity profile in this population. Odds ratios for these predictors were generally lower than in the other models, suggesting a lower impact of individual comorbidities, although it may be the case that the limited sample size available for this analysis is preventing the discovery of other risk factors. The impact of the lower sample size can be assessed easily using the bias/variance curves (Figure 33 and Figure 36), which although they

show convergence as the sample size increases, clearly indicates that the performance of the model is relatively unstable.

The relative contribution of each comorbidities to each hypothesis is illustrated in Figure 40. The groupings of rows and columns represent relative similarity between the rows and columns. For example, Hypotheses 1 and 3 are more similar in terms of comorbidities than Hypotheses 1 and 2. Similarly, individual comorbidities are sorted and grouped based on similarity of impact. The colors in the diagram signify the odds ratios of each comorbidity's contribution to the given hypothesis.

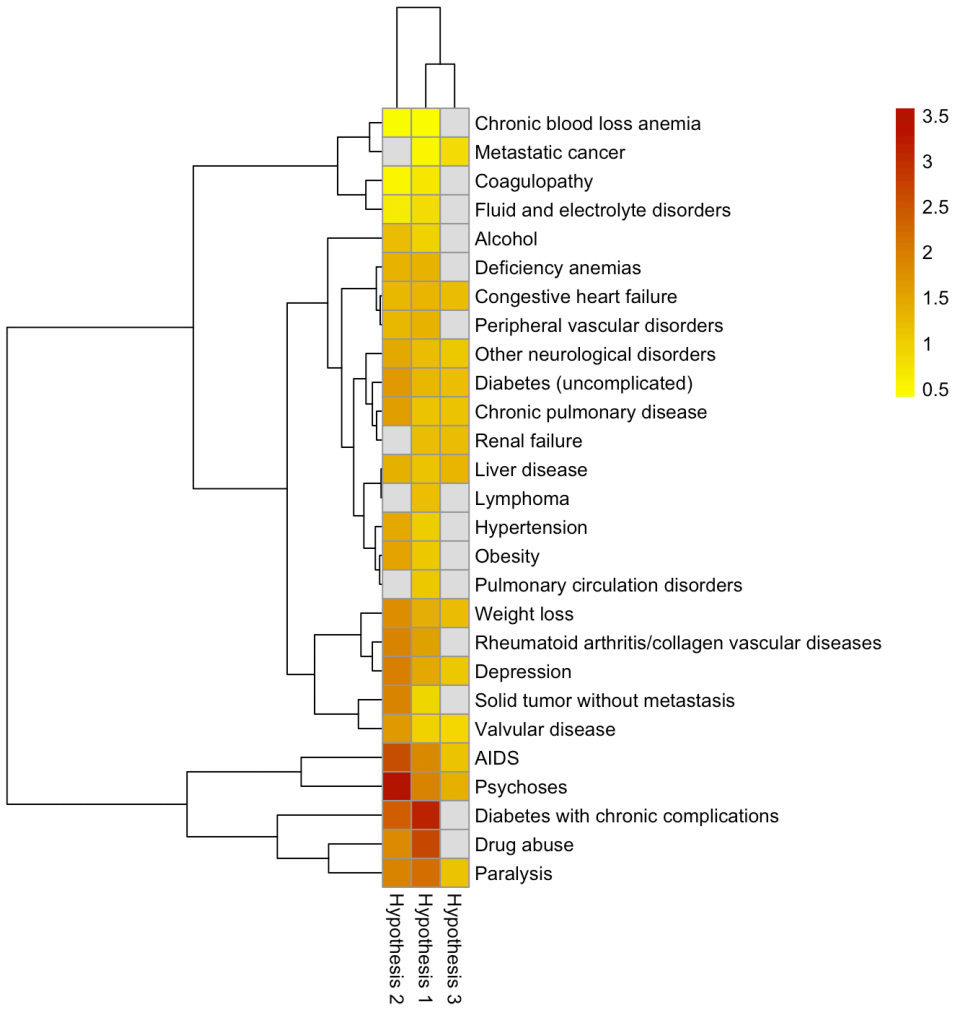

Figure 40: Relative Contribution of Comorbidities to Hypotheses

It can be seen from the heatmap that different comorbidities contribute to the outcome in different degrees based on the specific hypothesis. For example, psychoses has a relatively large contribution to Hypothesis 2 but is less predictive in Hypothesis 1 and 3. In general, for Hypothesis 3, which aims to separate MRSA-related SSI and MSSA-related SSI, there are fewer contributing comorbidities and their impact is generally lower than for Hypotheses 1 and 2. In general, the most significant comorbidities are AIDS, psychoses, diabetes with chronic complications, and paralysis. This is likely because MRSA-related SSI and MSSA-related SSI are very similar and difficult to differentiate clinically, whereas for Hypotheses 1 and 2, clinical differences are more pronounced, and thus the odds ratios of comorbidities are generally higher. This assertion is consistent with the performance of the predictive models, which exhibit relatively strong performance for Hypotheses 1 and 2 and significantly lower performance for Hypothesis 3.

While overall performance of the models is good it is important to understand whether they are good enough to be used in the clinical setting, however, to make this determination it is necessary to understand the pre-test probability in the population. Understanding the positive predictive value and negative predictive value of all these models is necessary to determine their utility when applied to a real clinical population.

5.6 Predictive Value and Clinical Validation

It is essential that a screening test or predictive model have excellent performance characteristics in order to be useful in clinical practice, however it is important to understand that performance metrics, such as sensitivity and specificity are characteristics

of the test itself, and the overall performance of the test is affected by the prevalence of the disease in the population.[212,215] Sensitivity measures the ability of the test (or model) to correctly identify positive cases and is formally defined as "…the proportion of individuals with the disease that have a positive test result..".[215] In contrast, specificity is the ability of the test (or model) to correctly identify negative cases and is formally defined as "…the proportion of individuals without the disease that have a negative test result".[215] In practice, the most critical question facing a clinician when administering a screening test or CDSS is, in the case whether the test/CDSS reports a positive result, what is the probability that the patient actually has the disease? Similarly, in the case where the test/CDSS reports a negative result, what is the probability that the patient actually does not have the disease? These concepts can be assessed by two further metrics: positive predictive value (PPV) and negative predictive value (NPV), both of which are substantially affected by the pre-test probability, which is equivalent to the prevalence of the disease.[212]

PPV, which is also known as the predictive value of a positive test result, is the probability of a patient having the disease given a positive test result. This is a Bayesian construct in which the pre-test probability, or prevalence impacts the final result. NPV, which is also known as the predictive value of a negative test results, is the probability of a patient not having the disease, given a negative test result. PPV and NPV are impacted by the a priori prevalence of the disease in the population. In the case of a rare disease, such as MRSA or MRSA-related SSI, PPV will invariably be very low while NPV will likely be very high. PPV and NPV are functions of sensitivity, specificity and

prevalence.[212,215] Prevalence, PPV and NPV for Hypothesis 1-3 are summarized in Table 45.

For Hypotheses 1 and 2, the population is defined as all patients undergoing an MSP, prevalence of MRSA in this population is approximately 1%, and prevalence of MRSA-related SSI is an order of magnitude lower, at 0.1%. In contrast, the population defined for Hypothesis 3 is those surgical patients with a *S. aureus* infection. The PPVs for Hypothesis 1 are 3.1% and 3.2% for the logistic regression and ANN models, respectively, while the NPVs are 99.7% and 99.8%. As would be expected, the PPVs for Hypothesis 2 are 0.3% and 0.6%, with the NPVs in both cases, 100%. For Hypothesis 3, the PPVs are 67.5% and 67.3% and the NPVs are 57.2% and 55.6%.

The low PPVs for Hypotheses 1 and 2 mean that it is likely not feasible to use the models for diagnostic purposes, however they may be useful indicators that further testing needs to be done. This assertion is supported by the high NPVs, which give an extremely high probability that if a test result is negative, the patient does not have the disease. Thus, the models are extremely effective at ruling out patients that do not have the disease, providing the opportunity for clinicians to perform more targeted testing on those identified as having the disease. Given the lack of routine testing for MRSA colonization or infection upon hospital admission, this approach may have high value for clinicians and could serve to reduce incidence and improve outcomes.[46,8,39,41,109,113] The higher PPV and lower NPV for Hypothesis 3 are due to the higher prevalence of MRSA-related SSI in patients with *S. aureus* infection. Despite generally lower performing models, the probability of the model prediction being correct are commensurately higher. This phenomenon suggests that one approach to maximizing the utility of a predictive

model for a low prevalence disease is to apply the model only to patients determined to be at high risk. Another approach is to use the high NPV of the models to exclude patients from consideration for further testing.

An alternative measure of model performance, which is independent of prevalence and also defined in terms of sensitivity and specificity, is the likelihood ratio. The likelihood ratio defines how much more likely it is that a person for which the model predicts the disease has the disease in comparison to a person that does not. Likelihood ratios for Hypotheses 1 to 3 are provided in Table 45. By this measure, it can be seen that the models for Hypotheses 1 and 2 perform well while the models for Hypothesis 3 relatively do not. Taken together, these results suggest that different metrics can be used to assess different aspects of a model's performance and thus form more a complete picture, however, it would be ideal to validate the model in a clinical setting.

Validation using a real clinical cohort would provide the opportunity to assess and understand the performance of the models in a realistic setting. Model parameters, such as the cut-off value used to determine whether the model predicts whether a patient has the disease can be adjusted to better reflect the true population. By validating the models in a real-world setting it will be possible to better understand their performance characteristics.

5.7 Clinical Decision Support System

The CDSS, illustrated in Figure 38, provides a wizard-like interface, which allows the clinical to select the outcome and modeling approach. Based on this selection, the user is presented with the appropriate predictors. Once the user has selected values for the

appropriate predictors, the system calculates the probability of the outcome and presents an appropriate message to the user. Predictors are grouped into demographics (age, sex, primary payer, race, income quartile), medical history (history of MRSA, long-term antibiotic use, suspected MRSA infection), hospital and procedural factors (length of stay, surgery type, number of chronic conditions, number of diagnoses, number of procedures performed, patient transferred in to hospital, hospital bed size, location/teaching status, region) and comorbidities (see discussion above). The user interface automatically adjusts to reflect the predictors required for each model. Numerical variables are presented as sliders and categorical variables are presented as dropdown lists. The user can navigate back and forth in the system to view the effect of changing the values of the predictors or the modeling approach.

While the CDSS is easy to use, in practice it would be important to limit the amount of data entry to the absolute minimum. This could easily be achieved through the presetting of hospital-level variables in the CDSS configuration, as these variables do not change for each patient — rather they are constant for the setting. Similarly, it may also be possible to connect the CDSS directly to the EHR system, and in this way automatically set the values of the key clinical variables, such as history of MRSA. If incorporated into the clinician's work flow in this way, CDSSs in general have been shown to reduce medical errors and increase compliance with guidelines.[115,147,160,166]

If integrated in this way, the CDSS would be able easily to identify patients without MRSA or MRSA-SSI and exclude these patients from further testing. In contrast, patients with a positive result from the CDSS could be referred for further testing. This approach addresses challenges with current MRSA and SSI treatment, including

inappropriate administration of antibiotics[53,54,100,175] and lack of uniformity in testing.[54,146,154] If integrated, validated, and used appropriately, the CDSS could have a significant impact on patients at high risk of MRSA and MRSA-related SSI. As an example, effective treatment of all MRSA infections could result in a reduction of up to 81,000 deaths annually and prevention of all SSIs would save the United States $10 billion per year and reduce the cost of individual treatment and length of stay by up to 50%.[8,21,22,24,34]

CHAPTER 6

CONCLUSIONS

6.1 Summary

Methicillin-resistant *Staphylococcus aureus* (MRSA) is the leading cause of antibiotic resistance related mortality in surgical patients, and MRSA-related surgical site infections (SSIs) remain a critical medical issue. Effective prediction of MRSA and MRSA-related SSI would facilitate the prophylactic use of appropriate antibiotics or application of other prevention techniques, which have been shown to improve clinical outcomes. While there is a range of patient, procedural and hospital level factors that have been shown to increase the risk of a surgical patient contracting a MRSA infection, research is less clear on the best approaches to developing predictive models for incorporation into a clinical decision support system. This study compared two common modeling approaches — logistic regression (LR) and artificial neural networks (ANN) — for the prediction of MRSA infection in patients undergoing major surgical procedures (MSPs) in the United States. Other research questions aimed to identify the prevalence of MRSA, SSI, MRSA-related SSI, and MSSA-related SSI in patients undergoing major surgical procedures in the US, and to identify and assess the significance of risk factors for these diseases.

Of the 10,660,420 surgical procedures that occur on average each year in the US, 114,181 (1.07%) had a confirmed MRSA infection, 112,534 had an SSI (1.07%), 10,983 had MRSA-related SSI (0.10%) and 8,962 on average had an MSSA-related SSI (0.10%).

Both the LR and ANN models performed well for the prediction of MRSA. The sensitivity of the LR model was 0.79, the specificity was 0.75, and the area under the curve (AUC) was 0.85. The sensitivity of the ANN model was 0.86, the specificity was 0.74, and the AUC was 0.87. Given the need to correctly identify positive patients, the ANN model is the optimal model in this case. Because of low prevalence, the positive predictive values (PPV) of these models were low, however the negative predictive values (NPV) were high, which suggests that the models should be used to exclude negative patients from further testing, thereby supporting the efficient use of resources, and supporting the testing of only high-risk patients.

Similar results were obtained for the prediction of MRSA-related SSI. The sensitivity of the LR model was 0.77, the specificity was 0.76, and the AUC was 0.85. The sensitivity of the ANN model was 0.73, the specificity was 0.87, and the AUC was 0.86. Given the need to correctly identify positive patients, the LR model is the optimal model in this case. Similar to MRSA models, because of low prevalence, the positive predictive values (PPV) of the MRSA-related SSI models were low, however the negative predictive values (NPV) were high, which suggests that the models should be used to exclude negative patients from further testing, thereby supporting the efficient use of resources, and supporting the testing of only high-risk patients.

Finally, results for the prediction of MRSA-related SSI in the population of patients with *S. aureus* infections were not as good as the MRSA and MRSA-related SSI

models. The sensitivity of the LR model was 0.61, the specificity was 0.57, and the AUC was 0.68. The sensitivity of the ANN model was 0.57, the specificity was 0.66, and the AUC was 0.67. Given the need to correctly identify positive patients, the LR model is the optimal model in this case. In contrast to the other models, the PPV was higher due to the higher prevalence of MRSA in the *S. aureus* infected surgical population. This advantage is offset somewhat by lower sensitivity and specificity, which are predominantly due to a lower sample size.

All the models were incorporated into a web-based CDSS, which is available for review.[213]

6.2 Limitations of Research

This research study had a number of limitations primarily due to the relatively low prevalence of MRSA and MRSA-related SSI in patients undergoing a major surgical procedure. Although the NIS sample size was very large, when filtered for the outcomes being studied, the sample sizes were significantly smaller. This was particularly problematic for the prediction of MRSA-SSI in patients with *S. aureus* infections. These models were generally unstable and had lower performance than the larger models. Similarly, the low prevalence of the diseases resulted in low positive predictive values for the MRSA and MRSA-related SSI, which limited the utility of the models to ruling out patients without the disease.

Another limitation is that the NIS dataset, although fairly rich, is lacking in detailed clinical information and thus some key risk factors for MRSA and MRSA-related SSI were unable to be included in the analysis. Specific examples include lab

values, such as hematocrit, confirmation of MRSA colonization, and white blood cell count. In addition, the availability of surgical factors to the extent that they were known ahead of the procedure (e.g., anticipated duration), would also be useful predictors.

6.3 Recommendations for Future Research

There are a number of recommendations for future research, the most important of which is the performance of a validation study. In any CDSS model it is important to validate and tune the model in a real clinical cohort. In this way, it is possible to determine how best to implement the model and how the information obtained from the CDSS should inform clinical practice.

A second recommendation, particularly for the *S. aureus* model, is to develop a new model using a larger sample size, for example, with additional years' data from the NIS. This is important in order to determine whether an increased sample size would result in more stable models that exhibit better performance or if the expressive power of the model is inherently limited due to the similarity of MRSA- and MSSA-related SSI patients. Related to this recommendation is the need to assess further modeling approaches, such as Support Vector Machines or Random Forests, which have the potential to yield further performance improvements.

A final recommendation is to further study the predictors included in the models to determine if smaller, more parsimonious models could be developed. These models would be easier to administer and likely exhibit stable performance; however it is important to understand whether smaller models could achieve comparable levels of performance with the larger models.

References

1. Khan HA, Baig FK, Mehboob R. Nosocomial infections: Epidemiology, prevention, control and surveillance. *Asian Pacific Journal of Tropical Biomedicine.* 2017;7(5):478-482.

2. Wang FD, Chen YY, Chen TL, Lin YT, Fung CP. Risk factors and mortality of nosocomial infections of methicillin-resistant Staphylococcus aureus in an intensive care unit. *J Crit Care.* 2011;26(1):82-88.

3. Brooks HL, Hodson J, Richardson SJ, Stezhka L, Gill MJ, Coleman JJ. Improving the timeliness of meticillin-resistant Staphylococcus aureus antimicrobial decolonization therapy administration: a descriptive account. *J Hosp Infect.* 2014;86(3):209-215.

4. Yang S, Rider BB, Baehr A, Ducoffe AR, Hu DJ. Racial and ethnic disparities in health care-associated Clostridium difficile infections in the United States: State of the science. *Am J Infect Control.* 2016;44(1):91-96.

5. Nazir A, Kadri S. An overview of hospital acquired infections and the role of the microbiology laboratory. *International Journal of Research in Medical Sciences.* 2014;2(1).

6. Lake JG, Weiner LM, Milstone AM, Saiman L, Magill SS, See I. Pathogen Distribution and Antimicrobial Resistance Among Pediatric Healthcare-Associated Infections Reported to the National Healthcare Safety Network, 2011-2014. *Infect Control Hosp Epidemiol.* 2018;39(1):1-11.

7. Al-Tawfiq JA, Tambyah PA. Healthcare associated infections (HAI) perspectives. *J Infect Public Health.* 2014;7(4):339-344.

8. Sydnor ER, Perl TM. Hospital epidemiology and infection control in acute-care settings. *Clin Microbiol Rev.* 2011;24(1):141-173.

9. Centers for Disease Control and Prevention. Preventing Healthcare-Associated Infections. *CDC At Work* 2018; https://www.cdc.gov/washington/~cdcatwork/pdf/infections.pdf. Accessed 6/2/2018, 2018.

10. Centers for Disease Control and Prevention. Healthcare-associated Infections in the United States, 2006-2016: A Story of Progress. 2018; https://www.cdc.gov/hai/surveillance/data-reports/data-summary-assessing-progress.html, 2018.

11. Magill SS, Edwards JR, Bamberg W, et al. Multistate point-prevalence survey of health care-associated infections. *N Engl J Med.* 2014;370(13):1198-1208.

12. Centers for Disease Control and Prevention. HAI Data and Statistics. 2018; https://www.cdc.gov/hai/surveillance/index.html.

13. Rebecca R. Roberts a, R. Douglas Scott IIa, Ralph Cordell a, et al. The Use of Economic Modeling to Determine the Hospital Costs Associated with Nosocomial Infections. *Clinical Infectious Diseases.* 2003(11):1424.

14. Khan HA, Ahmad A, Mehboob R. Nosocomial infections and their control strategies. *Asian Pacific Journal of Tropical Biomedicine.* 2015;5(7):509-514.

15. Emori TG, Gaynes RP. An overview of nosocomial infections, including the role of the microbiology laboratory. *Clinical Microbiology Reviews.* 1993;6(4):428-442.

16. Allareddy V, Das A, Lee MK, et al. Prevalence, predictors, and outcomes of methicillin-resistant Staphylococcus aureus infections in patients undergoing major surgical procedures in the United States: a population-based study. *Am J Surg.* 2015;210(1):59-67.

17. Mangram AJ, Horan TC, Pearson ML, Silver LC, Jarvis WR. Guideline for Prevention of Surgical Site Infection. *Infection Control and Hospital Epidemiology.* 1999;20(4).

18. Gani F, Makary MA, Pawlik TM. The price of surgery: markup of operative procedures in the United States. *J Surg Res.* 2017;208:192-197.

19. Steiner CA, Karaca Z, Moore BJ, Imshaug MC, Pickens G. Surgeries in Hospital-Based Ambulatory Surgery and Hospital Inpatient Settings, 2014. 2018; https://www.hcup-us.ahrq.gov/reports/statbriefs/sb223-Ambulatory-Inpatient-Surgeries-2014.jsp. Accessed 6/2/2018, 2018.

20. Bhattacharya S, Pal K, Jain S, Chatterjee SS, Konar J. Surgical Site Infection by Methicillin Resistant Staphylococcus aureus- on Decline? *J Clin Diagn Res.* 2016;10(9):DC32-DC36.

21. Sganga G, Tascini C, Sozio E, et al. Focus on the prophylaxis, epidemiology and therapy of methicillin-resistant Staphylococcus aureus surgical site infections and a position paper on associated risk factors: the perspective of an Italian group of surgeons. *World J Emerg Surg.* 2016;11:26.

22. Humphreys H, Becker K, Dohmen PM, et al. Staphylococcus aureus and surgical site infections: benefits of screening and decolonization before surgery. *J Hosp Infect.* 2016;94(3):295-304.

23. Lyden JR, Dellinger EP. Surgical Site Infections. *Hospital Medicine Clinics.* 2016;5(2):319-333.

24. Bagdasarian N, Schmader KE, Kaye KS. The Epidemiology and Clinical Impact of Surgical Site Infections in the Older Adult. *Current Translational Geriatrics and Experimental Gerontology Reports.* 2013;2(3):159-166.

25. Onyedum CC, Alobu I, Ukwaja KN. Prevalence of drug-resistant tuberculosis in Nigeria: A systematic review and meta-analysis. *PLoS One.* 2017;12(7):e0180996.

26. Luszczynska A, Gunson KS. Predictors of asking medical personnel about handwashing: the moderating role of patients' age and MRSA infection status. *Patient Educ Couns.* 2007;68(1):79-85.

27. Agrawal R, Sharma D, Dhiman P, Patro DK. Clinical and haematological predictors of acute hematogenous Methicillin Resistant Staphylococcus aureus (MRSA) osteomyelitis & septic arthritis. *J Orthop.* 2015;12(3):137-141.

28. Akhtar A, Kadir H, Chandran P. Surgical site infection risk following pre-operative MRSA detection in elective orthopaedic surgery. *J Orthop.* 2014;11(3):117-120.

29. de Lencastre H, Oliveira D, Tomasz A. Antibiotic resistant Staphylococcus aureus: a paradigm of adaptive power. *Curr Opin Microbiol.* 2007;10(5):428-435.

30. Nordmann P, Naas T, Fortineau N, Poirel L. Superbugs in the coming new decade; multidrug resistance and prospects for treatment of Staphylococcus aureus, Enterococcus spp. and Pseudomonas aeruginosa in 2010. *Curr Opin Microbiol.* 2007;10(5):436-440.

31. Klein EY, Mojica N, Jiang W, et al. Trends in Methicillin-Resistant Staphylococcus aureus Hospitalizations in the United States, 2010-2014. *Clin Infect Dis.* 2017;65(11):1921-1923.

32. Rice LB. Antimicrobial resistance in gram-positive bacteria. *Am J Med.* 2006;119(6 Suppl 1):S11-19; discussion S62-70.

33. Jarvis WR, Martone WJ. Predominant pathogens in hospital infections. *J Antimicrob Chemother.* 1992;29 Suppl A:19-24.

34. Elixhauser A, Steiner CA. *Infections with Methicillin-Resistant Staphylococcus Aureus (MRSA) in U.S. Hospitals, 1993–2005: Statistical Brief #35.*: Healthcare Cost and Utilization Projecty;2007.

35. Gurusamy KS, Koti R, Toon CD, Wilson P, Davidson BR. Antibiotic therapy for the treatment of methicillin-resistant Staphylococcus aureus (MRSA) infections in surgical wounds. *Cochrane Database Syst Rev.* 2013(8):CD009726.

36. Weigelt J, Kaafarani HM, Itani KM, Swanson RN. Linezolid eradicates MRSA better than vancomycin from surgical-site infections. *Am J Surg.* 2004;188(6):760-766.

37. Baddour MM. *MRSA (methicillin resistant Staphylococcus Aureus) infections and treatment.* New York: Nova Science Publishers; 2010.

38. Robicsek A, Beaumont JL, Paule SM, et al. Universal surveillance for methicillin-resistant Staphylococcus aureus in 3 affiliated hospitals. *Ann Intern Med.* 2008;148(6):409-418.

39. Harbarth S, Fankhauser C, Schrenzel J, et al. Universal screening for methicillin-resistant Staphylococcus aureus at hospital admission and nosocomial infection in surgical patients. *JAMA.* 2008;299(10):1149-1157.

40. Ju KL, Zurakowski D, Kocher MS. Differentiating between methicillin-resistant and methicillin-sensitive Staphylococcus aureus osteomyelitis in children: an evidence-based clinical prediction algorithm. *J Bone Joint Surg Am.* 2011;93(18):1693-1701.

41. Pan A, Lee A, Cooper B, et al. Risk factors for previously unknown meticillin-resistant Staphylococcus aureus carriage on admission to 13 surgical wards in Europe. *J Hosp Infect.* 2013;83(2):107-113.

42. Iqbal HJ, Ponniah N, Long S, Rath N, Kent M. Review of MRSA screening and antibiotics prophylaxis in orthopaedic trauma patients; The risk of surgical site infection with inadequate antibiotic prophylaxis in patients colonized with MRSA. *Injury.* 2017;48(7):1382-1387.

43. Hatachi T, Tachibana K, Inata Y, et al. Risk Factors for Healthcare-Associated Infections After Pediatric Cardiac Surgery. *Pediatr Crit Care Med.* 2018;19(3):237-244.

44. Harbarth S, Huttner B, Gervaz P, et al. Risk factors for methicillin-resistant Staphylococcus aureus surgical site infection. *Infect Control Hosp Epidemiol.* 2008;29(9):890-893.

45. Harbarth S, Sax H, Uckay I, et al. A predictive model for identifying surgical patients at risk of methicillin-resistant Staphylococcus aureus carriage on admission. *J Am Coll Surg.* 2008;207(5):683-689.

46. Jung WJ, Kang YA, Park MS, et al. Prediction of methicillin-resistant Staphylococcus aureus in patients with non-nosocomial pneumonia. *BMC Infect Dis.* 2013;13:370.

47. Harbarth S, Huttner B, Gervaz P, et al. Risk Factors for Methicillin-Resistant Staphylococcus aureus Surgical Site Infection. *Infection Control and Hospital Epidemiology.* 2008;29(9):890-893.

48. Kavanagh KT, Abusalem S, Calderon LE. The incidence of MRSA infections in the United States: is a more comprehensive tracking system needed? *Antimicrob Resist Infect Control.* 2017;6:34.

49. Murphy E, Spencer SJ, Young D, Jones B, Blyth MJ. MRSA colonisation and subsequent risk of infection despite effective eradication in orthopaedic elective surgery. *J Bone Joint Surg Br.* 2011;93(4):548-551.

50. Olderog CK, Schmitz GR, Bruner DR, Pittoti R, Williams J, Ouyang K. Clinical and epidemiologic characteristics as predictors of treatment failures in uncomplicated skin abscesses within seven days after incision and drainage. *J Emerg Med.* 2012;43(4):605-611.

51. Goodman KE, Lessler J, Cosgrove SE, et al. A Clinical Decision Tree to Predict Whether a Bacteremic Patient Is Infected With an Extended-Spectrum beta-Lactamase-Producing Organism. *Clin Infect Dis.* 2016;63(7):896-903.

52. Health Care Utilization Project. NIS Database Documentation. 2018; https://www.hcup-us.ahrq.gov/db/nation/nis/nisdbdocumentation.jsp.

53. Macesic N, Polubriaginof F, Tatonetti NP. Machine learning: novel bioinformatics approaches for combating antimicrobial resistance. *Curr Opin Infect Dis.* 2017;30(6):511-517.

54. Huh K, Chung DR, Park HJ, et al. Impact of monitoring surgical prophylactic antibiotics and a computerized decision support system on antimicrobial use and antimicrobial resistance. *Am J Infect Control.* 2016;44(9):e145-152.

55. Chaudhary R, Kumar Thapa S, Chani Rana J, Shah P. *Surgical Site Infections and Antimicrobial Resistance Pattern.* Vol 152017.

56. Yao CM, Ziai H, Tsang G, et al. Surgical site infections following oral cavity cancer resection and reconstruction is a risk factor for plate exposure. *J Otolaryngol Head Neck Surg.* 2017;46(1):30.

57. Salkind AR, Rao KC. Antibiotic prophylaxis to prevent surgical site infections. *Am Fam Physician.* 2011;83(5):585-590.

58. Meijs AP, Ferreira JA, SC DEG, Vos MC, Koek MB. Incidence of surgical site infections cannot be derived reliably from point prevalence survey data in Dutch hospitals. *Epidemiol Infect.* 2017;145(5):970-980.

59. Stanic S, Bojanic J, Grubor P, Mijovic B, Maric V. Examination of Risk Factors for the Development of Surgical Site Infections. *Mater Sociomed.* 2017;29(2):134-137.

60. Calina D, Docea AO, Rosu L, et al. Antimicrobial resistance development following surgical site infections. *Mol Med Rep.* 2017;15(2):681-688.

61. Akhi MT, Ghotaslou R, Alizadeh N, pirzadeh T, Beheshtirouy S, Memar MY. High frequency of MRSA in surgical site infections and elevated vancomycin MIC. *Wound Medicine.* 2017;17:7-10.

62. Kassavin DS, Pascarella L, Goldfarb MA. Surgical site infections: incidence and trends at a community teaching hospital. *Am J Surg.* 2011;201(6):749-753.

63. Anderson DJ. Surgical site infections. *Infect Dis Clin North Am.* 2011;25(1):135-153.

64. Owens PL, Barrett ML, Raetzman S, Maggard-Gibbons M, Steiner CA. Surgical site infections following ambulatory surgery procedures. *JAMA.* 2014;311(7):709-716.

65. Rubin RH. Surgical wound infection: epidemiology, pathogenesis, diagnosis and management. *BMC Infect Dis.* 2006;6:171.

66. Dubrovsky G, Huynh N, Rouch JD, et al. Subcutaneous cefazolin to reduce surgical site infections in a porcine model. *J Surg Res.* 2018;224:156-159.

67. Adams A. Surgical Site Ingections - Another Look at Banishing Bugs. *Nursing2000.* 2000;30(6):50-51.

68. Guo K, Ren J, Li G, et al. Risk factors of surgical site infections in patients with Crohn's disease complicated with gastrointestinal fistula. *Int J Colorectal Dis.* 2017;32(5):635-643.

69. Piper KF, Tomlinson SB, Santangelo G, et al. Risk factors for wound complications following spine surgery. *Surg Neurol Int.* 2017;8:269.

70. Lawson EH, Hall BL, Ko CY. Risk Factors for Superficial vs Deep/Organ-Space Surgical Site Infections. *JAMA Surgery.* 2013;148(9).

71. Ikeshima R, Mizushima T, Takahashi H, et al. The efficacy of active drainage for preventing postoperative organ/space surgical site infections in patients with Crohn's disease. *Surg Today.* 2018;48(1):25-32.

72. Young PY, Khadaroo RG. Surgical site infections. *Surg Clin North Am.* 2014;94(6):1245-1264.

73. Scolari Childress KM, Gavard JA, Ward DG, Berger K, Gross GA. A barrier retractor to reduce surgical site infections and wound disruptions in obese patients undergoing cesarean delivery: a randomized controlled trial. *Am J Obstet Gynecol.* 2016;214(2):285 e281-285 e210.

74. Onyekwelu I, Yakkanti R, Protzer L, Pinkston CM, Tucker C, Seligson D. Surgical Wound Classification and Surgical Site Infections in the Orthopaedic Patient. *JAAOS Glob Res Rev.* 2017;2017(1):e022.

75. Edmiston CE, Jr., Spencer M. Patient care interventions to help reduce the risk of surgical site infections. *AORN J.* 2014;100(6):590-602.

76. Gomila A, Carratala J, Camprubi D, et al. Risk factors and outcomes of organ-space surgical site infections after elective colon and rectal surgery. *Antimicrob Resist Infect Control.* 2017;6:40.

77. Ke C, Jin Y, Evans H, et al. Prognostics of surgical site infections using dynamic health data. *J Biomed Inform.* 2017;65:22-33.

78. Levy SM, Holzmann-Pazgal G, Lally KP, Davis K, Kao LS, Tsao K. Quality check of a quality measure: surgical wound classification discrepancies impact risk-stratified surgical site infection rates in pediatric appendicitis. *J Am Coll Surg.* 2013;217(6):969-973.

79. Chemaly RF, Hachem RY, Husni RN, et al. Characteristics and outcomes of methicillin-resistant Staphylococcus aureus surgical-site infections in patients with cancer: a case-control study. *Ann Surg Oncol.* 2010;17(6):1499-1506.

80. Assawapalanggool S, Kasatpibal N, Sirichotiyakul S, Arora R, Suntornlimsiri W. Risk factors for cesarean surgical site infections at a Thai-Myanmar border hospital. *American Journal of Infection Control.* 2016;44(9):990-995.

81. Rennert-May E, Bush K, Vickers D, Smith S. Use of a provincial surveillance system to characterize postoperative surgical site infections after primary hip and knee arthroplasty in Alberta, Canada. *Am J Infect Control.* 2016;44(11):1310-1314.

82. Engemann JJ, Carmeli Y, Cosgrove SE, et al. Adverse Clinical and Economic Outcomes Attributable to Methicillin Resistance amongPatients with Staphylococcus aureus Surgical Site Infection. *Clin Infect Dis.* 2003;36(5):592-598.

83. Cohen B, Choi YJ, Hyman S, Furuya EY, Neidell M, Larson E. Gender differences in risk of bloodstream and surgical site infections. *J Gen Intern Med.* 2013;28(10):1318-1325.

84. Sanger PC, van Ramshorst GH, Mercan E, et al. A Prognostic Model of Surgical Site Infection Using Daily Clinical Wound Assessment. *J Am Coll Surg.* 2016;223(2):259-270 e252.

85. Goldrick BA. Surgical-Site Infections. *The American Journal of Nursing.* 2003;103(4):64AA-64AA.

86. Du M, Li M, Liu K, et al. A real-time surgical site infections surveillance mode to monitor surgery classification-specific, hospital-wide surgical site infections in a Chinese tertiary hospital. *Am J Infect Control.* 2017;45(4):430-432.

87. Bartella AK, Lemmen S, Burnic A, et al. Influence of a strictly perioperative antibiotic prophylaxis vs a prolonged postoperative prophylaxis on surgical site infections in maxillofacial surgery. *Infection.* 2018;46(2):225-230.

88. Finkelstein R, Eluk O, Mashiach T, et al. Reducing surgical site infections following total hip and knee arthroplasty: an Israeli experience. *Musculoskelet Surg.* 2017;101(3):219-225.

89. Thakkar V, Ghobrial GM, Maulucci CM, et al. Nasal MRSA colonization: impact on surgical site infection following spine surgery. *Clin Neurol Neurosurg.* 2014;125:94-97.

90. Lefebvre J, Buffet-Bataillon S, Henaux PL, Riffaud L, Morandi X, Haegelen C. Staphylococcus aureus screening and decolonization reduces the risk of surgical site infections in patients undergoing deep brain stimulation surgery. *J Hosp Infect.* 2017;95(2):144-147.

91. Rennert-May E, Manns B, Smith S, et al. Validity of administrative data in identifying complex surgical site infections from a population-based cohort after primary hip and knee arthroplasty in Alberta, Canada. *Am J Infect Control.* 2018.

92. Jeon CY, Furuya EY, Berman MF, Larson EL. The role of pre-operative and post-operative glucose control in surgical-site infections and mortality. *PLoS One.* 2012;7(9):e45616.

93. Wenzel RP. Minimizing Surgical-Site Infections. *New England Journal of Medicine.* 2010;362(1):75-77.

94. Habiba S, Nygaard OP, Brox JI, Hellum C, Austevoll IM, Solberg TK. Risk factors for surgical site infections among 1,772 patients operated on for lumbar disc herniation: a multicentre observational registry-based study. *Acta Neurochir (Wien).* 2017;159(6):1113-1118.

95. Alluri RK, Leland H, Heckmann N. Surgical research using national databases. *Ann Transl Med.* 2016;4(20):393.

96. Fawley J, Chelius TH, Arca MJ. Relationship between perioperative blood transfusion and surgical site infections in pediatric general and thoracic surgical patients. *J Pediatr Surg.* 2018;53(6):1105-1110.

97. Wang X, Zhang X, Lai X. Exploring an optimal risk adjustment model for public reporting of cesarean section surgical site infections. *J Infect Public Health.* 2018.

98. Daabiss M. American Society of Anaesthesiologists physical status classification. *Indian Journal of Anaesthesia.* 2011;55(2):111-115.

99. Koch AM, Nilsen RM, Dalheim A, Cox RJ, Harthug S. Need for more targeted measures - only less severe hospital-associated infections declined after

introduction of an infection control program. *J Infect Public Health.* 2015;8(3):282-290.

100. Davis JJ, Boisvert S, Brettin T, et al. Antimicrobial Resistance Prediction in PATRIC and RAST. *Sci Rep.* 2016;6:27930.

101. Allcock S, Young EH, Holmes M, et al. Antimicrobial resistance in human populations: challenges and opportunities. *Glob Health Epidemiol Genom.* 2017;2:e4.

102. Kim HR, Hwang SS, Kim EC, et al. Risk factors for multidrug-resistant bacterial infection among patients with tuberculosis. *J Hosp Infect.* 2011;77(2):134-137.

103. Gould IM. The epidemiology of antibiotic resistance. *Int J Antimicrob Agents.* 2008;32 Suppl 1:S2-9.

104. Mody L, Bradley SF, Galecki A, et al. Conceptual model for reducing infections and antimicrobial resistance in skilled nursing facilities: focusing on residents with indwelling devices. *Clin Infect Dis.* 2011;52(5):654-661.

105. Alanis AJ. Resistance to antibiotics: are we in the post-antibiotic era? *Arch Med Res.* 2005;36(6):697-705.

106. Jacoby G. *History of Drug-Resistant Microbes.* 2017.

107. Boswihi SS, Udo EE. Methicillin-resistant Staphylococcus aureus : An update on the epidemiology, treatment options and infection control. *Current Medicine Research and Practice.* 2018;8(1):18-24.

108. Rundensteiner E, Kong X, Teeple E, Brownell S, Sen C, Hartvigsen T. Early Prediction of MRSA Infections using Electronic Health Records. Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies; 2018.

109. Claeys KC, Zasowski EJ, Lagnf AM, Levine DP, Davis SL, Rybak MJ. Novel application of published risk factors for methicillin-resistant S. aureus in acute bacterial skin and skin structure infections. *Int J Antimicrob Agents.* 2018;51(1):43-46.

110. Dietrich NL, Reid SD, Doo GD, Fineberg RN, Khoury RJ, Gilbert RS. Predicting MSSA in Acute Hematogenous Osteomyelitis in a Setting With MRSA Prevalence. *Journal of Pediatric Orthopaedics.* 2015;35(4):426-430.

111. Otter JA, Herdman MT, Williams B, Tosas O, Edgeworth JD, French GL. Low prevalence of meticillin-resistant Staphylococcus aureus carriage at hospital admission: implications for risk-factor-based vs universal screening. *Journal of Hospital Infection.* 2013;83(2):114-121.

112. Pasricha J, Harbarth S, Koessler T, et al. Methicillin-resistant Staphylococcus aureus risk profiling: who are we missing? *Antimicrobial Resistance and Infection Control.* 2013;2(1):17.

113. Furuno JP, Harris AD, Wright M-O, et al. Prediction rules to identify patients with methicillin-resistant Staphylococcus aureus and vancomycin-resistant enterococci upon hospital admission. *American Journal of Infection Control.* 2004;32(8):436-440.

114. Harbarth S, Sax H, Fankhauser-Rodriguez C, Schrenzel J, Agostinho A, Pittet D. Evaluating the probability of previously unknown carriage of MRSA at hospital admission. *Am J Med.* 2006;119(3):275 e215-223.

115. Carman MJ, Phipps J, Raley J, Li S, Thornlow D. Use of a Clinical Decision Support Tool to improve guideline adherence for the treatment of methicillin-resistant Staphylococcus aureus: Skin and Soft Tissue Infections. *Adv Emerg Nurs J.* 2011;33(3):252-266.

116. Thomas S, Cantrill S, Waghorn DJ, McIntyre A. The role of screening and antibiotic prophylaxis in the prevention of percutaneous gastrostomy site infection caused by methicillin-resistant Staphylococcus aureus. *Aliment Pharmacol Ther.* 2007;25(5):593-597.

117. Boucher HW, Corey GR. Epidemiology of methicillin-resistant Staphylococcus aureus. *Clin Infect Dis.* 2008;46 Suppl 5:S344-349.

118. Teshome BF, Lee GC, Reveles KR, et al. Application of a methicillin-resistant Staphylococcus aureus risk score for community-onset pneumonia patients and outcomes with initial treatment. *BMC Infect Dis.* 2015;15:380.

119. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. *Sci Data.* 2016;3:160035.

120. Lodise TP, McKinnon PS, Rybak M. Prediction Model to Identify Patients With Staphylococcus aureus Bacteremia at Risk for Methicillin Resistance. *Infection Control and Hospital Epidemiology.* 2003;24(9):655-661.

121. Cosgrove SE, Qi Y, Kaye KS, Harbarth S, Karchmer AW, Carmeli Y. The Impact of Methicillin Resistance in Staphylococcus aureus Bacteremia on Patient Outcomes: Mortality, Length of Stay, and Hospital Charges. *Infection Control and Hospital Epidemiology.* 2005;26(2):166-174.

122. Bassetti M, Peghin M, Trecarichi EM, et al. Characteristics of Staphylococcus aureus Bacteraemia and Predictors of Early and Late Mortality. *PLoS One.* 2017;12(2):e0170236.

123. Klein EY, Sun L, Smith DL, Laxminarayan R. The changing epidemiology of methicillin-resistant Staphylococcus aureus in the United States: a national observational study. *Am J Epidemiol.* 2013;177(7):666-674.

124. Klein EY, Jiang W, Mojica N, et al. National Costs Associated with Methicillin-susceptible and Methicillin-resistant S. aureus Hospitalizations in the United States, 2010-2014. *Clin Infect Dis.* 2018.

125. Baker AW, Dicks KV, Durkin MJ, et al. Epidemiology of Surgical Site Infection in a Community Hospital Network. *Infect Control Hosp Epidemiol.* 2016;37(5):519-526.

126. Cadena J, Thinwa J, Walter EA, Frei CR. Risk factors for the development of active methicillin-resistant Staphylococcus aureus (MRSA) infection in patients colonized with MRSA at hospital admission. *Am J Infect Control.* 2016;44(12):1617-1621.

127. Zilberberg MD, Chaudhari P, Nathanson BH, et al. Development and validation of a bedside risk score for MRSA among patients hospitalized with complicated skin and skin structure infections. *BMC Infect Dis.* 2012;12:154.

128. Thompson P, Houston S. Decreasing methicillin-resistant Staphylococcus aureus surgical site infections with chlorhexidine and mupirocin. *Am J Infect Control.* 2013;41(7):629-633.

129. Chen TY, Anderson DJ, Chopra T, Choi Y, Schmader KE, Kaye KS. Poor functional status is an independent predictor of surgical site infections due to methicillin-resistant Staphylococcus aureus in older adults. *J Am Geriatr Soc.* 2010;58(3):527-532.

130. Abbas M, Aghayev E, Troillet N, et al. Temporal trends and epidemiology of Staphylococcus aureus surgical site infection in the Swiss surveillance network: a cohort study. *J Hosp Infect.* 2018;98(2):118-126.

131. Callejo-Torre F, Eiros Bouza JM, Olaechea Astigarraga P, et al. Risk factors for methicillin-resistant Staphylococcus aureus colonisation or infection in intensive care units and their reliability for predicting MRSA on ICU admission. *Le infezioni in medicina : rivista periodica di eziologia, epidemiologia, diagnostica, clinica e terapia delle patologie infettive.* 2016;24(3):201.

132. Malik A, Purushothaman B, Aparajit P, Dixon P, Berrington A. Prediction model to identify patients developing MRSA infection on an orthopaedic unit. *Injury Extra.* 2007;38(4).

133. Luhmann SJ, Smith JC. Preoperative MRSA Screening in Pediatric Spine Surgery: A Helpful Tool or a Waste of Time and Money? *Spine Deform.* 2016;4(4):272-276.

134. Lin S, Melki S, Lisgaris MV, Ahadizadeh EN, Zender CA. Post-operative MRSA infections in head and neck surgery. *Am J Otolaryngol.* 2017;38(4):417-421.

135. Anderson DJ, Kaye KS, Chen LF, et al. Clinical and financial outcomes due to methicillin resistant Staphylococcus aureus surgical site infection: a multi-center matched outcomes study. *PLoS One.* 2009;4(12):e8305.

136. Kalra L, Camacho F, Whitener CJ, et al. Risk of methicillin-resistant Staphylococcus aureus surgical site infection in patients with nasal MRSA colonization. *Am J Infect Control.* 2013;41(12):1253-1257.

137. Stevens DL. Treatments for skin and soft-tissue and surgical site infections due to MDR Gram-positive bacteria. *Journal of Infection.* 2009;59:S32-S39.

138. Manian F, Meyer P, Setzer J, Senkel D. Surgical Site Infections Associated with Methicillin-Resistant Staphylococcus aureus: Do Postoperative Factors Play a Role? *Clinical Infectious Diseases.* 2003;36(7):863-868.

139. Anderson DJ, Sexton DJ, Kanafani ZA, Auten G, Kaye KS. Severe surgical site infection in community hospitals: epidemiology, key procedures, and the changing prevalence of methicillin-resistant Staphylococcus aureus. *Infect Control Hosp Epidemiol.* 2007;28(9):1047-1053.

140. Ihejirika RC, Thakore RV, Sathiyakumar V, Ehrenfeld JM, Obremskey WT, Sethi MK. An assessment of the inter-rater reliability of the ASA physical status score in the orthopaedic trauma population. *Injury.* 2015;46(4):542-546.

141. Catapano A, Charlson M, Wells MT, Ullman R, King F, Shmukler C. The Charlson Comorbidity Index Can Be Used Prospectively to Identify Patients Who Will Incur High Future Costs. *PLoS ONE.* 2014;9(12):e112479.

142. Musen MA, Shahar Y, Shortliffe EH. Clinical Decision-Support Systems. In: Shortliffe EH, Cimino JJ, eds. *Biomedical Informatics: Computer Applications in Health Care and Biomedicine*. New York, NY: Springer New York; 2006:698-736.

143. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ.* 2005;330(7494):765.

144. Ledley RS, Lusted LB. Reasoning Foundations of Medical Diagnosis. *Symbolic logic, probability, and value theory aid our understanding of how physicians reason.* 1959;130(3366):9-21.

145. Warner HR, Toronto AF, Veasy LG. EXPERIENCE WITH BAYE'S THEOREM FOR COMPUTER DIAGNOSIS OF CONGENITAL HEART DISEASE*. *Annals of the New York Academy of Sciences.* 1964;115(2):558-567.

146. Holstiege J, Mathes T, Pieper D. Effects of computer-aided clinical decision support systems in improving antibiotic prescribing by primary care providers: a systematic review. *J Am Med Inform Assoc.* 2015;22(1):236-242.

147. Anakal S, Sandhya P. Clinical decision support system for chronic obstructive pulmonary disease using machine learning techniques. Vol 2018-: IEEE; 2017:1-5.

148. Nakayama M, Osaki S, Shimokawa H. Validation of mortality risk stratification models for cardiovascular disease. *Am J Cardiol.* 2011;108(3):391-396.

149. Ruland CM, Bakken S. Developing, implementing, and evaluating decision support systems for shared decision making in patient care: a conceptual model and case illustration. *Journal of Biomedical Informatics.* 2002;35(5-6):313-321.

150. Wright MO, Robicsek A. Clinical decision support systems and infection prevention: to know is not enough. *Am J Infect Control.* 2015;43(6):554-558.

151. Koposov R, Fossum S, Frodl T, et al. Clinical decision support systems in child and adolescent psychiatry: a systematic review. *Eur Child Adolesc Psychiatry.* 2017;26(11):1309-1317.

152. Stillman RC. Clinical Decision Support Tools Improving Cancer Care. *Semin Oncol Nurs.* 2018;34(2):158-167.

153. Forrest GN, Van Schooneveld TC, Kullar R, Schulz LT, Duong P, Postelnick M. Use of electronic health records and clinical decision support systems for antimicrobial stewardship. *Clin Infect Dis.* 2014;59 Suppl 3:S122-133.

154. Calloway S, Akilo HA, Bierman K. Impact of a clinical decision support system on pharmacy clinical interventions, documentation efforts, and costs. *Hosp Pharm.* 2013;48(9):744-752.

155. Shibl R, Lawley M, Debuse J. Factors influencing decision support system acceptance. *Decision Support Systems.* 2013;54(2):953-961.

156. Khalifa M. Clinical Decision Support: Strategies for Success. *Procedia Computer Science.* 2014;37:422-427.

157. Zuccotti G, Maloney FL, Feblowitz J, Samal L, Sato L, Wright A. Reducing risk with clinical decision support: a study of closed malpractice claims. *Appl Clin Inform.* 2014;5(3):746-756.

158. Berner ES, La Lande TJ. Overview of Clinical Decision Support Systems. In: Berner ES, ed. *Clinical Decision Support Systems: Theory and Practice*. New York, NY: Springer New York; 2007:3-22.

159. Kabachinski J. A Look at Clinical Decision Support Systems. *Biomedical Instrumentation & Technology.* 2013;47(5):432-434.

160. Garg AX, Adhikari NJ, McDonald H, et al. Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: A systematic review. *JAMA.* 2005;293(10):1223-1238.

161. Samore MH, Bateman K, Alder SC, et al. Clinical decision support and appropriateness of antimicrobial prescribing: A randomized trial. *JAMA.* 2005;294(18):2305-2314.

162. Buzaev IV, Plechev VV, Nikolaeva IE, Galimova RM. Artificial intelligence: Neural network model as the multidisciplinary team member in clinical decision support to avoid medical mistakes. *Chronic Dis Transl Med.* 2016;2(3):166-172.

163. Islam R, Weir CR, Jones M, Del Fiol G, Samore MH. Understanding complex clinical reasoning in infectious diseases for improving clinical decision support design. *BMC Med Inform Decis Mak.* 2015;15:101.

164. Fillmore CL, Bray BE, Kawamoto K. Systematic review of clinical decision support interventions with potential for inpatient cost reduction. *BMC Medical Informatics and Decision Making.* 2013;13:135-135.

165. Miller A, Moon B, Anders S, Walden R, Brown S, Montella D. Integrating computerized clinical decision support systems into clinical work: A meta-synthesis of qualitative research. *Int J Med Inform.* 2015;84(12):1009-1018.

166. Meulendijk MC, Spruit MR, Willeboordse F, et al. Efficiency of Clinical Decision Support Systems Improves with Experience. *J Med Syst.* 2016;40(4):76.

167. Moja L, Kwag KH, Lytras T, et al. Effectiveness of computerized decision support systems linked to electronic health records: a systematic review and meta-analysis. *Am J Public Health.* 2014;104(12):e12-22.

168. Moore LJ, Turner KL, Todd SR, McKinley B, Moore FA. Computerized clinical decision support improves mortality in intra abdominal surgical sepsis. *Am J Surg.* 2010;200(6):839-843; discussion 843-834.

169. Geurts D, de Vos-Kerkhof E, Polinder S, et al. Implementation of clinical decision support in young children with acute gastroenteritis: a randomized controlled trial at the emergency department. *Eur J Pediatr.* 2017;176(2):173-181.

170. Campion JTR, May AK, Waitman LR, Ozdas A, Lorenzi NM, Gadd CS. Characteristics and effects of nurse dosing over-rides on computer-based intensive insulin therapy protocol performance. *Journal of the American Medical Informatics Association.* 2011;18(3):251-258.

171. Campion TR, Waitman LR, Lorenzi NM, May AK, Gadd CS. Barriers and facilitators to the use of computer-based intensive insulin therapy. *International Journal of Medical Informatics.* 2011;80(12):863-871.

172. Feldstein A, Simon SR, Schneider J, et al. How to Design Computerized Alerts to Ensure Safe Prescribing Practices. *The Joint Commission Journal on Quality and Safety.* 2004;30(11):602-613.

173. Litvin CB, Ornstein SM, Wessell AM, Nemeth LS, Nietert PJ. Use of an electronic health record clinical decision support tool to improve antibiotic prescribing for acute respiratory infections: the ABX-TRIP study. *J Gen Intern Med.* 2013;28(6):810-816.

174. Wagholikar KB, MacLaughlin KL, Casey PM, et al. Automated recommendation for cervical cancer screening and surveillance. *Cancer Inform.* 2014;13(Suppl 3):1-6.

175. Bates DW, Kuperman GJ, Wang S, et al. Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality. *J Am Med Inform Assoc.* 2003;10(6):523-530.

176. Droney J, Riley J. Bertsche, et al., multidisciplinary pain management based on a computerized clinical decision support system in cancer pain patients. *Pain.* 2009;147(1-3):1-2.

177. Shen Y, Yuan K, Chen D, et al. An ontology-driven clinical decision support system (IDDAP) for infectious disease diagnosis and antibiotic prescription. *Artif Intell Med.* 2018;86:20-32.

178. Beaudoin M, Kabanza F, Nault V, Valiquette L. Evaluation of a machine learning capability for a clinical decision support system to enhance antimicrobial stewardship programs. *Artif Intell Med.* 2016;68:29-36.

179. Pombo N, Araujo P, Viana J. Knowledge discovery in clinical decision support systems for pain management: a systematic review. *Artif Intell Med.* 2014;60(1):1-11.

180. Kuperman GJ, Gardner RM, Pryor TA. *HELP: A dynamic hospital information system.* New York: Springer; 1991.

181. Shortliffe EH. *Computer-based medical consultations, MYCIN.* New York: Elsevier; 1976.

182. Barnett G, Cimino JJ, Hupp JA, Hoffer EP. Dxplain: An evolving diagnostic decision-support system. *JAMA.* 1987;258(1):67-74.

183. Hardin JM, Chhieng DC. Data Mining and Clinical Decision Support Systems. In: Berner ES, ed. *Clinical Decision Support Systems: Theory and Practice*. New York, NY: Springer New York; 2007:44-63.

184. Hastie T. *The Elements of Statistical Learning Data Mining, Inference, and Prediction.* 2nd ed. ed. New York, NY: Springer New York; 2009.

185. Gultepe E, Green JP, Nguyen H, Adams J, Albertson T, Tagkopoulos I. From vital signs to clinical outcomes for patients with sepsis: a machine learning basis for a clinical decision support system. *J Am Med Inform Assoc.* 2014;21(2):315-325.

186. Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *Journal of Biomedical Informatics.* 2002;35(5-6):352-359.

187. *Artificial Neural Network Modelling.* 1st ed. 2016. ed. Cham: Springer International Publishing; 2016.

188. Alshammari MHS. *Comparison of artificial neural network and logistic regression models for prediction of diabetes type II with complications.* Thesis (Ph.D.)--Rutgers University, 2016.; 2016.

189. Hsu C-C, Lin YE, Chen Y-S, Liu Y-C, Muder RR. Validation Study of Artificial Neural Network Models for Prediction of Methicillin-Resistant Staphylococcus aureus Carriage. *Infection Control and Hospital Epidemiology.* 2008;29(7):607-614.

190. de Ville B. Decision trees. *Wiley Interdisciplinary Reviews: Computational Statistics.* 2013;5(6):448-455.

191. Kass G. An exploratory technique for investigating large quantities of categorical data. *Applied Statistics.* 1980;29(2):119.

192. Gerald LB, Tang S, Bruce F, et al. A decision tree for tuberculosis contact investigation. *American journal of respiratory and critical care medicine.* 2002;166(8):1122.

193. Ghaddar B, Naoum-Sawaya J. High dimensional data classification and feature selection using support vector machines. *European Journal of Operational Research.* 2018;265(3):993-1004.

194. Boyle BH. *Support vector machines data analysis, machine learning and applications.* New York: Nova Science Publishers; 2011.

195. *Bayesian network technologies applications and graphical models.* Hershey, Pa.: IGI Global; 2007.

196. Weber P. *Benefits of Bayesian network models.* Hoboken, NJ: Wiley; 2016.

197. Receiver Operating Characteristic (ROC) Curve. 2009:953-958.

198. Tan P-N. Receiver Operating Characteristic. 2009:2349-2352.

199. Parker C. On measuring the performance of binary classifiers. *Knowledge and Information Systems.* 2013;35(1):131-152.

200. Lapidus J. Sensitivity and Specificity. 2008:946-949.

201. *shiny: Web Application Framework for R.* [computer program]. Version 1.2.02018.

202. Bruce N. *Quantitative methods for health research : a practical interactive guide to epidemiology and statistics.* 2nd ed ed: John Wiley & Sons, Inc.; 2017.

203. Abu-Mostafa YS. *Learning from data : a short course.* United States]: AMLBook.com; 2012.

204. Dupont WD. *Statistical modeling for biomedical researchers a simple introduction to the analysis of complex data.* 2nd ed. ed. Cambridge, UK ;: Cambridge University Press; 2009.

205. James G. *An Introduction to Statistical Learning with Applications in R.* New York, NY: Springer New York; 2013.

206. Akaike H. A new look at the statistical model identification. *Automatic Control, IEEE Transactions on.* 1974;19(6):716-723.

207. Krawczyk B. Learning from imbalanced data: open challenges and future directions. *Progress in Artificial Intelligence.* 2016;5(4):221-232.

208. Wei Q, Dunbrack RL, Jr. The Role of Balanced Training and Testing Data Sets for Binary Classifiers in Bioinformatics. *PLOS ONE.* 2013;8(7):e67863.

209. Kuhn M. Building Predictive Models in R Using the caret Package. *2008.* 2008;28(5):26.

210. Beerenwinkel N, Sander O, Lengauer T, Sing T. ROCR: visualizing classifier performance in R. *Bioinformatics.* 2005;21(20):3940-3941.

211. Günther F, Fritsch S. *neuralnet: Training of Neural Networks.* Vol 22010.

212. Lalkhen AG, McCluskey A. Clinical tests: sensitivity and specificity. *Continuing Education in Anaesthesia Critical Care & Pain.* 2008;8(6):221-223.

213. Wilson KA. Clinical decision support system for the prediction of MRSA and MRSA-related SSI. 2019; http://kevinwilson.shinyapps.io/cdss/.

214. Burnham KP, Anderson DR. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach.* Springer New York; 2003.

215. Ranganathan P, Aggarwal R. Common pitfalls in statistical analysis: Understanding the properties of diagnostic tests - Part 1. *Perspect Clin Res.* 2018;9(1):40-43.