2009

THE ECONOMIC IMPACT OF ANTIMICROBIAL RESISTANCE IN PATIENTS WITH NOSOCOMIAL STAPHYLOCOCCUS AUREUS BACTEREMIA

Suzanne Phillips
Virginia Commonwealth University

THE ECONOMIC IMPACT OF ANTIMICROBIAL RESISTANCE IN PATIENTS WITH NOSOCOMIAL *STAPHYLOCOCCUS AUREUS* BACTEREMIA

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

by

SUZANNE TOUSSAINT PHILLIPS
Bachelor of Arts, Washington and Lee University, 2001
Masters of Public Health, Virginia Commonwealth University, 2008
Doctor of Pharmacy, Virginia Commonwealth University, 2008

Director: RONALD E. POLK, PHARM.D.
PROFESSOR AND CHAIRMAN, DEPARTMENT OF PHARMACY

Director: SPENCER E. HARPE, PHARM.D, M.P.H., PH.D.
ASSISTANT PROFESSOR, DEPARTMENT OF PHARMACY

Virginia Commonwealth University
Richmond, Virginia
May 2009

Dedication

This dissertation is dedicated to the two most important men in my life;
My husband, Ryan Phillips,
Who has loved and supported me; apparently marshmallows and popcorn aren't really dinner.
To my father, Craig Toussaint,
Who taught me to always challenge myself.

Acknowledgements

This dissertation project would not have been possible without tremendous support. First, I would like to thank my husband Ryan Phillips. He witnessed all of my ups and downs, managing to stay optimistic and encouraging. I am very grateful for his patience and promise that I will take a school break.

There is no way that I could have produced this manuscript without the direction of Dr. Spencer Harpe. He always provided prompt feedback to all my questions from SAS issues to methodology to dissertation formatting. I feel so lucky that I was able to benefit from his knowledge -Thank you, Thank you, Thank you!!!

Dr. Ron Polk inspired me as a summer research student to pursue a graduate degree. I am extremely grateful for his mentorship throughout my studies. Thank you for encouraging me to attend professional meetings and for broadening my horizons.

I owe a great deal of thanks to Dr. Norman Carroll. I have appreciated all his feedback and comments. He has been a true teacher and role model. I have the utmost respect for him.

Dr. Amy Pakyz has provided steadfast support. I have appreciated both her professional advice and her friendship. I learned a lot from her on my infectious disease clerkship. I will never forget our San Francisco bike trip!

I need to thank Dr. Betsy Turf for her help with this project as well as my MPH project. Without her assistance, I would not even have the MPH! I have valued all of her insight and encouragement.

Dr. Mike Edmond provided much clinical guidance on this project. I feel very lucky that he agreed to be on my committee. His guidance and support are greatly appreciated.

Dr. Patty Slattum deserves an enormous thank you. She provided invaluable advice about negotiating the combined degree program. I could always count on her for support and an honest opinion. Her impact on the graduate program is enormous and I am sure all the other graduate students that have ever been though this program would agree with me.

My parents, Craig and Margaret Toussaint, have always inspired me to my fullest potential. They are both amazing people who have encouraged me throughout my life. I really appreciate their continued support.

My sister, Michelle Adams, also deserves thanks. Our frequent conversations help keep me grounded. I am so thankful for our friendship.

Finally, I would like to thank my dogs Missie and Leeroy. They have been with Ryan and me since the beginning of pharmacy school. Their antics have provided endless comic relief.

Tables of Contents

List of Tables

List of Figures

Abstract

THE ECONOMIC IMPACT OF ANTIMICROBIAL RESISTANCE IN PATIENTS WITH NOSOCOMIAL *STAPHYLOCOCCUS AUREUS* BACTEREMIA

By Suzanne Toussaint Phillips

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2009

Major Directors: Ron E. Polk, Pharm.D. & Spencer E. Harpe, Pharm.D., M.P.H., Ph.D.

Background: The proportion of nosocomial *Staphylococcus* infections caused by methicillin-resistant S*taphylococcus aureus* (MRSA) has increased from 22% in 1995 to 63% in 2004. Blood stream infections, more commonly referred to as bacteremias, represented the majority (75.5%) of hospital-onset MRSA cases. The economic impact of *Staphylococcus aureus* bacteremia merits investigation.

Methods: This was a retrospective cohort analysis within Cerner HealthFacts data warehouse. Eligible patients were those who had *Staphylococcus aureus* bacteremia and were discharged between January 1, 2000, and December 31, 2006. Inclusion criteria include age > 18 years old and onset of infection > 48 hours post admission. The crude association was measured by subtracting the total mean hospital charge for MSSA bacteremia from the MRSA charge. A generalized linear model using a gamma distribution and log link were used to determine the adjusted hospital charge and post-infection length of stay for the MRSA and MSSA groups. Path analysis was used to describe the relationships between infection susceptibility status, LOS and total hospital charge.

Results: During the study period, 930 patients meet all the inclusion and exclusion criteria. The overall total hospital charge was $111,636 (MRSA = $121,713, MSSA = $97,307.) The crude difference in mean charge was $24,406. The multivariable model included predicted a MRSA patient would have an increased total charge of $22,889. MRSA had a higher total charge but when patients were more severely ill, MRSA charges decreased while MSSA charges increased. The second multivariable model predicted a MRSA patient would have an increased post-infection LOS of 1.3 days. However, the magnitude of increased post-infection LOS based on pre-infection LOS was different for MRSA and MSSA patients. The path analysis model indicated the direct and indirect effects of susceptibility status on both post-infection LOS and total charge were relatively small.

Conclusion: This investigation was the first large multi-center investigation to examine the economic impact of MRSA and MSSA bacteremia. MRSA was associated with a higher total charge and longer post-infection LOS than MSSA patients. The path analysis model analyzed suggests the actually role of infection susceptibility status on post-infection LOS and total charge was minor.

CHAPTER 1

Introduction

Overview of the document

This dissertation describes a study designed to examine the relationship between *Staphylococcus aureus* bacteremia susceptibility status (methicillin resistant or susceptible) and hospital charges. This chapter provides background information necessary to understand the significance of the project. The second chapter systematically reviews the available literature and provides more extensive background on previous investigations, economic issues, confounding factors, propensity scores and path analysis. Chapter 3 discusses the preliminary investigation. Chapter 4 describes the methodology used for the dissertation project. The results are provided in Chapter 5, followed by a discussion and concluding remarks in Chapter 6.

Background

Ever since the discovery of antibiotics, resistance has been emerging. Antimicrobial resistance is an important public health issue. In 1999, the Centers for Disease Control and Prevention (CDC) and several other national agencies developed an Interagency Task Force dedicated to antimicrobial resistance. Each year since its inception, the Antimicrobial Task Force has published an Annual Report describing the current state of resistance in the US.[1] According to the CDC, the proportion of total nosocomial *Staphylococcus* species infections caused by methicillin-resistant *Staphylococcus aureus* (MRSA) has increased from 22% in 1995 to 63% in 2004. [2] Additionally, a recent investigation sought to describe the incidence and distribution of invasive MRSA in nine communities across the United States.[3] The results showed blood stream infections, more commonly referred to as bacteremias, represented the majority of hospital-onset MRSA cases (75.5%).

The scientific community and healthcare professionals have acknowledged the importance of understanding the economic impact of resistance. The "cost of resistance" has been defined as "the incremental cost of care for an infection due to a resistant isolate minus the care costs of infection with a susceptible strain of the same organism."[4] For this analysis, the "cost of resistance" can be approximated by subtracting the total hospital charges of the susceptible from the resistant groups. The difference in charges estimates the "cost of resistance."

Performing and/or interpreting an economic analysis requires further explanation. For example, the terms costs and charges must be defined. Some authors have used the

terms interchangeably, but this is incorrect. Costs and charges reflect different economic values. Specifically, charges always over-inflate cost. [5] For the purposes of this report, great care has been given to use the terms costs and charges appropriately. Previous investigations have used costs or charges as outcomes depending on study design and data availability. The economic background section provides a more thorough discussion of costs versus charges.

Any epidemiologic investigation attempting to describe the relationship between an exposure and outcome must consider potential confounders. Confounding factors are variables that (1) are associated with the outcome as well as the exposure, and (2) are not variables in the causal pathway. If confounding exists, an association may appear to be present when one does not exist or there may seem to be no association when a true association does exist.[6] Therefore, it is imperative to identify confounders and control for them. In most situations confounders are identified *a priori* based on previous investigations or expert knowledge.[7] For this investigation, underlying severity of illness and comorbid conditions were identified as confounders. [8] Additionally, hospital level factors (e.g. teaching status, bed size, geographic location) can also confound the relationship between susceptibility status and total charges.

Propensity scores are another way to control for confounding. Observational studies employ this method to eliminate bias from an unequal distribution of confounders thereby mimicking the purpose of randomization in a randomized, controlled clinical trial. Propensity scores are the probability of exposure given measured baseline variables.[9] This probability can then be used as a matching or stratification factor, as a

covariate in multivariable model or to perform inverse probability of exposure weighting.[10]

The relationship between susceptibility status and charge is further complicated by length of stay (LOS). LOS must be considered in two parts as it relates to the onset of a nosocomial infection. The pre-infection LOS refers to the number of days in the hospital before infection onset. The post-infection LOS refers to hospital stay after the infection onset. Post-infection LOS is an intermediary between susceptibility status and charges. Pre-infection LOS is related to infection susceptibility, post-infection LOS and total charge. Therefore, pre-infection LOS can be considered as confounding the relationship between infection susceptibility status and charges. A technique called path analysis can be used to provide estimates of the magnitude and significance of hypothesized relationships between variables when some of the variables lie in the proposed causal pathway.

Previously, single center investigations have suggested that MRSA bacteremia may increase hospitalization costs by 1.2- to 2-fold over methicillin-susceptible *Staphylococcus aureus* (MSSA) bacteremia.[11-15] This supports the hypothesis that MRSA bacteremia is associated with higher costs than MSSA bacteremia. However, one study performed in three German hospitals indicated that hospital costs between patients with MRSA and MSSA patients with blood stream infections were similar.[16] A large multi-center investigation may help to quantify the economic impact of MRSA vs. MSSA bacteremia.

Objectives

The current investigation has four primary objectives which are listed below.

1. Measure the crude difference in total hospital charges between methicillin-resistant *Staphylococcus aureus* (MRSA) vs. methicillin-susceptible *Staphylococcus aureus* (MSSA) bacteremia.

2. Determine the impact of methicillin resistance vs. methicillin susceptibility in *Staphylococcus aureus* bacteremia (SAB) on total hospital charges while adjusting for potential confounders.

3. Determine the impact of methicillin resistance vs. methicillin susceptibility in SAB on post-infection length of stay (LOS) while adjusting for potential confounders.

4. Describe the relationships between SAB methicillin susceptibility, LOS and total hospital charges.

Significance

As described above, the incidence of MRSA is increasing and the majority of hospital acquired MRSA infections are bacteremias.[3] These factors stimulate the need for further investigation into the relationship between MRSA and MSSA bacteremia.

Before a clinician prescribes a drug, the potential risks and benefits must be considered. The benefit of prescribing an antibiotic often outweighs the adverse consequences for an individual patient since the side effect profile of most antibiotics is reasonably mild. However, with each antibiotic administration antimicrobial resistance is a potential unintended consequence. By identifying any unintended financial consequences, strategies for minimizing the costs and thereby resistance can be more fervently pursued. For example, infection control strategies are expensive but they may seem more manageable if the "cost of resistance" is significantly more. Finding ways to minimize resistance through an economic approach will positively impact public health.

The scientific community and healthcare professionals recognize the importance of understanding the economic impact of resistance.[4] Understanding the financial implications associated with MRSA will better equip hospitals to manage their financial resources.

Also, the relationship between infection susceptibility status with respect to LOS and total hospital charges needs to be more appropriately defined. It is common knowledge that LOS greatly contributes to hospital charges. But the relationship between pre- and post-infection LOS, infection susceptibility status and charges has not been fully characterized. Previous investigations have explored charges/costs and LOS as independent outcomes. Path analysis will be used to characterize the relationship between infection susceptibility status, LOS and total charge.

Data from over forty hospitals will be considered making this a large multi-center investigation. This includes both academic teaching medical centers and community

hospitals. Previous investigations only explored the economic impact of SAB in teaching hospitals. A multi-center investigation will provide a larger sample size than previous single center reports. This investigation will have greater external validity than previous studies.

CHAPTER 2

Literature Review

MRSA Bacteremia Overview

Staphylococcus aureus is a significant cause of infectious disease in humans. More specifically, it is an important cause of bloodstream infections. MRSA accounts for more than half of all *S. aureus* infections in many institutions. MRSA rates have been reported as high as 70% in Intensive Care Units (ICUs).[17] These are serious infections with mortality rates ranging from 15 to 60%.[13, 14, 18-22] MSSA infections can also be fatal. A meta-analysis estimated the mortality rate of MSSA bacteremia to be 12%.[23]

Why do some patients have MRSA versus MSSA infections? The investigators at the Mayo Clinic list 4 risk factors for an MRSA infection (1) a current or recent hospitalization, (2) residing in a long term care facility, (3) invasive devices, and (4) recent antibiotic use with fluoroquinolones (ciprofloxacin, ofloxacin or levofloxacin) or

cephalosporins.[24] Published reports have expounded on these 4 risk factors by subdividing the categories into areas of greater risk. For example, in an investigation published by McHugh et al.[13], factors including cardiac surgery during hospitalization and venous or bladder catheter > 3 days during hospitalization were identified to increase the risk of MRSA over MSSA. However, both infections caused by MSSA or MRSA are a serious public health concern meriting additional attention.

Economic Analysis of MRSA Bacteremia Overview

Several authors have identified key factors to consider when estimating the cost of resistance. Within the context of these publications, the considerations discussed were within a theoretical framework. Each author discussed how an investigation into the "cost of resistance" should be designed. A compilation of these considerations is summarized below in Table 2.1. McGowan was the first to distinguish the various viewpoints which determine applicable costs.[25] Such perspectives include the physician, the patient, the healthcare business, the drug industry and society. Ultimately, the societal/public health perspective would reflect the most comprehensive economic impact of resistance. This viewpoint would consider resistance from the perspective of the social good. Defining the economic burden of resistance quickly becomes complicated when forced to consider the long-term implications of antimicrobial usage. For example, treating a patient's infection with an antibiotic may expedite that patient's recovery, which would decrease short-term expenses. But antibiotic usage indirectly increases resistance thus increasing the overall cost of resistance to society despite the

individual patient's short-term savings. It is less complicated to define the economic burden of resistance from the hospital perspective. This perspective only considers expenditures that directly impact the hospital. Defining the perspective is an essential first step in any economic analysis as it determines what "costs" will be considered. [26]

Table 2.1 Key factors to consider in estimating the cost of resistance

	Cosgrove & Carmeli [27]	Howard et al [4]	McGowan [25]
Patient Characteristics			
Severity of underlying illness		X	
Mortality	X	X	
Morbidity	X		
Length of stay (LOS)	X	X	
Intensive care unit (ICU) admission	X		
Need for surgery/other procedures	X		
Status at discharge (functional)	X		
Economic Considerations	X		
Hospital cost (fixed and marginal)	X		
Per day per bed by specialty		X	
Per day per bed ICU vs. general vs. other		X	
Antimicrobial acquisition costs		X	
Antimicrobial administration costs		X	X
Staff time increase (MD / nurse)		X	
Occurrence of other procedures		X	
Occurrence of other infections/complications		X	
Lab costs (screening and diagnosis)		X	X
Infection control staff		X	X
Treatment failure			X
Hospital charges (larger than costs)	X		
Resources utilized	X		
Supplies, housekeeping, waste disposal, etc.		X	X
Hospital Factors			
Infection Control Practices			X
Formularies/Protocols			X
Educational Programs			X

The "x" represents factors identified in each paper as important considerations.
LOS = length of stay
ICU = intensive care unit

The choice of control group was very important. One of the pioneer studies examining SAB compared resistant infections to both susceptible patients and to patients without blood stream infections.[14] Although uninfected hospitalized patients represent the true source population, this comparison assesses the burden of having a resistant infection rather than no infection. Comparing a resistant to a susceptible infection is appropriate when trying to determine the excess economic burden attributable to the resistant infection.[27] The rest of the investigations discussed used a susceptible control group.

Economic Background

Understanding costs, charges and reimbursements from a hospital's perspective is complex. In a strictly economic sense, a cost can be thought of as "the extra amount of resource consumption incurred for providing a service as compared to the costs of not providing that service." [5] In layman's terms, costs refer to the price a hospital pays for the resources it consumes. This is different than a charge, which is simply a list price that hospitals charge to their customers. [5] Charges are always higher than the actual hospital cost so that patients who can pay will cover the losses from those who cannot afford to pay. [5]

Hospital reimbursements are an entirely different level of complexity. The government, through Medicare and Medicaid, is the major payer of hospital services in the United States. In 1982, Medicare moved to a prospective payment system (PPS) for

hospital reimbursement to control costs by capping the allowable reimbursement.[28] A hospital's listed charge, therefore, is not the same as the insurance reimbursement.

PPS works by dividing admissions into diagnosis-related group (DRG) categories. A DRG is computed taking into account the affected organ system, up to nine ICD-9-CM diagnosis codes, up to 6 ICD-9-CM procedure codes, morbidity and gender. Each DRG has an associated DRG relative weight. DRG relative weights reflect the average level of resources a Medicare patient in a particular DRG will utilize. The weight can range from greater than 0 to less than 20. An average hospital stay would have weight of 1. Conditions with greater costs are assigned a higher DRG relative weight. Hospitals are then reimbursed a fixed rate depending on the relative weight of the DRG. Reimbursements are also adjusted for geographic differences in wage, hospital teaching status, proportion of low income individuals a hospital treats and cost outliers.[29]

Since charges are known to inflate the economic burden of hospitalization, cost-to-charge ratios have been used to better approximate actual cost from charges [11]; however, cost-to-charge ratios are a poor approximation of actual costs.[30] Additionally, cost-to-charge ratios are specific for a particular hospital. The identity of a hospital must be known to determine which cost-to-charge ratio is appropriate. Cost-to-charge ratios were not used in this analysis since the identity of each hospital was unknown.

As previously mentioned, hospital charges are considered to be a gross overestimation of the true hospitalization cost. But the difference between MRSA and MSSA bacteremia for charge and cost has been shown to be similar.[31] Although this observation has been noted in the literature, mathematics would suggest that a 50%

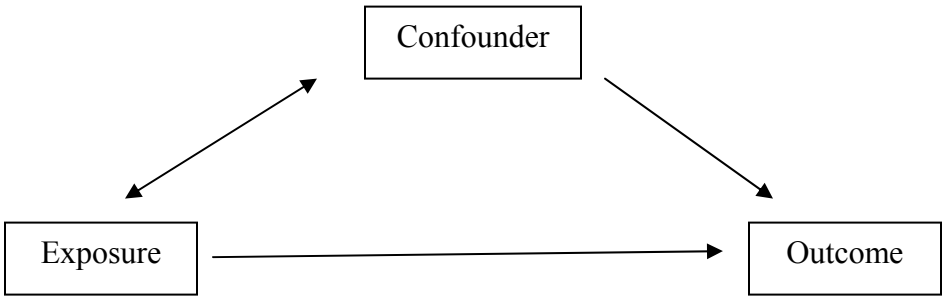
overestimation in charge would correspond to a 50% overestimation in the difference between charges. Therefore, it will be more useful to discuss differences as percent increase or relative change.

Descriptive statistics for variables that are non-normally distributed are conventionally expressed as median and interquartile range. Charges/costs are rarely normally distributed. It would seem logical to express the central tendency of charge/cost as a median, but averages are used since economists are interested solely in means.[32] This convention stemmed from the practical need to obtain annual budget data which could be obtained by multiplying the arithmetic mean (average) by the total number of patients. Therefore, means are the central tendency measures reported for charge/cost data.[33]

Confounding Factors

Confounding factors are variables that are associated with the outcome as well as the exposure. However, they cannot be variables in the causal pathway (i.e., intermediary) between the exposure and outcome. A directed acyclic graph (or a causal diagram) depicting the relationship between exposure, outcome and confounder can be found below in Figure 2.1.

Figure 2.1. Mechanism of confounding



If confounding exists, an association may appear to be present when one does not exist or there may seem to be no association when a true association does exist.[6] This is because the confounder distorts the effect of exposure on outcome. It is important to identify and control for confounding since it can lead to a misinterpretation of study results.

Confounders should be identified from the base population, not the study sample. This means that confounders are identified *a priori* based on previous investigations or expert knowledge.[7] Since almost all investigations examine a small subset of a larger population, it is possible that a confounding effect within the population may not be present within the sample. Known confounders should be included regardless of their "statistical significance" in the sample.

Once confounders are identified they must be controlled. This can be done during the design phase through restriction, matching or randomization or during analysis via stratification-, or multivariable analysis.[34] By definition, randomization should equalize the distribution of all known and unknown confounders between the groups.

The technique is frequently used in prospective, experimental study designs. However, the retrospective nature of database projects does not allow for this technique. Retrospective investigations more commonly use restriction or matching. Restriction involves limiting the scope of design to one stratum of the confounder but, this limits the investigation's external validity. Matching is used to make the groups comparable with respect to the confounder. This technique requires more complicated analysis since the matched nature of the data must be taken into consideration. Additionally, "over-matching" can become a concern if the groups become too similar as to disguise as actual effect.[35]

As stated above, confounding can be controlled for at the analysis level through stratification or multivariable analysis. When confounding exists, the magnitude of effect will be the same between the strata but the crude estimate of effect of the exposure and outcome will be different.[36] Stratification quantifies the relationship between exposure and outcome as a pooled estimate with respect to the confounder. Stratification becomes more complicated when multiple confounders exist because the statistical power to detect a difference decreases as the number of observations within a stratum decreases. Mathematical modeling during the analysis can account for many confounders simultaneously. Randomization, restriction, matching, stratification, and multivariable analysis are all methods for reducing confounding. Except for randomization, each technique requires the confounder to be identified and cannot control any unmeasured confounding effects.[37]

Several variables have been acknowledged as known confounders in the literature regarding the relationship between infection susceptibility and charge. Specifically, underlying severity of illness and comorbid conditions have been identified.[38] Underlying severity of illness and comorbid conditions are both gauges of baseline health status not actual disease acuity. There is an important distinction between underlying severity of illness and severity of illness during the infection. Severity of illness during the infection is considered an intermediary, not a confounder since a sicker patient would require a longer length of stay and would increase total charges.

There is currently no universally recognized severity of illness score for infectious disease outcomes.[27] Recently, a comorbidity risk-adjustment measure was developed for MRSA infections.[39] The Chronic Disease Score (CDS) was modified to include more co-morbidities to create a CDS-MRSA. Similar to the original CDS, patient medications were used to identify the co-morbidities. This method has not yet been well validated and it requires a complete medication history. Confounding attributable to co-morbid conditions has been most commonly estimated using the Acute Physiology and Chronic Health Evaluation (APACHE) score [40, 41], McCabe/Jackson score [42] and the Charlson Comorbidity score.[43] The APACHE score is intended for use with ICU patients and requires clinical parameters not generally available in administrative databases.[40, 41] The McCabe/Jackson score has been evaluated for non-ICU patients and has been used for patients with gram-negative infections. MRSA is a gram- positive organism.[42] The Charlson Comorbidity Index was designed to measure the 1-year mortality risk in a general population of hospitalized patients.[43] It has been modified

to use multiple International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis codes to determine the score.[44] DRG categories have also been used as a surrogate for severity of illness.[45] Although not well validated, DRGs have been used as a surrogate for severity of underlying illness in MRSA bacteremia.[13] DRGs were originally developed to accurately assess the cost of hospitalization by accounting for severity of illness relative to other DRG classifications. [46, 47] A more complete discussion of DRGs can be found in the economic background section.

Data from a meta-analysis demonstrated a significant increase in mortality associated with MRSA bacteremia relative to MSSA bacteremia.[20] Mortality is most typically reported as discharged alive or expired. Differences in mortality are especially important in an economic analysis since patients who die during their hospitalization have truncated costs.

Length of stay (LOS) is known to increase hospital charges; however, LOS must be considered in 2 parts as it relates to the onset of a nosocomial infection. The pre-infection LOS refers to the number of hospital days before infection onset. The post-infection LOS refers to days after the infection onset. The relationship between infection susceptibility status and total hospital charges is complicated by LOS. Post-infection LOS is an intermediary between susceptibility status and charges. As discussed above, it should not be included as a confounder in a model measuring the association between total susceptibility and charges. Pre-infection LOS is related to infection susceptibility,

post-infection LOS and total charges. Therefore, pre-infection LOS can be considered as confounding the relationship between infection susceptibility status and charge.

Hospital level factors can also confound the relationship between susceptibility status and total charges. As previously mentioned, DRG payments are calculated by adjusting for hospital specific factors. Income wage index, hospital teaching status, and percentage of low income patients are also confounders to be considered. These factors are confounders since they directly impact hospital charges and are related to susceptibility status.

Previous Investigations

As previously mentioned, there have been prior investigations attempting to quantify the economic burden of MRSA bacteremia. They suggest MRSA bacteremia increases hospitalization costs by 1.2-to 2-fold over MSSA bacteremia.[11-15] Table 2.2 below outlines investigations that examined MRSA vs. MSSA bacteremia. Only one of the investigations used hospital charge as an outcome variable.[11] This investigation also reported costs as calculated using Medicare cost-to-charge ratios. The estimated difference in median charge attributable to MRSA vs. MSSA was 1.36 fold greater (1.44 for median attributable cost). Regarding the different estimates in hospitalization costs between investigations, it has been hypothesized that the differences are due to disparities in study populations and differences in calculating costs.[27] With the exception of one investigation [16], all previous reports support the hypothesis that infections caused by MRSA are associated with higher costs compared to MSSA infections.

Table 2.2. Investigations comparing the economic impact of MRSA vs. MSSA bacteremia

Author	Setting	Sample Size	Data Source	Economic variable	Outcome Measures	Result	Notes
Abramson MA., et al. [14]	University tertiary care center	8 MRSA 11 MSSA 19 Controls	Pairwise-matched case control study	Total & variable direct costs of hospitalization	• LOS • Cost	• Attributable median LOS: 12 vs. 4 days (MRSA vs. MSSA) • Attributable median total costs: $27,083 vs. $9,661 (MRSA vs. MSSA)	MRSA and MSA groups compared to non-infected controls, not each other.
Cosgrove, S.E., et al. [11]	University tertiary care center	348 (96=MRSA)	Cohort study	Hospital charges were used to approximate costs using the Medicare cost-to-charge ratio	• LOS • Hospital charges	• MRSA had a median attributable LOS of 2 days • MRSA had a median attributable hospital charge of $6,916 ($3,836 median attributable hospital cost)	Charges reported were post-infection only.
Greiner W, et al. [16]	3 University teaching centers in Germany ESRD on hemodialysis	109	Retrospective cohort study	Hospital costs attributable to bacteremia and costs of other bacteremia related medical services after discharge.	• Mortality • Cost	• No difference in duration of stay, outcome or mortality between the groups.	Investigation examined community-acquired and nosocomial infections separately. Largest component of total costs was the initial hospitalization (93%).

(continued)

Table 2.2 (continued)

Author	Setting	Sample Size	Data Source	Economic variable	Outcome Measures	Result	Notes
Lodise, T.P., et al. [12]	University tertiary care center	353 (174 hospital onset)	Retrospective cohort study	Microcosting structure: fixed indirect costs, variable direct costs, & fixed direct costs	• LOS • Costs	• Adjusted mean LOS: 19.1 vs. 14.2 days (MRSA vs. MSSA) • Adjusted mean hospital cost: $21,577 vs. $11,688 (MRSA vs. MSSA)	Included both hospital and community onset SAB. Adjusted model did include a variable for hospital-acquired SAB.
McHugh, C.G., et al. [13]	Tertiary-care hospital	60 (20=MRSA, 42=MSSA)	Retrospective case-control study	All hospital costs accrued by the patient during hospitalization	• Cost	• Cost per day for MRSA ($5,878) vs. for MSSA ($2,073)	Used CMI for severity of illness.
Reed SD., et al. [15]	105 hospital patients ESRD on hemodialysis	143 (54=MRSA, 89=MSSA)	Prospective cohort study	Hospital costs for index hospitalization and for rehospitalization	• Cost • Costs at 12 wks	• Initial hospitalization costs: $21,251 vs. $13,978 (MRSA vs. MSSA) • 12 week costs: $25,518 vs. $17,354 (MRSA vs. MSSA)	Used propensity scores in multivariable regression.

MRSA = methicillin-resistant S*taphylococcus aureus*
MSSA = methicillin-susceptible S*taphylococcus aureus*
LOS = length of stay
ESRD = end stage renal disease

There have been several single center investigations that have quantified the costs of MSSA-SAB and MRSA-SAB.[11-15, 31, 48-52] Each report is briefly summarized below. The first two investigation presented [11, 12] are the most similar to the current study.

Cosgrove et al. [11] examined mortality, length of stay and hospital charges as outcome variables using a cohort design. They concluded that median attributable hospital charge for infection caused by MRSA bloodstream infections was $6,916 per patient. Costs were also estimated using hospital charges adjusted by the Medicare cost-to-charge ratio from onset of infection until discharge. There was no significant difference in mortality between the resistant and susceptible groups, but there was an increased median attributable length of stay for MRSA of two days. This analysis did not match patients with MRSA and MSSA bacteremia. Potential confounders were controlled during statistical analysis. A multivariable model was constructed that adjusted for whether a patient was receiving dialysis, involvement of prosthetic material, comorbidities, surgical wound source, bone and joint source and a severity of illness score.

Lodise et al. [12] characterized hospital costs in a retrospective cohort investigation. Patients with infection caused by MRSA and MSSA were not matched, rather baseline characteristics (e.g., hospital stay before infection onset, APACHE II, age, source of bacteremia) were considered during statistical analysis. The total costs of an infection caused by MRSA and MSSA was $21,577 and $11,668, respectively. All cost data were log-transformed. The cost information reported was the actual cost according

to the hospital where the investigation took place, the Detroit Receiving Hospital. A 1.4-fold (4.9 days) longer post-infection LOS was observed in MRSA patients. This investigation was limited to ICU patients.

Greiner et al. [16] evaluated the costs of nosocomial MRSA versus MSSA bacteremia in patients undergoing hemodialysis. This retrospective analysis considered patients who were hospitalized in one of three German centers. A German refined DRG system was used to calculate costs which were reported in Euros. This investigation is especially important since it found no difference in MRSA vs. MSSA costs and it included data for three hospitals. However, there were only 49 patients between the three hospitals. This investigation was limited to patients with end stage renal disease.

McHugh et al.[13] performed a case-control study to identify risk factors for developing MRSA bacteremia. They did not initially match patients according to severity of illness. However, patients were stratified into two groups based on their case-mix index (CMI). CMI is a hospital level average of individual patient DRGs. Therefore, CMI was used as a surrogate for severity of illness. They determined that the cost of MRSA bloodstream infection was higher by nearly 120% as compared to MSSA bloodstream infections when severity of illness was controlled. Cost data appear to reflect costs from the entire hospitalization, not just the costs from infection onset to discharge. Similar to the report by Lodise et al. [12] cost information appears to be institution specific.

Abramson et al.[14] performed a pairwise-matched nested case-control study to examine the impact of MSSA and MRSA blood stream infections on length of hospital

stay, total costs and variable direct costs attributable to the infection. MRSA infections had a higher median attributable cost than MSSA infections, $27,083 vs. $9,661 respectively. Costs were calculated using actual hospital costs (not charges) for the entire hospitalization. The study matched MSSA and MRSA patients to non-infected hospital controls. In a case-control design it is appropriate to match patients with an outcome (MRSA or MSSA) to the entire at risk population (non-infected hospital controls). This design does not allow for direct comparison between the MRSA and MSSA groups while controlling for confounding.

Reed et al.[15] conducted a cost analysis of MRSA vs. MSSA bacteremia, but their study was limited to patients receiving hemodialysis. This study compared initial hospitalization costs and costs at 12 weeks after initial hospitalization. Initial hospitalization costs were $21,251 vs. $13,978 for MRSA and MSSA respectively. Twelve week costs were $25,518 vs. $17,354 for MRSA and MSSA respectively.

The following investigations are not included in Table 2.2 above. Their methodology was not as similar to the current study. However, they will still be discussed as relevant background information.

Kim et al. [50] performed an analysis to determine the economic burden of MRSA in a university-affiliated, tertiary-care hospital. This analysis was not limited to MRSA bacteremia, other infection sites included soft tissue, surgical sites, pneumonia and osteomylitis. Additionally, patients were not matched. A chart review process was used to determine the attributable days of hospitalization for each MRSA infection.

Costs were then calculated for only attributable days of hospitalization. All costs were reported in Canadian dollars.

Similar to Kim et al., investigations by Capitano et al. [51], Kopp et al. [31] and Rubin et al. [52] were not limited to patients with SAB. Even though Kopp et al. [31] did stratify based on infection location, cost information was not delineated based on location. Also, patients were not matched according to severity of illness. Rubin et al. [52] also stratified by type of infection, but did not match patients with MRSA to patients with MSSA. An additional difference in the paper by Capitano et al. [51] was the long-term care facility setting. Data from long-term care facilities is not necessarily generalizable to hospitals. However, all of these investigations reported higher costs/charges for treating MRSA infections.

Chaix et al. [48] performed a cost-benefit analysis of an MRSA control program in an ICU with endemic MRSA. They concluded that the mean cost attributable to MRSA infection was $9,275. An unmatched case control study by Lepelletier et al.[49] also examined ICU costs. The authors concluded that MRSA involved extra cost due to antimicrobial treatment and quantified that cost in euros. Both of these French studies have limited external validity due to their ICU focus.

The methodologies used in each of the above investigations have advantages and limitations. The current study attempted to incorporate the strengths of these previous investigations whenever possible.

CHAPTER 3

Preliminary Study

Objectives

1. Estimate the crude hospital charges for treating methicillin-resistant *Staphylococcus aureus* (MRSA) bacteremia.

2. Measure the association between MRSA bacteremia and total hospital charges.

3. Estimate the adjusted hospital charges for treating MRSA bacteremia.

Methods

Study population

Eligible patients were those who had *Staphylococcus aureus* bacteremia (SAB) and were discharged from the Virginia Commonwealth University Health System

(VCUHS) between May 1, 2006, and April 31, 2007. If a patient had more than one episode of SAB during their admission, only the first episode was considered.

Study Design

This was a retrospective cohort analysis within VCUHS. Data were extracted for each patient from electronic medical records. The exposure of interest was infection susceptibility, and the outcome was total hospital charges. Patients with *Staphylococcus aureus* were assessed to be either resistant or susceptible to methicillin by the microbiology laboratory at VCUHS. Susceptibility testing was performed and interpreted according to the Clinical and Laboratory Standards Institute (formerly the NCCLS) guidelines.[53] The dependent variable, total hospital charge, was obtained from the hospital accounting department.

Data Collection

Data were collected from 2 databases within the hospital and integrated using medical record numbers. First, the pathology database was electronically queried for patients with *Staphylococcus aureus* bacteremia (SAB). Second, hospital accounting provided charge information for each SAB patient.

Inclusion criteria

Patients were eligible for inclusion if they were > 18 years old and had a blood culture obtained > 48 hours post admission. The onset of infection was defined as the

time the blood culture was obtained. The length of stay from admission to onset had to be at least 48 hours to infer hospital acquisition. Patients who were transferred from another hospital, as noted by their admission source, were not required to be in the hospital for > 48 hours for inclusion in the analysis.

Matching

Frequency matching on pre-infection LOS was used. This ensured the groups had a similar duration of hospitalization before infection onset. The confounding effects of pre-infection LOS on infection susceptibility and charge were controlled by matching.

Statistical Analysis

Descriptive statistics were reported as median and inter-quartile range (IQR) since the outcome variable was not normally distributed. Proportions between the groups were compared using Pearson's chi-squared test. Continuous variables were compared using the Wilcoxon Rank Sum test. A two-sided alpha of 0.05 was considered significant. Statistical analysis was performed using JMP (version 7; SAS Institute, Cary, NC) and SAS (version 9.1.3; SAS Institute, Cary, NC) software.

The median crude hospital charge was approximated from the data for the overall sample and by infection susceptibility group. The median adjusted hospital charge was calculated using a generalized linear model (GLM) with a gamma distribution and a log link. This method employs maximum-likelihood estimation. The model's predicted values were used to estimate the median hospital charge by infection susceptibility group.

The association between total hospital charge and infection susceptibility status was measured by examining the infection susceptibility coefficient. Exponentiated coefficients provide a ratio of means which can be re-expressed as the percentage increase in mean cost per unit increase in the covariate.[54]

Before building the multivariable model, potential predictor variables were assessed in univariate models. Variables had to have a p-value < 0.25 for consideration in the multivariable model.

Since the association between susceptibility status and total hospital charge was being examined, susceptibility status remained in the model without attention to statistical significance. Homogeneity of slope was assessed by evaluating the interaction terms between susceptibility status and each covariate. All one-way interactions were assessed. The likelihood ratio test was used to determine if the interaction terms belonged in the model.

Known confounders (i.e., DRG weight and discharge status) were included in the model. Potential covariates not involved in effect modification were assessed for confounding. A change of more than 10% in the coefficient for susceptibility status was considered significant for confounding.

Influential diagnostics were performed to identify observations that could greatly influence the multivariable model. The standardized Pearson residual and h_{ii} were used to identify leverage, outlier and highly influential points. The affected coefficients were identified using Cook's D and DFBETAS. A correlation matrix was used to assess the potential for pairwise multicollinearity.

Results

During the study period, 219 episodes of SAB were identified. One hundred and thirty-seven patients failed to meet the inclusion criteria. Of the remaining 82 patients, 72 patients met the frequency matching criterion (MRSA = 45, MSSA = 27). No MSSA patients were excluded by the frequency distribution matching.

The overall median total hospital charge was $140,396. The charge for each group can be found in Table 3.1. The crude difference in median charge was $80,771.

Table 3.1 Crude hospital charge in US dollars

	Median	IQR	Minimum	Maximum
Overall	140,396	63,236 to 308,847	11,896	1,169,816
MRSA	166,901	73,526 to 316,101	11,896	1,169,816
MSSA	86,130	52,675 to 232,370	32,498	761,555

IQR = interquartile range
MRSA = methicillin-resistant *Staplylococcus aureus*
MSSA = methicillin-susceptible *Staplylococcus aureus*

The median ages for the MRSA and MSSA groups were 55 years old and 53 years old, respectively. The MRSA group was 53.3% male while 51.8% of the MSSA group was male. The susceptible and resistant groups appear balanced with respect to the measured covariates.

The results of the univariate GLMs are available in Table 4. ICU status, age and DRG weight were significant at $p<0.25$. Variables not significant in the univariate

analysis were not considered for multivariable model inclusion, with the exception of discharge status. Discharge status was included since it was a known confounder.

All two-way interaction effects between infection susceptibility status and the four potential modifiers (ICU status, DRG weight, age and discharge status) were assessed by including an interaction term for each in the multivariable model. Likelihood ratio tests were used to compare the full and reduced models with different combinations of the interaction terms. There were no significant interactions with infection susceptibility status.

All possible two-way interactions between age, ICU status, DRG weight and discharge status were assessed using likelihood ratio tests as described above. There were two significant interactions; DRG weight and ICU status as well as ICU status and age. These interactions were retained in the model along with the two associated lower-order terms. No variables were assessed for confounding. All model variables were either involved in effect modification or were known confounders.

The final model included organism (infection susceptibility status), DRG weight, ICU, age, discharge status, an interaction between DRG weight and ICU and an interaction between ICU and age. The parameter estimates were exponentiated for interpretation (Table 3.2.)

Table 3.2. Multivariable model exponentiated coefficients (hospital charge = dependent variable)

Parameter	Exponentiated Coefficient	Standard Error	Wald 95% Confidence Limits	
Organism	1.276	1.16	0.95	1.71
DRG weight	1.455	1.11	1.18	1.79
ICU status	1.437	1.79	0.46	4.47
Age	0.985	1.01	0.97	1.00
Discharge status	0.845	1.22	0.57	1.25
Age x ICU status	1.026	1.01	1.01	1.05
DRG weight x ICU status	0.743	1.11	0.60	0.94

Organism = infection susceptibility (MRSA vs. MSSA)
DRG = diagnosis related group
ICU = intensive care unit

No pairwise multicollinearity was found. There were no leverage points but six observations were identified as outliers. The analysis was rerun without the outliers and the parameter estimates were within the original model's confidence intervals; thus, removing the outliers did not alter the regression coefficients. Cook's D identified seven observations that may be affecting the regression coefficient estimates. The DFBETAS for these observations indicated the regression coefficient most influenced was the scale parameter.

The multivariable model predicted an overall charge of $163,811 (IQR $117,029 to $230,954). The MRSA and MSSA median charges are presented in Table 3.3.

Table 3.3. Predicted charges

	Overall	MRSA	MSSA
Median ($)	163,811	186,559	145,345
IQR ($)	117,029 to 230,954	125,324 to 308,003	101,440 to 187,123

IQR = inter-qurtile range
MRSA = methicillin-resistant *Staphylococcus aureus*
MSSA = methicillin-susceptible *Staphylococcus aureus*

According to the model, a case of MRSA bacteremia is expected to cost $41,214 more than a case of MSSA bacteremia. This represents a mean increase of 28% for MRSA over MSSA bacteremias.

Discussion

According to the multivariable model, a case of MRSA bacteremia will have 28% higher costs than an MSSA bacteremia. The 95% confidence interval for this estimate ranges from 5% decrease to a 71% increase. Since the confidence interval included no difference (i.e., 1), there is not a statistically different mean charge between the 2 groups. The point estimate was, however, consistent with the 44% increase published in another report.[55] Being discharged alive from the hospital (as opposed to expiring during admission) is associated with a decrease in charge by 15%.

DRG weight, ICU status and age are all involved in effect modification and therefore must be interpreted in light of that interaction. The impact of age while not in the ICU would be a decrease in charge by 1.5% for each additional year of life. In the ICU, each additional year increases charges by 1%. The impact of a one unit increase in

DRG weight when not in the ICU would be an increase in charge of 46%. In the ICU, each unit increase in DRG weight would only increase charges by 8%.

The literature has reported differences in charge as ranging from $5,000 to $40,000 [11-14, 56-59] The crude difference in charge between MRSA and MSSA, $80,771, was much greater than previously reported values. The adjusted difference in charge, $41,214, was still slightly higher than values reported in other investigations. This study found a 1.28-fold increase in charges associated with MRSA bacteremia as compared to MSSA bacteremia. This is consistent with previous reports which report a 1.2-to 2-fold increase.[12, 13]

Previous methods for calculating the economic burden of resistance have varied largely. Many single center investigations were able to collect actual cost from their institution.[12, 14, 15] Another investigation was able to identify only costs associated with the SAB.[16] And one of these investigations was able to collect costs attributable to SAB after hospital discharge.[16] Charges were not able to be sub-categorized in this analysis since only an overall aggregate charge was available for each subject.

Since charges are known to inflate the economic burden of hospitalization, cost-to-charge ratios have been used to better approximate actual cost from charges [11]; however, cost-to-charge ratios are a poor approximation of actual costs [30], which is why they were not used in this investigations. Hospital charges are considered to be a gross overestimation of the true hospitalization cost but, the difference between MRSA and MSSA bacteremia for charge and cost has been shown to be similar.[31]

As previously stated, adjusting for underlying severity of illness was a major concern in this investigation. There is currently no well-validated illness severity score for infectious disease outcomes.[27] Other investigations have used a variety of techniques including APACHE score [12, 18], McCabe/Jackson score [11] and the Charlson comorbidity score.[60] The APACHE score is intended for use with ICU patients while the McCabe/Jackson score has been evaluated for non-ICU patients. The calculation of the Charlson comorbidity score required parameters not available in this analysis (i.e. multiple ICD-9-CM diagnosis codes). This lack of information precluded the calculation of a Charlson comorbidity score. DRGs have been used as a surrogate for severity of underlying illness in MRSA bacteremia.[13] This approach was originally developed to accurately assess the cost of hospitalization by adjusting for severity of illness within the DRG classification scheme.[46, 47] This investigation used DRG weights to adjust for severity of underlying illness.

This analysis is based on two fundamental assumptions. First, the charges prior to infection onset are comparable between the MRSA and MSSA groups. The VCUHS billing department provided one charge for the entire hospitalization. Charges were not available as pre- and post-infection charges. Differences in pre-infection charges could bias the study results. Specifically, the difference in total hospital charge would be overestimated if pre-infection charges were higher in the MRSA group. The difference would be underestimated if the pre-infection charges were higher in the MSSA group. Second, the charges are assumed to be a result of the resistant or susceptible bacteremia. Hospital charges unrelated to the bacteremia introduce bias into the investigation and also

inflate the difference in hospital charges between the MRSA and MSSA and overestimate charge. There is no indication that either assumption was violated.

There were several additional limitations in this investigation. The study relied on previously collected data. Retrospective data can be convenient since the researcher does not have to wait for the data to be prospectively collected. However, records must be complete and accurate or the results will be biased.[61] This dataset contained no missing data.

The small sample size of this investigation limited its precision. This and all previous investigations are plagued by their small sample size. The size of this investigation was comparable to the size of previous reports. A multi-center investigation is needed to adequately power an investigation to achieve accurate and precise estimates.[4]

No attempt was made to estimate the charges associated with infection control. It was assumed that any infection control measures would be the same for all patients within the study timeframe since all patients were from the same hospital. It should be noted that the incidence of MRSA infections can be reduced substantially through prevention.[62]

Finally, the modeling technique used also greatly influenced the results. The use of GLMs is not without limitations. The choice of link function and distribution can greatly impact the model's success. The distribution and link function were appropriately defined prior to analysis; however, the exploration of different link functions and distributions may have altered the parameter estimates. Although the impact of outliers is

minimized, extreme charge values can still influence parameter estimates.[63] The impact of outliers in this investigation is thought to be minimal since the removal of outliers did not significantly change the parameter estimates. The last modeling limitation could probably be more accurately referred to as investigator naivety. Many unanticipated problems arose using the specified GLM. For example, many of the influential diagnostics that are computed automatically for linear regression were not available using PROC GENMOD in SAS. Fortunately the SAS Institute has published some macros that compute some of these diagnostics, but adapting the macro to the study data was less than intuitive.

The results of this investigation are similar in generalizability to the previous investigations. The study was performed in a large tertiary care teaching hospital where patient acuity is relatively high. These results may not extend to a community hospital. To increase external validity, any future investigations should include a variety of hospitals with differing characteristics.

Conclusion

The purpose of this investigation was to quantify the additional hospitalization charge attributable to methicillin-resistant vs. methicillin-susceptible SAB. This objective was achieved by estimating the crude hospital charge for treating MRSA and MSSA bacteremia, measuring the association between MRSA bacteremia and total hospital charges, and estimating the adjusted hospital charge for treating MRSA bacteremia.

The crude "cost of resistance" for treating MRSA bacteremia was estimated and the association between MRSA bacteremia and total hospital charges was explored. The crude difference in hospital charges was $80,771 and the adjusted difference was $41,214. The model accounted for around half of the disparity in cost. The association between bacteremia susceptibility status and hospital charges was a 1.28-fold increase for resistant infections.

Future research in this area is still needed. A large, multi-center investigation needs to be performed to more precisely estimate the association between MRSA/MSSA bacteremia and hospital charges. Also, investigation is needed to explore the relationship between susceptibility status, pre- and post-infection length of stay and hospital charges.

CHAPTER 4

Methods

Data Source

Cerner HealthFacts (Cerner Corporation, Kansas City, MO) is a national data warehouse that represents the electronic patient charts from millions of inpatient admissions and emergency department and outpatient visits at U.S. healthcare organizations.[64]

Within Cerner HealthFacts, patient records (UB-92/UB-04 standard format) contain detailed information on inpatient care. This includes: principle and secondary diagnosis codes (in *International Classification of Diseases, 9th Revision, Clinical Modification* [ICD-9-CM] format), inpatient procedure codes (in ICD-9-CM format), patient demographic information (age, insurer and gender) and hospital demographic information (teaching status and urban/rural). The database also contains admission and discharge dates as well as time stamped microbiologic susceptibility to methicillin for

bacteremia. The analysis was conducted from a hospital perspective. Therefore, only data elements which were important from a hospital perspective were considered.

DRG relative weights from 2006 were used as listed by the Centers for Medicare and Medicaid Services (CMS) in the Acute Inpatient files for download.[65] The weights were integrated into the project database. As previously discussed, each DRG code was assigned a relative weight. DRG weights from 2006 were used since all charges were adjusted to their 2006 value. These relative weights are publically available through CMS.

Study population

The data warehouse was electronically queried for patients with *Staphylococcus aureus* bacteremia (SAB). Eligible patients were those who had SAB and were discharged by a hospital whose data were collected by the Cerner HealthFacts between January 1, 2000, and December 31, 2006. So that every observation within a hospital was independent, only the first episode of SAB per patient was considered.

Inclusion and Exclusion Criteria

Inclusion criteria include: age > 18 years old and onset of infection > 48 hours post admission. The length of stay from admission to onset must have been at least 48 hours to infer hospital acquisition. Patients who were transferred from another hospital as their admission source were excluded.

Study Design

This was a retrospective cohort analysis using the Cerner HealthFacts data warehouse. The exposure of interest was infection susceptibility, and the outcomes were post-infection LOS and total hospital charges. Patients were identified as having SAB according to microbiologic susceptibility determined by the hospital's microbiologic laboratory. Infection susceptibility standards were developed by the Clinical and Laboratory Standards Institute (formerly the NCCLS) guidelines[53].

Matching

Frequency matching is a technique that can be used to achieve reasonable efficiency between two groups. This investigation used frequency matching by pre-infection LOS to exclude observations with an extremely long pre-infection LOS. The MRSA and MSSA groups should have similar hospital lengths of stays before infection onset. However, the confounding effects of pre-infection LOS on infection susceptibility and charge cannot be completely controlled by frequency matching. Therefore, pre-infection LOS was also used as a variable in the multivariable model.[66]

Outcomes

There were two outcomes of interest in this analysis. First, post-infection LOS in days was evaluated. For the purposes of this analysis, LOS was divided into two parts. The pre-infection LOS is the LOS from admission to the time when the initial positive blood culture was drawn. For this analysis, the time of blood culture is referred to as the

infection onset. The post-infection LOS is the LOS from infection onset to discharge. This distinction attempted to designate the hospital stay attributed to the SAB from the stay attributed to the primary reason for hospitalization.

Second, total charges in US dollars were assessed. Charges were adjusted for inflation using the 2006 Consumer Price Index for hospitals.[67] When comparing economic values over multiple years the relative value of money must be considered. Normally, money is worth less in the future than it is worth today. Therefore, past dollars must be adjusted for inflation.[68]

The consumer price index (CPI) measures the average change over time of goods and services and is generally used as a measure of inflation. The reference index for the CPI is set at 100 which represents the average price level for the 36 month period between 1982 and 1984. The reported annual CPI reports a change relative to the reference index. For example, an index of 120 would mean that there has been a 20% increase in price since the reference period. An index less than 100 would reflect a decrease in price. Movements of the index from one date to another can be expressed as the difference between index levels. But, it is more useful to express the movements as percent changes. The CPI allows for comparisons of consumer costs over time. Table 4.1 below shows the CPI of medical care services that pertain to hospital and related services.

Table 4.1. Consumer Price Indexes for Hospital and Related Services: 2000 to 2006

	CPI for Hospital and Related Services	Percent change from previous year
2000	317.3	
2001	338.3	6.62
2002	367.8	8.72
2003	394.8	7.34
2004	417.9	5.85
2005	439.9	5.26
2006	468.1	6.41

CPI = Consumer Price Index

Statistical Analysis by Objective

Statistical analysis was performed using JMP (version 7; SAS Institute, Cary, NC) and SAS (version 9.1.3; SAS Institute, Cary, NC) software. JMP was primarily used for descriptive statistics while SAS was used for the GLM and path analysis. Proportions between the groups were compared using Pearson's chi-square. Continuous variables were compared using a Wilcoxon Rank Sum test. A two-sided alpha of 0.05 was considered significant. The following statistical methods were used to address each objective.

Modeling economic data is less than straightforward. The most commonly used regression technique, ordinary least squares (OLS), is not generally appropriate for the following reasons.[69] First, the dependent variable, total hospital charge, is rarely normally distributed. The typical distribution is bounded by zero with a long right tail (most closely resembling a gamma distribution). OLS requires the error terms to be normally distributed for statistical inference. The second problem with OLS is the

homogeneous variance assumption. With economic data the variance often increases as the mean increases making OLS inappropriate. Several statistical solutions have been proposed to more appropriately model charges.

A log transformation can be performed to make the dependent variable more normally distributed. However, resulting estimates are in terms of "log charges." This often leads to problems with interpretation. Smearing factors have been used to transform economic data from logarithmic back to natural units, but these factors can introduce substantial bias in the presence of heteroscedasticity.[70]

A generalized linear model (GLM) can be used to create virtually any model. The most appropriate distributions for cost data would be the gamma or inverse Gaussian distribution.[71] Both of these distributions are appropriate for non-zero continuous outcomes. However, most economic data analysis utilizes the gamma distribution.[54] The gamma distribution is also appropriate since it assumes that variance is proportional to the square of the mean. Economic data is non-zero, continuous and usually has a variance which increases with the mean. The negative binomial distribution would not be appropriate since it assumes a categorical outcome. The link function is not a transformation on the data, but a transformation of the population mean. The most commonly used link function for economic analysis is the log link.[54] Although the log link is not the canonical link function for the gamma distribution, it has been used to analyze cost data because it will not predict a negative value. Using the log link, an exponentiated coefficient provides a ratio of the means.[54]

Observational investigations attempt to estimate the effect of an exposure by comparing outcomes for subjects not randomly assigned to the exposure of interest.[72] In randomized clinical trials, random assignment serves to balance covariates so that the study groups are comparable with respect to the distribution of their covariates.[73] The presence of dissimilar groups can introduce systematic error into an observational study.[61]

The theory of counterfactuals contemplates the outcome of the exposed group if there was no exposure or conversely the outcome of the unexposed if there had been exposure. The term "counterfactuals" indicates that at least one of the two circumstances is contrary to fact.[6] There is no way to estimate the counterfactuals, but the bias of the estimates can be corrected through the predicted probability of exposed vs. unexposed using observed predictors. The predicted counterfactual can be estimated using propensity scores.

Propensity scores are the probability of exposure given measured baseline variables.[9] Observational studies employ this method to adjust for observable bias with the goal being to eliminate bias from unequal distribution of confounders. To do this, a group of likely confounders is converted into one scalar score through a two-stage process. The actual propensity score is unknown and therefore must be estimated using a logit or probit regression model where the dependent variable is exposure status and the independent variables are the potential confounders.[74] The score is the probability ($0 \leq$ propensity score ≤ 1) of receiving the treatment (i.e., exposure) based on the set of identified covariates as predicted by the first stage regression model. The score can be

used as a matching or stratification factor, as a covariate in multivariable model or to perform inverse probability of exposure weighting.[10]

Propensity scores can be used to match patients between exposure groups on multiple confounders. However, the possibility of over-matching must be considered. This occurs when matching is done incorrectly or unnecessarily. The more variables used to calculate the propensity score, the greater the likelihood of overmatching. To avoid overmatching, the propensity score should preferably include only established confounders.[75]

As previously discussed, stratification becomes difficult when many confounders are present since there are not enough observations in each stratum. Stratification on propensity score limits the number of stratum thus making stratification a more robust method to control for confounding.[76]

If the score is used in a multivariable model, the second part of the two-stage regression process is the traditional model where the dependent variable is the outcome of interest and the propensity score is used as a covariate. The advantage of including propensity scores over traditional regression is in not over-parameterizing the model. [77] Normally each confounder would require one degree of freedom (df) while the propensity score, which could be comprised of many confounders, only requires 1 df.

Several papers have been published exploring propensity score methodology.[9, 74, 78] Of particular interest, an investigation examining the economic impact of MRSA bacteremia used propensity scores.[15] The propensity score was calculated using demographic characteristics, comorbidities, risk factors, clinical characteristics and

infection related factors. It was then used as an independent variable in the multivariable model. This investigation will be discussed more fully later in this report.

Propensity scores are not without limitation. They do not balance uncorrelated, unmeasured characteristics and confounders.[79] Unmeasured confounding can be dealt with using other methods. Additionally, covariates cannot be used that are directly affected by the exposure of interest.[80] If the covariate is directly affected by the exposure, it may be an intermediary not a confounder. Including an intermediary in the propensity score would introduce bias into the investigation. Lastly, the models used to generate the propensity score rely on the same assumptions as logistic regression. If the model uses the propensity score as a continuous variable, the assumption of a (log-) linear association with the dependent variable must be tested using categories. The covariates must be balanced across the groups.[81]

Objective 1: Measure the crude difference between infection susceptibility status MRSA vs. MSSA bacteremia and total hospital charges.

This aim was accomplished by subtracting the mean MSSA charge from the mean MRSA charge. The charges were compared using a Wilcoxon Rank Sum test. Additionally, trends in hospital charge over the study period were examined using repeated measures analysis of variance (ANOVA). Tukey's HSD was performed if a significant difference was found.

Objective 2: Determine the impact of methicillin resistance vs. methicillin susceptibility in SAB on total hospital charges while adjusting for potential confounders.

The impact of infection type on hospital charge was calculated using a GLM with a gamma distribution and a log link while adjusting for various factors. This method employs maximum-likelihood estimation. Variables with a univariate p-value < 0.25 were eligible for inclusion in the multivariable model. Known confounders (e.g. severity of illness, DRG weight) and potential covariates were eligible for inclusion. A propensity score was used to include hospital level factors (e.g., hospital teaching status, urban/rural status). Propensity scores require a two stage regression procedure. In the first stage, the outcome variable was infection susceptibility and the independent variables were the hospital level factors. The second stage was the multivariable model where the predicted value from the first model was used as an independent variable. The covariates were checked for balance. Using the predicted values from the multivariable regression model, the mean charge of MSSA bacteremia was subtracted from the mean charge of MRSA bacteremia.

Objective 3: Determine the impact of methicillin resistance vs. methicillin susceptibility in SAB on post-infection LOS while adjusting for potential confounders.

The impact of infection type on post-infection LOS was calculated using a GLM with a gamma distribution and a log link while adjusting for various factors. This method employs maximum-likelihood estimation. Variables with a univariate p-value < 0.25 were eligible for inclusion in the multivariable model. Known confounders (e.g., severity

of illness, DRG weight) and potential covariates were eligible for inclusion. A propensity score was used to include hospital level factors (e.g., hospital teaching status, urban/rural status). Propensity scores require a two stage regression procedure. In the first stage, the outcome variable was infection susceptibility and the independent variables were the hospital level factors. The second stage was the multivariable model where the predicted value from the first model was used as an independent variable. The covariates were checked for balance. Using the predicted values from the multivariable regression model, the mean LOS of MSSA bacteremia was subtracted from the mean LOS of MRSA bacteremia.

Objective 4: Describe the relationship between SAB methicillin susceptibility, LOS and total hospital charges using path analysis.

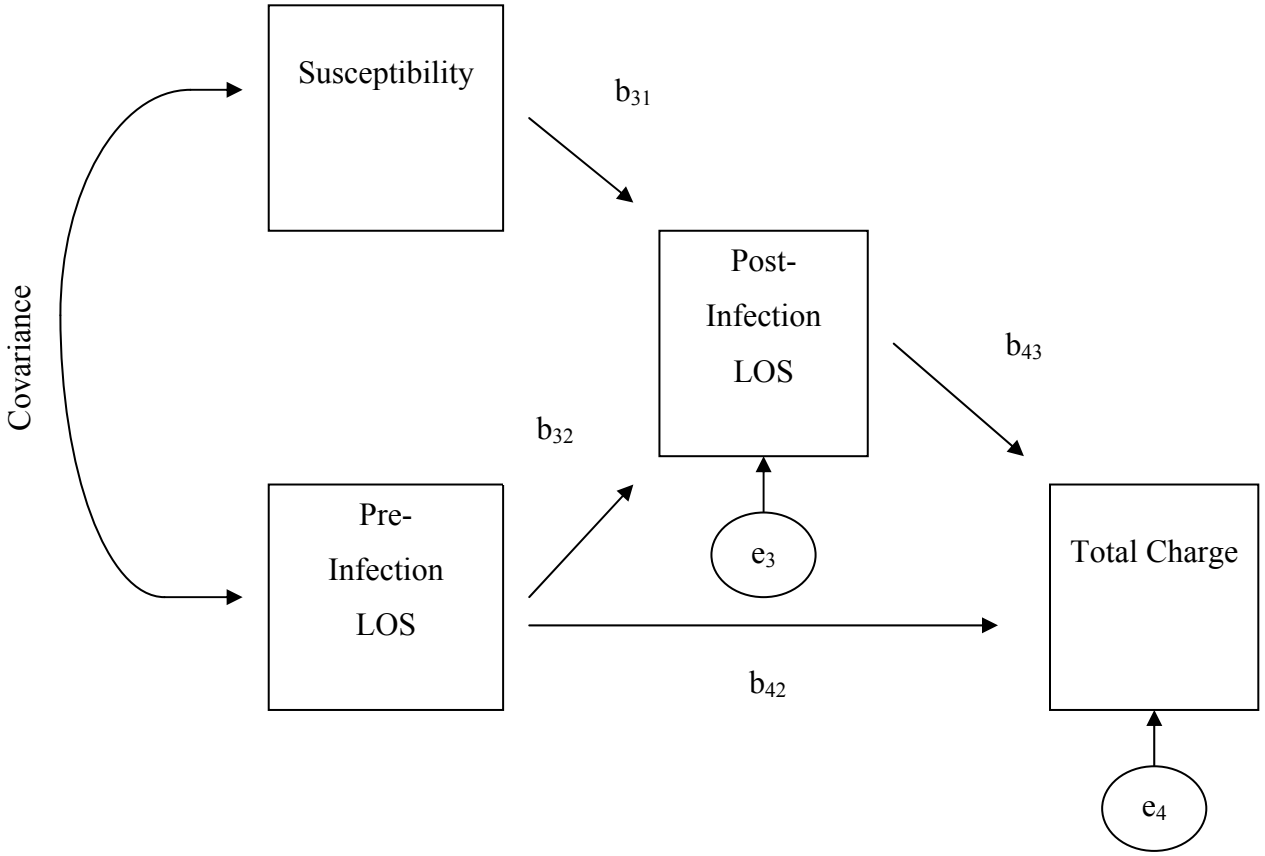
Path Analysis is a type of analysis that uses multiple regression modeling to explore complex relationships. The purpose is to provide estimates of the magnitude and significance of hypothesized connections between variables. As previously discussed, the relationship between hospital LOS and total hospital charges is not simple. Post-infection LOS is an intermediate between infection susceptibility and hospital charges. Although pre-infection LOS confounds the relationship between infection susceptibility and charge, there is also a relationship between pre infection and post-infection LOS. The complex relationships involving LOS can be explored using path analysis.

There are 5 general steps in path analysis.[82] First, the model must be specified, which involves formally stating the proposed model. Specification is probably the most

important step because if the model is incorrect the results will be meaningless. Of note, variables in path analysis need to be continuous except if the variable represents group membership or is an exogenous variable. Categorical variables cannot be endogenous.[82] Second, the model must be identified. Identification refers to verifying the specific model assumptions. Similar to regression, observations need to be independent, normally distributed and have uniform variances.[82] Third, the model is estimated. Logistically, this occurs using a series of regression models. Many statistical software packages are available that perform this function. Fourth, the fit of the model is tested. The fit of the model is good if the fitting function is close to zero. If the ratio between chi-square and the degrees of freedom is less than 2, then the model fit is good.[83] Finally, the model can be manipulated. This can be useful especially if the model fit is not good.

The proposed path analysis model is diagramed below in Figure 4.1. The direct and indirect impact of one variable on another can be estimated.[82] Direct effects are represented when a single arrow connects two variables. Indirect effects are defined when no single line directly connects two variables, but instead, the variables are connected through one or more other variables along their path.[84] Indirect effects measure the impact of intermediates or mediator variables.[82] The endogenous structural equations used that correspond to Figure 4.1 are described below in Equations 4.1 to 4.2. The model will be tested using SAS's PROC CALIS. The model fit was assessed using chi square.

Figure 4.1. Path Analysis diagram



$$\text{Post-infection LOS} = b_{31} \times \text{susceptibility status} + b_{32} \times \text{pre-infection LOS} + e_3 \qquad \text{Eq. 4.1}$$

$$\text{Charge} = b_{43} \times \text{post-infection LOS} + b_{42} \times \text{pre-infection LOS} + e_4 \qquad \text{Eq. 4.2}$$

Human Subjects Protection and Data Privacy

The largest potential risk for the subjects was exposure of medical information. To ensure minimal risk to the patients, the data were coded and encrypted. The data did not contain patients' medical record numbers. Access to the dataset was restricted to those individuals listed on this protocol, and the dataset was centrally maintained in a password-protected environment. Multiple copies of the data were minimized. In the event that additional copies of the dataset were required for the conduct of the study, the principle investigator maintained a list of those copies and ensured that all extra copies were appropriately destroyed once analysis was finished so that only one copy was retained to satisfy university record keeping policies relevant to research data. Disclosure of information did not take place without the expressed written permission of Cerner HealthFacts or required by law. Data within HealthFacts is compliant with the Health Insurance Portability and Accountability Act of 1996 (HIPAA). Results will be published in such a way that no subject will be individually identifiable. This study qualified for exemption according to 45 CFR 46.101(b) Category 4 at Virginia Commonwealth University internal review board (IRB). (VCU IRB#: HM11841). A copy of the IRB Approval form can be found in Appendix A.

CHAPTER 5

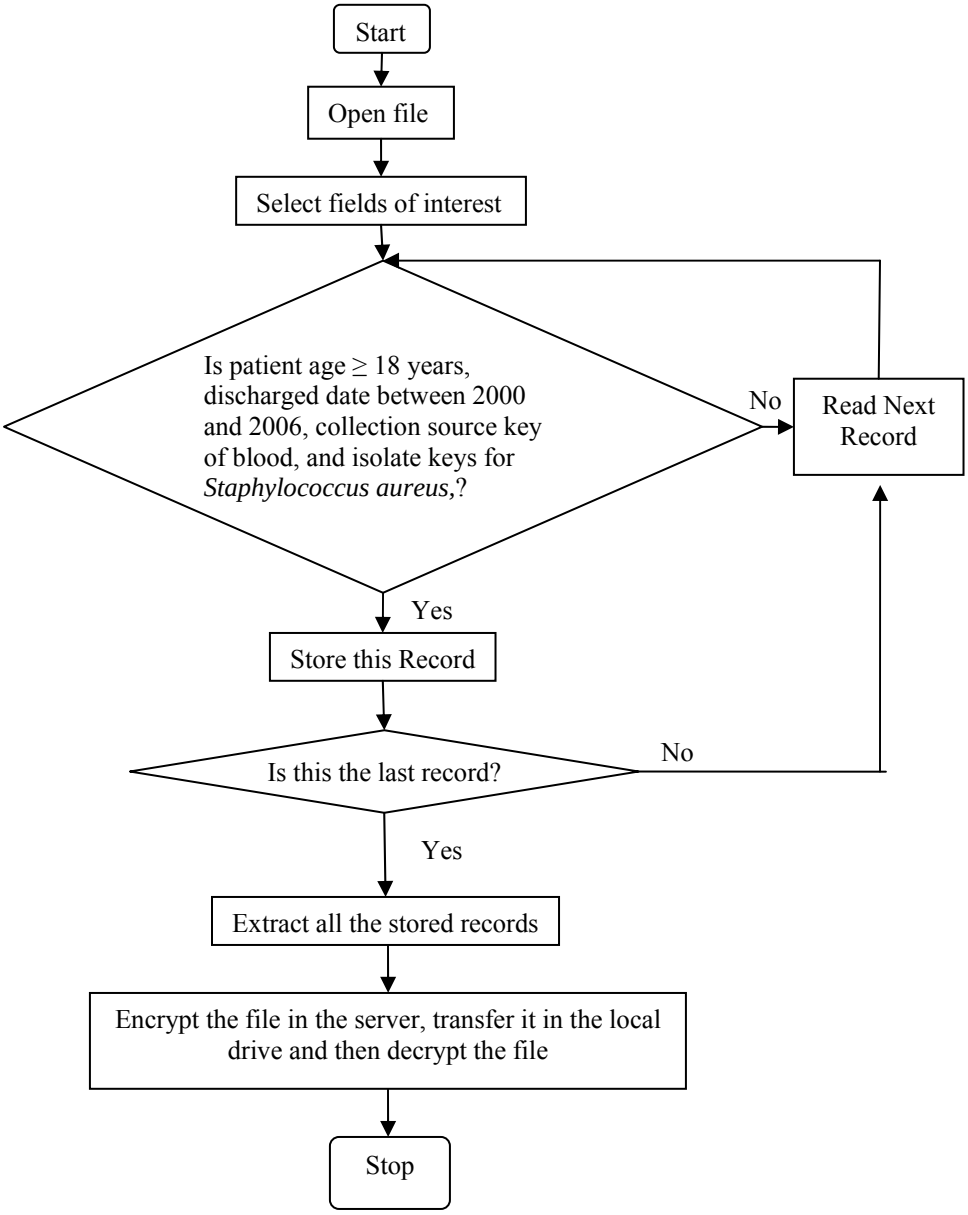
Results

Data Manipulation

Data from the Cerner HeathFacts data warehouse were received as numerous files that were cleaned by a programmer. The flow chart below (Figure 5.1) outlines the general methodology the programmer used to extract and combine the data. Data fields were extracted that contained the variables of interest when: (1) the patient age was greater than 18 years old, (2) discharge date was between 2000 and 2006, (3) collection source key of blood, and (4) isolate key for *Staphylococcus aureus*. Files that contained the variables of interest were joined so that all the necessary data elements were contained in two datasets. The first dataset contained the microbiologic information (referred to as Micro) and the second dataset included all the other variables including time stamped dates (referred to as Large). The programmer also deleted duplicate observations. Only observations that contained identical information for all data

elements were excluded here. Data were provided to the investigator in a much improved form, but some manipulation was still required.

Figure 5.1. Flow chart used by programmer for data retrieval

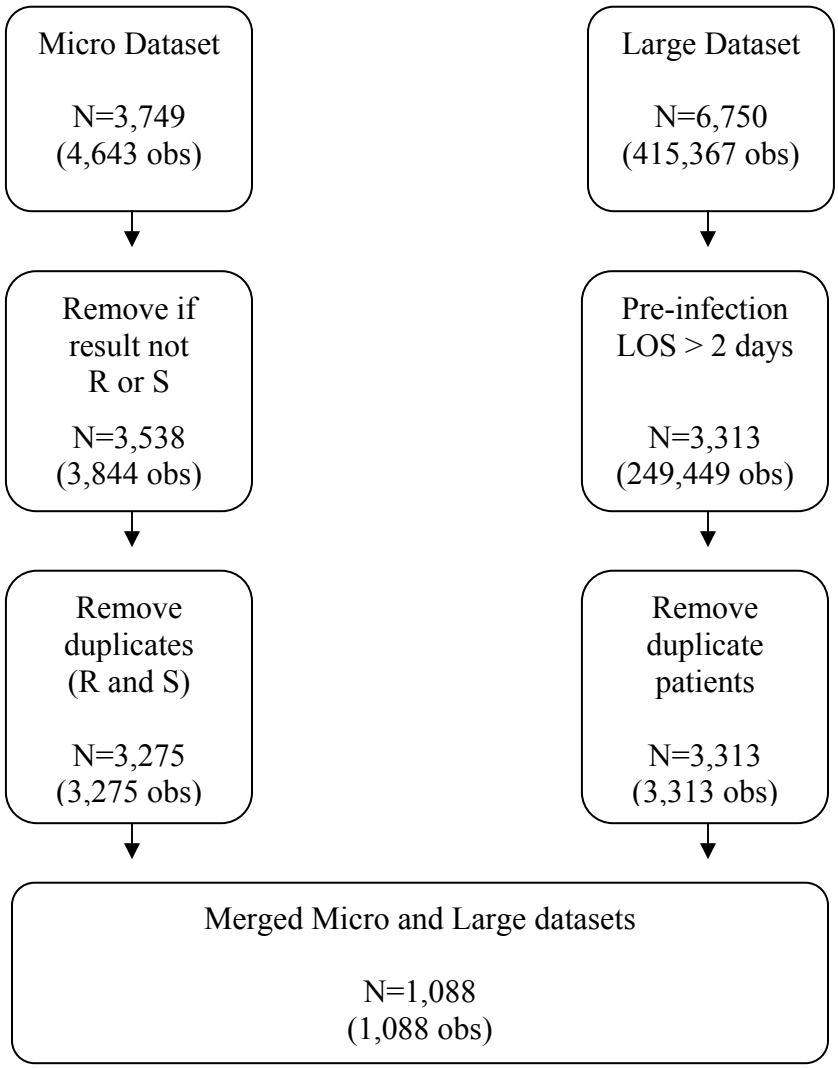


The Micro dataset contained a unique patient identified and recorded isolate susceptibilities. These susceptibilities contained a variety of values, such as resistant, susceptible, null and moderately sensitive. Only observations where susceptibilities were reported as resistant or susceptible were retained in the dataset. All microbiologic susceptibility testing was done with oxacillin or methicillin. Some patients had both a resistant and a susceptible isolates (n=263). A flowchart describing this cleaning process is available below (Figure 5.2).

The Large dataset contained all other data elements. Patients were excluded if their pre-infection length of stay (LOS) was less than two days. Duplicate patients were removed for a final Large dataset n=3,313. The Micro and Large datasets were merged creating one dataset with 1,088 patients. A flowchart is provided below in Figure 5.2.

A substantial number of patients in both the Micro and Large datasets were excluded during the merge since the datasets contained observations from different hospitals. Table 5.1 below shows the number of excluded hospitals per year.

Figure 5.2 Flow chart used by investigator for data manipulation



obs =observations
LOS=length of stay
R= Resistant
S=Susceptible

Table 5.1 Number of Hospitals excluded by the merge of Micro and Large datasets

	Merged Dataset (Micro + Large)	Large Dataset	Excluded Hospitals
2000	9	9	0
2001	15	16	1
2002	13	18	5
2003	17	20	3
2004	15	19	4
2005	14	26	12
2006	22	32	10

Relative DRG weights, publically available through CMS, for 2006 were added into the dataset.[65] Some DRGs from earlier years were no longer valid in 2006, meaning the codes were no longer used. The patients with the invalid DRG codes were removed (n=966). A Charlson comorbidity score was calculated for each patient.

After examining the pre-infection LOS distributions for MRSA (Figure 5.3) and MSSA (Figure 5.4), frequency matching was performed. Only one MSSA observation had a pre-infection LOS longer than 62 days (circled below in Figure 5.4). There were eight MRSA patients excluded. Therefore, patients with a pre-infection LOS greater than 62 days were excluded from the analysis (n=957).

Figure 5.3 Distribution of pre-infection LOS (MRSA)

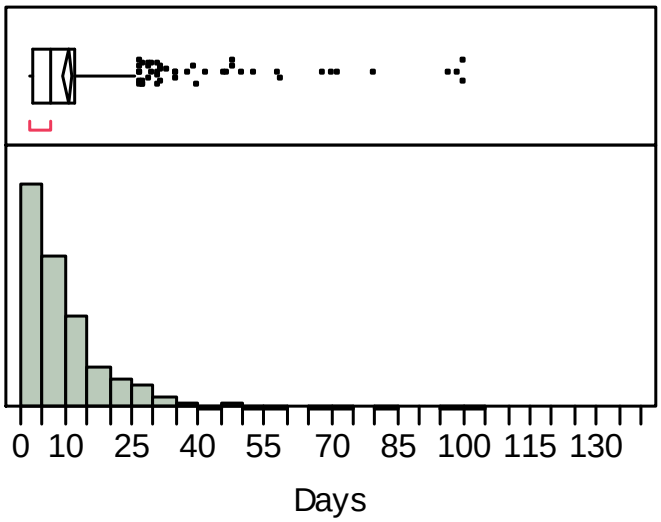
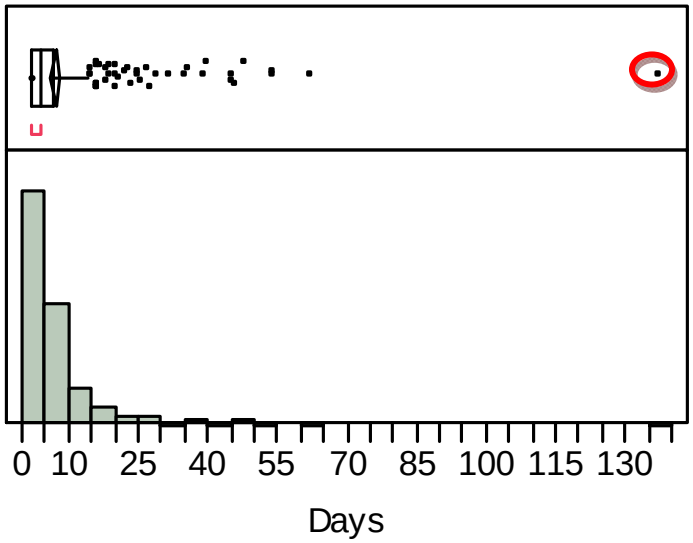
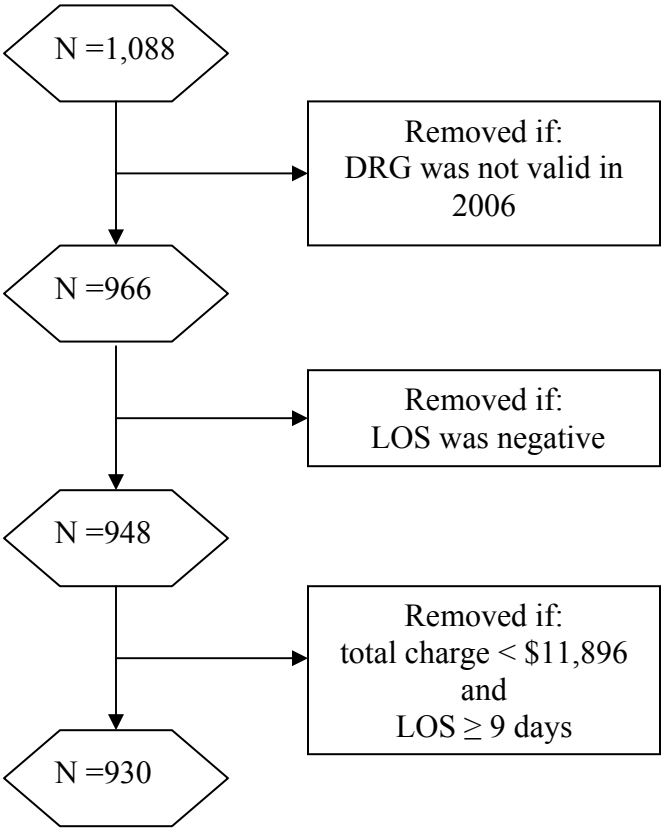


Figure 5.4 Distribution of pre-infection LOS (MSSA)



The dataset was then examined for erroneous data points. Some observations existed with a negative LOS. These observations were removed. Some observations reported incredibly low charges (less than $100). There was not a previously defined method to deal with this scenario. The data were explored but there was no apparent pattern to the low charges. The investigator chose a reasonably prudent criterion for inclusion that would exclude patients with an unrealistically low charge. The pilot investigation had a minimum charge of $11,896. Patients with charges less than $11,896 and with LOS $\geq$ 9 days were assumed spurious and were removed. The final dataset contained 930 observations. See Figure 5.5 below.

Figure 5.5 Flowchart leading to final dataset



Hospital charges were inflated to their 2006 value. CPI adjusted rates (compounded annually) were calculated for each year. Table 5.2 below assumed a $1.00 reference value. Total hospital charges were adjusted to the 2006 by multiplying the reported charge by the CPI adjusted rate. For example, 2000 values were increased by 48%, and 2001 values by 38% to estimate their 2006 value. Figure 5.6 below shows step by step how the CPI adjusted rates were calculated. The inflation rates can be found in Table 4.1.

Table 5.2 CPI adjusted rates

	2000 to 2006	2001 to 2006	2002 to 2006	2003 to 2006	2004 to 2006	2005 to 2006
2000	1					
2001	1.07	1				
2002	1.16	1.09	1			
2003	1.24	1.17	1.07	1		
2004	1.32	1.24	1.14	1.06	1	
2005	1.39	1.30	1.20	1.11	1.05	1
2006	1.48	1.38	1.27	1.19	1.12	1.06

CPI = Consumer Price Index

Figure 5.6 Calculation of CPI adjusted rate

2000 to 2001 had an inflation rate of 6.62%
2001 to 2002 had an inflation rate of 8.72%

Assuming the value in 2000 was $1, the value in 2001 would be:
$1.00 x (1 + 0.0662) = $1.07

The value in 2002 would be: $1.07 x (1 + 0.0872) = $1.16.

This means that the adjusted inflation rate for 2000 to 2002 was 16%.

Descriptive Statistics

Demographic data for the categorical variables (admission source, payer type, and discharge disposition) are described below in Tables 5.3. There were more categories for each categorical variable that were collapsed. Additional categories can be found in Appendix B. Of the 930 patients, 546 were MRSA (58.7%) and 384 were MSSA (41.3%). There was a significant difference in admission source ($\chi^2 = 10.43$, df = 3, p-value = 0.0152). More MSSA were admitted from the Emergency Room and more MRSA patients were transferred to the hospital. There was also a significant difference in discharge status ($\chi^2 = 19.55$, df = 5, p-value = 0.0015). More MSSA patients were discharged and more MRSA patients expired in the hospital. Discharge status was examined strictly as alive, dead or not available; 30.8% of MRSA patients expired in the hospital compared to 21.4% of MSSA patients ($\chi^2 = 11.92$, df = 2, p-value = 0.0026). A p-value was not calculated for the payer status since a majority of the patients had data classified as not available. The groups were balanced with respect to gender; the MRSA group was 53.5% male while 53.4% of the MSSA group was male ($\chi^2 = 0.008$, df = 1, p-value = 0.9773).

Demographic data for the continuous variables can be found in Table 5.4. Data is reported as medians and interquartile ranges since the variables were not normally distributed. The Wilcoxon Rank Sum test indicated all continuous variables were significantly different between the MRSA and MSSA groups except post-infection LOS.

Table 5.3 Descriptive statistics for categorical variables (n=930)

	MRSA n (column %)	MSSA n (column %)	Total n
Admission Source			p-value = 0.0152
Referral	96 (17.6)	68 (17.7)	164
Transfer	55 (10.1)	19 (4.9)	74
Emergency Room	261 (47.8)	212 (55.2)	473
Not available	134 (24.5)	85 (22.1)	219
	546	384	
Discharge Status			p-value = 0.0015
Discharged	163 (29.9)	151 (39.3)	314
Transferred	172 (31.5)	120 (31.3)	292
Hospice	16 (2.9)	9 (2.3)	25
Expired	168 (30.8)	82 (21.4)	250
Other	5 (0.9)	11 (2.9)	16
Not available	22 (4.0)	11 (2.9)	33
	546	384	
Payer Type			p-value= not calculated
Insured	33 (6.0)	35 (9.1)	68
Medicare	20 (3.7)	15 (3.9)	35
Medicaid	145 (26.6)	85 (22.1)	230
Self Pay	9 (1.6)	3 (0.8)	12
Not available	339 (62.1)	246 (64.1)	585
	546	384	

MRSA = methicillin-resistant *Staphylococcus aureus*
MSSA = methicillin-susceptible *Staphylococcus aureus*

Table 5.4 Descriptive statistics for continuous variables (n=930)

	MRSA		MSSA		
Variable	Median	IQR	Median	IQR	p-value
Age [1]	72	59 to 80	67	53 to 78	0.0003
Charlson comorbidity score	1.00	0 to 2.00	1.00	0 to 2.00	0.0138
DRG weight	1.68	1.21 to 3.07	1.35	1.04 to 2.39	0.0004
Pre-infection LOS[2]	6	3 to 12	6	2 to 7	<0.0001
Post-infection LOS[2]	6	2 to 14	6	3 to 12	0.6039
Total LOS[2]	15	9 to 25	12	8 to 19	<0.0001

MRSA = methicillin-resistant *Staphylococcus aureus*
MSSA = methicillin-susceptible *Staphylococcus aureus*
IQR = interquartile range
DRG = diagnosis related group
LOS = length of stay
[1] Age was measured in years
[2] LOS was measured in days

Objective 1:

Crude difference in hospital charge by infection susceptibility

The total mean hospital charge was $111,636. The charge for each group can be found below in Table 5.5. The crude difference in mean charge was $24,406. Using the Wilcoxon Rank Sum, there was no difference between the MRSA and MSSA groups (p-value < 0.0001).

Table 5.5 Crude hospital charge in US dollars

	Mean	SD	Median	IQR		Minimum	Maximum
Overall	111,636	241,211	59,764	30,751	121,956	3,323	3,392,801
MRSA	121,713	252,465	68,013	33,247	131,060	4,035	3,392,801
MSSA	97,307	223,781	49,199	27,338	98,898	3,323	2,957,732

US = United States
SD = standard deviation
IQR = interquartile range
MRSA = methicillin-resistant *Staplylococcus aureus*
MSSA = methicillin-susceptible *Staplylococcus aureus*

The intention was to perform repeated measures ANOVA. However, only four hospitals had data every year from 2000 through 2006. Since four hospitals represent a small fraction of the datasets, repeated measures ANOVA was not appropriate. Instead, descriptive statistics were analyzed graphically to see if a difference appeared within the study period.

Descriptive statistics for total charge by discharge year are presented below (Tables 5.6 and 5.7). "N" represents the number of observations per year. The mean for

2001 was much higher than any other year. The maximum values for 2001 were $3,392,801 and $2,957,732 for MRSA and MSSA respectively. There were five extreme values in 2001 which artificially inflated the 2001 mean. All the other observations appeared within one standard deviation of one another. Starting in 2002, there did appear to be a slight downward trend in the MSSA mean charge. However, this trend was not echoed by the MRSA data.

Boxplots for the total charges per year are below in Figure 5.7. The boxplot is labeled with the number of outliers per year. For the purposes of this figure, an outlier was defined as above the 75th percentile. Of note are five extreme outlier observations for 2001. They represent both MRSA and MSSA cases. In Figure 5.8, the interquartile ranges have been enlarged to more closely examine the means which are depicted by red triangles. The only mean that appeared different was 2001. This mean was extreme due to five extreme outliers that will be addressed later in the analysis. An alternative statistical test, instead of repeated measures ANOVA, was not pursued further since there was no visually apparent difference in the means.

Table 5.6 Descriptive statistics for total hospital charge by discharge year for MRSA

Discharge year	N	Mean	Std Dev	Min	Max
2000	35	102,313	105,190	10,746	487,359
2001	115	222,773	502,148	8,641	3,392,801
2002	93	116,864	123,483	6,141	574,895
2003	62	103,628	111,777	4,339	461,235
2004	49	66,633	45,819	14,946	182,601
2005	67	94,264	106,543	4,035	675,738
2006	125	83,050	90,009	4,710	683,193
Overall	546	121,713	252,465	4,035	3,392,801

Std Dev = standard deviation
Min = minimum
Max = maximum

Table 5.7 Descriptive statistics for total hospital charge by discharge year for MSSA

Discharge year	N	Mean	Std Dev	Min	Max
2000	24	68,359	46,113	4,414	180,160
2001	89	168,224	431,715	9,888	2,957,732
2002	69	103,714	108,640	10,074	435,997
2003	27	79,214	65,329	5,260	229,330
2004	24	76,372	93,130	3,323	397,363
2005	53	63,210	82,007	7,826	517,451
2006	98	64,030	79,107	4,324	445,316
Overall	384	97,307	223,781	3,323	2,957,732

Std Dev = standard deviation
Min = minimum
Max = maximum

Figure 5.7 Boxplot of Total Charge by Discharge Year

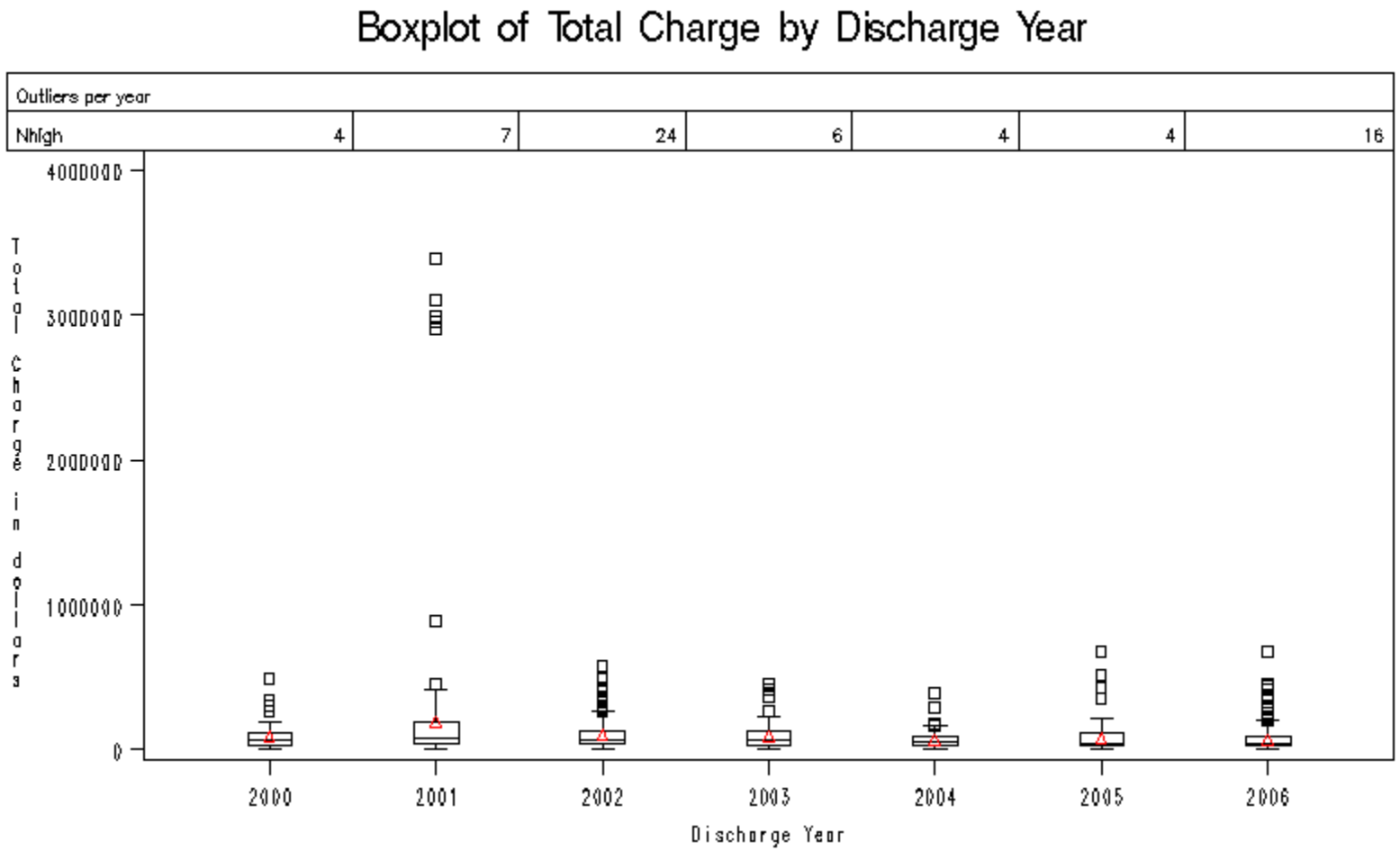
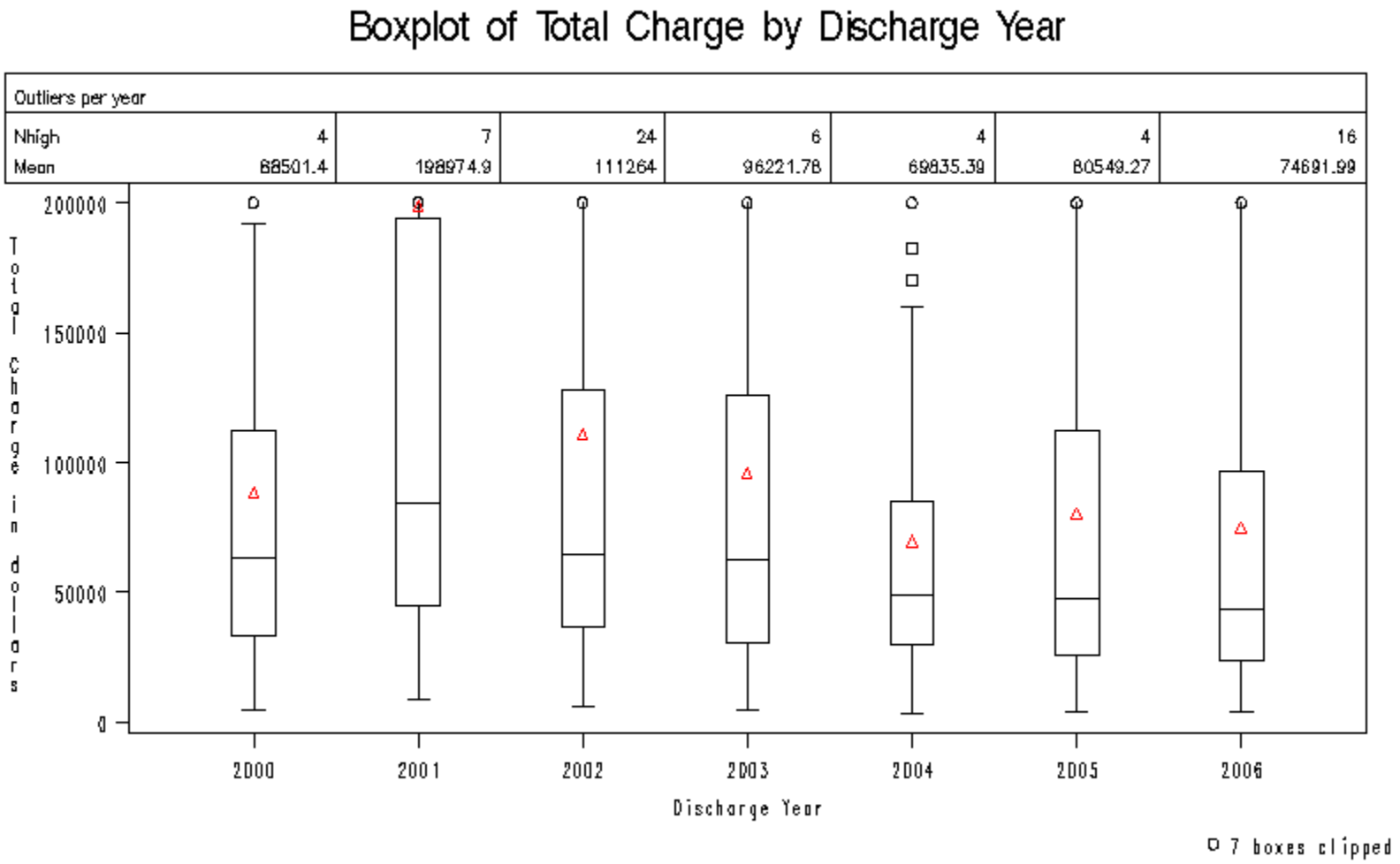


Figure 5.8 Boxplot of Total Charges by Discharge Year with outliers omitted



Objective 2:

Adjusted difference in hospital charge by infection susceptibility

A GLM utilizing a gamma distribution and logarithmic link was used to estimate total charge adjusting for potential confounders. Before a multivariable model could be analyzed, each potential covariate was evaluated for inclusion in the model. Using total charge as the dependent variable, the results of the univariate GLMs are available in Table 5.8. Admission source was not considered since data were missing for 219 patients. All variables were significant at p-value < 0.25, except for discharge status, and thus eligible for inclusion in the multivariable model. Discharge status was still included in the multivariable model since it has been previously established as a known confounder.

Table 5.8 Univariate analysis (hospital charge = dependent variable)

Variable	Coefficient	Exponentated Coefficient	Standard Error	Wald 95% Confidence Limits		p-value
Gender	-0.227	0.7969	1.0684	0.7001	0.9073	0.0006
Discharge Status	0.0413	1.042	1.078	0.899	1.208	0.5838
Age	-0.0033	0.997	1.002	0.992	1.001	0.1464
DRG weight	0.1596	1.173	1.020	1.129	1.219	<0.0001
Charlson comorbidity score	-0.0548	0.947	1.015	0.920	0.975	0.0002
Pre-Infection LOS	0.0428	1.044	1.004	1.036	1.052	<0.0001

DRG = diagnosis related group
LOS = length of stay

A propensity score was calculated for the hospital level factors (i.e. bed size, urban/rural, teaching status.) The independent variable was infection susceptibility (MRSA or MSSA) and the dependent variables were the hospital level factors. Propensity scores were only calculated for 900 patients since 30 patients had missing data. Goodness of fit was assessed using the Hosmer-Lemeshow goodness of fit test (χ^2 = 1.6365, df = 5, p-value = 0.8968). This indicated good fit. Balance between the MRSA and MSSA groups was reached with respect to the confounders used to calculate the propensity score (Table 5.9). The distribution of propensity scores is below in Figures 5.9 and 5.10.

Table 5.9 Propensity score by confounders included in the score

	MRSA	MSSA
Bed size		
6 to 99	0.65	0.62
100 to 199	0.41	0.44
200 to 299	0.38	0.39
300 to 499	0.41	0.43
500+	0.41	0.42
Urban/Rural Status		
Rural	0.24	0.32
Urban	0.41	0.43
Teaching Status		
Non-teaching	0.38	0.39
Teaching	0.51	0.51

MRSA = methicillin-resistant *Staphylococcus aureus*
MSSA = methicillin-susceptibile *Staphylococcus aureus*

Figure 5.9 Distribution of propensity scores for MRSA

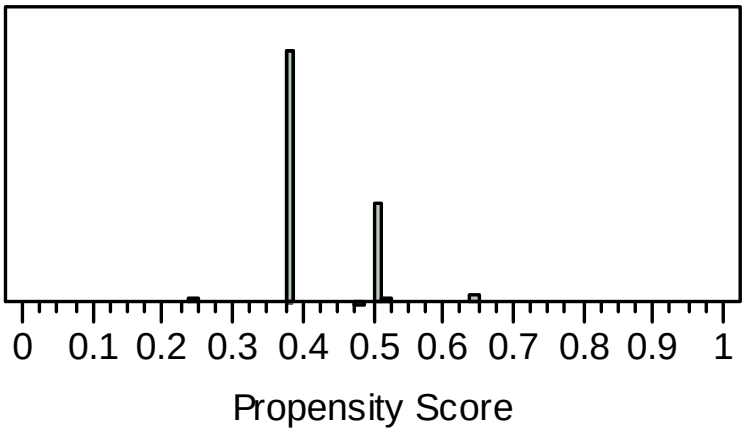
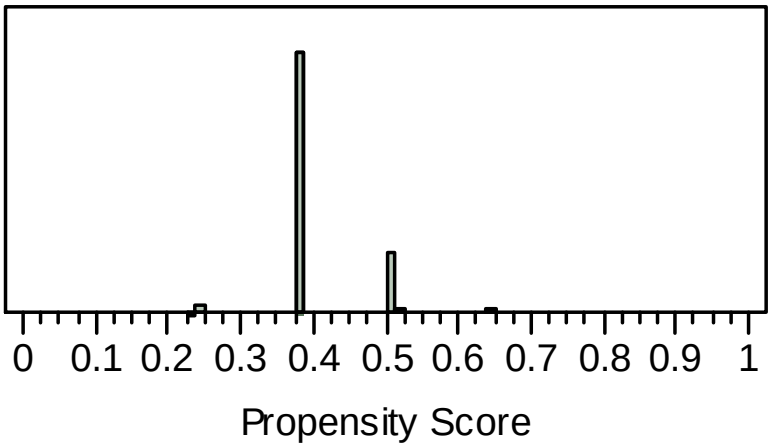


Figure 5.10 Distribution of propensity scores for MSSA



All variables in Table 5.8 were used in the initial multivariable model as well as the calculated propensity score. Age was then removed from the model since it had a multivariable p-value > 0.05. The results of the multivariable model, excluding age, are below in Table 5.10. The model contained 867 observations. Propensity scores were only calculated for 900 observations. Another 33 observations in the final sample had an unavailable discharge status. In this and all subsequent models the variable susceptibility refers to MRSA vs. MSSA.

Table 5.10 Multivariable model (hospital charge = dependent variable)

Parameter	Coefficient	Exponentated Coefficient	Standard Error	Wald 95% Confidence Limits for Exponentated Coefficent		p-value
Susceptibility	0.0512	1.05	1.07	0.93	1.19	0.4277
Gender	-0.2179	0.80	1.06	0.71	0.91	0.0005
Propensity Score	-0.9575	0.38	1.62	0.15	0.99	0.0479
DRG weight	0.1135	1.12	1.02	1.08	1.16	<.0001
Charlson comorbidity score	-0.0472	0.95	1.01	0.93	0.98	0.0014
Pre-infection LOS	0.0413	1.04	1.00	1.03	1.05	<.0001
Discharge Status	0.1004	1.11	1.07	0.96	1.27	0.1566

Log likelihood = -10833.3650, $\chi^2 = 2633$, df = 859, n = 867, p-value < 0.0001
DRG = diagnosis related group
LOS = length of stay

All two-way interaction effects between infection susceptibility status and the five potential modifiers (gender, DRG weight, Charlson comorbidity score, discharge status and pre-infection LOS) were assessed by including an interaction term for each in the multivariable model. Likelihood ratio tests were used to compare the full and reduced models with different combinations of the interaction terms. The results are below in Table 5.11. No significant interactions between susceptibility status and discharge status, gender or pre-infection LOS were found. However, there were interactions between susceptibility and DRG weight as well as susceptibility and Charlson comorbidity score.

Table 5.11 Potential interaction terms with susceptibility (dependent variable = hospital charge)

Variable	2LL	df	Chi-Square	p-value
Main Effects Model	-21666.73			
Susceptibility x Charlson comorbidity score	-21658.10	1	8.63	0.0033
Susceptibility x DRG weight	-21646.17	1	11.93	0.0006
Susceptibility x Gender	-21643.33	1	2.84	0.092
Susceptibility x Pre-infection LOS	-21640.54	1	2.79	0.095
Susceptibility x Discharge status	-21640.20	1	0.34	0.5601

LL = Log likelihood
df = degrees of freedom
DRG = diagnosis related group
LOS = length of stay

All possible interactions between the Charlson comorbidity score, pre-infection LOS, DRG weight, gender and discharge status were assessed using likelihood ratio tests as described above. The interaction terms were tested by adding each one sequentially in the model. There were two significant interactions; DRG weight and pre-infection LOS as well as gender and the Charlson comorbidity score. These interactions were retained in the model along with the two associated lower-order terms.

Table 5.12 Potential interaction terms among the covariates (dependent variable = hospital charge)

Variable	2LL	DF	Chi-Square	P-value
Main Effects Model	-21666.73			
Susceptibility x Charlson comorbidity score	-21658.10	1	8.63	0.0033
Susceptibility x DRG weight	-21646.17	1	11.93	0.0006
DRG weight x pre-infection LOS	-21633.75	1	12.42	0.0004
Gender x Charlson comorbidity score	-21626.203	1	7.54	0.006
DRG weight x Charlson comorbidity score	-21624.43	1	1.77	0.1835
DRG weight x discharge status	-21623.51	1	0.92	0.3374
DRG weight x gender	-21622.72	1	0.79	0.373
gender x discharge status	-21622.02	1	0.7	0.4038
Gender x pre-infection LOS	-21621.88	1	0.14	0.7066
Charlson comorbidity score x pre-infection LOS	-21621.71	1	0.17	0.6784
Charlson comorbidity score x discharge status	-21621.53	1	0.18	0.6727
pre-infection LOS x discharge status	-21621.53	1	0	0.9461

LL = Log likelihood
DF = degrees of freedom
LOS = length of stay
DRG = diagnosis related group

The final model included susceptibility, gender, the propensity score, DRG weight, the Charlson comorbidity score, pre-infection LOS, discharge status, and four interaction terms (susceptibility and DRG weight, susceptibility and the Charlson comorbidity score, gender and Charlson comorbidity score, and DRG weight and pre-infection LOS). The parameter estimates were exponentiated for interpretation (Table 5.13). No pairwise multicollinearity was found.

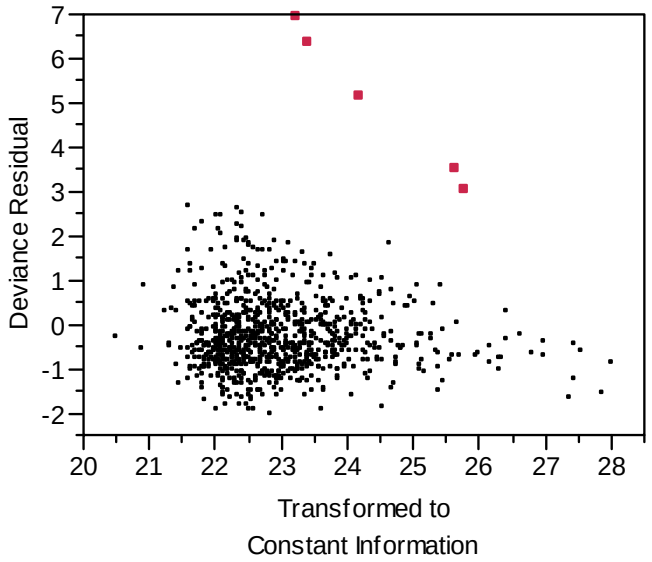
Table 5.13 Parameter estimates (dependent variable = hospital charge)

Parameter	Coefficient	Exponentiated Coefficient	Standard Error	Wald 95% Confidence Limits of Exponentiated Coefficient		p-value
Susceptibility	0.4521	1.57	1.11	1.27	1.94	<.0001
Gender	0.3306	1.39	1.08	1.19	1.62	<.0001
Propensity Score	0.7105	0.49	1.61	0.19	1.25	0.1372
DRG weight	0.2487	0.55	1.29	0.33	0.90	0.0174
Charlson comorbidity score	0.0529	0.46	1.24	0.30	0.71	0.0004
Pre-infection LOS	0.0548	1.06	1.01	1.05	1.07	<.0001
Discharge status	0.1062	1.11	1.07	0.97	1.27	0.127
Susceptibility x Charlson comorbidity score	-0.103	1.11	1.03	1.05	1.18	0.0005
Susceptibility x DRG weight	-0.106	1.11	1.03	1.04	1.19	0.002
DRG weight x pre-infection LOS	-0.005	1.00	1.00	0.99	1.00	<.0001
Gender x Charlson comorbidity score	-0.079	0.92	1.03	0.87	0.98	0.006

Log likelihood = -10813.1013, χ^2 = 2338, df = 855, n = 867, p-value < 0.0001
DRG = diagnosis related group
LOS = length of stay

The deviance residuals were plotted against the fitted values transformed to constant information (Figure 5.11). The transformation to constant information is transforming the fitted values to a constant variance scale or constant information of the error distribution. The transformation formula for the gamma distribution is below in Equation 5.1. This scatterplot can be interpreted analogously to a residual by predicted plot in linear regression. Extreme values from Figure 5.11 (highlighted in red) are described below in Table 5.14. Excluding these values from the deviance residuals by fitted values plot yielded a random scatter.

Figure 5.11 Deviance residuals by the fitted values transformed to constant information



Transformation to constant information = 2 x log(fitted value) Eq. 5.1

Table 5.14 Extreme values

MRSA or MSSA	Gender	Age	Pre-Infection LOS	Post-Infection LOS	Discharge Year	Charge ($)
MSSA	Male	72	8	25	2001	2,909,322.44
MSSA	Male	55	6	19	2001	2,957,731.55
MRSA	Male	67	4	57	2001	3,111,884.28
MRSA	Female	61	18	52	2001	2,993,126.42
MRSA	Male	55	29	95	2001	3,392,801.31

LOS = length of stay

The above procedures were repeated without the five extreme observations (n=925).Variables significant at p-value < 0.25 during univariate analysis were: DRG weight, pre-infection LOS and mortality. Gender was no longer significant and not included in the model. The Charlson comorbidity index was included even though the univariate p-value = 0.6430 since severity of illness was a known confounder. The multivariable model included: susceptibility, DRG weight, charlson comorbidity index, pre-infection LOS, mortality, and the propensity score. All two-way interactions with infection susceptibility were assessed. Only the interaction between susceptibility and the charlson comorbidity index was significant. Interactions were assessed between all other variables. The interaction between DRG weight and pre-infection LOS was significant. The final model included susceptibility, the propensity score, DRG weight, the Charlson comorbidity score, pre-infection LOS, discharge status, and two interaction terms (susceptibility and the Charlson comorbidity score, and DRG weight and pre-infection LOS). The parameter estimates were exponentiated for interpretation (Table 5.15).

Table 5.15 Parameter estimates without extreme values (dependent variable = hospital charge)

Parameter	Coefficient	Exponentated Coefficient	Standard Error	Wald 95% Confidence Limits of Exponentiated Coefficent		p-value
Susceptibility	0.20	1.23	1.08	1.06	1.42	0.0055
Propensity Score	-0.09	1.10	1.54	0.47	2.56	0.8265
DRG weight	0.13	1.14	1.02	1.09	1.19	<.0001
Charlson comorbidity score	0.04	1.04	1.02	1.00	1.09	0.0783
Pre-infection LOS	0.05	1.05	1.00	1.04	1.06	<.0001
Discharge Status	-0.05	0.95	1.06	0.84	1.07	0.4174
Susceptibility x Charlson comorbidity score	-0.09	0.92	1.03	0.87	0.97	0.0018
DRG weight x pre-infection LOS	-0.003	1.00	1.00	1.00	1.00	0.0004

Log likelihood = -10621.4681, χ^2 = 836.06, df = 853, n = 862, p-value =0.7312
DRG = diagnosis related group
LOS = length of stay

Since susceptibility status was involved in an interaction with the Charlson comorbidity score, it must be interpreted in that context. The median Charlson comorbidity score was one and the maximum score was 12. Higher Charlson scores are associated with a higher probability of in-patient death. Table 5.16 below outlines the percent increase/decrease in hospital charge for infection susceptibility taking into account the Charlson comorbidity score. As patients get more severely ill, the total charge for MSSA patients increases while the total charge for MRSA patients decreases. These results are also displayed in Figure 5.12. The results of the interaction term DRG weight by pre-infection LOS are below in Table 5.17. As the pre-infection LOS increases, the effects of the DRG weight on total charge decrease.

Table 5.16 Change in hospital charge adjusting for Charlson comorbidity score

Charlson comorbidity score	MRSA	MSSA
1	16%	4%
2	11%	8%
6	-9%	27%
12	-33%	62%

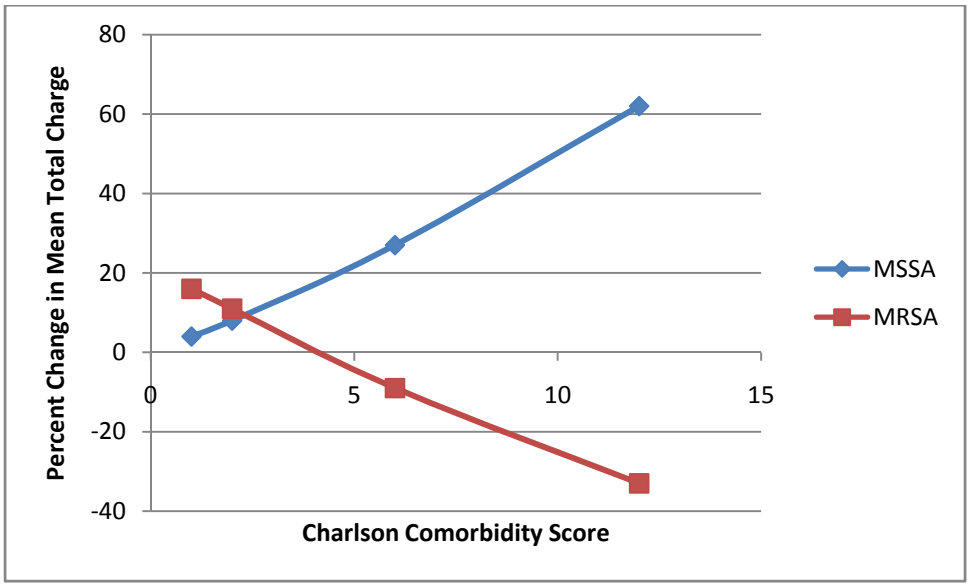
MRSA = methicillin-resistant *Staphylococcus aureus*
MSSA = methicillin-susceptible *Staphylococcus aureus*

Table 5.17 Effect modification of DRG weight and pre-infection LOS

DRG weight	Pre-infection LOS			
	3	6	9	15
1	1.31	1.39	1.62	2.21
5	1.92	1.09	1.01	1.00
10	3.67	1.18	1.02	1.00
15	7.03	1.29	1.03	1.00

DRG = diagnosis related group
LOS = length of stay

Figure 5.12 Change in hospital charge adjusting for Charlson comorbidity score



The multivariable model predicted an overall mean charge of $116,404. The MRSA and MSSA mean charges are presented in Table 5.16. According to the model, a case of MRSA bacteremia has an affiliated charge of $22,889 more than a case of MSSA bacteremia.

Table 5.18 Predicted charges in US dollars

	Overall	MRSA	MSSA
Mean	116,404	125,910	103,021
Standard Deviation	107,898	99,535	117,524

MRSA = methicillin-resistant *Staphylococcus aureus*
MSSA = methicillin-susceptible *Staphylococcus aureus*

Objective 3:
Adjusted difference in LOS by infection susceptibility

A GLM utilizing a gamma distribution and logarithmic link was used to estimate LOS adjusting for potential confounders. Before a multivariable model could be analyzed, each potential covariate was evaluated for inclusion in the model. Objective 3 was similar to Objective 2 but used post-infection LOS as the dependent variable in the modeling process. There were 48 observations that had a post-infection LOS equal to zero. Zero is an invalid response value for the gamma distribution. These observations were excluded from this objective.

Admission source was not considered since information was missing for 219 patients. The five extreme observations identified in Objective 2 were also excluded from this analysis. The results of the univariate GLMs are available in Table 5.19. All variables were significant at p-value < 0.25 except for gender. Age, DRG weight, Charlson comorbidity score, pre-infection LOS and discharge status were all eligible for inclusion in the multivariable model.

Table 5.19 Univariate analysis (post-infection LOS = dependent variable)

Parameter	Coefficient	Exponentated Coefficient	Standard Error	Wald 95% Confidence Limits of Exponentated Coefficent		p-value
Gender	-0.029	0.97	1.07	0.86	1.10	0.6427
Age	-0.005	1.00	1.00	0.99	1.00	0.0133
DRG weight	0.086	1.09	1.02	1.06	1.12	<.0001
Charlson comorbidity score	-0.049	0.95	1.02	0.92	0.98	0.0019
Pre-infection LOS	0.009	1.01	1.00	1.00	1.02	0.0083
Discharge Status	0.419	1.52	1.07	1.32	1.75	<.0001

DRG = diagnosis related group
LOS = length of stay

The same propensity score calculated above for Objective 2 was used again to account for differences in hospital level factors (i.e. bed size, urban/rural, teaching status.) All variables in Table 5.19, excluding gender, were used in the initial multivariable model as well as the calculated propensity score. Age was then removed from the model since it had a multivariable p-value >0.05 (p=0.0991). The results of the multivariable model, excluding age, are below in Table 5.20.

Table 5.20 Multivariable model (post-infection LOS = dependent variable)

Parameter	Coefficient	Exponentated Coefficient	Standard Error	Wald 95% Confidence Limits of Exponentiated Coefficient		p-value
Susceptibility	0.10	1.11	1.07	0.97	1.26	0.12
Propensity Score	1.26	3.54	1.63	1.36	9.19	0.01
DRG weight	0.07	1.07	1.02	1.04	1.11	<.0001
Charlson comorbidity score	-0.04	0.96	1.02	0.93	1.00	0.03
Pre-infection LOS	0.01	1.01	1.00	1.00	1.02	0.01
Discharge Status	0.38	1.47	1.07	1.27	1.69	<.0001

Log likelihood = -2633.4724, χ^2 = 1078, df = 807, n = 814, p-value < 0.0001
DRG = diagnosis related group
LOS = length of stay

All two-way interaction effects between infection susceptibility status and the four potential modifiers (DRG weight, Charlson comorbidity score, discharge status and pre-infection LOS) were assessed by including an interaction term for each in the multivariable model. Likelihood ratio tests were used to compare the full and reduced models with different combinations of the interaction terms. The results are below in Table 5.21. Neither discharge status, DRG weight nor the Charlson comorbidity score had significant interactions with susceptibility status. However, there was a significant interaction between susceptibility and pre-infection LOS.

Table 5.21 Potential effect modifiers of interaction susceptibility

Variable	2LL	DF	Chi-Square	P-value
Main Effect Model	-5266.9			
Susceptibility x pre-infection LOS	-5261.2	1	5.73	0.0167
Susceptibility x DRG weight	-5260.8	1	0.39	0.5333
Susceptibility x Charlson comorbidity score	-5258.6	1	2.18	0.1395
Susceptibility x discharge status	-5258.6	1	0.03	0.8684

LL = Log likelihood
DF = degrees of freedom
DRG = diagnosis related group
LOS = length of stay

All possible interactions between the Charlson comorbidity score, pre-infection LOS, DRG weight, and discharge status were assessed using likelihood ratio tests as described above. There was one significant interaction; susceptibility and pre-infection LOS. The interaction was retained in the model along with the associated lower-order terms.

Table 5.22 Identification of interactions (dependent variable = post-infection LOS)

Variable	2LL	DF	Chi-Square	P-value
Main Effects Model	-5266.9			
Susceptibility x pre-infection LOS	-5261.2	1	5.73	0.0167
Charlson comorbidity score x pre-infection LOS	-5255.1	1	6.09	0.0136
DRG weight x discharge status	-5254.3	1	0.83	0.3625
Charlson comorbidity score x discharge status	-5253.1	1	1.18	0.2775
DRG weight x Charlson comorbidity score	-5252.8	1	0.29	0.5934
Pre-infection LOS x discharge status	-5252.5	1	0.31	0.5806

LL = Log likelihood
DF = degrees of freedom
DRG = diagnosis related groups
LOS = length of stay

The final model included susceptibility, the propensity score, DRG weight, the Charlson comorbidity score, pre-infection LOS, discharge status, and two interaction terms (susceptibility and pre-infection LOS and pre-infection LOS and Charlson comorbidity score). The parameter estimates were exponentiated for interpretation (Table 5.23). Pairwise multicollinearity was assessed. The only multicollinearity that existed was between pre-infection LOS and the interaction term between susceptibility and pre-infection LOS. This collinearity was expected since both terms take into account pre-infection LOS.

Table 5.23 Parameter estimates (dependent variable = post-infection LOS)

Parameter	Coefficient	Exponentiated Coefficient	Standard Error	Wald 95% Confidence Limits of Exponentiated Coefficients		P-value
Susceptibility	-0.05	0.95	1.09	0.80	1.13	0.5715
Propensity Score	-1.285	3.61	1.62	1.40	9.31	0.0078
DRG weight	0.0716	1.07	1.02	1.04	1.11	<.0001
Charlson comorbidity score	0.0037	1.00	1.02	0.96	1.05	0.8707
Pre-infection LOS	0.1697	1.18	1.06	1.05	1.33	0.0044
Discharge Status	0.3625	1.44	1.07	1.25	1.66	<.0001
Susceptibility x pre-infection LOS	-0.021	0.98	1.01	0.96	0.99	0.0091
Charlson comorbidity score x pre-infection LOS	-0.004	1.00	1.00	0.99	1.00	0.0093

Log likelihood = -2627.5613, χ^2 = 832, df = 805, n = 814, p-value = 0.2476
DRG = diagnosis related groups
LOS = length of stay

There was an interaction between infection susceptibility and pre-infection LOS. As pre-infection LOS increased, the post-infection LOS for MSSA patients increased more than MRSA patients (Table 5.24). There was also an interaction between pre-infection LOS and Charlson comorbidity score (Table 5.25). Figure 5.13 depicts the interaction between susceptibility status and pre-infection LOS.

Table 5.24 Relative increase in post-infection LOS adjusting for pre-infection LOS

Pre-infection LOS	MRSA	MSSA
2	1.28	1.40
6	2.32	2.77
9	3.63	4.61
12	5.66	7.66
20	18.59	29.78

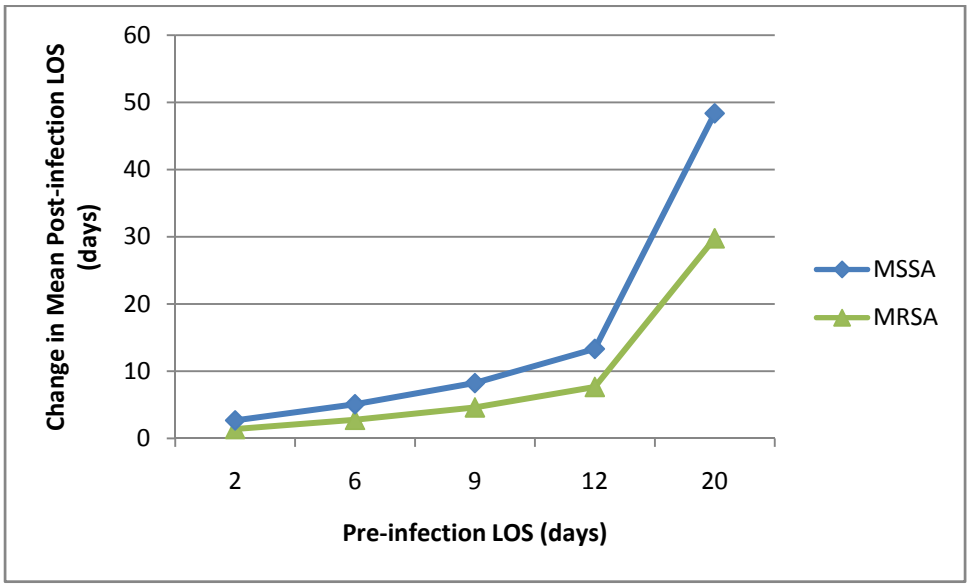
LOS = length of stay
MRSA = methicillin-resistant *Staphylococcus aureus*
MSSA = methicillin-susceptible *Staphylococcus aureus*

Table 5.25 Effect modification of Charlson comorbidity score and pre-infection LOS

Charlson Comobidity score	Pre-infection LOS			
	3	6	9	15
1	1.65	2.74	4.46	11.72
3	1.01	1.00	1.00	1.00
6	1.02	1.00	1.00	1.00
12	1.05	1.00	1.00	1.00

LOS = length of stay

Figure 5.13 Relative increase in post-infection LOS adjusting for pre-infection LOS



The multivariable model predicted an overall mean post-infection LOS of 9.86 days. The MRSA and MSSA mean post-infection LOSs are presented in Table 5.26. According to the model, a case of MRSA bacteremia has an affiliated post-infection LOS of 1.30 days more than a case of MSSA bacteremia.

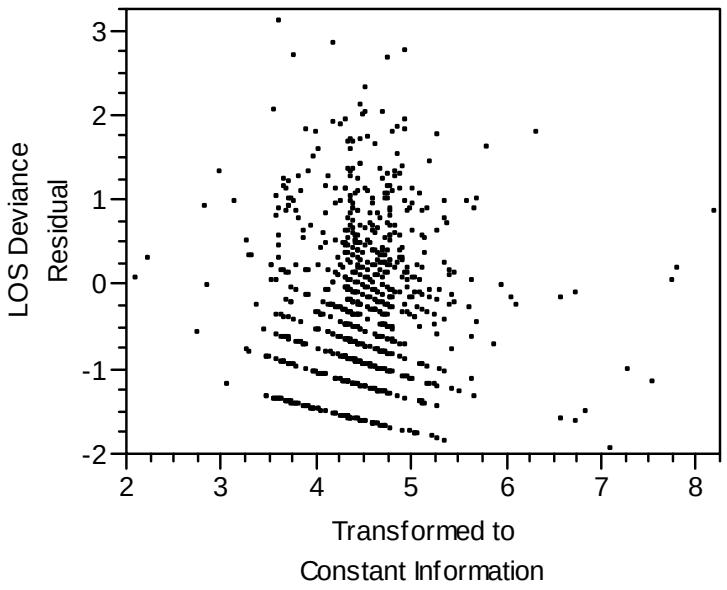
Table 5.26 Predicted post-infection LOS in days

	Overall	MRSA	MSSA
Mean	9.86	10.40	9.10
Standard Deviation	4.16	4.99	2.39

LOS = length of stay
MRSA = methicillin-resistant *Staphylococcus aureus*
MSSA = methicillin-susceptible *Staphylococcus aureus*

A deviance residual against fitted values transformed to a constant scale plot was examined. As stated in the above objective, this plot can be interpreted similarly to a residual by predicted plot in regression. The points should look randomly scattered. In Figure 5.14, there appeared to be a striped pattern. No changes were made to the model but the model implications will be addressed in the discussion.

Figure 5.14 Deviance residuals by the fitted values transformed to constant information



Objective 4:

Using path analysis to explore the relationship between susceptibility, LOS and total hospital charge

PROC CALIS was used to assess the proposed path analysis model in Figure 4.1. Post-infection LOS was modeled as an intermediate step and total hospital charges were analyzed. Maximum likelihood estimation for model fitting was used. The convergence criterion was satisfied indicating starting values were successfully calculated by the software. The five extreme observations identified in the previous objectives were excluded from this analysis as well (n = 925). A correlation matrix was calculated to verify the association between the variables (Table 5.27). Post-infection LOS and total charge were the most highly correlated variables.

Table 5.27 Correlation matrix

	Susceptibility	Pre-infection LOS	Post-infection LOS	Total Charge
Susceptibility	1			
Pre-infection LOS	0.1468	1		
Post-infection LOS	0.0598	0.0772	1	
Total Charge	0.1072	0.4234	0.4872	1

LOS = length of stay

There were two exogenous variables (susceptibility and pre-infection LOS). They were assumed to be correlated. Therefore, a covariance was modeled and calculated for susceptibility and pre-infection LOS. A covariance between exogenous variables means

the variables arose from common causes not modeled in the path diagram. There were two endogenous variables (post-infection LOS and charge) and their variances were also modeled.

The standardized estimates are below in Table 5.28. Unstandardized coefficients were not calculated since SAS used a correlation matrix for data input. Also in Table 5.28 are the R^2 values and the proportion of unexplained variance ($1-R^2$) for each outcome variable. The proportion of explained variance ranged from 1% for post-infection LOS to 39% for total charge.

The variances and covariance for the exogenous variables are in Table 5.29. The variances for susceptibility and pre-infection LOS were one by convention. This allowed SAS to estimate just the variance of the endogenous variables. (The variances of endogenous variables are considered exogenous by definition.) The covariance between susceptibility and pre-infection LOS was 0.14682. This was the correlation between the two variables and can be interpreted similar to a Pearson correlation.

Table 5.28 Path Analysis: Regression Coefficients

Outcome	Predictors	Standardized Coefficient	R^2	1-R^2
Post-infection LOS	Susceptibility	0.0495	0.01	0.99
	Pre-infection LOS	0.0699		
Total charge	Post-infection LOS	0.4573	0.39	0.61
	Pre-infection LOS	0.3881		

LOS = length of stay

Table 5.29 Variances of exogenous variables

Variance	Estimate	Standard Error
Susceptibility	1	0.04652
Pre-infection LOS	1	0.04652
Post-infection LOS	0.99165	0.04614
Total charge	0.61286	0.02851
Covariance		
Susceptibility & Pre-infection LOS	0.14682	0.03325

LOS = length of stay

Table 5.30 illustrates the direct, indirect and total effects of each causal variable. A direct effect is one variable's influence on another not taking into account other factors. An indirect effect is one variable's influence on another through a mediating variable. Total effects are the sum of the direct and indirect effects.[85] The total effect of pre-infection LOS on total charge was 0.42. The total effect of post-infection LOS on charge was 0.46. Susceptibility only had a 0.02 total effect on charge and a 0.05 total effect on post-infection LOS. Pre-infection LOS had a 0.07 total effect on total charge.

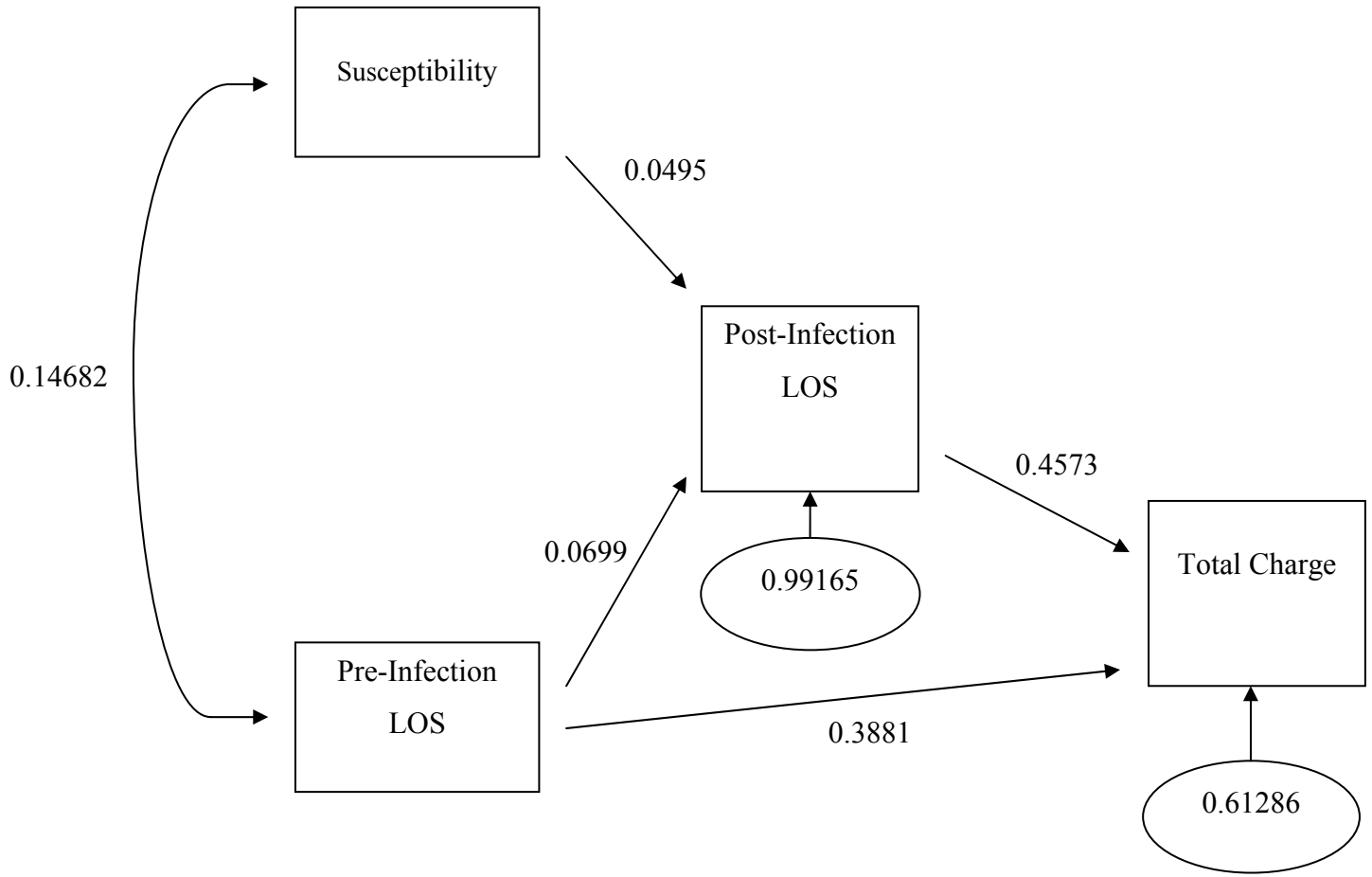
Table 5.30 Decomposition of effects

	Endogenous variable	
Causal Variable	Post-infection LOS	Total Charge
Susceptibility		
Direct effect	0.0495	
Indirect via post-infection LOS		0.0226
Indirect via pre-infection LOS	0.0103	0.05698
Total effect	0.0598	0.0796
Pre-infection LOS		
Direct effect	0.0699	0.3881
Indirect via post-infection LOS		0.032
Indirect via susceptibility	0.0073	0.0033
Total effect	0.0772	0.4234
Post-infection LOS		
Direct effect		0.4573
Total effect		0.4573

LOS = length of stay

Figure 5.15 depicts the standardized estimates within the structural equation model. By convention, the variances of the exogenous variables are not shown in the diagram of the path model. These are the same estimates above in Table 5.28.

Figure 5.15 Path Analysis model with standardized estimates



The above model had adequate fit ($\chi^2 = 0.8115$, df = 1, p-value = 0.3677). Because pre-infection LOS, post-infection LOS and charge are all skewed, significance tests that take into account non-normal distributions were used. Other methods to assess model fit can be found below in table 5.27. Specifically, the elliptic correlated chi-square and the Bentler and Bonnett's non-normed index adjust for non-normal distributions. Table 5.26 also contains other commonly used significance tests to evaluate goodness of fit, their value for this model, and the interpretation of that statistic.[86] All the statistical tests concur that the path analysis model tested had adequate fit.

Table 5.31 Path Analysis: Assessment of Fit

Statistical Test	Model Statistic	Interpretation
Goodness of Fit Index (GFI)	0.9996	Values close to 0.90 represent a good fit
GFI Adjusted for Degrees of Freedom (AGFI)	0.9956	Value adjusted for df, with 0.90 a good model fit
Parsimonious GFI	0.1666	Used to compare between models
Chi-Square	0.8115	
Chi-Square DF	1	p-value greater than 0.05 indicates good fit
p-value > Chi-Square	0.3677	
RMSEA Estimate	0	Values less than 0.05 indicates a good model fit
Elliptic Corrected Chi-Square	0.3101	p-value greater than 0.05 indicates good fit - adjusts for kurtosis
p-value > Elliptic Corrected Chi-Square	0.5776	
Bentler & Bonett's Non-normed Index	1.0024	Value close to 0.9 reflects a good model fit

.RMSEA = root mean squared error approximation

CHAPTER 6

Discussion

Review of Study Objectives

There were four general objectives of this investigation. The crude difference in total hospital charges was determined for bacteremias caused by MRSA versus those caused by MSSA. The impact of methicillin resistance in *Staphylococcus aureus* bacteremia was examined separately for dependent variables (1) total hospital charges and (2) post-infection LOS while adjusting for confounding. Finally, the relationships between infection susceptibility, LOS, and total hospital charges were examined.

Summary of Findings

Objective 1 examined the crude difference in total hospital charge. The overall mean hospital charge was $111,636. The mean charge for MRSA was $121,713 and $97,307 for MSSA. Using the Wilcoxon signed rank test, the difference between the total MRSA and MSSA charge was significant. The median overall charge was $59,764, which can be separated into a median MRSA charge of $68,013 and a median MSSA

charge of $27,338. The mean values are at the upper end of the interquartile range which affirms the highly skewed nature of charge data.

According to the multivariable model in Objective 2, the overall mean charge was $116,404, which can be divided into $125,910 for MRSA and $103,021 for MSSA. The difference in MRSA charge over MSSA charge was $22,889. There was an interaction between susceptibility and Charlson comorbidity score. This means that susceptibility results must be interpreted in the with respect to the Charlson comorbidity score. The above estimates are for the mean Charlson score of 1.72. This means the above estimates are for the average patient who had a low chance of dying within the year.

Objective 3 was very similar to Objective 2 except the dependent variable was post-infection LOS instead of total charge. The overall mean LOS was 9.9 day which can be broken down into 10.4 and 9.1 days for MRSA and MSSA respectively. The difference was 1.3 days. The crude post-infection LOS estimates were not significantly different. There was an interaction between susceptibility and pre-infection LOS. Therefore, the susceptibility results must be interpreted with pre-infection LOS. The above estimates are for the mean pre-infection LOS of 8.6 days.

In Objective 4, a path analysis model examined more closely the relationship between infection susceptibility, pre-infection LOS and post-infection LOS. The overall model fit was good ($\chi^2 = 0.8115$, df = 1, p-value = 0.3677). The model predicted 39% of the variability in total charge but only 1% of the variability on post-infection LOS. The standardized direct and indirect effects of each variable were considered. Susceptibility status was not highly correlated with total charge or post-infection LOS. This implies

that susceptibility status may not strongly influence charge or post-infection LOS through the proposed model.

Discussion of Results by Objective

Objective 1

The crude charges for MRSA and MSSA were $121,713 and $97,307 respectively. This represents a mean crude difference of $24,406, which is higher than previously reported data. Most of these investigations reported costs. As previously discussed, charges are known to overestimate actual costs. Mean charges are reported in this investigation since economists are interested in means by convention.[32] Economists care about means since they can be extrapolated from a sample to the population. However, medians are also reported since it is the measure of central tendency that corresponds to skewed data.

The median overall charge was $59,764, which can be separated into a median MRSA charge of $68,013 and a median MSSA charge of $27,338. The median MRSA charge was over twice the median MSSA charge. Cosgrove et al.[11] reported only a 1.36-fold increase in median MRSA charge over MSSA.

The pilot investigation that analyzed data from one hospital had an overall crude median charge of $140,396, which was broken down as $166,901 for MRSA and $86,130 for MSSA. These values are over twice the median total charge reported in this investigation. This difference in estimates is substantial. One possible explanation is that the pilot included only patients from a tertiary care facility. These patients may be more

severely ill than patients in other facilities that were included in this investigation. Another possible reason for the increased charge from the pilot investigation stems from the study period. This investigation adjusted total charge to 2006 dollars while the pilot included 2007 values. Inflating this investigation's total charges to 2007 dollars would increase the estimates, but not enough to account for the entire charge disparity. However, both investigations indicate that MRSA bacteremias are more expensive than MSSA bacteremias.

It was intended to use repeated measures ANOVA to examine trends in hospital charge over the study period. However, only four hospitals had data for every study year. SAS used only these four hospitals to calculate the repeated measures ANOVA. Extrapolating these results to the entire study population would be inappropriate. Instead the data were examined descriptively; 2001 was the only year that seemed drastically different. This was due to five extreme observations that were excluded from the analysis during objective two. (See Objective 2 for a more complete discussion of the extreme observations.)

A post-hoc analysis was performed using a generalized estimating equations (GEE) approach to compare total hospital charge over the study period while accounting for the hospital groupings within the dataset. The GEE model used a gamma distribution and log link. No difference was seen in total charge per year during the study period. An interaction term was used to test whether there was a difference in MRSA vs. MSSA total charge throughout the study period. No difference was found. Therefore, it can be

concluded that the total charges and the total charges for MRSA and MSSA were not statistically different during each year of the investigation.

Objective 2

A multivariable model was used to determine the role susceptibility played on total hospital charges. Along with potential confounders, a propensity score was included in the model as a way to account for hospital differences (e.g., teaching status, bed size, and urban/rural). A recent meta-analysis identified key components necessary for any investigation using propensity scores.[87] Important components included sufficient events per variable (EPV), continuous variable conformity with linear gradient, interactions, collinearity, assessment of model fit, discrimination of the model, balance achieved between the confounders, and adjustment methodology. In a logistic model, there need to be at least 10 observations per variable.[88-90] This rule of thumb has also been generalized to propensity scores to assure sufficient EPV. This investigation used only 3 variables for more than 900 observations. Continuous variable conformity relates to continuous variables used to create the propensity score. All the variables used to create the propensity score were categorical. Therefore, it was not possible to evaluate collinearity. Fit was assessed using the Hosmer-Lemeshow goodness of fit statistic. A non-significant p-value was obtained indicating a lack of evidence suggesting the model did not fit the data well. Assessment of fit relates closely to balance between the treatment groups. The MRSA and MSSA groups appeared to be balanced by the variables used to calculate the propensity score (Table 5.9). Propensity scores from a

poorly fit model and without balance between the treatment groups could lead to biased estimates of treatment effect. The propensity score used in this analysis had adequate fit and was balanced between the treatment groups.

Propensity scores can either be used as continuous variables or stratified into quintiles. Since there was a lot of overlap between the MRSA and MSSA groups with respect to propensity scores, the scores were used as a continuous variable in the second stage regression model. In a post hoc fashion, propensity score quintiles were run with the final multivariable model. Less than a 1% change occurred in the susceptibility parameter estimate. This indicates using the propensity score as a linear variable was not inappropriate. Of note, when the observations were separated into quintiles based on propensity score, it was observed that no observations existed in the 3rd quintile. Although unusual, this was thought to be a function of the cluster distribution of scores. Most likely this was caused by the propensity score being composed entirely of categorical variables with few categories.

After the model was initially fit, an analysis of residuals was conducted. Residual analysis is used in regression to identify a model specification problem or nonhomogeneous variance.[91] For the gamma distribution, it is recommended that deviance residuals be plotted against the fitted values transformed to the constant scale.[92] This plot should look like a random splatter of points. There were five observations that stood out from the rest as extreme (Figure 5.11).

Extreme values identified by the deviance residual plot were examined more closely. They were a mixture of MRSA and MSSA observations all from the same

hospital and same study year (2001). Removing all data points from that hospital was considered but there were many observations from various study years that were not extreme. Leaving the five observations in the dataset was also considered. As previously discussed, the charge data was expected to have a long right tail. Having a handful of large charge data points was anticipated. But, the magnitude of the five extreme observations was astronomical. It was decided to remove these five observations for the remainder of the analysis. The model was re-run excluding these observations.

Adjusting for confounders, the overall mean charge was $116,404, which can be divided into $125,910 for MRSA and $103,021 for MSSA. The difference in MRSA charge over MSSA charge was $22,889. These estimates adjusted for propensity score, DRG weight, the Charlson comorbidity score, pre-infection LOS, discharge status and the interaction between susceptibility and the Charlson comorbidity score as well as the interaction between DRG weight and pre-infection LOS.

Since susceptibility status was involved in an interaction with the Charlson comorbidity score, it must be interpreted in that context. The median Charlson comorbidity score was one and the maximum score was 12. The interquartile range was from 0 to 2. As a reminder, higher Charlson scores are associated with a higher probability of in-patient death. Figure 5.12 graphically depicts the relationship between susceptibility status and the Charlson comorbidity score. For most patients (Charlson comorbidity score 0 to 2) MRSA bacteremias have a higher total charge than MSSA bacteremias. But as the Charlson comorbidity score increases, the total charge for MSSA infections increases while the total charge for MRSA infections decreases. One possible

explanation is that MRSA patients may be expiring pre-maturely in the hospital causing them to have a truncated charge.

There was also an interaction between DRG weight and pre-infection LOS (Table 5.17). Interpretation of this interaction showed that as pre-infection LOS increased, the total charge was less dependent on the DRG weight. For example, a patient with a pre-infection LOS of three days had highly variable charges depending on DRG weight. (A higher DRG weight had a higher total charge.) But the effects of DRG weight diminished as pre-infection LOS increased. Since DRG weight was one method that controlled for severity of illness it can be suggested that severity of illness was less important to total hospital charge as pre-infection LOS increased.

Objective 3

Objective 2 demonstrated that pre-infection LOS was more influential on total charge as it increased than severity of illness. Susceptibility status was examined to see if it could predict post-infection LOS similarly to total charge. The same main effects were used in the LOS multivariable model as the model with total charge as the outcome variable. Although the main effect of susceptibility status was not significant in the multivariable model, there was a significant interaction with pre-infection LOS. For MRSA and MSSA patients the post-infection LOS increased as pre-infection LOS increased but, the post-infection LOS increased more for MSSA patients than MRSA patients as pre-infection LOS increased. The mean post-infection LOS for MRSA patients was not much different than for the MSSA bacteremia patients. However, a

difference in post-infection LOS becomes apparent as pre-infection LOS increases (Figure 5.13).

There was also an interaction between pre-infection LOS and the Charlson comorbidity score (Table 5.24). As the Charlson comorbidity score increased, the effect of pre-infection LOS decreased. The median Charlson score was 1 (IQR = 0 to 2). For most patients, pre-infection LOS has a large impact on post-infection LOS. But for patients whose Charlson score was outside the IQR, pre-infection LOS does not have much effect on post-infection LOS.

The multivariable model adjusted for the propensity score, DRG weight, the Charlson comorbidity score, pre-infection LOS, discharge status and the interaction between susceptibility and pre-infection LOS as well as the interaction between Charlson comorbidity score and pre-infection LOS. The overall mean LOS was 9.86 day, which can be broken down into 10.4 and 9.1 days for MRSA and MSSA respectively. The difference was 1.3 days. These estimates are for the mean pre-infection LOS of 8.59 days.

After the multivariable model was calculated, an analysis of residuals was conducted. The deviance by residual plot could have looked better (Figure 5.14). There appeared to be a stripped pattern. This could indicate model misspecification or nonhomogeneous variance. Most likely the problem related to the distribution and link functions used for the model. A logarithmic transformation of the data did not make post-infection LOS look "normal." It has been suggested that the Inverse Gaussian

distribution, with a high initial peak and long right tail may be more appropriate for LOS data.[93] Different distributions and link functions were not explored.

Objective 4

Based on the results from Objectives 2 and 3, the question then becomes whether the increase in total hospital charge for MRSA bacteremia patients is purely a function of LOS. Path analysis was used to evaluate the joint effects of susceptibility, pre-infection LOS, post-infection LOS and charge. The standardized regression coefficients can be interpreted as correlation coefficients.[82] Susceptibility status had a direct effect of 0.0485 on post-infection LOS and an indirect effect on total charge of 0.0226. Pre-infection LOS had a direct effect on post-infection LOS of 0.0699. Pre-infection LOS had a direct effect of 0.3881 on charge and an indirect effect of 0.032 through post-infection LOS for a total effect of 0.4201. Post-infection LOS had a direct effect on charge of 0.4573.

Standardized path coefficients with an absolute value less than 0.10 are considered to have a small effect. Values around 0.30 have a medium effect and values greater than 0.5 are considered to have a large effect. [94] The path between susceptibility and post-infection LOS and the path between pre-infection LOS and post-infection LOS both have small effects. This implies that the susceptibility status (MRSA vs. MSSA) had little impact on post-infection LOS or total charge. The path from post-infection LOS to total charge had a large effect. Also, the total effect of pre-infection

LOS on charge was medium to large. This suggests (as Objectives 2 and 3 indicated) that pre-infection and post-infection LOS have a large impact on total charge.

SAS output produced only standardized results. The biggest advantage of standardized results is that they can be directly compared across variables. Unfortunately, by standardizing the regression coefficients the results lose their original units. Unstandardized results yield estimates in their original metrics but cannot be directly compared across variables. SAS does not produce unstandardized estimates because the path analysis input was based on the correlation matrix. Since the input (the correlation matrix) is already standardized, unstandarized estimates cannot be produced. Epidemiologists are most often concerned with the magnitude of association where unstandardized estimates are preferred.[82] However, this analysis was designed to estimate how much each variable contributed to the other variables in the model. In this case, interpretation of standardized regression coefficients was appropriate.

Like other statistical techniques, path analysis relies on certain assumptions. SAS used maximum likelihood estimation to calculate the regression coefficients. This technique requires the data to be normally distributed. Transformation of the data was considered. Although a logarithmic transformation made the charge data normal, it did not have the same effect on LOS. Another approach, which was executed, was to move forward with the non-normally distributed but carefully consider the implications on the model. First, the parameter estimates could be biased. With a large sample size, however, parameter estimates are assumed to be fairly accurate.[94] The bigger problem is that significance tests tend to be significant too often. Put another way, the true model

may be rejected too frequently (Type I error).[95] The significance testing problem was addressed by using corrected test statistics such as the elliptic correlated chi-square. If the goodness-of-fit tests had indicated poor fit, this would have meant the model did not explain the associated correlation well.

Another problem with this analysis is the independence assumption. Path analysis assumes independence. However, the observations in this analysis came from different hospitals. Patients within the same hospital are not independent. Previous research indicates that having dependent data leads to goodness-of-fit tests that are too conservative. In the same investigation it was noted that the parameter estimates appear minimally biased despite their dependence.[96]

As previously discussed, the last step of path analysis is to manipulate the model. The fit was good, so this model was not manipulated. One possibility would have been to trim the model. Small effect path coefficients can be essentially removed from the model and the model can be evaluated. Susceptibility had the smallest effect and would be the first variable trimmed from the model. This implies that having a MRSA vs. a MSSA bacteremia had a small impact total charge.

The model accounted for 39% of the variability in total hospital charge and it accounted for only 1% of the variability on post-infection LOS. Although the model fit was good, much of the variability in charge and post-infection LOS was unexplained. This implies the absence of important factors in the relationship between charge, susceptibility and LOS. The two most obvious factors that were missing are severity of illness and discharge status/mortality. Including these factors would probably have

accounted for more variability. Also, including these confounders may have changed the correlation coefficient between susceptibility status and post-infection LOS and total charge. Future path models should include these confounders.

Path analysis allows for the testing of structural models developed based on *a priori* assumptions about the direct and indirect relationship between variables. However, the structural model being tested must be based on theoretical assumptions. The model assumed susceptibility status had an effect on total charge. The model was not rejected but this does not prove that having MRSA vs. MSSA leads to higher charges. Failure to reject a path model does not prove that it is correct.

Practical Implications

This investigation confirms previous single-center analyses that reported MRSA bacteremia is more expensive than MSSA bacteremia. These results were found in both the crude analysis and the multivariable model. The multivariable model controlled for known confounders (i.e., DRG weight, Charlson comorbidity score, discharge status, pre-infection LOS and hospital level factors). While controlling for confounders, the predicted hospital charge was still higher for MRSA than MSSA.

Why would MRSA bacteremias be associated with higher total charges as compared to MSSA bacteremias? As previously discussed several single center reports have reported MRSA bacteremia being associated with a higher cost/charge than MSSA bacteremia. There are several possible explanations for this phenomenon: MRSA and MSSA patients are different, the MRSA and MSSA organisms are different, or their

treatments have very different charges. These three possibilities are discussed more fully in the following paragraphs.

The most obvious difference between MRSA and MSSA patients is severity of illness. Although this investigation attempted to control for severity of illness using the Charlson comorbidity score and DRG weights, there may still have been uncontrolled disparity in patients underlying disease status. For example, a patient with endocarditis is not necessarily comparable to a patient with a simple bacteremia. The patient with endocarditis will most likely be sicker and require more resources. However, the use of DRG weights should help control for this dissimilarity since the patient with endocarditis should have a higher relative weight.

Another potential explanation is that the MRSA and MSSA organisms are different. The inequality in total hospital charge could be that MRSA is more virulent than MSSA. If MRSA bacteremia were more virulent, these patients would require longer hospitalizations and more resources. This would make their hospital stays more expensive. However, there is currently no evidence in the literature to suggest that hospital-onset MRSA infections are more virulent than their MSSA counterparts.[97, 98] Host differences are not thought to be the cause of charge disparity.

Alternatively, MRSA and MSSA bacteremia patients may have different treatments factors, which cause MRSA bacteremia patients to have a higher total charge. These factors include (1) decreased effectiveness of pharmacologic treatment (2) a delay in effective treatment while waiting for microbiologically effective antimicrobials (3) an increased need for surgery or other procedures. Vancomycin has been the drug of choice

for treating MRSA bacteremia. However, vancomycin has shown to have decreased effectiveness in treating MRSA bacteremia.[99] Newer antibiotics are available (e.g., quinupristin–dalfopristin, linezolid, daptomycin and tigecycline) but there are limited data regarding their efficacy as compared to vancomycin in treating MRSA infections.[100] Lodise et al. found that delayed treatment was associated with a longer duration of hospitalization (20 vs. 14 days, p-value = 0.05).[101] This is important since previous investigations have reported that roughly 35% of patients with MRSA bacteremia do not receive appropriate empiric therapy.[102, 103] MRSA patients also have an increased need for surgery and other procedures resulting from the resistant infection.[55] However, the use of DRG weights should account for differences in surgeries and procedures as long as the complication was severe enough to merit a DRG. The exception to this would be procedures not captured by an individual DRG. For example, vancomycin requires blood levels to be taken to monitor the drug's trough concentrations. Repeated blood levels are an extra charge that MSSA patients would not incur. There is not a DRG for blood draws. Additional surgeries should be accounted for by DRG weight but minor procedures may not, which could result in a discrepancy of total charge.

Another consideration about treatment differences relates to inter-hospital variability. Although guidelines advise standards of care, different hospitals have their own standard operating procedures. For example, researchers at Duke Medical Center recommend using transesophageal echocardiography (TEE) to determine the length of treatment for uncomplicated catheter-associated *Staphylococcus aureus* bacteremia.[104]

This diagnostic step is not universally practiced to determine duration of treatment. Also, different hospitals have different infection control policies.[4] Even within a hospital, infection control policies may have changed during the study period. The impact of inter- and intra-hospital variability should be minimal since there was not a significant difference in total charge (overall or by susceptibility status) over the study period as evidenced by the GEE model.

From a hospital administrator's perspective, the results of this investigation are important. The increased total charge for MRSA vs. MSSA bacteremia was roughly $23,000 per case. Charges are known to over-inflate costs; a conservative increase of hospital cost could be approximated at $11,000 per case. If a hospital had 100 MRSA bacteremias per year, this represents an additional $1,100,000 spent on MRSA bacteremias. The cost of an infection control program can be compared to this figure. If an infection control program costs $500,000, the hospital would have a net savings. If the infection control program costs $1,200,00, more factors may need to be considered. When considering whether to spend hospital resources on an infection control program, the cost of the program must be weighed against the potential savings of preventing a resistant infection.

Furthermore, the argument has been made that the calculated "cost of resistance" underestimates the true burden of resistance.[105] Additional costs may include home intravenous therapy, an extended care facility and/or costs of rehabilitation.[106] Providing outpatient intravenous therapy would involve the medication costs, nursing time, supplies, laboratory tests, intravenous line placement and management. Since this

study was conducted from the hospital's perspective, the additional cost of managing MRSA vs. MSSA bacteremia outside of the hospital was not considered.

The effects of mortality on total charge merit further discussion. In this investigation the 30.8% of MRSA bacteremia patients expired in the hospital as compared to 21.4% of MSSA patients ($\chi^2 = 11.92$, df = 2, p-value = 0.0026). A meta-analysis demonstrated a significant increase in mortality associated with MRSA bacteremia relative to MSSA bacteremia (OR = 1.93, p-value < 0.001).[20] However increased mortality could either increase or decrease total hospital charge. If the patient died sooner the hospital stay charge would be truncated. However, the patient could have complications prior to death which would lengthen the hospital stay. A drawn out hospital stay ending in death could increase total charge, especially if significant treatments and procedures were conducted surrounding the end-of-life period. Since this investigation had a large sample size, the overall effects of extended and truncated LOS was assumed to be minimal.

The results of this study should have good external validity. This was a multi-hospital investigation that included hospitals of various size and teaching status. However, the study population was refined to make the MRSA and MSSA groups as similar as possible with respect to known confounders. The purpose of this data manipulation was to isolate the effect of infection susceptibility on total charge and post-infection LOS. By excluding outliers the generalizability does decrease. These exclusions were necessary to maximize internal validity. For example, MRSA patients with pre-infection LOS longer than 62 days were excluded from the analysis. These data

points were not excluded because they were believed to be erroneous. They were excluded because MRSA patients had a pre-infection LOS only this long (except for 1 MSSA patient.) In reality, patients do have pre-infection LOSs longer than 62 days before contracting MRSA. Excluding these patients illustrates the trade-off between internal and external validity. However, the effects of excluding these patients should bias the results toward the null. Since a significant difference was found between total charges, including these patients would have made the difference in charge greater. By decreasing the pre-infection LOS for the MRSA group, the difference in overall total charge was conservative.

Limitations

Admittedly, there were limitations to this investigation. Different hospitals have varying infection control policies, formularies/protocols, and/or provide dissimilar levels of additional education regarding proper selection of initial antibiotic therapy. This was a limitation since the costs of these programs would affect each hospital's total charges differently. Additionally, the incidence of MRSA infections can be reduced substantially through prevention.[62] The role of infection control was not addressed since that information was not available within the Cerner HealthFacts data warehouse. Hospital level factors were used in an attempt to control for inter-hospital variability.

This analysis was based on two fundamental assumptions. First, the charges prior to infection onset were comparable between the MRSA and MSSA groups. Only one charge was provided for the entire hospitalization. Charges were not available as pre-

and post-infection charges. Differences in pre-infection charges could bias the study results. Specifically, the difference in total hospital charge would be overestimated if pre-infection charges were higher in the MRSA group. The difference would be underestimated if the pre-infection charges were higher in the MSSA group. Second, the charges are assumed to be a result of MRSA or MSSA bacteremia. Hospital charges unrelated to the bacteremia that were unequal between the groups could introduce bias into the investigation and also inflate the difference in hospital charges between the MRSA and MSSA and overestimate charge. There is no way to discern which charges relate to the bacteremia from the aggregate total charge provided for this analysis. However, controlling for severity of illness should help account for differences between the MRSA and MSSA groups. It was assumed that all differences in post-infection charge were attributed to the bacteremia. Given the study design, this limitation was unavoidable.

Previous methods for calculating the economic burden of resistance have varied largely. Many single center investigations were able to collect actual cost from their institution.[12, 14, 15] Another investigation was able to identify only costs associated with the SAB.[16] And one of these investigations was able to collect costs attributable to SAB after hospital discharge.[16] Charges were not able to be sub-categorized in the current study. Only an overall aggregate charge was available for each subject since the data came from a large multi-hospital database.

The pilot investigation included transferred patients in the patient population. However, transferred patients were excluded in this investigation. In both investigations

there was no way to know the duration of hospitalization before the transfer. Including these patients could lead to the inclusion of non-nosocomial infections. Excluding transferred patients could underestimate the number of actual cases. Since the pilot included only a tertiary care facility it was assumed that patients transferred there had been at another facility previously. This investigation included all types of hospitals. Therefore, it seemed more appropriate to exclude transferred patients. This means that potential cases may have been excluded. The effects of excluding these patients were thought to be minimal since only 55 MRSA and 19 MSSA patients were transferred from another facility.

Three data manipulation steps led to the exclusion of potential patients. First, several hospitals' data were excluded when the Micro and Large datasets were merged. This was most likely a result of hospitals subscribing to different Cerner services -- an unavoidable circumstance that resulted in a smaller sample size. Cerner is a for-profit company that provides information technology services to its clients. Each client customizes the services it contracts for through Cerner which means that not all hospitals manage the same data elements within their system. Second, only 2006 DRG codes were used. This was done for consistency since all charges were inflated to their 2006 value. Unfortunately, 122 observations were discarded. Almost all of them were surgical DRGs. When DRGs are updated, their weights change. For the purposes of this project, all the weights needed to be consistent. Third, patients with charges less than $11,896 and a LOS greater than nine days were excluded. It became obvious from looking at the data that some observations were erroneous. Some charges were too low based on the

total LOS. The pilot investigation has a minimum charge of $11,896. Any observation with a total charge lower than this was excluded. In the pilot investigation, the median pre-infection LOS for MRSA was 10 days. A LOS greater that nine days should have a total hospital charge greater than the minimum which was $11,896; therefore, any charge less than $11,896 with a LOS greater than 9 days was excluded. This data manipulation step excluded patients with extremely low charges and a long LOS.

Adjusting for underlying severity of illness was a major concern in this investigation. There is currently no one well-validated, universally accepted illness severity score for infectious disease outcomes.[27] One investigation did explore a comorbidity risk-adjustment measure specifically for MRSA.[39] However, this measure has not been widely used in the literature. Another investigation comparing two severity of illness indices created for non-infectious disease related indications, the Charlson comorbidity score and the Chronic Disease Score, found both to be indicators of increased risk for a nosocomial infection based on preexisting comorbidities.[107] Other investigations have used a variety of techniques including APACHE score [12, 18], McCabe/Jackson score [11] and the Charlson comorbidity score.[60] The APACHE score is intended for use with ICU patients while the McCabe/Jackson score has been evaluated for non-ICU patients. DRGs have been used as a surrogate for severity of underlying illness in MRSA bacteremia.[13] This approach was originally developed to accurately assess the cost of hospitalization by adjusting for severity of illness within the DRG classification scheme.[46, 47] This investigation used both DRG weights and the Charlson comorbidity score to adjust for severity of underlying illness. However, it

should be mentioned that the Charlson comorbidity score might not have been the ideal method for controlling comorbid conditions. The adapted Charlson comorbidity score uses ICD-9-CM information, which is based on discharge data. Therefore, additional comorbidities that arose after the onset of infection would be intermediaries between infection susceptibility and charge, thereby potentially biasing the study results.[107]

This investigation did not assess the appropriateness of a patient's antibiotic therapy. One previous investigation found that 32.9% of patients with a MRSA bacteremia did not initially receive appropriate antibiotic therapy.[102] As previously discussed, charges are higher for patients who have delayed initiation of appropriate therapy.[108] Additionally, differences in antibiotics play a role in patient outcomes for MRSA and MSSA infections.[109] For example, whether an antibiotic is bacteriostatic or bacteriocidal could change the effectiveness and duration of treatment. Although researchers are beginning to question the effectiveness of vancomycin for treating MRSA bacteremia,[99] this investigation did not evaluate the efficacy of vancomycin. Pharmacotherapy information was not included in the data provided by Cerner.

This investigation did not examine any patient clinical sub-populations in the multivariable analysis. *Staphylococcus aureus* infections are a serious and frequent complication of hemodialysis.[110] Two reports have been published specifically evaluating the economic impact of SAB in end-stage renal disease patients undergoing hemodialysis.[15, 16] This investigation deliberately did not focus on any particular diagnosis sub-populations in an attempt to increase external validity.

Finally, this study relied on previously collected data. Retrospective data can be convenient since the researcher does not have to wait for the data to be prospectively collected. However, records must be complete and accurate or the results could be biased.[61] The final sample for this investigation was complete. However, many potential observations were excluded from the sample because their data were not complete. The advantage of the database used was the size. Even though many observations were removed, the final sample was still large.

Future Research

Future research in this area should be focused on path analysis. Path analysis is ideal for modeling an exposure, confounders, intermediaries and outcomes simultaneously. This method is a suitable technique to tease out why MRSA infections seem to cost more than MSSA infections. However, this technique requires much forethought. The model is only as good as the *a priori* assumptions it is based on. Developing a more complete path analysis model that includes all potential confounders as well as post-infection LOS would be helpful. The model from this analysis suggests the impact of susceptibility status on total charge was small. Creating a more comprehensive path model would either support or refute this finding. Severity of illness and mortality should be added to the model. This investigation should serve as a baseline of comparison for future path analysis models.

Conclusions

The broad purpose of this investigation was to explore the relationships between susceptibility status (MRSA vs. MSSA), total hospital charge and LOS (pre- and post-infection). MRSA was associated with a $24,406 increase in total hospital charge. A multivariable model took into account potential confounders and estimated a $22,889 adjusted increase. However, the multivariable indicated that the effect of infection susceptibility was different based on the Charlson comorbidity score. For the majority of patients, MRSA had higher total charges. But, as patients became more severely ill, MRSA charges decreased while MSSA charges increased.

In the multivariable model with post-infection LOS as the outcome, the model predicted a mean post-infection LOS increase of 1.3 days for MRSA over MSSA patients. However, the magnitude of increased post-infection LOS based on pre-infection LOS was different for MRSA and MSSA patients. For the majority of patients, there was no difference in post-infection LOS based on susceptibility. But as pre-infection LOS increased, post-infection LOS for MSSA patients became notably longer than MRSA patients.

Path analysis incorporated all the variables of interest in one model. Susceptibility status was not highly correlated with total charge or post-infection LOS. This implies that susceptibility status may not highly influence charge or post-infection LOS through the proposed model. However, two important confounders were not included in the analysis (severity of illness and mortality). Future research should incorporate these confounders in the structural equation models. The role of

susceptibility status may still have an indirect effect on total charge or post-infection LOS through confounders not included in this investigations path analysis.

Literature Cited

Literature Cited

(1) Centers for Disese Control and Prevention. A Public Health Action Plan to Combat Antimicrobial Resistance. http://www.cdc.gov/drugresistance/actionplan/index htm 2007 June 12 [cited 2008 Apr 1];

(2) Centers for Disease Control and Prevention. MRSA in healthcare settings. http://www cdc gov/ncidod/dhqp/ar_MRSA_spotlight_2006 html 2007 October 3 [cited 2007 Dec 18];

(3) Klevens RM, Morrison MA, Nadle J, et al. Invasive methicillin-resistant Staphylococcus aureus infections in the United States. JAMA **2007 Oct 17**; 298(15):1763-71.

(4) Howard D, Cordell R, McGowan JE, Jr., Packard RM, Scott RD, Solomon SL. Measuring the economic costs of antimicrobial resistance in hospital settings: summary of the Centers for Disease Control and Prevention-Emory Workshop. Clin Infect Dis **2001 Nov 1**; 33(9):1573-8.

(5) Finkler SA. The distinction between cost and charges. Ann Intern Med **1982 Jan**; 96(1):102-9.

(6) Greenland S, Rothman KJ. Measures of Effect and Measures of Association. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott-Raven, **1998**:47-64.

(7) Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. Epidemiology **1999 Jan**; 10(1):37-48.

(8) Cosgrove SE. The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs. Clin Infect Dis **2006 Jan 15**; 42 Suppl 2:S82-S89.

(9) Mansson R, Joffe MM, Sun W, Hennessy S. On the estimation and use of propensity scores in case-control and case-cohort studies. Am J Epidemiol **2007 Aug 1**; 166(3):332-9.

(10) Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Stat Med **2004 Oct 15**; 23(19):2937-60.

(11) 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. Infect Control Hosp Epidemiol **2005 Feb**; 26(2):166-74.

(12) Lodise TP, McKinnon PS. Clinical and economic impact of methicillin resistance in patients with Staphylococcus aureus bacteremia. Diagn Microbiol Infect Dis **2005 Jun**; 52(2):113-22.

(13) McHugh CG, Riley LW. Risk factors and costs associated with methicillin-resistant Staphylococcus aureus bloodstream infections. Infect Control Hosp Epidemiol **2004 May**; 25(5):425-30.

(14) Abramson MA, Sexton DJ. Nosocomial methicillin-resistant and methicillin-susceptible Staphylococcus aureus primary bacteremia: at what costs? Infect Control Hosp Epidemiol **1999 Jun**; 20(6):408-11.

(15) Reed SD, Friedman JY, Engemann JJ, et al. Costs and outcomes among hemodialysis-dependent patients with methicillin-resistant or methicillin-susceptible Staphylococcus aureus bacteremia. Infect Control Hosp Epidemiol **2005 Feb**; 26(2):175-83.

(16) Greiner W, Rasch A, Kohler D, Salzberger B, Fatkenheuer G, Leidig M. Clinical outcome and costs of nosocomial and community-acquired Staphylococcus aureus bloodstream infection in haemodialysis patients. Clin Microbiol Infect **2007 Mar**; 13(3):264-8.

(17) National Nosocomial Infections Surveillance (NNIS) System Report, data summary from January 1992 through June 2004, issued October 2004. Am J Infect Control **2004 Dec**; 32(8):470-85.

(18) Blot SI, Vandewoude KH, Hoste EA, Colardyn FA. Outcome and attributable mortality in critically Ill patients with bacteremia involving methicillin-susceptible and methicillin-resistant Staphylococcus aureus. Arch Intern Med **2002 Oct 28**; 162(19):2229-35.

(19) Chang FY, MacDonald BB, Peacock JE, Jr., et al. A prospective multicenter study of Staphylococcus aureus bacteremia: incidence of endocarditis, risk factors for mortality, and clinical impact of methicillin resistance. Medicine (Baltimore) **2003 Sep**; 82(5):322-32.

(20) Cosgrove SE, Sakoulas G, Perencevich EN, Schwaber MJ, Karchmer AW, Carmeli Y. Comparison of mortality associated with methicillin-resistant and methicillin-susceptible Staphylococcus aureus bacteremia: a meta-analysis. Clin Infect Dis **2003 Jan 1**; 36(1):53-9.

(21) Fowler VG, Jr., Olsen MK, Corey GR, et al. Clinical identifiers of complicated Staphylococcus aureus bacteremia. Arch Intern Med **2003 Sep 22**; 163(17):2066-72.

(22) Harbarth S, Rutschmann O, Sudre P, Pittet D. Impact of methicillin resistance on the outcome of patients with bacteremia caused by Staphylococcus aureus. Arch Intern Med **1998 Jan 26**; 158(2):182-9.

(23) Whitby M, McLaws ML, Berry G. Risk of death from methicillin-resistant Staphylococcus aureus bacteraemia: a meta-analysis. Med J Aust **2001 Sep 3**; 175(5):264-7.

(24) Mayo Clinic. http://www mayoclinic com/health/mrsa/DS00735/DSECTION=4 2007 June 3 [cited 7 A.D. Oct 4];

(25) McGowan JE, Jr. Economic impact of antimicrobial resistance. Emerg Infect Dis **2001 Mar**; 7(2):286-92.

(26) Carroll NV. Pharmacoeconimcs. Financial Management for Pharmacists. 2nd ed. Philadelphia: Lippincott Williams & Wilkins, **1998**:254-5.

(27) Cosgrove SE, Carmeli Y. The impact of antimicrobial resistance on health and economic outcomes. Clin Infect Dis **2003 Jun 1**; 36(11):1433-7.

(28) US Department of Health and Human Services. Prospective Payment System. http://www cms hhs gov/HomeHealthPPS/ 2008 April 10 [cited 2008 Apr 25];

(29) Centers for Medicare and Medicaid Services. Overview of Acute Inpatient PPS. http://www cms hhs gov/acuteinpatientpps/ 2008 December 8

(30) Ashby JL. The accuracy of cost measures derived from Medicare cost report data. Washington DC; **1993**. Report No.: GPO item number 1061-H-03.

(31) Kopp BJ, Nix DE, Armstrong EP. Clinical and economic analysis of methicillin-susceptible and -resistant Staphylococcus aureus infections. Ann Pharmacother **2004 Sep**; 38(9):1377-82.

(32) Nixon RM, Thompson SG. Parametric modelling of cost data in medical studies. Stat Med **2004 Apr 30**; 23(8):1311-31.

(33) Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? BMJ **2000 Apr 29**; 320(7243):1197-200.

(34) Greenland S, Morgenstern H. Confounding in health research. Annu Rev Public Health **2001**; 22:189-212.

(35) Rothman KJ, Greenland S. Matching. In: Rothman KJ, Greenland S, eds. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott-Raven, **1998**:147-61.

(36) Csizmadi I, Collet JP. Bias and Confounding in Pharmacoepidemiology. In: Strom BL, Kimmel SE, eds. Textbook of Pharmacoepidemiology.England: John Wiley & Sons, Ltd, **2006**:262-75.

(37) Rothman KJ, Greenland S. Accuracy and Consideration in Study Design. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott-Raven, **1998**:135-45.

(38) Cosgrove SE. The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs. Clin Infect Dis **2006 Jan 15**; 42 Suppl 2:S82-S89.

(39) McGregor JC, Perencevich EN, Furuno JP, et al. Comorbidity risk-adjustment measures were developed and validated for studies of antibiotic-resistant infections. J Clin Epidemiol **2006 Dec**; 59(12):1266-73.

(40) Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med **1985 Oct**; 13(10):818-29.

(41) Knaus WA, Wagner DP, Draper EA, et al. The APACHE III prognostic system. Risk prediction of hospital mortality for critically ill hospitalized adults. Chest **1991 Dec**; 100(6):1619-36.

(42) McCabe WR, Jackson GG. Gram-negative bacteremia, I:etiology and ecology. Arch Intern Med **1962**; 110:847-55.

(43) Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis **1987**; 40(5):373-83.

(44) Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol **1992 Jun**; 45(6):613-9.

(45) Iezzoni LI. The risks of risk adjustment. JAMA **1997 Nov 19**; 278(19):1600-7.

(46) Anderson GF, Lave JR, Russe CM. Providing Hospital Services: The Changing Financial Environment. Baltimore, MD: Johns Hopkins University Press, **1989**.

(47) Feinglass J, Scherubel JC, Swartz JA. Comparison of two systems to measure the severity of illness. J Am Med Rec Assoc **1988 Jul**; 59(7):34-8.

(48) Chaix C, Durand-Zaleski I, Alberti C, Brun-Buisson C. Control of endemic methicillin-resistant Staphylococcus aureus: a cost-benefit analysis in an intensive care unit. JAMA **1999 Nov 10**; 282(18):1745-51.

(49) Lepelletier D, Ferreol S, Villers D, Richet H. [Methicillin-resistant Staphylococcus aureus nosocomial infections in ICU: risk factors, morbidity and cost]. Pathol Biol (Paris) **2004 Oct**; 52(8):474-9.

(50) Kim T, Oh PI, Simor AE. The economic impact of methicillin-resistant Staphylococcus aureus in Canadian hospitals. Infect Control Hosp Epidemiol **2001 Feb**; 22(2):99-104.

(51) Capitano B, Leshem OA, Nightingale CH, Nicolau DP. Cost effect of managing methicillin-resistant Staphylococcus aureus in a long-term care facility. J Am Geriatr Soc **2003 Jan**; 51(1):10-6.

(52) Rubin RJ, Harrington CA, Poon A, Dietrich K, Greene JA, Moiduddin A. The economic impact of Staphylococcus aureus infection in New York City hospitals. Emerg Infect Dis **1999 Jan**; 5(1):9-17.

(53) Clinical and Laboratory Standards Institute. Performance Standards for Antimicrobial Susceptibility Testing. http://www clsi org/ 2008Available from: URL: http://www.clsi.org/

(54) Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. J Health Serv Res Policy **2004 Oct**; 9(4):197-204.

(55) Cosgrove SE. The relationship between antimicrobial resistance and patient outcomes: mortality, length of hospital stay, and health care costs. Clin Infect Dis **2006 Jan 15**; 42 Suppl 2:S82-S89.

(56) Graffunder EM, Venezia RA. Risk factors associated with nosocomial methicillin-resistant Staphylococcus aureus (MRSA) infection including previous use of antimicrobials. J Antimicrob Chemother **2002 Jun**; 49(6):999-1005.

(57) Wakefield DS, Helms CM, Massanari RM, Mori M, Pfaller M. Cost of nosocomial infection: relative contributions of laboratory, antibiotic, and per diem costs in serious Staphylococcus aureus infections. Am J Infect Control **1988 Oct**; 16(5):185-92.

(58) Engemann JJ, Carmeli Y, Cosgrove SE, et al. Adverse clinical and economic outcomes attributable to methicillin resistance among patients with Staphylococcus aureus surgical site infection. Clin Infect Dis **2003 Mar 1**; 36(5):592-8.

(59) Pittet D, Tarara D, Wenzel RP. Nosocomial bloodstream infection in critically ill patients. Excess length of stay, extra costs, and attributable mortality. JAMA **1994 May 25**; 271(20):1598-601.

(60) Shurland S, Zhan M, Bradham DD, Roghmann MC. Comparison of mortality risk associated with bacteremia due to methicillin-resistant and methicillin-susceptible Staphylococcus aureus. Infect Control Hosp Epidemiol **2007 Mar**; 28(3):273-9.

(61) Woodward M. Cohort Study. Epidemiology: Study Design and Data Analysis.Boca Raton, Florida: Chapman & Hall/CRC, **1999**:191-241.

(62) Shorr AF. Epidemiology and economic impact of meticillin-resistant Staphylococcus aureus: review and analysis of the literature. Pharmacoeconomics **2007**; 25(9):751-68.

(63) Blough DK, Ramsey SD. Using generalized linear models to assess medical care costs. Health Srevices and Outcomes Research Methodology **2000**; 1(2):185-202.

(64) Cerner Corporation. Cerner Corporation: HealthFacts. http://www cerner com/public/Cerner_3 asp?id=3568 2008 [cited 2008 Dec 15];

(65) Centers for Medicare and Medicaid Services. DRG Relative Weights. http://www cms hhs gov/AcuteInpatientPPS/FFD/list asp?listpage=2 2008 December 15

(66) Elwood, JM. Counfounding. Critical Appraisal of Epidemiological Studies and Clinical Trials. 3rd ed. New Yourk: Oxford University Press, **2007**:157-220.

(67) Centers for Medicare and Medicaid Services. Consumer Price Indexes of Medical Care Prices: 1980 to 2006. http://www census gov/compendia/statab/tables/08s0132 xls#Notes!A1 2008

(68) Drummond MF SMTGOBSG. Critical Assessment of Economic Evaluation. Methods for the Economic Evaluation of Health Care Programmes. 3rd ed. New York: Oxford University Press, **2005**:27-50.

(69) Briggs A, Gray A. The distribution of health care costs and their statistical analysis for economic evaluation. J Health Serv Res Policy **1998 Oct**; 3(4):233-45.

(70) Skrepnek GH. Regression methods in the empiric analysis of health care data. J Manag Care Pharm **2005 Apr**; 11(3):240-51.

(71) Evans M, Hastings N, Peacock B. Statistical Distributions. New York: Wiley, **1993**.

(72) Woodward M. Fundemental Issues. Epidemiology: Study Design and Data Analysis.Boca Raton, Florida: Chapman & Hall/CRC, **1999**:1-29.

(73) Page RM, Cole GE, Timmreck TC. Epidemiologic Research. In: Boston, ed. Basic Epidemiological Methods and Biostatistics: A Practical Guidebook.Boston: Jones and Bartlett Publishers, **2005**:81-144.

(74) Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. Am J Epidemiol **1999 Aug 15**; 150(4):327-33.

(75) Mitchell H.Gail. Propensity Scores. In: Mitchell H.Gail JB, ed. Encyclopedia of Epidemiologic Methods.New York: John Wiley and Sons, **2000**:738-41.

(76) D'Agostino RB Jr. Adjustment Methods. In: D'Agostino RB, ed. Tutorials in Biostatistics, Statistical Methods in Clinical Studies.New York: John Wiley and Sons, **2004**:67-84.

(77) Rubin D.B. Using multivariate matched sampling and regression adjustments to control bias in observational studies. Journal of the American Statistical Association **1979**; 74:318-24.

(78) Oakes JM, Church TR. Invited commentary: advancing propensity score methods in epidemiology. Am J Epidemiol **2007 May 15**; 165(10):1119-21.

(79) Lapane K. Analytic Strategies for the Evaluation of Pharmacoepidemiologic Studies. In: Hartzema AG, Porta M, Tilson HH, eds. Pharmacoepidemiology. 3rd ed. Cincinnati, Ohio: Harvey Whitney Books Company, **2009**:235-66.

(80) Winkelmayer WC, Kurth T. Propensity scores: help or hype? Nephrol Dial Transplant **2004 Jul**; 19(7):1671-3.

(81) Sturmer T, Schneeweiss S, Avorn J, Glynn RJ. Adjusting effect estimates for unmeasured confounding with validation data using propensity score calibration. Am J Epidemiol **2005 Aug 1**; 162(3):279-89.

(82) Kline RB. Basic Statistical Concepts. In: Kenny DA, ed. Principles and Practice of Structural Equation Modeling.New York: The Guilford Press, **1998**:15-46.

(83) Ullman JB. Structural equation modeling. In: Tabachnick B.G., Fidell L.S., eds. Using Multivariate Statistics. 3rd ed. New York: HarperCollins College Publishers, **1996**:709-819.

(84) Tomarken AJ, Waller NG. Structural equation modeling: strengths, limitations, and misconceptions. Annu Rev Clin Psychol **2005**; 1:31-65.

(85) Schumacker RE, Lomax RG. Developing Structural Equation Models: Part 2. A Beginner's Guide to Structural Equation Modeling.Mahwah, New Jersey: Lawrence Erbaum Associates, **1996**:76-97.

(86) Schumacker RE, Lomax RG. Goodness-of-Fit Criteria. A Beginner's Guide to Structural Equation Modeling.Mahwah, New Jersey: Lawrence Erbaum Associates, **1996**:119-37.

(87) Weitzen S, Lapane KL, Toledano AY, Hume AL, Mor V. Principles for modeling propensity scores in medical research: a systematic literature review. Pharmacoepidemiol Drug Saf **2004 Dec**; 13(12):841-53.

(88) Bagley SC, White H, Golomb BA. Logistic regression in the medical literature: standards for use and reporting, with particular attention to one medical domain. J Clin Epidemiol **2001 Oct**; 54(10):979-85.

(89) Greenland S. Modeling and variable selection in epidemiologic analysis. Am J Public Health **1989 Mar**; 79(3):340-9.

(90) Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol **1996 Dec**; 49(12):1373-9.

(91) Raymond Myers. The Simple Linear Regression Model. Classical and Modern Regression with Applications. 2nd ed. Pacific Grove, California: Duxbury Thomson Learning, **1990**:57-62.

(92) Raymond Myers, Douglas Montgomery, G.Geoffrey Vining. The Family of Generalized Linear Models. Generalized Linear Models.New York: John Wiley & Sons Inc, **2002**:171-90.

(93) Moran JL, Solomon PJ, Peisach AR, Martin J. New models for old questions: generalized linear models for cost prediction. J Eval Clin Pract **2007 Jun**; 13(3):381-9.

(94) Kline RB. Structural Models with Observed Variables and Path Analysis: I. Fundementals, Recursive Models. In: Kenny DA, ed. Principles and Practice of Structural Equation Modeling.New York: The Guilford Press, **1998**:95-152.

(95) Kline RB. Measurement Models and Confirmatory Factor Analysis. In: Kenny DA, ed. Principles and Practice of Structural Equation Modeling.New York: The Guilford Press, **1998**:189-243.

(96) McGue M, Wette R, Rao DC. Evaluation of path analysis through computer simulation: effect of incorrectly assuming independent distribution of familial correlations. Genet Epidemiol **1984**; 1(3):255-69.

(97) Baba T, Takeuchi F, Kuroda M, et al. Genome and virulence determinants of high virulence community-acquired MRSA. Lancet **2002 May 25**; 359(9320):1819-27.

(98) Naimi TS, LeDell KH, Como-Sabetti K, et al. Comparison of community- and health care-associated methicillin-resistant Staphylococcus aureus infection. JAMA **2003 Dec 10**; 290(22):2976-84.

(99) Sakoulas G, Moellering RC, Jr., Eliopoulos GM. Adaptation of methicillin-resistant Staphylococcus aureus in the face of vancomycin therapy. Clin Infect Dis **2006 Jan 1**; 42 Suppl 1:S40-S50.

(100) Chambers HF, Hegde SS. Combating the growing problem of methicillin-resistant Staphylococcus aureus: do the newer antibiotics represent a better alternative to vancomycin? Expert Rev Anti Infect Ther **2007 Jun**; 5(3):333-5.

(101) Lodise TP, McKinnon PS, Swiderski L, Rybak MJ. Outcomes analysis of delayed antibiotic treatment for hospital-acquired Staphylococcus aureus bacteremia. Clin Infect Dis **2003 Jun 1**; 36(11):1418-23.

(102) Ibrahim EH, Sherman G, Ward S, Fraser VJ, Kollef MH. The influence of inadequate antimicrobial treatment of bloodstream infections on patient outcomes in the ICU setting. Chest **2000 Jul**; 118(1):146-55.

(103) Leibovici L, Shraga I, Drucker M, Konigsberger H, Samra Z, Pitlik SD. The benefit of appropriate empirical antibiotic treatment in patients with bloodstream infection. J Intern Med **1998 Nov**; 244(5):379-86.

(104) Rosen AB, Fowler VG, Jr., Corey GR, et al. Cost-effectiveness of transesophageal echocardiography to determine the duration of therapy for intravascular catheter-associated Staphylococcus aureus bacteremia. Ann Intern Med **1999 May 18**; 130(10):810-20.

(105) Howard DH, Scott RD, Packard R, Jones D. The global impact of drug resistance. Clin Infect Dis **2003 Jan 15**; 36(Suppl 1):S4-10.

(106) Tice AD, Hoaglund PA, Nolet B, McKinnon PS, Mozaffari E. Cost perspectives for outpatient intravenous antimicrobial therapy. Pharmacotherapy **2002 Feb**; 22(2 Pt 2):63S-70S.

(107) McGregor JC, Kim PW, Perencevich EN, et al. Utility of the Chronic Disease Score and Charlson Comorbidity Index as comorbidity measures for use in epidemiologic studies of antibiotic-resistant organisms. Am J Epidemiol **2005 Mar 1**; 161(5):483-93.

(108) Roghmann MC. Predicting methicillin resistance and the effect of inadequate empiric therapy on survival in patients with Staphylococcus aureus bacteremia. Arch Intern Med **2000 Apr 10**; 160(7):1001-4.

(109) Lodise TP, Jr., McKinnon PS. Burden of methicillin-resistant Staphylococcus aureus: focus on clinical and economic outcomes. Pharmacotherapy **2007 Jul**; 27(7):1001-12.

(110) Marr KA, Kong L, Fowler VG, et al. Incidence and outcome of Staphylococcus aureus bacteremia in hemodialysis patients. Kidney Int **1998 Nov**; 54(5):1684-9.

Appendix A

IRB Approval Form

VCU Memo

Virginia Commonwealth University

Office of Research Subjects Protection
BioTechnology Research Park
BioTech One, 800 E. Leigh Street, #114
P.O. Box 980568
Richmond, Virginia 23298-0568
(804) 828-0868, fax (804) 827-1448

DATE: September 25, 2008

TO: Spencer Harpe, PharmD, MPH, PhD
Pharmacy
Box 980533

FROM: Lloyd Byrd, MS LB/DA
Chairperson, VCU IRB Panel E
Box 980568

RE: **VCU IRB #: HM11841**
Title: The Economic Impact of Antimicrobial Resistance in Patients with Nonsocomial *Staphylococcus aureus* Bacteremia

On September 22, 2008 the following research study ***qualified for exemption*** according to 45 CFR 46.101(b) Category 4. This approval includes the following items reviewed by this Panel:

RESEARCH APPLICATION/PROPOSAL: NONE

PROTOCOL: The Economic Impact of Antimicrobial Resistance in Patients with Nonsocomial *Staphylococcus aureus* Bacteremia, version 9/4/08, received 9/15/08

ADDITIONAL DOCUMENTS:
- None

The Primary Reviewer assigned to your research study is Stephen Auerbach, PhD. If you have any questions, please contact Dr. Auerbach at sauerbac@vcu.edu and 828-1172; or you may contact Donna Gross, IRB Coordinator, VCU Office of Research Subjects Protection, at dsgross@vcu.edu or 827-2261.

Attachment – Conditions of Approval

Conditions of Approval:

In order to comply with federal regulations, industry standards, and the terms of this approval, the investigator must *(as applicable)*:

1. Conduct the research as described in and required by the Protocol.

2. Obtain informed consent from all subjects without coercion or undue influence, and provide the potential subject sufficient opportunity to consider whether or not to participate (unless Waiver of Consent is specifically approved or research is exempt).

3. Document informed consent using only the most recently dated consent form bearing the VCU IRB "APPROVED" stamp (unless Waiver of Consent is specifically approved).

4. Provide non-English speaking patients with a translation of the approved Consent Form in the research participant's first language. The Panel must approve the translated version.

5. Obtain prior approval from VCU IRB before implementing any changes whatsoever in the approved protocol or consent form, unless such changes are necessary to protect the safety of human research participants (e.g., permanent/temporary change of PI, addition of performance/collaborative sites, request to include newly incarcerated participants or participants that are wards of the state, addition/deletion of participant groups, etc.). Any departure from these approved documents must be reported to the VCU IRB immediately as an Unanticipated Problem (see #7).

6. Monitor all problems (anticipated and unanticipated) associated with risk to research participants or others.

7. Report Unanticipated Problems (UPs), including protocol deviations, following the VCU IRB requirements and timelines detailed in VCU IRB WPP VIII-7):

8. Obtain prior approval from the VCU IRB before use of any advertisement or other material for recruitment of research participants.

9. Promptly report and/or respond to all inquiries by the VCU IRB concerning the conduct of the approved research when so requested.

10. All protocols that administer acute medical treatment to human research participants must have an emergency preparedness plan. Please refer to VCU guidance on http:/www.research.vcu.edu/irb/guidance.htm.

11. The VCU IRBs operate under the regulatory authorities as described within:
    a) U.S. Department of Health and Human Services Title 45 CFR 46, Subparts A, B, C, and D (for all research, regardless of source of funding) and related guidance documents.
    b) U.S. Food and Drug Administration Chapter I of Title 21 CFR 50 and 56 (for FDA regulated research only) and related guidance documents.
    c) Commonwealth of Virginia Code of Virginia 32.1 Chapter 5.1 Human Research (for all research).

Appendix B

Codebook for Categorical Variables

Variable Collected	Variable Categories & Subcategories	Variable Name	Variable Code
Susceptibility	MRSA MSSA	susceptibility	MRSA = 1 MSSA = 0
Gender	Gender	gender1	Male = 1 Female = 0
Admission Source	Referral Physician referral Clinic referral HMO referral Transfer Transfer from a hospital Transfer from a skilled nursing facility (SNF) Transfer from another health care facility Emergency Room N/A Not available Null Not mapped	admission_source	Referral =1 Transfer = 2 Emergency Room = 7 N/A = 3

(continued)

Appendix B: (continued)

Variable Collected	Variable Categories & Subcategories	Variable Name	Variable Code
Status at Discharge	Discharged Home	discharge	Discharged Home = 1
	Discharged to home		
	Discharged/transferred to home under care of Home IV provider		
	Discharged/transferred to home with home health service		
	Transferred		Transferred = 2
	Discharged/transferred to another short term hospital		
	Discharged/transferred to SNF		
	Discharged/transferred to ICF		
	Discharged/transferred to another type of inpatient care institution		
	Discharged/transferred to another rehabilitation facility		
	Discharged/transferred to a long term care hospital		
	Hospice		Hospice = 6
	Hospice / home		
	Hospice / medical facility		
	Expired		Expired = 3
	Other		Other = 4
	Left AMA		
	Still patient or expected to return for outpatient services		
	Discharged/transferred within this institution to Medicare approved swing bed		
	N/A		N/A = 5
	NULL		
	Not Mapped		

(continued)

Appendix B: (continued)

Variable Collected	Variable Categories & Subcategories	Variable Name	Variable Code
Discharge Status	Discharged	mortality	Discharged = 1
	Expired		Expired = 0
	N/A		N/A=2
Type of Payer	Insured	payer	Insured = 1
	BCBS		
	Other Commercial Payer		
	Other Government		
	Other Non-govt		
	Self-Insured		
	CHAMPUS (military dependents)		
	HMO/Managed care		
	Worker's Compensation		
	Medicare		Medicare = 5
	Medicaid		Medicaid = 4
	Self Pay		Self Pay = 2
	N/A		N/A = 5
	Unknown/Missing/Invalid		
	Null		
	Not mapped		

Appendix C

Codebook for Continuous Variables

Variable Collected	Unit of Measurement	Variable Code
Age	year	age
DRG weight	n/a	weights
Charlson comorbidity score	n/a	charlson
Propensity score	n/a	prob
Total hospital charge	dollars	charge
Total length of stay	day	total_LOS
Pre-infection LOS	day	pre_infection_LOS
Post-infection LOS	day	post_infection_LOS

DRG = diagnosis related group
LOS = length of stay

Appendix D

SAS Code

```
*Entire SAS code;
libname phd 'C:\Documents and Settings\Suzanne Phillips\Desktop\SAS PhD
files';
run;

libname phdfinal 'C:\Documents and Settings\Suzanne Phillips\My
Documents\PhD';
run;

*Micro data;
PROC IMPORT OUT= phd.phdmicro
            DATAFILE= "C:\Documents and Settings\Suzanne
Phillips\Deskto
p\Suzanne_MRSA\all_criteria_ms_sep1.txt"
            DBMS=DLM REPLACE;
     DELIMITER='7C'x;
     GETNAMES=YES;
     DATAROW=2;
RUN;

ods html file = "phd.phdmicro";
libname phd 'C:\Documents and Settings\Suzanne Phillips\Desktop\SAS PhD
files';
proc print data=phd.phdmicro;
format id comma30.0;
run;
ods html close;

data phd.phdmicro1;
      set phd.phdmicro;
      format id comma31.0;
      format SUSCEPTIBILITY comma3.0;
      format microbial_code comma5.0;
run;
```

```
*micro results with all observations;
proc freq data=phd.phdmicro1;
      tables susceptibility;
      run;

Proc sort data = phd.phdmicro1;
      by id;
      run;

proc freq data = phd.phdmicro1 noprint;
      table id / out=phd.phdmicro1__1;
      run;


*remove observations if susceptibility = 2,3,4,10,11;
data phd.phdmicro2;
      set phd.phdmicro1;
      if SUSCEPTIBILITY <7 then delete;
      if SUSCEPTIBILITY > 9 then delete;
      run;

proc sort data=phd.phdmicro2;
      by id ascending susceptibility;
      run;

proc sort data=phd.phdmicro2 out=phd.phdmicro_nodup
dupout=phd.phdmicrodups nodupkey;
      by id;
run;

Proc sort data = phd.phdmicro_nodup;
      by id;
      run;

proc freq data = phd.phdmicro_nodup noprint;
      table id / out=phd.phdmicronodup1__1;
      run;


data phd.phdmicrodups (rename = (susceptibility = suscep));
set phd.phdmicrodups;
run;

*how to sort and merge.;
PROC SORT Data=phd.phdmicro_nodup;
   BY id;
  RUN;

  PROC SORT Data=phd.phdmicrodups;
   BY id;
  RUN;
```

```
  DATA phd.phdmicroremerge;
   MERGE phd.phdmicro_nodup phd.phdmicrodups;
   BY id;
   run;
*only 7 and 8 duplicates;
data phd.phdmicroremerge1;
      set phd.phdmicroremerge;
      if suscep = "." then delete;
      run;

*no duplicates - final dataset;
data phd.phdmicrofinal;
      set phd.phdmicroremerge;
      if SUSCEPTIBILITY = 8 and suscep = 7 then delete;
      if SUSCEPTIBILITY =7 and suscep = 8 then delete;
      run;

*(micro masacure is complete);

Proc sort data = phd.phdmicrofinal;
      by id;
      run;

proc freq data = phd.phdmicrofinal noprint;
      table id / out=phd.phdmicrofinal1__1;
      run;

*Descriptive stats for just micro;
proc freq data=phd.phdmicrofinal;
      tables susceptibility*microbial_code;
      run;

proc freq data=phd.phdmicrofinal;
      tables susceptibility;
      run;

*large dataset;
PROC IMPORT OUT= phd.PHDlarge
            DATAFILE= "C:\Documents and Settings\Suzanne
Phillips\Desktop\Suzanne_MRSA\emhdipipa_flag_icu_sep1.txt"
            DBMS=DLM REPLACE;
     DELIMITER='7C'x;
     GETNAMES=YES;
     DATAROW=2;
       RUN;

data phd.phdlarge1;
      set phd.phdlarge;
      admit= INPUT(ADMITTED_DT_TM, DATE9.);
      discharge= INPUT(DISCHARGED_DT_TM,DATE9.);
      tolab= INPUT(MICRO_LAB_RECEIVED_DT_TM, DATE9.);
      LOS_total= discharge- admit;
      Pre_infection_LOS = tolab- admit;
```

```
      Post_infection_LOS = discharge- tolab;
      Format admitdischargetolabdate9.;
      format id comma32.;
run;

Proc sort data = phd.phdlarge1;
      by id;
      run;

proc freq data = phd.phdlarge1 noprint;
      table id / out=phd.phdlarge1__1;
      run;
* represents 6750 patients;

/* exclude if pre infection LOS < 2 days. */

data phd.phdlarge2;
      set phd.phdlarge1;
      if pre_infection_LOS >=2;
run;


Proc sort data = phd.phdlarge2;
      by id;
      run;

proc freq data = phd.phdlarge2 noprint;
      table id / out=phd.phdlarge2__1;
      run;

* this dataset has 3,313 obs;


*an aside, pre-infection LOS > 2 n=165,918 ;
data phd.phdlarge2_1;
      set phd.phdlarge1;
      if pre_infection_LOS < 2;
run;

proc sort data=phd.phdlarge2_2 out=phd.phdlarge2_4 noduprecs;
by id pre_infection_LOS;
run;

data phd.phdlarge2_4firstobs phd.dupobs2_4;
set phd.phdlarge2_4;
by id;
if first.id then output phd.phdlarge2_4firstobs;
else output phd.dupobs2_4;
run;

proc freq data= phd.phdlarge2_4firstobs;
      tables ad_source;
      run;
```

```
/*After excluding these observations, n= 249,449*/

*phd.phdlarge3 has 249,449 obs;
/*Remove duplicate observations */

proc sort data=phd.phdlarge3 out=phd.phdlarge4 noduprecs;
by id pre_infection_LOS;
run;

*phd.phdlarge4 had 5799 obs but there are still some dup where multiple
cultures were taken in the same admission;
data phd.phdlargefirstobs phd.dupobs;
set phd.phdlarge4;
by id;
if first.id then output phd.phdlargefirstobs;
else output phd.dupobs;
run;

*phd.phdlargefirstobs has 3313 obs;

Proc sort data = phd.phdlargefirstobs;
      by id;
      run;

proc freq data = phd.phdlargefirstobs;
      table id / out=phd.charlson_1;
      run;



*Descriptive Statistics;

*large dataset hospitals by year;
data phd.phdlarge3_1;
      set phd.phdlarge3;
format tolabyear4.;
run;

proc sort data=phd.phdlarge3_1 out=phd.phdlarge3_2;
by into_microlab_date;
run;

proc freq data=phd.phdlarge3_2;
      tables hospital;
      by into_microlab_date;
run;

*smaller dataset hospital by year;
data phd.phdlarge4_1;
      set phd.phdlarge4;
format tolabyear4.;
run;
```

```
proc sort data=phd.phdlarge4_1 out=phd.phdlarge4_2;
by into_microlab_date;
run;

proc freq data=phd.phdlarge4_2;
      tables hospital;
      by into_microlab_date;
run;

*final dataset;
data phd.phdlargefirstobs_1;
      set phd.phdlargefirstobs;
format tolabyear4.;
run;

proc sort data=phd.phdlargefirstobs_1 out=phd.phdlargefirstobs_2;
by into_microlab_date;
run;

proc freq data=phd.phdlargefirstobs_2;
      tables hospital;
      by into_microlab_date;
run;

*merge micro and large file;

/*    micro data = phd.phdmicrofinal - 3275
      descriptive data = phd.phdlargefirstobs - 3313
      */

  PROC SORT Data=phd.phdmicrofinal;
   BY id;
  RUN;

  PROC SORT Data=phD.phdlargefirstobs;
   BY id;
  RUN;

  DATA phd.phd1;
   MERGE phd.phdmicrofinal phD.phdlargefirstobs;
   BY id;
   run;

  /* merged dataset (phd.phd1) contains 4855 obs */

 data phd.phd2;
      set phd.phd1;
      if SUSCEPTIBILITY <7 then delete;
      if SUSCEPTIBILITY > 9 then delete;
      run;
```

```
  /* merged dataset (phd.phd2) missing micro data eliminated contains
3275 obs */

 data phd.phd3;
      set phd.phd2;
      if Age <18 then delete;
      run;

 /* merged dataset (phd.phd3) contains 1733 obs */

*Calculate Charlson score;

data phd.charlson;
set phd.diagnosis_codes;
length uniqueid $ 35;
dx=icd9_diagnosis_code;
uniqueid = cat (encounterid);
run;

proc sort data=phd.charlson;
by encounterid;
run;

data a b c d e f g h i j k l m n o p q;
set phd.charlson;
if ((substr(dx,1,3) >= '531') & (substr(dx,1,3) <= '534')) then output
a;
if substr(dx,1,4) in ('2504','2505','2506','2507') then output b;
if ((substr(dx,1,3) >= '196') & (substr(dx,1,3) <= '199')) then output
c;
if substr(dx,1,3) in ('410','412') then output d;
if substr(dx,1,3) in ('428') or substr(dx,1,4) in
('4254','4255','4257','4258','4259') or substr(dx,1,5) in
('39891','40201','40211','40291','40401','40403','40411','40413','40491
','40493') then output e;
if substr(dx,1,3) in ('441','440') or substr(dx,1,4) in
('4439','V434','0930','4373','4431','4432','4438','4471','5571','5579')
then output f;
if substr(dx,1,3) in
('430','431','432','433','434','435','436','437','438') or
substr(dx,1,5)='36234' then output g;
if substr(dx,1,3)='290' or substr(dx,1,4) in ('2941','3312') then
output h;
if substr(dx,1,3) in
('490','491','492','493','494','495','496','500','501','502','503','504
','505') or substr(dx,1,4) in
('5064','4168','4169','5064','5081','5088') then output i;
if substr(dx,1,4) in
('4465','7100','7102','7103','7101','7104','7140','7141','7142','7148')
or substr(dx,1,3)='725' or substr(dx,1,5)='71481' then output j;
if substr(dx,1,4) in ('0706','0709','5733','5734','5738','5739','V427')
or substr(dx,1,5) in ('07022','07023','07032','07033','07044','07054')
or substr(dx,1,3) in ('570','571') then output k;
```

```
if substr(dx,1,4) in ('2500','2501','2502','2503','2508','2509') then
output l;
if substr(dx,1,3) in ('342','343') or substr(dx,1,4) in
('3341','3440','3441','3442','3443','3444','3445','3446','3449') then
output m;
if substr(dx,1,3) in ('582','585','586','V56') or substr(dx,1,4) in
('5830','5831','5832','5834','5836','5837','5880','V420','V451') or
substr(dx,1,5) in
('40301','40311','40391','40402','40403','40412','40413','40492','40493
') then output n;
if substr(dx,1,3) in
('140','141','142','143','144','145','146','147','148','149','150','151
','152','153','154','155','156','157','158','159','160','161','162','16
3','164','165','166','167','168','169','170','171','172','174','175','1
76','177','178','179','180','181','182','183','184','185','186','187','
188','189','190','191','192','193','194','195','200','201','202','203',
'204','205','206','207','208') or substr(dx,1,4)='2386' then output o;
if substr(dx,1,4) in ('5722','5723','5724','5728','4560','4561','4562')
then output p;
if substr(dx,1,3) in ('042','043','044') then output q;
run;

proc freq data=a noprint;
table uniqueid / out=afreq;
proc freq data=b noprint;
table uniqueid / out=bfreq;
proc freq data=c noprint;
table uniqueid / out=cfreq;
run;
proc freq data=d noprint;
table uniqueid / out=dfreq;
run;
proc freq data=e noprint;
table uniqueid / out=efreq;
run;
proc freq data=f noprint;
table uniqueid / out=ffreq;
run;
proc freq data=g noprint;
table uniqueid / out=gfreq;
run;
proc freq data=h noprint;
table uniqueid / out=hfreq;
run;
proc freq data=i noprint;
table uniqueid / out=ifreq;
run;
proc freq data=j noprint;
table uniqueid / out=jfreq;
run;
proc freq data=k noprint;
table uniqueid / out=kfreq;
run;
```

```
proc freq data=l noprint;
table uniqueid / out=lfreq;
run;
proc freq data=m noprint;
table uniqueid / out=mfreq;
run;
proc freq data=n noprint;
table uniqueid / out=nfreq;
run;
proc freq data=o noprint;
table uniqueid / out=ofreq;
run;
proc freq data=p noprint;
table uniqueid / out=pfreq;
run;
proc freq data=q noprint;
table uniqueid / out=qfreq;
run;

data phd.ulcer;
set afreq;
ulcer = 1;
drop count percent;
data phd.dmorgan;
set bfreq;
dmodamage = 2;
drop count percent;
data phd.mtumor;
set cfreq;
mtumor = 6;
drop count percent;
run;
data phd.myocard;
set dfreq;
myocard = 1;
drop count percent;
run;
data phd.chf;
set efreq;
chf = 1;
drop count percent;
run;
data phd.pvd;
set ffreq;
pvd = 1;
drop count percent;
run;
data phd.cerebro;
set gfreq;
cerebro = 1;
drop count percent;
run;
data phd.dementia;
```

```
set hfreq;
dementia = 1;
drop count percent;
run;
data phd.copd;
set ifreq;
copd = 1;
drop count percent;
run;
data phd.rheum;
set jfreq;
rheum = 1;
drop count percent;
run;
data phd.mildliver;
set kfreq;
mildliver = 1;
drop count percent;
run;
data phd.dbnocomp;
set lfreq;
dbnocomp = 1;
drop count percent;
run;
data phd.hemiplegia;
set mfreq;
mtumor = 6;
drop count percent;
run;
data phd.renal;
set nfreq;
renal = 2;
drop count percent;
run;
data phd.cancer;
set ofreq;
cancer = 2;
drop count percent;
run;
data phd.msliver;
set pfreq;
msliver = 3;
drop count percent;
run;
data phd.hiv;
set qfreq;
hiv = 6;
drop count percent;
run;

data phd.phd3beforecharlson;
set phd.phd3;
length uniqueid $ 35;
```

```
uniqueid = cat (id);
run;

proc sort data=phd.phd3beforecharlson;
by uniqueid;
run;


data phd.phd3_aftercharlson;;
merge phd.phd3beforecharlson phd.ulcer phd.dmorgan phd.mtumor
phd.myocard phd.chf phd.pvd phd.cerebro phd.dementia phd.copd phd.rheum
phd.mildliver phd.dbnocomp phd.hemiplegia phd.renal
phd.cancer phd.msliver phd.hiv;
by uniqueid;
if ulcer = . then ulcer = 0;
if dmodamage = . then dmodamage = 0;
if mtumor = . then mtumor = 0;
if myocard = . then myocard= 0;
if chf = . then chf=0;
if pvd = . then pvd=0;
if cerebro = . then cerebro=0;
if dementia = . then dementia = 0;
if copd = . then copd = 0;
if rheum = . then rheum = 0;
if dbnocomp = . then dbnocomp = 0;
if hemiplegia = . then hemiplegia = 0;
if renal = . then renal = 0;
if cancer = . then cancer = 0;
if msliver = . then msliver = 0;
if mildliver = . then mildliver = 0;
if hiv = . then hiv = 0;

charlson = sum(ulcer, dmodamage, mtumor, myocard, chf, pvd, cerebro,
dementia, copd, rheum, dbnocomp, hemiplegia, renal, cancer, msliver,
mildliver, hiv);
run;

 data phd.mergedwithcharlson;
      set phd.phd3_aftercharlson;
      if id = . then delete;
      run;


proc univariate data = phd.mergedwithcharlson;
class charlson;
var charlson;
run;


*add in DRG relative weights;

PROC IMPORT OUT= WORK.DRG
```

```
            DATAFILE= "C:\Documents and Settings\Suzanne Phillips\My
Doc
uments\PhD\2006drg_relative_weights.xls"
            DBMS=EXCEL REPLACE;
     SHEET="drgimport1$";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

data phd.drg;
      set drg;
      run;

 data phd.mergedwithcharlson1;
      set phd.mergedwithcharlson;
      if total_charges <=0 then delete;
      drg = drg_id;
      if drg <=0 then delete;
run;

*data set now contains 1088 obs;

*add drg weights (phd.drg) into dataset (phd.mergedwithcharlson);

  PROC SORT Data= phd.drg;
   BY drg;
  RUN;

  PROC SORT Data=phd.mergedwithcharlson1;
   BY drg;
  RUN;

  DATA phd.drg_charlson;
   MERGE phd.drg phd.mergedwithcharlson1;
   BY drg;
   run;

  data phd.drg_charlson_1088;
      set phd.drg_charlson_1088;
      if id = . then delete;
      run;

  data phd.drg_charlson_drg_weight;
      set phd.drg_charlson_1088;
      if weights = . then delete;
      if weights <= 0 then delete;
      run;

*dataset contains 966 pts;
```

```
*must adjust for inflation;

data phd.drg_charlson1;
      set phd.drg_charlson_drg_weight;
      discharge_year = year (discharge_date);
      if discharge_year = 2000 then cpi=1.475;
      if discharge_year = 2001 then cpi=1.384;
      if discharge_year = 2002 then cpi=1.273;
      if discharge_year = 2003 then cpi=1.186;
      if discharge_year = 2004 then cpi=1.196;
      if discharge_year = 2005 then cpi=1.064;
      if discharge_year = 2006 then cpi=1;
      charge = (total_charges * cpi);
run;

*re-visit LOS delete if pre-infection LOS is < 62 days;

data phd.drg_charlson2;
      set phd.drg_charlson1;
      if Pre_infection_LOS >62 then delete;
      run;

 *tweak the dataset - look for low cost payers;
data phd.drg_charlson3;
      set phd.drg_charlson2;
      if post_infection_los < 0 then delete;
      run;

*sample size = 948 b/c 9 people has post-infection LOS < 0;

data phd.drg_charlson4;
      set phd.drg_charlson3;
      if charge < 11896 and LOS_total>= 9 then delete;
      run;

*sample size = 930;
libname phdfinal 'C:\Documents and Settings\Suzanne Phillips\My
Documents\PhD';
run;

data phdfinal.final;
      set phd.drg_charlson4;
run;

data phdfinal.final_1;
      set phdfinal.final;
if ad_source = 1 then admission_source = 1;
if ad_source = 2 then admission_source = 1;
if ad_source = 3 then admission_source = 1;
if ad_source = 4 then admission_source = 2;
if ad_source = 5 then admission_source = 2;
if ad_source = 6 then admission_source = 2;
if ad_source = 10 then admission_source = 2;
```

```
if ad_source = 7 then admission_source = 7;
if ad_source = 8 then admission_source = 8;
if ad_source = 17 then admission_source = 3;
if ad_source = 20 then admission_source = 3;
if ad_source = 9 then admission_source = 3;
if payer= 14 then payer = 1;
      if payer= 15 then payer = 1;
      if payer= 16 then payer = 1;
      if payer= 18 then payer = 1;
      if payer= 13 then payer = 1;
      if payer= 21 then payer = 1;
      if payer=1 then payer = 1;
      if payer=2 then payer = 1;
      if payer=4 then payer = 1;
      if payer=17 then payer = 2;
      if payer=22 then payer = 3;
      if payer=20 then payer = 3;
      if payer=23 then payer = 3;
      if payer=10 then payer = 4;
      if payer=11 then payer = 4;
      if payer=6 then payer = 5;
      if payer=7 then payer = 5;
If DISCHARGE_CODE= 1 then discharge =1;
If DISCHARGE_CODE= 8 then discharge =1;
If DISCHARGE_CODE= 6 then discharge =1;
If DISCHARGE_CODE= 2 then discharge =2;
If DISCHARGE_CODE= 3 then discharge =2;
If DISCHARGE_CODE= 4 then discharge =2;
If DISCHARGE_CODE= 5 then discharge =2;
If DISCHARGE_CODE= 22 then discharge =2;
If DISCHARGE_CODE= 23 then discharge =2;
If DISCHARGE_CODE= 11 then discharge =3;
If DISCHARGE_CODE= 7 then discharge =4;
If DISCHARGE_CODE= 12 then discharge =4;
If DISCHARGE_CODE= 15 then discharge =4;
If DISCHARGE_CODE= 18 then discharge =5;
If DISCHARGE_CODE= 25 then discharge =5;
If DISCHARGE_CODE= 13 then discharge =6;
If DISCHARGE_CODE= 14 then discharge =6;
If DISCHARGE_CODE= 1 then discharge_status =1;
If DISCHARGE_CODE= 8 then discharge_status =1;
If DISCHARGE_CODE= 6 then discharge_status =1;
If DISCHARGE_CODE= 2 then discharge_status =1;
If DISCHARGE_CODE= 3 then discharge_status =1;
If DISCHARGE_CODE= 4 then discharge_status =1;
If DISCHARGE_CODE= 23 then discharge_status =1;
If DISCHARGE_CODE= 22 then discharge_status =1;
If DISCHARGE_CODE= 5 then discharge_status =1;
If DISCHARGE_CODE= 11 then discharge_status =0;
If DISCHARGE_CODE= 7 then discharge_status =1;
If DISCHARGE_CODE= 12 then discharge_status =1;
If DISCHARGE_CODE= 15 then discharge_status =1;
If DISCHARGE_CODE= 18 then discharge_status =2;
```

```
If DISCHARGE_CODE= 25 then discharge_status =2;
If DISCHARGE_CODE= 13 then discharge_status =1;
If DISCHARGE_CODE= 14 then discharge_status =1;
If discharge_status =0 then mortality = 0;
If discharge_status =1 then mortality = 1;
run;
* Admission Source 1=referral, 2=transfer, 7=ER, 8=Court/Law
enforcement, 3=NA;
* Payer 1= Insured, 2=self pay, 3=N/A, 4=Medicaid, 5=Medicare;
*Discharge home=1, transfer = 2, expired =3, other = 4, N/A = 5,
Hospice = 6;
*Discharge status alive =1, dead = 0 n/a=2;



*categorical data frequency table;

proc freq data=phdfinal.final_1;
      tables susceptibility * gender / chisq;
      run;

proc freq data=phdfinal.final_1;
      tables admission_source * susceptibility / chisq;
      run;

proc freq data=phdfinal.final_1;
      tables discharge * susceptibility  / chisq;
      run;

proc freq data=phdfinal.final_1;
      tables discharge_status * susceptibility  / chisq;
      run;

proc freq data=phdfinal.final_1;
      tables payer * susceptibility / chisq;
      run;

*continuous data analysis;
proc npar1way data=phdfinal.final_1 wilcoxon;
      class susceptibility;
      var age;
run;

proc npar1way data=phdfinal.final_1 wilcoxon;
      class susceptibility;
      var charlson;
run;

proc npar1way data=phdfinal.final_1;
      class susceptibility;
      var weights;
run;
```

```
proc npar1way data=phdfinal.final_1;
      class susceptibility;
      var pre_infection_LOS;
run;

proc npar1way data=phdfinal.final_1;
      class susceptibility;
      var post_infection_LOS;
run;

proc npar1way data=phdfinal.final_1;
      class susceptibility;
      var total_LOS;
run;

proc npar1way data=phdfinal.final_1;
      class susceptibility;
      var charge;
      run;

*repeated measures ANOVA - interested in mean charge but must take into
account suseptibility status & hospital #;

Proc sort data = phdfinal.final_1;
      by hospital susceptibility;
      run;

proc transpose data=phdfinal.final_1
                        let
                        label  = discharge_year
                        name = charge
                        out = phdfinal.final_1a;
by hospital susceptibility;
id discharge_year;
var charge;
run;

ODS graphics on;
proc glm data = phdfinal.final_1a;
      class susceptibility hospital;
      model _2000 _2001 _2002 _2003 _2004 _2005 _2006 = susceptibility
hospital;
      repeated year 7 / printe summary;
      run;

proc means data= phdfinal.final_1a;
      by susceptibility;
      var _2000 _2001 _2002 _2003 _2004 _2005 _2006;
      run;

proc plot data = phdfinal.final_1;
      by descending susceptibility;
      plot charge*discharge_year;
```

```
run;

Proc sort data = phdfinal.final_1;
      by discharge_year susceptibility;
      run;

  proc gplot data= phdfinal.final_1;
     PLOT charge*discharge_year=susceptibility ;
RUN;

symbol1 v=triangle c = r;
proc boxplot data=phdfinal.final_1;
plot charge*discharge_year /
            boxstyle=schematic
            vaxis=axis2
      cboxes   = bl
      clipfactor = 2;
insetgroup nhigh mean/ header = 'Outliers per year';
 label charge='Total Charge in dollars';
label discharge_year='Discharge Year';
run;


*propensity score - hospital size, hospital teaching status,
urban/rural status;

data phdfinal.final_1a;
      set phdfinal.final_1;
      if susceptibility = 7 then susceptibility = 1;
      if susceptibility = 8 then susceptibility = 0;
      run;


proc logistic data = phdfinal.final_1a descending;
      class bedteach urban_rural_status_ind;
      model susceptibility = bed teach urban_rural/ lackfit corrb;
      OUTPUT OUT= phdfinal.final_2 prob=prob;
      RUN;


proc univariate data=phdfinal.final_2 plot;
      var prob;
      run;

Proc sort data = phdfinal.final_2;
      by susceptibility;
      run;

symbol1 v=triangle c = r;
proc boxplot data=phdfinal.final_2;
plot prob*susceptibility /
            boxstyle=skeletal
```

```
            vaxis=axis2
      cboxes   = bl;
 label prob ='Propensity Score';
label susceptibility ='Susceptibility';
run;

proc rank data=phdfinal.final_2 groups = 5 out = phdfinal.rank;
      ranks ranks;
      var prob;
      run;

data phdfinal.final_2a;
      set phdfinal.rank;
      quintile = ranks + 1;
      if quintile = 2 then ps2 =1; else ps2 = 0;
      if quintile = 3 then ps3 =1; else ps3 = 0;
      if quintile = 4 then ps4 =1; else ps4 = 0;
      if quintile = 5 then ps5 =1; else ps5 = 0;
      run;

proc freq data=phdfinal.final_2a;
      tables quintile*susceptibility;
      run;

proc univariate data=phdfinal.final_2a plot;
      var prob quintile;
      run;

*propensity score = prob;

*Univariate analysis (obj 2)each variable by charge;
proc genmod data = phdfinal.final_2a;
      model charge = susceptibility age / dist=gamma link=log obstats;
      run;

proc genmod data = phdfinal.final_2a;
      model charge = susceptibility weights / dist=gamma link=log
obstats;
      run;

proc genmod data = phdfinal.final_2a;
      model charge = susceptibility charlson / dist=gamma link=log
obstats;
      run;

proc genmod data = phdfinal.final_2a;
      model charge = susceptibility pre_infection_LOS / dist=gamma
link=log obstats;
      run;

proc genmod data = phdfinal.final_2a;
      model charge = susceptibility mortality / dist=gamma link=log
obstats;
```

```
      run;

proc genmod data = phdfinal.final_2a;
      class gender;
      model charge = susceptibility gender / dist=gamma link=log
obstats;
      run;


*multivarable model - obj 2;
proc genmod data = phdfinal.final_2a;
      class gender;
      model charge = susceptibility gender prob weights charlson age
pre_infection_LOS mortality / dist=gamma link=log obstats;
      run;

*remove age;
proc genmod data = phdfinal.final_2;
      class gender;
      model charge = susceptibility gender prob weights charlson
pre_infection_LOS mortality / dist=gamma link=log obstats;
      run;

*test homogeneity of slope assumption;
data phdfinal.final_3;
set phdfinal.final_2a;
if GENDER = 'Female' then sex = '0';
Else sex = '1';
gender1=input(sex,comma4.);
X3 = susceptibility * gender1;
x2 = susceptibility * weights;
x1 = susceptibility * charlson;
x4 = susceptibility * pre_infection_LOS;
x5 = susceptibility * mortality;
X6 = weights*pre_infection_LOS;
X7 = gender1 * charlson;
X8 = weights * charlson;
X9 = weights*mortality;
X10 = weights* gender1;
X11 = gender1 * mortality;
X12 = gender1 * pre_infection_LOS;
X13 = charlson * pre_infection_LOS;
X14 = charlson*mortality;
X15 = pre_infection_LOS * mortality;
run;



proc genmod data = phdfinal.final_3;
      class gender;
      model charge = susceptibility gender prob weights charlson
pre_infection_LOS mortality x1 x2 x3 x4 x5
/ dist=gamma link=log type1;
```

```
      run;


*2-way interactions;
proc genmod data = phdfinal.final_3;
      model charge = susceptibility gender1 prob weights charlson
pre_infection_LOS mortality x1 x2 x6 x7 x8 x9
      x10 x11 x12 x13 x14 x15/ dist=gamma link=log type1;
      run;

*final model obj 2;

 data phdfinal.final_3a;
      set phdfinal.final_3;
      if susceptibility = 7 then susceptibility = 1;
      if susceptibility = 8 then susceptibility = 0;
      run;

proc genmod data = phdfinal.final_3a;
      model charge = susceptibility gender1 prob weights charlson
pre_infection_LOS mortality x1 x2 x6 x7
      / dist=gamma link=log covb corrb obstats type1 type3 waldci;
      output out=phdfinal.gencook resraw=resraw reschi=reschi
stdreschi=stdreschi pred=pred resdev=resdev;
      run;

proc univariate data = phdfinal.gencook;
var pred;
run;

proc sort data = phdfinal.gencook;
by susceptibility;
run;

proc univariate data = phdfinal.gencook;
by susceptibility;
var pred;
run;

*reschi are pearson residuals;

proc plot data=phdfinal.gencook;
plot charge*pred;
plot stdreschi * pred;
plot resraw * pred;
plot reschi * pred;
plot resdev * pred;
run;

*influential diagnostics ;

ods output covb=covb parameterestimates=parameterestimates;
proc genmod data=phdfinal.final_3;
```

```
      model charge = susceptibility gender1 prob weights charlson
pre_infection_LOS mortality x1 x2 x6 x7
      / dist=gamma link=log r covb corrb obstats type1 type3 waldci;
      output out=phdfinal.gencook2 resraw=resraw reschi=reschi
stdreschi=stdreschi pred=pred;
run;

proc iml;
use phdfinal.gencook2;
read all var {susceptibility gender1 prob weights charlson
pre_infection_LOS mortality x1 x2 x6 x7} into x;
read all var {charge} into y;
read all var {resraw} into resraw;
read all var {reschi} into reschi;
read all var {stdreschi} into stdreschi;

use covb;
read all var {prm1 prm2 prm3 prm4 prm5 prm6 prm7 prm8 prm9 prm10 prm11
prm12 scale} into covb;

use parameterestimates;
read all var {estimate} into estimate;

p=ncol(x);
n=nrow(x);

/*scale = estimate[p+1.]##2;*/
scale = estimate[p+2.]##2;

rr=(reschi/stdreschi)##2;
invrr = 1/rr;

stddev=sqrt(vecdiag(covb));

add=j(n,1,1);
newx=add||x||add;
/*tx = t(x);*/
tx = t(newx);

lev = J(n,1.1)- rr/scale;
D=(1/p)#(lev/(1-lev))#(stdreschi##2);
dfbeta = covb * tx * diag(invRR#resraw);
dfbetas = dfbeta#(stddev##-1);

print lev;
print D;
print dfbeta;
print dfbetas;
run;

proc univariate data=phdfinal.gencook2;
var pred;
      run;
```

```
proc univariate data=phdfinal.gencook2;
by susceptibility;
var pred;
      run;


*removed if dev residuals > 3 - 5 obs were deleted (n=925);
data phdfinal.gencook2a;
      set phdfinal.gencook;
if resdev > 3 then delete;
run;

*rerun analysis;
proc genmod data = phdfinal.gencook2a;
      model charge = susceptibility age / dist=gamma link=log obstats;
      run;

proc genmod data = phdfinal.gencook2a;
      model charge = susceptibility weights / dist=gamma link=log
obstats;
      run;

proc genmod data = phdfinal.gencook2a;
      model charge = susceptibility charlson / dist=gamma link=log
obstats;
      run;

proc genmod data = phdfinal.gencook2a;
      model charge = susceptibility pre_infection_LOS / dist=gamma
link=log obstats;
      run;

proc genmod data = phdfinal.gencook2a;
      model charge = susceptibility mortality / dist=gamma link=log
obstats;
      run;

proc genmod data = phdfinal.gencook2a;
      class gender;
      model charge = susceptibility gender / dist=gamma link=log
obstats;
      run;


*rerun multivarable model - obj 2;
proc genmod data = phdfinal.gencook2a;
      model charge = susceptibility prob weights charlson
pre_infection_LOS mortality / dist=gamma link=log obstats;
      run;

*test homogeneity of slope assumption;
data phdfinal.gencook2aa;
```

```
set phdfinal.gencook2a;
x2 = susceptibility * weights;
x1 = susceptibility * charlson;
x4 = susceptibility * pre_infection_LOS;
x5 = susceptibility * mortality;
X6 = weights*pre_infection_LOS;
X8 = weights * charlson;
X9 = weights*mortality;
X10 = weights* gender1;
X13 = charlson * pre_infection_LOS;
X14 = charlson*mortality;
X15 = pre_infection_LOS * mortality;
run;

proc genmod data = phdfinal.gencook2aa;
      class gender;
      model charge = susceptibility gender prob weights charlson
pre_infection_LOS mortality x1 x2 x4 x5
/ dist=gamma link=log type1;
      run;


*2-way interactions;
proc genmod data = phdfinal.gencook2aa;
      model charge = susceptibility gender1 prob weights charlson
pre_infection_LOS mortality x1 x6 x8 x9
      x10 x13 x14 x15/ dist=gamma link=log type1;
      run;

*keep only x1 and x6;

*final model obj 2;

 data phdfinal.gencook2aaa;
      set phdfinal.gencook2aa;
      if susceptibility = 7 then susceptibility = 1;
      if susceptibility = 8 then susceptibility = 0;
      run;

proc genmod data = phdfinal.gencook2aaa;
      model charge = susceptibility prob weights charlson
pre_infection_LOS mortality x1 x6
      / dist=gamma link=log covb corrb obstats type1 type3 waldci;
      output out=phdfinal.gencook resraw=resraw reschi=reschi
stdreschi=stdreschi pred=pred resdev=resdev;
      run;

proc univariate data = phdfinal.gencook2aaa;
var pred;
run;

proc sort data = phdfinal.gencook2aaa;;
by susceptibility;
```

```
run;

proc univariate data = phdfinal.gencook2aaa;
by susceptibility;
var pred;
run;

*rerun analysis with propensity scores as quintiles;
proc genmod data = phdfinal.gencook2aaa;
      model charge = susceptibility ps2 ps3 ps4 ps5 weights charlson
pre_infection_LOS mortality x1 x6
      / dist=gamma link=log covb corrb obstats type1 type3 waldci;
      output out=phdfinal.gencook1 resraw=resraw reschi=reschi
stdreschi=stdreschi pred=pred resdev=resdev;
      run;


*Univariate analysis (obj 3)each variable by post-infection LOS;

*propensity score = prob;

*Univariate analysis (obj 3)each variable by charge;
proc genmod data = phdfinal.gencook2aaa;
      model post_infection_LOS = susceptibility age / dist=gamma
link=log obstats;
      run;

proc genmod data = phdfinal.gencook2aaa;
      model post_infection_LOS = susceptibility weights / dist=gamma
link=log obstats;
      run;

proc genmod data = phdfinal.gencook2aaa;
      model post_infection_LOS = susceptibility charlson / dist=gamma
link=log obstats;
      run;

proc genmod data = phdfinal.gencook2aaa;
      model post_infection_LOS = susceptibility pre_infection_LOS /
dist=gamma link=log obstats;
      run;

proc genmod data = phdfinal.gencook2aaa;
      model post_infection_LOS = susceptibility mortality / dist=gamma
link=log obstats;
      run;

proc genmod data = phdfinal.gencook2aaa;
      class gender;
      model post_infection_LOS = susceptibility gender / dist=gamma
link=log obstats;
      run;
```

```
*multivarable model - obj 3;
proc genmod data = phdfinal.gencook2aaa;
      model post_infection_LOS = susceptibility prob weights charlson
age pre_infection_LOS mortality / dist=gamma link=log obstats;
      run;

*remove age;
proc genmod data = phdfinal.gencook2aaa;
            model post_infection_LOS = susceptibility prob weights
charlson pre_infection_LOS mortality / dist=gamma link=log obstats;
      run;

*test homogeneity of slope assumption;
data phdfinal.gencook2aaaa;
set phdfinal.gencook2aaa;

x2 = susceptibility * weights;
x1 = susceptibility * charlson;
x4 = susceptibility * pre_infection_LOS;
x5 = susceptibility * mortality;
X6 = weights*pre_infection_LOS;
X8 = weights * charlson;
X9 = weights*mortality;
X13 = charlson * pre_infection_LOS;
X14 = charlson*mortality;
X15 = pre_infection_LOS * mortality;
run;

proc genmod data = phdfinal.gencook2aaaa;
      model post_infection_LOS = susceptibility prob weights charlson
pre_infection_LOS mortality x4 x2 x1 x5
/ dist=gamma link=log type1;
      run;


*2-way interactions;
proc genmod data = phdfinal.gencook2aaaa;
      model post_infection_LOS = susceptibility prob weights charlson
pre_infection_LOS mortality x4 x13 x9
      x14 x8 x15/ dist=gamma link=log type1;
      run;

*final model obj 3;
proc genmod data = phdfinal.gencook2aaaa;
      model post_infection_LOS = susceptibility  prob weights charlson
pre_infection_LOS mortality x4 x13
      / dist=gamma link=log covb corrb obstats type1 type3 waldci;
      output out=phdfinal.LOSgencookaa resraw=resraw reschi=reschi
stdreschi=stdreschi pred=pred resdev=resdev;
      run;

proc univariate data = phdfinal.LOSgencookaa;
```

```
var pred2;
run;

proc sort data = phdfinal.LOSgencook;
by susceptibility;
run;

proc univariate data = phdfinal.LOSgencook;
by susceptibility;
var pred2;
run;


*path analysis;
proc calis data=phdfinal.gencook2aaaac corr residual pall toteff;
var susceptibility pre_infection_LOS Post_infection_LOS charge;
lineqs
Post_infection_LOS = b32 pre_infection_LOS + b31 susceptibility + e3,
charge = b43 Post_infection_LOS + b42  pre_infection_LOS + e4;
std
susceptibility = var_susc,
pre_infection_LOS = var_preLOS,
e3 = var_e3,
e4 = var_e4;
cov
susceptibility pre_infection_LOS = c_spre;
run;
```

VITA

Name: Suzanne Toussaint Phillips

Birth date/place: August 4, 1980, Silver Spring, Maryland

Citizenship: United States of America

Education:

Bachelor's degree, Chemistry
Washington and Lee University, 2001

Masters of Public Health
Virginia Commonwealth University, 2008
Magna Cum Laude

Doctor of Pharmacy
Virginia Commonwealth University, 2008
Magna Cum Laude

Publications:

Phillips S, MacDougall C, Holdford DA. Analysis of empiric antimicrobial strategies for cellulitis in the era of methicillin-resistant *Staphylococcus aureus*. Ann Pharmacother. 2007 Jan;41(1):13-20

Pharmacoepidemiology of Colistin Use in US Teaching Hospitals
Poster presented at Society of Healthcare Epidemiologists of America conference 2005

Teaching experience:
Virginia Commonwealth University, Department of Pharmacy
PHAR 745: Drug Literature Evaluation, Fall 2007, Fall 2008
PHAR 744: Integrated Therapeutics, Spring 2007, Spring 2008

Clarke County High School, Berryville, Virginia
Chemistry
Honors Chemistry

Clifton Middle School, Covington, Virginia
Sixth grade math
Seventh grade math
Eighth grade math