Summer 2015

Clostridium difficile infection as a novel marker for hospital quality, efficiency and other factors associated with prolonged inpatient length of stay

Aaron Christopher Miller
University of Iowa

CLOSTRIDIUM DIFFICILE INFECTION AS A NOVEL MARKER FOR HOSPITAL QUALITY, EFFICIENCY AND OTHER FACTORS ASSOCIATED WITH PROLONGED INPATIENT LENGTH OF STAY

by

Aaron Christopher Miller

A thesis submitted in partial fulfillment of the
requirements for the Doctor of Philosophy
degree in Pharmacy
in the Graduate College of
The University of Iowa

August 2015

Thesis Supervisor: Assistant Professor Linnea Polgreen

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH.D. THESIS

This is to certify that the Ph.D. thesis of

Aaron Christopher Miller

has been approved by the Examining Committee for the thesis requirement for the Doctor of Philosophy degree in Pharmacy at the August 2015 graduation.

Thesis Committee: _____
Linnea Polgreen, Thesis Supervisor

_____
Barry Carter

_____
Joseph Cavanaugh

_____
Philip Polgreen

_____
Mary Schroeder

_____
Alberto Segre

ABSTRACT

Excess inpatient length of stay (LOS) varies between hospitals and is burdensome to patients and the overall healthcare system. Variation in LOS has often been associated with hospital-level factors, such as hospital efficiency and quality. *Clostridium difficile* infection (CDI) is an increasingly common hospital-acquired (HA) infection. This thesis explores the connection between hospital incidence of CDI and excess LOS in patients without a CDI. It is hypothesized that HA-CDI incidence may act as a "proxy variable" to capture unobserved hospital characteristics, such as hospital quality or efficiency, associated with prolonged LOS. In addition, hospitals with longer LOS may tend to observe more HA-CDI cases prior to discharge. This thesis analyzes the ability of CDI incidence to capture excess LOS variation across hospitals, while controlling for CDI cases that occur after discharge.

We use data on hospital inpatient visits, spanning the years 2005-2011, from three data sources distributed by the Healthcare Cost and Utilization Project: the Nationwide Inpatient Sample (NIS), and the State Inpatient Databases (SID) for California and New York. The NIS provides discharge records from a nationwide sampling of hospitals in a given year. The SIDs are longitudinal populations of inpatient records in each state, and patient records can be linked across stays. We compute a variety of different measures of hospital CDI incidence and identify HA-CDI cases that occur after a patient is discharged.

Various multivariable regression models are analyzed to predict LOS at an

individual patient level. A generalized linear modeling approach is used, and different distributions and link functions are compared using the Akaike information criterion. A multilevel modeling approach is also used to estimate the amount of between-hospital variation in LOS that can be explained by HA-CDI incidence.

We find CDI incidence to be a strong predictive factor for explaining a patient's LOS and is one of the strongest predictive variables we identified. Moreover, CDI incidence appears to primarily capture between-hospital variation in excess LOS. Although we find evidence that present-on-admission indicators may underreport cases of HA CDI, our findings suggest the connection between CDI incidence and excess LOS is driven primarily by CDI cases that are HA. In addition, when we account for HA-CDI cases that occur post-discharge, the relationship between CDI incidence and LOS appears even stronger. Our results suggest that CDI incidence may be a powerful tool for making comparisons of excess LOS across hospitals.

PUBLIC ABSTRACT

Excess inpatient length of stay (LOS) is burdensome to patients and the overall healthcare system. Inpatient LOS varies dramatically among patients and hospitals. Excess LOS has often been associated with hospital-level factors, such as hospital efficiency and quality. *Clostridium difficile* infection (CDI) is an increasingly common healthcare-associated infection, and can be community or hospital acquired (HA). We hypothesize that a hospital's incidence of HA CDIs may serve as a "proxy variable" for unmeasured hospital characteristics related to excess LOS.

This thesis evaluates the connection between CDI incidence and LOS in patients without CDIs. We use data on hospital inpatient visits, spanning the years 2005-2011, from three data sources distributed by the Healthcare Cost and Utilization Project: the Nationwide Inpatient Sample, and the State Inpatient Databases for California and New York. Various statistical models are used to explain LOS variation while controlling for patient and hospital characteristics. We compare different measures of CDI incidence and identify HA-CDI cases that occur after a patient is discharged.

We find CDI incidence to be a strong predictive factor for explaining a patient's LOS, and CDI incidence works well to explain variation in the average LOS among hospitals. Our findings also suggest this connection is driven primarily by CDI cases that are HA. In addition, when we account for HA-CDI cases that occur post-discharge, the relationship between CDI incidence and LOS appears even

stronger. Our results suggest that CDI incidence may be a powerful tool for making comparisons of excess LOS across hospitals.

TABLE OF CONTENTS

LIST OF TABLES viii

CHAPTER

1 INTRODUCTION 1

2 LITERATURE REVIEW 13

2.1 Inpatient Length of Stay 14
2.1.1 Trends in LOS 18
2.1.2 LOS as an Outcome Measure: Excess LOS, Efficiency and Quality 21
2.2 *Clostridium difficile* Infection 24
2.2.1 General Overview: Epidemiology, Measurement, Risk factors, and Outcomes 24
2.2.2 Classification of CDI By Source and Occurrence of Symptoms 28
2.2.3 Hospital Reporting and Comparison of CDI Incidence . . 32
2.3 CDI Incidence and LOS in Uninfected Individuals 34
2.3.1 Hospital Efficiency and Quality 35
2.3.2 Other Unmeasured Factors and The Post-Discharge Effect 38

3 METHODS 41

3.1 Design 41
3.2 Data Source 41
3.3 Study Sample 43
3.4 Variables 45
3.4.1 Dependent Variables 45
3.4.2 Primary Independent Variable 45
3.4.3 Secondary Independent Variables 48
3.5 Model 55
3.5.1 Analyzing HA CDI from Order of Diagnosis in The NIS . 59
3.5.2 Identifying and Analyzing Post-Discharge CDI 61
3.5.3 Multilevel Hospital Fixed Effects Model 64
3.6 Statistical Analysis 66

4 RESULTS 68

4.1 Nationwide Inpatient Sample 68

4.2 State Inpatient Sample: California 80
4.3 State Inpatient Sample: New York 93

5 DISCUSSION 107

5.1 Limitations 115
5.2 Areas for Future Research 119
5.3 Conclusions 121

APPENDIX 125

A LIST OF ABBREVIATIONS 125

B RESULTS USING QUARTERLY RATES 126

REFERENCES 130

LIST OF TABLES

Table

2.1 Inpatient LOS Across Time 19

3.1 Patient-Level Explanatory Variables 52

3.2 Hospital-Level Explanatory Variables 55

4.1 NIS - CDI Case Counts by CDI Type 69

4.2 NIS - Yearly CDI Incidence Summary Statistics 71

4.3 NIS - Mean LOS and CDI Incidence Correlation 72

4.4 NIS - OLS Results By Year 73

4.5 NIS - GLM (Gamma) Results By Year 74

4.6 NIS - Using Diagnosis Order to Control for CA CDI 76

4.7 NIS - Results of Multilevel Model 79

4.8 California SID - CDI Case Counts by CDI Type 81

4.9 California SID - Yearly CDI Incidence Summary Statistics 83

4.10 California SID - Mean LOS and CDI Incidence Correlation 84

4.11 California SID - Regression Results 86

4.12 California SID - Regression Results Split By Policy Change Point 88

4.13 California SID - Results From Second-Level Multilevel Model 90

4.14 California SID - Controlling For Post-Discharge Effect 92

4.15 New York SID - CDI Case Counts by CDI Type 95

4.16 New York SID - Yearly CDI Incidence Summary Statistics 97

4.17 New York SID - Mean LOS and CDI Incidence Correlation 98

4.18 New York SID - Regression Results 100

4.19 New York SID - Regression Results Split By Policy Change Point 102

4.20 New York SID - Results From Second-Level Multilevel Model 104

4.21 New York SID - Controlling For Post-Discharge Effect 106

A.1 Table of Abbreviations 125

B.1 NIS OLS Quarterly Results By Year 127

B.2 NIS GLM (Gamma) Quarterly Results By Year 128

B.3 SID Quarterly Regression Results in California 129

B.4 SID Quarterly Regression Results in New York 129

CHAPTER 1
INTRODUCTION

Inpatient length of stay (LOS) is one of the main drivers of hospital costs and, as a result, a significant contributor to overall health expenditure [124, 125]. Efforts to reduce healthcare spending have often targeted LOS as an area where costs may be contained [142]. Consequently, reimbursement strategies on the part of Medicare, as well as managed care organizations, have been responsible for decreasing tends in LOS and shifts in care from inpatient to ambulatory settings [72, 136, 108, 52]. However, increased LOS is not only costly from a reimbursement perspective, but also because it may increase a patient's risk of exposure to harmful events such as hospital-acquired infections (HAIs) or other hospital-associated (HA) adverse events [65]. Hospital administrators must also consider the 'opportunity costs' associated with a patient's LOS, insofar as each day a patient remains hospitalized represents a potential missed opportunity for the treatment of another patient [62]. Because of these burdens, LOS has become one of the most commonly used outcome measures in clinical research. LOS is frequently used to measure healthcare costs and other burdens attributable to specific diseases and healthcare events. Studies of healthcare quality have also used LOS as an indirect outcome measure, and prolonged LOS has been shown to be associated with other markers of care quality [168, 143, 146]. Finally, LOS is one of the primary measures of hospital output in studies of healthcare efficiency [68, 99].

LOS varies dramatically across both patients and hospitals [19], and such

variations have often been considered a marker for unnecessary or excess healthcare use [99, 153]. However, ordinary variations in LOS, either across patients, providers or hospitals, are expected as patient disease states and treatment resources also vary. First of all, LOS can be expected to vary across different categories of diagnoses as well as procedures [25, 27, 57]. Patient characteristics such as age, disease severity or comorbidity have also been associated with variations in LOS [126, 92, 70]. Hospital and provider-level characteristics, such as bed size and teaching status, have also been found to be related to LOS [99, 166, 57]. Thus, when using LOS as a measure of the healthcare burden associated with specific conditions, it is important to properly control for all of these sources of LOS variation. Failing to control for any of the factors associated with LOS may lead to inaccurate estimation of the LOS attributable to specific events or conditions. In fact, studies have shown that when estimating the excess LOS attributable to HAIs, one can obtain drastically different estimates when different sets of control variables are used [157, 106]. Given the abundance of research that has used LOS as an outcome measure, it is important that researchers have an understanding of all the factors that contribute to variations in LOS.

LOS has been frequently used as an outcome measure in the study of HA adverse events (AEs). AEs occurring in hospitalized patients have generally been associated with increased LOS. Examples include hospital falls [47, 163], adverse drug events [25], and surgical complications [69, 79]. Patient safety indicators (PSIs), which are specific indicators of patient safety and hospital quality developed by the Agency for Healthcare Research and Quality (AHRQ), have also been analyzed using

LOS [168, 133]. HAIs are, perhaps, the most frequently studied set of hospital related AEs. Some of the various HAIs that have been studied using LOS include surgical-site infections [161, 75], bloodstream infections [121, 123] and methicillin-resistant *Staphylococcus aureus* (MRSA) [96, 29].

Clostridium difficile infection (CDI) is a common HAI that has often been studied using LOS. CDIs are the leading cause of nosocomial diarrhea and colitis, and have now become one of the most common HAIs [5, 84, 45]. In general, CDIs can be either community associated (CA CDI) or healthcare/hospital associated (HA CDI), though HA CDI's are most common [148]. Much effort has been expended to classify CDIs according to these specific sources [83, 87, 46]. Patients who acquire CDIs in hospital settings tend to experience much longer LOS than uninfected patients, although estimates of the increased LOS attributable to CDIs have varied quite dramatically [45, 58]. Part of the variation in the estimates of LOS attributable to CDI likely stems from the fact that such estimates have been obtained using a variety of study designs and statistical methodologies [109]. Moreover, there is an interdependent connection between CDI and LOS, since patients that experience longer stays are at greater risk of acquiring a CDI. Certain methodological approaches may be better suited to handle this type of endogenous relationship [110, 62]. Consequently, increasing awareness has been drawn to the complex relationship between CDI and LOS in infected patients.

While previous research has analyzed the excess LOS that can be attributed to HAIs such as CDI, the occurrence of these conditions at a hospital level has rarely

been studied as a factor associated with LOS in uninfected patients. In the case of CDI, it is conceivable that a hospital's CDI incidence rate may act as a "proxy variable" capturing unmeasured factors associated with increased LOS in all hospitalized patients, even those who do not become infected. It is possible that hospitals with a higher incidence of CDIs may also have factors that lead to increased LOS in all patients, or vice versa. Two basic theoretical links may establish a connection between CDI incidence and LOS in uninfected patients.

First, a number of hospital-level factors that are associated with increased LOS may directly or indirectly lead to a greater number of HA CDI cases. Many of the frequently used measures of hospital quality have been associated with increased LOS [168, 160]. Some dimensions of quality may also be associated with the occurrence of HA CDI. For example, hospital cleaning and sanitation practices, antimicrobial stewardship and hand-hygiene are all factors that have been cited for reducing the spread and occurrence of CDI in hospital settings [39]. It is plausible that hospitals that are of better quality, and have shorter LOS, may also have better infectious disease practices leading to lower rates of HA CDI. Additionally, other hospital characteristics, such as the availability of private rooms to treat infected patients, which may be difficult to assess from commonly available hospital measures (e.g., teaching status or bed-size) may also be directly related to LOS and CDI incidence [39, 31].

Factors that are associated with increased LOS may also be indirectly associated with CDI, since increased LOS is itself a risk factor for HA CDI. LOS has often been viewed as a measure of hospital efficiency, and hospitals that have longer LOS

may be deemed to be inefficient [99, 9, 17]. If such hospitals have discharge or care-coordination inefficiencies that result in longer LOS, they may also have a greater occurrence of HA CDIs as a result. Thus, hospital efficiency, or quality, and other factors not directly associated with CDI, may be indirectly related to CDI incidence because of prolonged patient exposure.

While this first theoretical link between CDI incidence and LOS is driven by an increased occurrence of HA CDI cases, the second theoretical connection is associated with an increased number of CDI cases that are captured by hospital discharge records. The patient records at hospitals with a higher average LOS may tend to capture a greater number of HA CDI cases, insofar as HA CDI cases may be more likely to be observed prior to discharge at such hospitals. A growing body of literature suggests a significant number of HA CDI cases do not occur until after a patient has been discharged from a health care setting [87, 23, 40, 35]. Studies have also suggested that the number of post-discharge CDI cases may be increasing [113, 37]. Therefore, hospitals that have longer average LOS, either because of quality, efficiency or other factors, may appear to have a greater CDI incidence simply because their discharge records may be more likely to capture HA CDIs that occur later in patient stays. Evidence for this effect has been provided by one study that found post-discharge cases of HA CDI to be more common at hospitals having longer average LOS [113].

For both of these theoretical connections, CDI incidence acts as a proxy variable for unmeasured hospital factors that are associated with increased LOS. Thus,

if CDI incidence can be shown to be highly correlated with a patient's LOS, after accounting for other patient and provider characteristics, it may serve as a potential confounding factor that should be controlled for when making LOS comparisons across hospitals. Future studies that use LOS as an outcome measure will need to take steps to control for CDI incidence or the factors for which it serves as a proxy. Moreover, if CDI incidence can be shown to account for LOS variations across hospitals, then hospital CDI incidence represents not only a useful measure for infectious disease surveillance but also for hospital-resource utilization.

Organizations such as the Centers for Medicare and Medicaid Services (CMS) and the Centers for Disease Control and Prevention (CDC) have already required or called for the reporting and surveillance of the rates of HAIs across hospitals [150, 113]. Consequently, in many states, hospitals are now required by law to report their incidence of HAIs, including CDI [2, 129, 66]. At least 20 states in the US now have laws mandating hospitals to report cases of HA CDI [129]. A motivation of such efforts has been to measure and compare healthcare quality and disease presence across hospitals. Studies have already begun using HA CDI incidence, as well as incidence rates for other HAIs, to create rankings across hospitals [30, 151]. If CDI incidence also serves as a marker for excess LOS and resource use between hospitals, this information may be useful for policy makers wishing to reduce overall healthcare expenditure. Similarly, the reporting of CDI incidence may be helpful for patients wishing to avoid unnecessary LOS and hospital costs.

A preliminary study was conducted to analyze the link between LOS and CDI-

incidence; this preliminary research serves as the impetus for this thesis [107]. This previous work was carried out using data from the Healthcare Cost and Utilization Project (HCUP) Nationwide Inpatient Sample (NIS), for the years 2009-2011. In this study, LOS was analyzed at both an individual patient level and an average hospital level, using only patients that did not have a CDI. We find that CDI incidence is strongly associated with increased LOS even after controlling for a large number of patient and hospital characteristics. An increase in a hospital's CDI incidence by one percentage point was found to be associated with an increase in a patient's LOS between 4.37% and 7.47%. Moreover, CDI incidence was not just found to be a significant predictor of inpatient LOS, but was also reported to be one of the "strongest" explanatory variables out of a set of over 400 different variables included. Of all the hospital-level variables (e.g., bed-size or teaching status) that this study analyzed, CDI incidence was found to be the strongest hospital-related factor for explaining variation in inpatient LOS among hospitals.

While this preliminary study has suggested a link between CDI incidence and LOS, and that CDI incidence may be one of the strongest explanatory variables for LOS, there are two important issues this initial research was unable to address. First, the preliminary study was unable to distinguish between CDI cases that were HA versus those that were CA. Many of the theoretical connections between CDI incidence and LOS, mentioned above, are suggested to occur via HA CDI. Administrative databases that allow for LOS comparisons among hospitals, such as the NIS used in the preliminary study, often do not contain information on HA status. Thus, the

preliminary research was conducted using all CDI cases to calculate a hospital's CDI incidence, which has been previously validated as means of assessing overall CDI burden within a hospital [140, 41, 38]. As a sensitivity analysis, the preliminary study also used secondary diagnostic codes (i.e. CDI was not the primary reason for admission) to suggest that the effect was being driven by HA CDI cases. However, CDI indicated by a secondary diagnosis has been shown to be a poor marker for HA CDI [38]. If, as has been theorized, HA CDI is the primary factor linking CDI incidence to LOS, then CDI incidence should be calculated and analyzed using only these cases. Therefore, if future studies are to use CDI incidence as a marker for unmeasured confounding associated with LOS, research is first needed to determine which CDI cases are primarily driving this effect.

Second, the preliminary study was unable to account for any CDI cases that may have occurred post discharge. The other theoretical link connecting CDI incidence to LOS is that hospitals with longer LOS may simply capture more CDI cases that would have otherwise occurred post-discharge. It is important to consider how much of the connection between CDI incidence and LOS is driven by this effect. Understanding the magnitude of this effect is necessary for both future researchers and policy makers wishing to use CDI incidence to make LOS comparisons between hospitals. For example, if the post-discharge effect is substantial, it is imperative to include post-discharge CDI cases in the hospital's rate. Failure to capture HA CDI cases that occur post discharge, when calculating a hospital's HA CDI rate, may lead to a type of attenuation bias when using this rate as an explanatory variable. More-

over, changes have been occurring over time that may alter the connection between CDI incidence and LOS. The number of post-discharge CDI cases has been increasing, and it has been suggested that this may be due in part to pressure on hospitals to reduce LOS [113]. In addition, LOS has been decreasing while the number of CDI cases has been increasing over time [3, 148, 37]. Given these trends, the connection between CDI incidence and LOS that is due to post-discharge CDI cases may also be changing across time. If a significant portion of the connection between CDI incidence and LOS is driven by the effect of post-discharge CDI, these changes may alter the ability of CDI incidence to be used a proxy measure for LOS.

Given the potential for CDI incidence to serve as a proxy variable for unmeasured hospital factors associated with LOS, and the strong connection suggested by the preliminary work, this thesis is intended to extend the preliminary research while addressing its main limitations. This thesis uses inpatient records across hospitals, states and time to study the connection between HA CDI incidence and increased LOS in patients who do not experience a CDI. In doing so, this study extends the previous analysis in the following dimensions. First, a wider range of years is studied to analyze trends in the connection between CDI incidence and LOS. Second, alternative data sources and statistical techniques are used to estimate the effect of HA CDI incidence as opposed to overall CDI incidence in a hospital. Finally, data are used where discharge records may be linked for patients across time in order to help control for, and estimate the effect of, post-discharge CDI cases. This study was designed to address the following two objectives:

- **Objective 1:** *Estimate the effect of HA CDI incidence on inpatient LOS across hospitals and analyze if, and how, this effect has changed across time.*
- **Objective 2:** *Determine the extent to which the effect described by Objective 1 is attributable to post-discharge cases of HA CDI.*

In order to meet these objectives this study relies on two main sources of inpatient data, and analyses were conducted separately for each. First, the HCUP NIS, used in the preliminary study, is expanded to cover the seven-year period from 2005-2011. Second, HCUP State Inpatient Databases (SID) from the states of California and New York, also covering the years 2005-2011, were analyzed. While the NIS is a sample of selected hospitals across the entire US, the SID represents a complete set of all inpatient records within a state. Each of these data sources has strengths and weaknesses when it comes to studying LOS and CDI incidence. The NIS contains a wider range of hospital- and patient-level variables that may be related to LOS but lacks indicators for whether a diagnosis is present on admission (POA) or HA. Additionally, hospitals and patient records are unable to be linked across time in the NIS. On the other hand, patient records can be linked in the New York and California SID, and diagnoses are assigned POA indicators. However, the SID also lacks many of the variables that were found to be associated with LOS in the preliminary study. By conducting two separate analyses using these different data sources, this study provides a much better understanding of the connection between CDI incidence and LOS, and it overcomes some of the limitations associated with a single set of data.

This study found consistent evidence of a strong relationship between CDI incidence and LOS that reinforces and expands upon the findings of our preliminary work. This connection was also observed across the entire study period and in each of the data sources that were analyzed. The results of this thesis provide strong evidence that the connection between CDI incidence and LOS is driven by HA cases of CDI. Moreover, this relationship does not appear to be driven by the effect of CDI cases occurring post discharge. If anything, controlling for post discharge CDI made the relationship between CDI incidence and LOS appear even stronger. This research also uncovered interesting secondary findings, including evidence that hospital reporting of HA CDI, using POA indicators, may be dramatically underutilized and is heavily influenced by reporting mandates and other differences across states. Moreover, this thesis provides foundational motivation for future studies to better understand the specific factors driving the relationship between HA CDI incidence and LOS, along with the hospital reporting mandates and various CDI measurements that can allow one to capture this relationship.

The remainder of this thesis is organized in the following manner: Chapter 2 reviews the pertinent body of previous literature. This review summarizes the literature relevant to inpatient LOS, CDI, and the proposed link between CDI incidence and LOS. Chapter 3 outlines the methodological approach that was used to study the connection between inpatient LOS and hospital CDI incidence; this includes the data sources, variables, theoretical models and statistical methods that were employed. Chapter 4 summarizes the results that were obtained, which are broken into results

obtained in the NIS, the SID for California and the SID for New York. Finally, chapter 5 provides a discussion of the findings, along with a summary of this study's major limitations, areas for future work and overall conclusions.

CHAPTER 2
LITERATURE REVIEW

As described in the introduction, a large body of literature exists that has either studied LOS directly or has used LOS as an outcome when analyzing other healthcare-related events. However, when analyzing LOS associated with such events it is important to control for all confounding factors associated with variations in LOS. One HA event that has commonly been studied using LOS is CDI; LOS has been studied as both an outcome and a risk factor for HA CDIs. Such studies have focused on the connection between LOS and CDI only in infected patients. However, on a theoretical level, it is that conceivable that CDI incidence may capture factors associated with increased LOS, even in uninfected patients. Although literature in this regard is limited, some studies provide evidence that such a link may exist. Below I review some of the previous literature that is relevant to the study of CDI incidence as a proxy measure for prolonged inpatient LOS. I begin by providing a general discussion of inpatient LOS, some of its drivers, and why it has become a commonly used outcome measure. I then provide a description of CDI, its epidemiology, risk factors, and outcomes. I also highlight some of the intricacies of identifying, measuring and classifying cases of CDI. Finally, I provide a discussion of the factors that may be suggestive of a theoretical link between the occurrence of CDI at a hospital level and the LOS in patients who do not experience a CDI.

2.1 Inpatient Length of Stay

Inpatient LOS has received a great deal of attention and is one of the most commonly used outcome measures in clinical research. The primary reason for this attention to LOS is the connection with cost. Inpatient LOS is generally treated as a measure for hospital costs and resource utilization. Publications of hospital costs and overall healthcare expenditure often report dollar figures associated with spending a day in the hospital. For example, the Kaiser Family Foundation, which publicly provides many types of healthcare statistics, often reports hospital expenses per inpatient day as one measure for comparing hospital costs across states and time [56]. They report that in 2012, the average hospital expense in the US was around $2,090 per inpatient day. While these values reflect one dimension of hospital-related costs, they may not perfectly reflect the true costs or burdens associated with inpatient LOS.

As a starting point, many hospital costs are directly associated with a patient's LOS. The resources required to house and care for patients on a daily basis are directly tied to a patient's LOS. Such costs may include the staffing of nurses and physicians, the cleaning and maintenance of rooms, and any routine procedures and testing that are conducted on a daily basis. These direct LOS costs are likely to depend on where a patient is located in the hospital and the conditions being treated. For example, Fine et al. (2000) find that room costs make up 59% of the median daily cost for patients admitted with community-acquired pneumonia. Similarly, Plowman et al. (2001) find the majority of costs associated with HAIs to be driven by factors directly

tied to LOS, such as nursing care or medical time. On the other hand, direct LOS costs have also been reported to be much greater for certain patients. Coello et al. (1993) report that over 90% of the increased costs in surgical patients that have HAIs are directly associated with increased LOS.

In addition to these direct costs, LOS is also highly correlated with hospital costs in general. Patients who have greater disease severity [88] or have greater comorbidity [63] are also likely to have longer LOS and greater overall costs. LOS and overall hospital costs may also be correlated across various diagnostic related groups (DRGs) [25]. Even at an average hospital level, those hospitals that use a greater number of overall resources have been shown to have a higher average LOS [114]. As a result of the correlation between LOS and overall costs, a number of studies have used LOS as a stand-alone surrogate measure for a patient's resource consumption and cost [20, 61, 7, 127, 12]. However, because LOS is correlated with overall hospital costs, estimating the relationship between overall costs and those directly associated with LOS can be challenging. Polverejan et al. (2003) use a two-model approach to estimate the mean hospital cost associated with LOS. The authors highlight the difficulty of separating overall costs from those specifically associated with LOS. Perhaps as a consequence, it has been fairly common for studies to use both LOS and total cost simultaneously, as separate outcome measures, rather than estimating the interdependence between LOS and overall costs [10, 24, 115, 28, 32].

Costs associated with LOS are also highly dependent on which days in a patient's stay are being observed. Several studies have reported that daily costs are

much greater at the beginning of a patient's stay, where resource use may be more intensive, than at the end of the stay, where room and board costs become most prevalent. Fine et al. (2000) find that, while daily room costs were fairly constant across a patient's stay, daily non-room costs were significantly concentrated at the beginning of the stay. They report that the daily non-room cost on the first day of a stay was 282% greater than the average daily non-room cost over the entire stay. Similarly, Taheri et al. (2000) found that the majority of daily costs were concentrated in the first 3 days, with the last full day accounting for around 2.4% of a patient's mean total cost. Polverejan et al. (2003) also find that average patient charges increase less rapidly across patients with longer total LOS. These results imply that reducing a patient's overall LOS may not have a proportional effect on costs.

While the costs associated with LOS are often framed in terms of resource consumption, there are many indirect costs associated with LOS. Increased LOS may also be associated with increased potential exposure to harmful events, which can lead to additional resource costs. Increased LOS is an established risk factor for many HA adverse events. HAIs such as CDI, MRSA, and Vancomycin-resistant Enerococcus (VRE) have all been associated with increased LOS [13, 60, 155]. Hauck & Zhao (2011), find that a patient's risk of having an adverse drug reaction, an infection, or an ulcer increase by 0.5%, 1.6%, and 0.5%, respectively, for each day a patient stays in the hospital. Thus, increased LOS may expose a patient to additional harm that may further burden their level of health and lead to additional costs.

However, if a shorter LOS is due to a patient being discharged too early, then

the indirect costs associated with LOS might be inversely related to LOS. When patients are discharged too soon, many of the healthcare responsibilities that would otherwise be undertaken by providers in the inpatient setting may be shifted to individuals outside the hospital. Jönsson & Lindgren (1980) note that a patient's family members and friends often assume a role of helping to care for a patient when they are discharged. Such costs are often not considered from a reimbursement perspective but are seen by patients and are indirectly tied to LOS. Similarly, a study by Yaldo et al. (2001) surveyed physicians to analyze factors associated with the decision to discharge patients infected with MRSA. This study found that physicians considered a patient's level of social support, along with the availability of oral or outpatient therapy when deciding whether to discharge a patient. These findings suggest, at least from the perspective of physicians, that self-care can be a substitute for LOS. Therefore, the timing of a patient's discharge and, consequently overall LOS, may play a role in who bears the costs associated with LOS.

Not only is LOS costly from a patient or reimbursement perspective, it can also be costly from a hospital and administrator perspective. A growing body of literature has analyzed LOS from a hospital's perspective. Graves et al. (2010) note that one of the potential benefits of reducing the LOS associated with HAIs may come from freeing up bed-days that can be used to treat other patients. Thus, from an economic perspective, increased LOS represents a type of opportunity cost to hospitals in the form of lost patient volume, with each additional day a patient stays representing one less bed-day available to treat other patients. However, Graves et al.

(2010) also note that, from an accounting perspective, many hospital costs are fixed and, in the short-run, if demand for bed-days is low, decreased LOS may actually be more costly. Roberts et al. (1999) make a similar argument, noting that around 84% of a hospital's budget is associated with fixed costs. The authors suggest that efforts to shift inpatient stays to ambulatory or observational settings, which reduce inpatient LOS, might actually be costly to hospitals where fixed costs may not be easily reduced. Therefore, from a hospital perspective, the connection between LOS and costs may be much less clear than from a reimbursement standpoint.

2.1.1 Trends in LOS

Although some costs may be inversely related to LOS, increased LOS has generally been treated as a factor associated with increased hospital costs. Consequently, many efforts to reduce rising healthcare spending have targeted unnecessary or extended LOS as a potential area where costs may be contained. Many of these cost containment efforts took root in the early 1980s and, since then, the number of inpatient days and average inpatient LOS have been trending downwards. Table 2.1 summarizes average inpatient LOS, total inpatient days and the total number of discharges from 1980 to 2010 [3]. From 1980 to 2010, average, age-adjusted, inpatient LOS has decreased from 7.5 days to 4.8 days, while total inpatient days decreased from 13,027 days to 5,369. While LOS and inpatient stay declined fairly rapidly throughout the 1980s and 1990s, the declines appear to have leveled off in the early 2000s. These decreases were driven not only by a decline in the duration of inpatient

Table 2.1: Inpatient LOS Across Time

Year	Average LOS	Inpatient Days	Discharges
1980	7.5	13,027.0	1,744.5
1985	6.6	10,017.9	1,522.3
1990	6.5	8,189.3	1,252.4
1995	5.4	6,386.2	1,180.2
2000	4.9	5,576.8	1,132.8
2005	4.8	5,541.7	1,162.4
2010	4.8	5,369.2	1,125.1

Source: CDC National Center for Health Statistics [3]

stays, but also by a shift of inpatient care to outpatient settings [72, 136]. For example, Schwartz and Mendelson (1991) estimated that between 1981 and 1988 the number of inpatient days decreased by 28.1%, after accounting for the existing trend of growth in inpatient days. They found that 4.6 percentage points of this reduction was due to a decline in average LOS while 23.6 percentage points were due to reduced inpatient admissions. The substitution of outpatient for inpatient care is also reflected in in Table 2.1 by the declining trend in total discharges.

Two primary factors have been cited for contributing to the deceasing trend in inpatient stays and the shift to outpatient care. First, beginning in 1983, Medicare introduced the Prospective Payment System (PPS), which fundamentally changed the way hospitals were reimbursed for services provided to Medicare patients. Under the PPS, hospitals were no longer reimbursed for the actual costs associated with a patient's care, but instead were reimbursed a fixed rate based on a patient's specific diagnosis-related group (DRG). The PPS was motivated, in part, because of the belief that it would incentivize hospitals to deliver more efficient care by eliminating any

unnecessary, but costly, care practices (e.g., excess LOS) [134]. However, a secondary result of the PPS was that hospitals had a financial incentive to diversify into outpatient care facilities, which were exempt from the PPS system, and to shift patients into such facilities [136, 52]. Second, the growth of managed care further shifted patient care away from outpatient settings and created financial incentives to lower inpatient LOS. Managed care plans have increasingly contracted with non-hospital providers and directed patients to such outpatient settings. Moreover, managed care has often relied on the same type of PPS established by Medicare [136, 14, 108].

While the PPS and managed care have been cited as the primary drivers of the diminished trend in LOS, a number of advances in both treatment and clinical knowledge have also made this trend possible. For example, studies have been conducted to show that outpatient settings, such as emergency diagnostic and treatment units or emergency department observation units, can be as, or more, effective than inpatient care for treating certain types of conditions [138, 100]. Similarly, improvements in disease screening, such as accelerated diagnostic protocols, have been shown to reduce the number of patients needing to be admitted to inpatient care [152, 135]. A variety of improvements in inpatient care itself, such as clinical care pathways or early rehabilitation, have been shown to lessen the total amount of time that inpatients may be required to stay [147, 111]. In addition, the rise of hospitalists, within inpatient care settings, has been credited with improved efficiency and reductions in LOS [158].

2.1.2 LOS as an Outcome Measure: Excess LOS, Efficiency and Quality

Given the connection between LOS and costs, and the fact that LOS has been shown to be somewhat malleable, LOS has become one of the most commonly used measures of healthcare burden. In fact, it is fairly common to see total costs, length of stay and mortality as a 'generic' set of outcomes employed in clinical research [88, 29, 123, 133, 25]. Two broad categories of research have utilized LOS as an outcome measure. First, studies have looked at the LOS that is attributable to patient-level characteristics, such as disease or treatment characteristics. Second, studies have analyzed the role that hospital- and provider-level characteristics have on LOS.

Studies where LOS has been used as an outcome have frequently analyzed hospital AEs. HAIs are, perhaps, the most common set of AEs studied using LOS as an outcome measure. Some of these HAIs include surgical-site infections [161, 75], bloodstream infections [121, 123], MRSA [96, 29] and CDI [88, 59, 36]. Studies have also estimated the effect of aggregated HAIs on LOS [63]. Other hospital AEs that have been studied using LOS as an outcome include hospital falls [47, 163], adverse drug events [25], and surgical complications [69, 79]. Increased LOS has also been analyzed across aggregated AEs [69].

In addition to hospital-related AEs, studies have also analyzed how the characteristics of a hospital stay influence LOS. First, LOS has also been used as an outcome measure in studies of various patient and disease-specific characteristics. For example, researchers have analyzed the LOS associated with specific disease or

treatment characteristics received during a hospital stay: these include specific DRGs [57], acute kidney injury [24], malnutrition [28], major elective surgery [27], reasons for admission [98], disease severity and comorbidity [92, 70, 63, 88, 167]. Studies have also analyzed the relationship between LOS and patient characteristics such as age, race, sex [88, 144, 19], or primary payer [166]. The LOS attributable to these types of factors highlights the importance of properly risk-adjusting a patient's LOS when attempting to estimate the excess LOS attributable to factors such as hospital AEs.

Studies have also analyzed how hospital- and provider-level characteristics can influence a patient's LOS. Hospital characteristics such as teaching status, ownership, patient volume, bed size, region, and rural vs. urban location have been studied as contributors to variation in LOS [57, 99, 166, 19, 77, 167]. Similarly, other provider characteristics such as daily rounding of an ICU physician [32] or a physician's years in practice [146] have also been found to be associated with LOS. Any increased LOS attributed to such hospital/provider characteristics may be considered to be in excess of ordinary care, insofar as such additional LOS is not directly linked to a patient's initial disease state. Thus, variation in LOS associated with such factors has generally been labeled as a marker for hospital/provider quality or efficiency.

Inpatient days and LOS have frequently been used as measures of health care productivity and, consequently, variation in LOS across hospitals/providers has often been attributed to differences in health care efficiency. Average LOS, across different hospitals or time, has frequently been used as a measure of a hospital's level of productivity, or output [99, 9]. Healthcare efficiency is generally assessed by the

amount of output that is produced from a given amount of input and, thus, LOS has been assumed to be a direct marker for efficiency. LOS has been used to assess efficiency at both a hospital level [17] and at a physician level [18]. Hollingsworth (2008) surveys the health care efficiency literature and finds that over 50% of studies were applied to hospitals, with most studies using output measures, such as inpatient days, to measure productivity.

Although LOS has frequently been viewed as a measure of healthcare productivity and efficiency, a number of studies have used it as a marker of quality. Some studies have used LOS as a direct measure of healthcare quality. For example, Scott et al. (2004) use the presence of long stays, namely patient stays with a LOS exceeding the 90th percentile for a particular diagnostic group, as a marker of hospital quality. Southern et al. (2011) use both LOS and mortality as markers of quality in comparing physicians with more or fewer years of practice [146]. Similarly, Edwards et al. (1991) and Munoz et al. (1990) use LOS and mortality as outcomes when evaluating the quality of high- and low-volume surgeons [112, 49].

While LOS has often been used as marker for healthcare quality, much of the healthcare quality literature has focused on the occurrence of specific quality-related events. Thus, LOS is often not considered a direct measure of quality. However, many quality markers have been found to be associated with increased LOS, and LOS has often been used to validate such measures of quality. Since the late 1990s AHRQ has worked to develop and maintain a series of quality indicators that can be used to assess patient safety and healthcare quality. One popular set of indicators are the patient

safety indicators (PSIs), which are a series of indicators of potential complications and errors that may occur in hospital settings. A number of studies have used LOS, along with inpatient mortality and/or charges, as outcome measures to assess the validity of these PSIs [168, 137, 133, 22]. In each of these studies, various PSIs were found to be correlated with increased LOS. Similarly, beginning in 2008 the Centers for Medicaid and Medicare Services (CMS) identified a series of HA conditions, termed 'never events,' for which hospitals would receive zero reimbursement when treating such events. These events are said to be the result of poor healthcare quality and the National Quality Foundation has determined that such events "should never occur in a healthcare setting" [91]. Similar to the case with PSIs, studies have also evaluated the performance of such never events as predictors of increased LOS [160].

2.2 *Clostridium difficile* Infection

2.2.1 General Overview: Epidemiology, Measurement, Risk factors, and Outcomes

Clostridium difficile (CD) is a gram-positive bacterium and a nosocomial pathogen that is the leading cause of hospital-associated diarrhea and colitis [84, 35]. Since the early 2000s CDIs have become much more frequent, severe and the disease itself has become more virulent [11]. CDIs have now become one of the most common types of HAIs [5, 84, 45]. Between 2000 and 2005 the incidence of CDIs nearly doubled, and over this period the number of cases per year increased from 134,361 to 291,303 [71]. One of the factors associated with such increasing trends was the emergence of a hypervirulent and drug-resistant strain of CD [101]. Much of the increase

in CDIs occurred among older adults, and CDIs have most frequently been thought of as a condition primarily affecting the elderly [84]. However, recent evidence has also found CDIs to be increasing among infants and children [116, 169, 86]. CDIs now pose a significant burden to the overall healthcare system and the costs associated with CDIs in the US have been estimated between $500 million to nearly $3.2 billion annually [103, 117].

One important issue that arises in the surveillance of CDI is that a variety of tests may be used to diagnose an infection, and various tests may be used across different studies [94, 80, 81]. Cell culture cytotoxicity assay (CYT) has historically been considered the gold standard test used to diagnose CDI. However, CYT can be time and labor intensive and, consequently, enzyme immunoassay (EIA) tests, which are faster and easier to use, have frequently been adopted. A limitation of EIA tests is they have also been shown to have a low sensitivity. Recently, polymerase chain reaction (PCR) assays have been increasingly adopted and have been shown to be highly sensitive. When comparing studies that rely on different tests to identify CDI, it is important to realize that cases identified by one test may not be able to be identified by another. Studies have shown that around 50% more CDI cases are identified using the more recently adopted PCR tests than CYT and EIA tests, which were used more frequently in the past [94, 80]. Additionally, research has suggested that CDI cases identified by PCR alone may have a differing degree of transmissibility than those identified by both PCR and CYT tests [81]. Because of changes in diagnostic testing, study results may differ across time as different

CDI patients are identified; this fact is especially important to consider when using administrative databases where the type of testing often cannot be identified.

The leading risk factor for CDI is antibiotic use. CD is commonly found to exist in the intestinal tract of many healthy individuals and is especially common in adults with recent healthcare exposure, residents in long-term care facilities, newborns and healthy infants [148]. Treatment with antibiotics may disrupt normal intestinal flora and allow pre-existing CD in such colonized individuals to grow and form into a CDI [90]. The duration of use and the usage of multiple antibiotics have also been associated with increased risk for CDI [13]. In addition to antibiotics, exposure risks that may increase a patient's likelihood of becoming colonized with CD have also been linked to increased risk for CDI; these include prolonged LOS, stay within an ICU, colonization pressure, transfer from another hospital, and recent hospitalization [95, 44, 43, 89, 16, 88].[1] Moreover, a patient's overall level of health has also been established as a risk factor for CDI, with advanced age, disease severity and comorbidity all linked to increased CDI risk [44, 88]. Other patient level factors that have been associated with increased risk of CDI include gastrointestinal procedures, cancer chemotherapy, enteral feeding, mechanical ventilation, and use of gastric acid suppressors [95, 44, 88, 15, 82]. Hospitals that have slower turnover rates, offer transplant services or are located in urban areas have been found to have greater incidence of CDI [130].

[1]Colonization pressure is a measure of the proportion of patients colonized with an organism in a defined location (e.g., hospital ICU) for a given period of time (e.g., a day, week or month).

CD is endemic in healthcare settings and transmission has generally been perceived to occur primarily in such settings. However, the spread and the transmission of CD has recently become a matter of some controversy. On the one hand, it has been estimated that over 20% of hospitalized patients and 50% of long-term care patients, are colonized with CD, as opposed to around 5% of healthy adults [131, 53, 102]. These numbers would seem to suggest that the potential for transmission is much greater in healthcare settings than in the community. Moreover, the previously mentioned risk factors associated with increased hospital exposure, such as colonization pressure, prolonged LOS or recent hospitalization, also suggest that CD is primarily transmitted in healthcare settings.

On the other hand, the notion that CDI is primarily transmitted in healthcare settings has been called into question by recent studies that have analyzed the source of CDI using whole-genome sequencing. Walker et al. (2012) performed gene sequencing on CD isolates from the stool samples of hospitalized patients that were either symptomatic or received oral vancomycin. The authors construct a network of ward-based contacts and find that only around 25% of CDI cases can be linked to a previous ward-based source. Similarly, Eyre et al. (2013) performed gene sequencing on all symptomatic cases of CDI that occurred in the Oxfordshire (England) community over a 3-year period. The authors found that only 55% of the CDI cases could be genetically linked to a previous case; with 23% being linked through hospital contact, 7% through community contact and 25% without any known hospital or community contact. These results may imply that healthcare-associated transmission of CDI

may be less common than was previously believed. However, these findings may also suggest that much of the transmission of CD within healthcare settings simply occurs via asymptomatic carriers. Indeed, previous research has shown that a majority of CD carriers are asymptomatic [131, 102].

The most common outcomes associated with CDIs are increased costs and LOS. A variety of methods have been used to analyze LOS and costs associated with CDI [109] and a wide range of estimates has been reported [45, 58]. Estimates of increased LOS associated with CDIs have been cited as high as 11 days [118, 122], 18 days [132], 21.3 days [162] and even greater than 3 weeks [97]. However, studies that have used more rigorous statistical methods, have controlled for more confounding factors, and have used larger sample sizes, tend to find much lower estimates of attributable LOS ranging from around 2 or 3 days [36, 6, 117, 88] to around 5 or 6 days [55, 145]. Cost estimates also mirror the variation in LOS estimates, and have varied quite dramatically across studies [58]. Dubberke et al. (2008) also find long-term increases in inpatient costs at 180 days following a CDI [42]. In addition to increased costs and LOS, other outcomes associated with CDIs include increased risk of mortality [119, 59, 85], increased risk for CDI recurrence [117], and an increased need for care or rehabilitation following discharge [139, 90].

2.2.2 Classification of CDI By Source and Occurrence of Symptoms

CDIs can be acquired in both community and healthcare settings. Because of this, and because CDIs are a commonly-cited HAI, there is an increased awareness

of both where CDIs are *acquired* and where the *onset* of symptoms first occur. Not all CDIs that present in a hospital stay are hospital associated and some hospital-associated cases may not present until after a patient is discharged from a hospital [87, 23, 40, 35]. Surawicz et al. (2013) provide a thorough summary of the standard classifications of CDI and summarize four main classes of non-recurrent CDIs applicable to a hospital setting.[2] CDIs that are associated with a hospital stay (i.e. HA CDI) may be labeled as either hospital-onset hospital-associated (HO-HA) CDI, if the infection occurs within a healthcare facility and symptoms begin at least 3 days after admission, or community-onset hospital-associated (CO-HA) CDI, if the symptoms begin outside a hospital but occur within 4 weeks of being discharged. On the other hand, CDIs are not considered to be associated with a hospitalization if they occur more than 12 weeks after a previous hospital discharge, and are labeled as community associated (CA) CDI. Finally, CDI cases that develop between 4 and 12 weeks after being discharged from a hospital are labeled as having an indeterminate or unknown source, since such cases may be HA or CA. Studies have indicated that HO-HA is most common type of CDI followed by CO-HA and CA CDI [46, 87, 83] .

It is important to note that, although much attention is given to distinguishing between the different types of CDI, individual studies and actual hospital surveillance practices may use different classification criteria to define types of CDI, and terminol-

[2]Because Surawicz et al. (2013) classify CDIs associated with various healthcare settings, they use a slightly different terminology and abbreviation (e.g. Health-care-facility-onset health-care-facility associated; HO-HCFA). However, I retain the same designations of "onset" and "associated."

ogy is often inconsistent. For example, studies have used cutoff windows of 72 hours [55, 156, 122], 48 hours [8, 42] and even 24 hours [89] after admission, for determining HA status. Moreover, it has been shown that changing this hospital-associated risk window can dramatically alter the calculated rates of HA CDI [64]. Even when studies use similar inclusion criteria to define CDI patients, the terminology used to describe them may differ. For example, the terms hospital-onset and hospital-associated (or hospital-acquired) are often used interchangeably. A study by Dubberke et al. (2010) defines CDI cases occurring 48 hours after admission to be hospital-*onset*, whereas Kutty et al., (2008) defines such cases as health-care-facility-*associated*. Similarly, the variation of different CDI definitions within the literature is also reflected in surveillance methods used in practice to measure incidence of HA CDI. Research has shown that traditional surveillance methods, which use a 48 hour window to define HA cases, and the laboratory-identified method used by the National Healthcare Safety Network, which uses a 72 hour window, can produce dramatically different results when ranking hospitals based on their incidence of HA CDI [48]. Therefore, when analyzing and comparing results across studies and hospital surveillance reporting, it is important to pay attention to the various inclusion criteria used when defining different CDI cases.

In addition to the classifications of CDI based on the onset of symptoms and source of infection, CDIs have also been classified based on where the infection is placed on a diagnostic record. Administrative discharge databases typically contain both a principal diagnosis along with a series of secondary diagnoses, often coded

using the International Classification of Disease, Ninth Revision, Clinical Modification (ICD-9-CM) diagnostic codes. In such databases, CDIs can be identified in patient records by the ICD-9-CM code 008.45, and previous research has found this be a valid means for measuring overall CDI (i.e. HA and CA) burden within a hospital [38, 41, 140]. However, one limitation of such datasets is that the timing of a CDI diagnosis often cannot be established, making it difficult to classify CDI cases as HA versus CA. Consequently, some authors have used a secondary diagnosis of CDI as a marker for being HO-HA CDI, since secondary diagnoses are generally not considered the principal reason for an admission [73, 118, 117]. Some of these studies also use a secondary factor, such as total LOS, antibiotic use or previous admission, to help establish hospital association.

However, using a secondary diagnosis to classify cases of CDI as HO-HA can be problematic because secondary diagnoses are not necessarily HA. While Dubberke et al. (2010) find ICD-9-CM codes to be useful for identifying overall CDI burden; they find secondary diagnostic codes to be a poor marker for identifying HO-HA CDI. The authors analyzed CDI cases at five different hospitals and found that slightly under half of the CDI cases identified by a secondary diagnosis were HO-HA. Many of the secondary CDI cases they identified were either community-onset, recurrent CDI[3] or not CDI at all. These results imply that while CDI cases identified by a primary diagnosis are likely to be CA, those identified as secondary may be either HO-HA or

[3]Recurrent CDI is defined as the recurrence of CDI symptoms within some time window following the resolution of symptoms for a previous CDI diagnosis. This time window is generally set at 8 to 12 weeks following the resolution of previous symptoms.

CA. Consequently, studies have also relied on present on admission (POA) indicators when using administrative databases to determine if an infection was HA [74, 93].

2.2.3 Hospital Reporting and Comparison of CDI Incidence

Due to the increasing incidence and growing costs associated with HAIs, many policy makers have targeted the public reporting of such infections as one means by which to improve hospital quality. By 2013, over half of all US states had enacted laws mandating the reporting of at least one type of HAI; the majority of which went into affect in the period around 2006 to 2007. Occurring slightly later, around 2008, states began requiring hospitals to report cases of HA CDI [66]. Since 2008, at least 20 states have adopted some type of legal mandate requiring hospitals to report occurrences of HA CDI, the majority of which began in 2013 [129]. California and New York both implemented reporting mandates that took effect in 2009. Additionally, in January of 2013 CMS made reporting of HA CDI mandatory as part of the Inpatient Quality Reporting Program (IQR), and hospitals in the National Healthcare Safety Network also began reporting CDI events [129, 48]. Given this increased availability of public information on hospital incidence of HAIs, studies have now begun using hospital incidence rates of HAIs, such as MRSA and CDI, as measures for making hospital-level comparisons [30, 151].

A number of factors must be taken into consideration when using hospital incidence of HA CDI to make comparisons across hospitals. First, hospital CDI incidence is likely influenced by patient and hospital risk factors for CDI that may

not be directly associated with poor hospital quality or outcomes. One study found that controlling for factors related to CDI, such as hospital bed size, teaching status, CDI test type, patient age and comorbidities, resulted in substantial variation in the ranking of hospitals by CDI incidence [30]. Similar findings have been reported when using MRSA incidence to rank hospitals [151]. These findings suggest care should be taken to control for other factors associated with CDI when using CDI incidence to rank and compare hospitals.

A second factor to consider is that policy mandates for reporting of HA CDI have been enacted, and gone into effect, at various points in time across different states. From 2008 through 2011, only six states enacted mandatory reporting requirements for HA CDI, in 2012 four more states enacted such laws, and in 2013 an additional 10 states did the same [129]. Assuming that such mandates influence hospital reporting practices, differences in hospital CDI incidence across states and time may simply reflect changes in the amount of HA CDI that gets reported, rather than differences in actual CDI incidence. Moreover, policies for the reporting of various HAIs have often been enacted at different points in time. Reporting mandates for central line-associated bloodstream infections generally preceded mandates for catheter-associated urinary tract infections, which preceded those for MRSA and CDI [66]. It is conceivable that prior reporting mandates for one disease may influence the effect of reporting mandates for other diseases that are enacted later. For example, early mandates for one type of HAI could lead hospitals to enact systems and processes for reporting that are also applicable to other HAIs. In this sense, early

investments made to satisfy to one reporting requirement may lessen the cost of satisfying to later requirements. Thus, comparisons across hospitals in different states and different times should attempt to account for differences in reporting requirements.

Finally, not only do different states have various reporting mandates, which are enacted in different time periods, but the manner in which mandates are implemented can vary across states. Research has found that the implementation of reporting mandates can vary across states in terms of: how HAI data is required to be submitted by hospitals, if and how the data are reported to the public, and if individual healthcare facilities can be identified in such reports [128]. Thus, even if the timing of reporting mandates can be accounted for, the type of data available on HA CDI incidence may not always be directly comparable across states. In addition, reporting policies across states have been enacted in different forms and through different legal authority. Some states have enacted HA CDI mandates independently and others have simply incorporated the CMS IQR requirements into state law [129]. Similarly, in some states, legal authority for reporting is imposed through a state statute while in others it is imposed through administrative mechanisms. Such differences also seem likely to influence hospital incentives and the impact of such mandates on hospital reporting.

2.3 CDI Incidence and LOS in Uninfected Individuals

In the previous two sections, it was described how a patient's LOS may be a risk factor for acquiring CDI and that CDI is associated with excess LOS in patients

that become infected. The interdependent relationship between LOS and CDI has been frequently studied, and a variety of studies have noted the technical difficulties of estimating these relationships from a statistical perspective [110, 62]. However, the relationship between CDI and LOS has always been approached from the perspective of patients that are infected with CDI. No previous study, aside from the preliminary research, has attempted to uncover a connection between a hospital CDI incidence and patient LOS in patients that do not have a CDI. While one may not typically think that a hospital's CDI incidence would be associated with LOS in uninfected patients, a number of theoretical connections may exist. Below I describe three broad categories of theoretical factors that may provide a link between CDI and LOS; these include (1) hospital quality, (2) hospital efficiency, and (3) other factors associated with LOS. In each of these cases CDI may act as a proxy variable for unmeasured confounding factors associated with both LOS and CDI. Moreover, in each of these cases, the theorized link between CDI and LOS occurs primarily through the incidence of HO-HA CDI.

2.3.1 Hospital Efficiency and Quality

As previously described, LOS has often been used as a maker for hospital quality, and many of the factors associated with poor quality have also been linked to prolonged LOS. Because LOS is itself a risk factor for HO-HA CDI, it is possible that any unmeasured quality-related factors that are associated with increased LOS may also be indirectly related to CDI incidence, insofar as longer LOS leads to more HO-

HA CDI. Thus, LOS may indirectly capture any unmeasured dimensions of hospital quality. However, there are also a number of theoretical components of quality that may be directly related to both CDI incidence as well as LOS.

Some of the factors related to hospital quality that might also directly influence CDI incidence include hospital cleanliness and maintenance, hand-hygiene practices, antibiotic/antimicrobial stewardship, CDI screening practices, and staff education of infectious disease. Each of these factors has been linked to the occurrence or spread of CDI, and each has been suggested as an area where hospitals can potentially work to prevent CDI [39, 148]. It is conceivable that hospitals of greater quality may have better cleaning and maintenance practices. Higher quality hospitals might also implement policies to promote, or employ providers and staff that practice better antimicrobial stewardship, hand hygiene, CDI screening and infectious disease awareness. While these factors have been shown to be related to CDI, their connection to quality and LOS is largely uncharted territory in the existing literature. However, for some of these factors, there is some evidence that is suggestive of a connection.

Both antibiotic stewardship and proper hand hygiene represent types of "guideline recommended practices," that may capture a dimension of healthcare quality [148, 34]. Proper adherence to guideline recommended practices and treatment has been a commonly used measure of healthcare quality; some examples include: the timing of antibiotic therapy following pneumonia; misuse of antibiotics with the common cold or other upper respiratory infections; use of aspirin, ACE inhibitors, and beta-blockers with acute myocardial infarction; appropriate treatment for depression;

along with patient counseling and disease screening practices [141, 104]. If a hospital's or provider's failure to adhere to proper antimicrobial stewardship and/or hand hygiene is representative of a broader failure to adhere to guidelines, then CDI incidence may represent a proxy measure for this dimension of quality.

Similarly, diagnostic testing and screening for CDI is another guideline-recommended practice that may be related to a hospital's reported CDI incidence [148]. Screening practices, such as screening for smoking, alcohol abuse, or various diseases, have also been used as measures of healthcare quality. Although poor quality is most-often related to underuse of such practices [104], in the case of CDI, overuse of screening may be associated with poor quality. It has been recommended that hospitals should only test the stools of patients with diarrhea, avoid repeated testing, and avoid screening for CDI in patients without diarrhea [148, 39]. Hospitals and providers that do not properly adhere to these screening guidelines and over-test for CDI may also over-report the number of CDI cases by including false positives or asymptomatic carriers. Consequently, such hospitals may have a higher perceived incidence of CDI (both CA and HA). If such a failure to adhere to guidelines associated with CDI screening is indicative of a failure to adhere to other treatment guidelines, then CDI incidence may also be a proxy for quality in this regard.

Physician and staff knowledge along with their level of infectious disease awareness may also play a role in both the quality of care that is delivered and the occurrence of CDI. While a provider's skills and level of knowledge is not a direct dimension of quality care, it has frequently been suggested as a factor associated with the struc-

ture, and may shape the process, of delivering quality care [21, 33]. Attempts to define healthcare quality have often mentioned the expected benefits of care relative to expected harm [141]. Thus, a basic component of care quality is the avoidance of unnecessary exposure to harm. In the case of CDI, a number of 'best-practices' have been described as ways to reduce a patient's likelihood of acquiring CDI. For example, avoiding the use of electronic thermometers, remembering to use dedicated equipment and cleaning equipment between use, or remembering not to prescribe certain antibiotics to patients at a high risk of CDI have been suggested as ways of preventing CDI [39]. It seems likely that providers' knowledge of these practices along with their awareness of CDI, may improve their adherence to such practices. Indeed, the connection between disease awareness/knowledge and quality care practices has been highlighted by Pavese et al. (2009).[120] In this study, the authors found that an educational session led by an infectious disease physician was successful in reducing inappropriate antibiotic prescribing. This example also illustrates that other factors associated with quality care, such as appropriate antibiotic use, may also depend on the level of provider knowledge.

2.3.2 Other Unmeasured Factors and The Post-Discharge Effect

In addition to hospital quality or efficiency, there may be other unmeasured factors associated with LOS that may be captured by CDI incidence. One of these factors may be hospital structural characteristics, such as the availability of private rooms to treat infected patients. Treating patients with CDI in private rooms has

been cited as one of the ways to prevent the spread of CDI within a hospital [39]. Although administrative databases often contain measures of hospital capacity or bed size, they do not contain measures of internal structure of a hospital and they cannot be used to assess patient proximity. Thus, the availability of private rooms for housing infected individuals or the proximity of infected patients, may represent unmeasured factors related to CDI incidence. Given that many hospital structure variables, such as bed-size, volume and teaching status have been shown to be related to LOS [57, 99, 166, 19, 77], its conceivable that other structural variables may be as well. There is some evidence, at least from articles citing expert opinions, that use of private rooms may be associated with decreased LOS [31].

One major factor that may allow CDI incidence to capture variation in LOS between hospitals has to do to with HA CDIs that do not occur until after discharge, i.e. CO-HA CDI. As previously described, CO-HA CDIs are cases that are associated with a hospital stay but do not appear, and are not diagnosed, until after a patient is discharged. Various time windows have been used to identify such cases, including 30, 60 and 90 days along with 4 or 12 weeks following discharge; although, 4 weeks post discharge seems to have become the standard length of time that is now most frequently applied [48]. There is now a large body of evidence that suggests a significant number of CDI cases occur as CO-HA, after a patient has been discharged from a healthcare setting [87, 23, 40, 35]. Moreover, studies have found the number of CO-HA cases to be increasing over time [113, 37]. Murphy et al. (2012) even find that a majority of hospital associated-CDI cases may now occur as CO-HA within 12

weeks following discharge [113].

The reason the occurrence of CO-HA CDI may further link CDI incidence and LOS is that hospitals with longer average LOS, either as a result of quality, efficiency or other factors, will be more likely to observe HA CDI cases before a patient is discharged. For example, if a given case of CDI were to occur 5 days after a patient was admitted to a hospital, the case would be considered HO-HA at hospital where the patient stayed 6 days but would be considered CO-HA at a hospital where the patient stayed only 4 days. Thus, CDI cases associated with hospitals that have longer average LOS are more likely to be observed as HO-HA and, subsequently, recorded in discharge records. I refer to this effect as the *post-discharge effect*. Murphy et al. (2012) find that HO-HA CDI cases were more likely to be associated with a high-length-of-stay hospital than CDI cases occurring within 12 weeks post discharge [113]. While these findings support the possible existence of this effect, the size, and existence, of this effect across hospitals and time is unknown. Because of this post-discharge effect, it may be difficult to fully distinguish between unmeasured LOS factors that are directly related to CDI and those that are indirectly related. However, the post-discharge effect may still cause CDI to be an important marker for unmeasured factors associated with LOS variation across hospitals, regardless of the source of variation.

CHAPTER 3
METHODS

3.1 Design

This study is carried out using inpatient records at both a national and state level for the seven year study period from 2005-2011. Because the national-level data represent a different sampling of hospitals each year, and because hospitals and patients cannot be linked, the national-level data are not pooled across years. Thus, the analyses conducted using national-level data are carried out as separate cross-sectional studies for each year from 2005-2011. On the other hand, state-level data from California and New York can be linked by hospitals and patients. Therefore, the analyses carried out at the state-level are pooled across years and can be viewed as a type of "quasi" longitudinal design.[1]

3.2 Data Source

The data used for this study come from two databases that are a part of the Healthcare Cost and Utilization Project (HCUP). HCUP databases are constructed through a federal-state-industry partnership, which is sponsored by AHRQ [4]. Data contained in these databases is collected through a variety of sources that include state and federal data collection agencies along with hospital and other private data

[1]It is worth noting that the data are highly unbalanced and irregular. Over the seven year period many patients have only one visit and very few have more than one visit per year. Moreover, for most patients with multiple visits, revisits tended to be relatively clustered within time. Thus, many longitudinal methods may be inapplicable or difficult to perform.

organizations. Together HCUP databases contain the largest collection of longitudinal hospital data in the US. Two HCUP sources of inpatient data, which span the years 2005-2011, will be used for this study. Both of these data sources have been used previously to study LOS and CDI.

The first HCUP database that will be used is the National Inpatient Sample (NIS). The NIS is a nationally representative database of inpatient records, and is the largest all-payer inpatient database that is publicly available in the US. It contains records of roughly 8 million inpatient stays each year, and includes sample weights that allow for nationally representative estimates applicable to more than 36 million hospital stays. Through 2011, the NIS provides complete discharge records from each hospital that is sampled. Over the study period, from 2005-2011, the NIS contains complete discharge records from around 1,000 different hospitals annually, spanning between 37 and 46 participating states.

The second HCUP data source that will be used are State Inpatient Databases (SIDs) for the states of California and New York. The SID contains the universe of hospital discharge records for all inpatient stays at community hospitals within a state. Over the study period, the SID contains roughly 4 million discharges per year in California, and roughly 2.6 million discharges per year in New York. These data represent nearly 400 different hospitals per year in California and nearly 200 different hospitals per year in New York. In both New York and California, patient records include a linking variable that allow records from the same patient to be linked across stays and for the number of days between visits to be calculated. Records within these

states also contain POA indicators, which can be used to determine whether specific diagnoses were HA.

In addition to HCUP SIDs, the HCUP State Emergency Department Databases (SEDDs) from California and New York will also be used to identify CDI diagnoses that occur following a previous inpatient stay. Similar to the SID, the SEDD contains records for the universe of emergency department visits within a state. In both California and New York, visits can be linked across time and between the SID and SEDD for a given patient.[2] This allows cases of CO-HA CDI to be identified, where the infection occurred during an inpatient stay and the diagnosis occurred during a later emergency department visit. The SEDD data will only be used to identify cases of CO-HA CDI that occurred in a previous inpatient stay, and none of the SEDD data will be used directly in this analysis.

3.3 Study Sample

As described in the Data Source section above, the NIS and SID contain all inpatient records from a selected hospital, regardless of age or payer type. The study sample used for this research includes all inpatient records contained in the databases described above, subject to three different inclusion criteria. The first requirement for inclusion will be the absence of a CDI. Patient records with a CDI diagnosis, either principal or secondary, during a hospital stay will be excluded. Because increased LOS is an outcome associated with a CDI, and because CDI incidence may be a risk factor

[2]The SEDD is not available for 2005 in the state of New York.

for acquiring a CDI, CDI incidence may be endogenously related to LOS in patients with an infection. Thus, in order to avoid a potential endogeniety problem associated with estimating the effect of CDI incidence on LOS in patients with CDI, such patients will be excluded from the analysis. Although CDI patients will be excluded from the LOS analysis, these patients will be used to calculate each hospital's CDI incidence.

Second, patient records were only analyzed from hospitals that contained at least one case of CDI in the time window used for analysis (i.e. either discharge quarter or year). It is assumed that all inpatient settings have at least some likelihood of a CDI occurring. This requirement ensures that all hospitals analyzed have a CDI incidence that could be estimated and removes those hospitals where an estimated CDI rate could not be obtained. Moreover, hospitals with zero CDI cases over the course of a year (or quarter) are likely to either be very small, in terms of patient discharges, or have a very specific patient population not applicable to this study.

Finally, patient records were excluded that had a LOS equal to zero. The purpose of this study was to estimate the marginal increase in LOS that could be captured by CDI incidence, and the effect of interest is not observable for patients with a LOS of zero. Moreover, patients requiring an inpatient hospitalization of less than a day are less likely to be exposed to the effect being captured by CDI incidence.

3.4 Variables

3.4.1 Dependent Variables

Two basic dependent variables are used in this study. The primary dependent variable is individual inpatient LOS, measured in number of days. Individual patient-level LOS will first be estimated in a model that contains CDI incidence along with patient and hospital characteristics. Let $LOS_{i,j,t}$ denote the LOS of patient i, who is hospitalized at hospital j over time period t. A secondary analysis will be conducted using a multilevel model to explain between-hospital variation in LOS; this model is described in section 3.5.3. In the first stage of this model $LOS_{i,j,t}$ will be estimated using only patient characteristics along with a hospital-specific fixed effect. In the second stage of this model the hospital-specific fixed effects, estimated in the first stage, will become the dependent variable. These hospital fixed effects are the second dependent variable of interest and can be thought of as the portion of a patient's LOS that can be attributed to a specific hospital. Let $\hat{\beta}_{FE,j}$ represent the estimated hospital fixed effect for hospital j. The hospital fixed effects will then be estimated in the second stage using CDI incidence along with other hospital characteristics.

3.4.2 Primary Independent Variable

The main explanatory variable of interest in this study is the incidence of CDI within a hospital. A number of specifications of this incidence rate will be used to control for HA CDI cases versus overall CDI cases, as well as CO-HA CDI cases that occur post-discharge. A description of how each of these various CDI incidence rates

will be used to capture a different dimension of this problem is discussed in further detail section 3.5.

The first set of CDI incidence rates used are those that are calculated using CDI cases indicated by the order of a patient's CDI diagnosis. Separate CDI incidence rates are calculated using the number of CDI cases that are assigned a principal, secondary or any CDI diagnosis. The following equations describe the notation and calculations used for each of these incidence rates, corresponding to hospital j over time period t:

$$CDI^1_{j,t} = \frac{\text{Principal CDI Discharges}}{\text{All Discharges}} \tag{3.1}$$

$$CDI^2_{j,t} = \frac{\text{Secondary CDI Discharges}}{\text{All Discharges}} \tag{3.2}$$

$$CDI^{All}_{j,t} = \frac{\text{All CDI Discharges}}{\text{All Discharges}} = CDI^1_{j,t} + CDI^2_{j,t} \tag{3.3}$$

The overall CDI incidence rate, $CDI^{All}_{j,t}$, was the primary explanatory variable used in the preliminary study. Because the NIS does not contain POA indicators, the national-level analysis in this study uses the principal, $CDI^1_{j,t}$, and secondary, $CDI^2_{j,t}$, incidence rates to help control for CA CDI cases receiving a secondary diagnosis; this is further discussed in Section 3.5.1, below.

The state-level analysis also compares CDI incidence rates corresponding to the number of HO-HA CDI cases and the total number of HA CDI cases (i.e. HO-HA CDI cases plus CO-HA cases occurring post-discharge). POA indicators are used to define hospital CDI incidence rates corresponding to HO-HA CDI; those CDI cases

where the diagnosis is not marked as POA are labeled as HO-HA. Additionally, patient revisits are linked across time to identify any CDI cases occurring post-discharge. CDI diagnoses that are coded as POA and occur less than 4 weeks following a previous inpatient discharge are labeled as CO-HA. Together a hospital's cases of HO-HA CDI and CO-HA CDI comprise its overall number of HA CDI cases, and each hospital's overall HA CDI incidence was also calculated corresponding to these cases. The following equations describe the notation and calculations used for each of these rates, corresponding to hospital j over time period t:

$$CDI_{j,t}^{\text{HO-HA}} = \frac{\text{HO-HA CDI Discharges}}{\text{All Discharges}} \tag{3.4}$$

$$CDI_{j,t}^{\text{CO-HA}} = \frac{\text{CO-HA CDI Discharges}}{\text{All Discharges}} \tag{3.5}$$

$$\begin{aligned} CDI_{j,t}^{\text{HA}} &= \frac{\text{HO-HA + CO-HA CDI Discharges}}{\text{All Discharges}} \\ &= CDI_{j,t}^{\text{HO-HA}} + CDI_{j,t}^{\text{CO-HA}} \end{aligned} \tag{3.6}$$

By comparing estimates that are obtained using these two different rates, this study analyzes the degree to which the post-discharge effect contributes to the connection between hospital CDI incidence and LOS; this is described in further detail in section 3.5.2, below.

Because cases of CDI may be categorized in a variety of different ways (e.g., primary, secondary, or not POA) the range of calculated CDI incidence values will vary dramatically among the different measures. For example, overall CDI incidence,

$CDI^{All}_{j,t}$, will always be greater than secondary CDI incidence, $CDI^{2}_{j,t}$, because only a portion of CDI cases are coded as secondary. Thus, the coefficient estimates from regression models using different CDI incidence measures will not be directly comparable in terms of their relative magnitude. In order make the coefficient estimates comparable, this analysis is conducted using standardized CDI incidence values. The following equation was used to standardize all CDI incidence measures:

$$\tilde{CDI}_j = \frac{CDI_j - \mu_{CDI}}{\sigma_{CDI}} \tag{3.7}$$

where $\tilde{CDI}_j$ is the standardized value of the CDI incidence at hospital j, and μ_{CDI} and σ_{CDI} are the mean and standard of the deviation of CDI incidence across all hospitals in a given time period. The values μ_{CDI} and σ_{CDI} are calculated for the discharge quarter or year, corresponding to the time period in which CDI incidence is measured. It is important to notice that this standardization preserves both hospital ordering in terms of CDI incidence as well as the relative magnitude of the differences. This standardization is used for all the CDI rates described by equations (3.1)-(3.6). From this point forward, all models and analyses referring to a CDI incidence measure will be in this standardized form, unless otherwise noted.

3.4.3 Secondary Independent Variables

In addition to CDI incidence, two basic sets of explanatory variables will be analyzed in connection with LOS. The first set of explanatory variables are hospital and provider characteristics. Many hospital factors such as teaching status, bed size,

rural vs. urban location and type of ownership have been shown to be related to LOS [57, 99, 166, 19, 77]. Similarly, a hospital's nurse-to-patient ratio and the percentage of nurses that are registered have been suggested as markers of quality and shown to be related to LOS [22, 154]. Each of these variables are included as potential explanatory variables. In addition, a hospital's regional location is used to control for any potential regional variations in LOS that may exist. These hospital-level characteristics were only available in the NIS and were only used for this analysis.

The second category of explanatory variables is patient, or disease, level characteristics. First, a patient's demographics such as age, gender, and primary payer have often been used to explain variation in LOS across patients [144, 19, 166, 88, 119, 85, 139]. In addition, general hospital stay characteristics such as the type of admission, the discharge disposition of a patient, whether the patient died during a stay, if the admission occurred over a weekend, and if there was a major operating room procedure performed are also factors that are likely to be related to LOS [10, 67]. Since all patient records will be included, indicators for whether a record corresponds to neonatal or maternal admission are also included to control for LOS variations associated with hospital births. Indicators for the discharge quarter and year are used to control for potential seasonal variation and trends in LOS over time.

In addition to demographics and general hospital stay characteristics, a patient's disease state, as indicated by diagnostic types, disease severity, and comorbidity has been shown to be associated with LOS [63, 88, 167]. The 30 Elixhauser comorbidity indicators, which have been developed for use with administrative databases,

are used to control for a patient's comorbidity [50]. The total number of procedures (NPR), number of diagnoses (NDX), and number of chronic conditions (NCHRON) are also used to control for a patient's severity and the complexity of a hospital stay. Patients with more diagnoses, procedures, and chronic conditions are likely to be more severe and require longer LOS, and these variables were found to be highly significant in the preliminary study.

The NIS contains All Patient Refined Diagnostic Related Group (APR-DRG) information. The APR-DRG system is an inpatient-classification system developed by 3M Health Information Systems that is intended to improve and expand the traditional Medicare DRG system [1]. The APR-DRG system creates a three dimensional representation of a patient's disease state based on the following components: APR-DRG, severity, and mortality. Each patient first receives an APR-DRG corresponding to his or her diagnoses. Then, corresponding to this APR-DRG, each patient is assigned a value between 1-4 for both their severity of illness and risk of dying. In the national-level analysis these alternative DRG specifications were analyzed in comparison to the standard DRG indicators and were found to offer far greater explanatory power. The SID analysis lacks the APR-DRG variables, but does contain standard DRGs. However, the DRG coding system changed over the course of the SID study window and revisions in DRG coding have been shown to alter the way in which hospitals assign such codes [105, 76]. Therefore, indicators for a patient's primary diagnosis clinical classification code, as assigned by the HCUP Clinical Classifications Software single level diagnosis code, were used as a replacement for APR-DRG

indicators in the SID analysis.[3]

Tables Tables 3.1 and 3.2 describe all of the hospital- and patient-level variables, respectively, that were analyzed as potential explanatory variables. These tables also provide a description of how each variable is specified. For some variables multiple specifications were analyzed in order to select the one with the best fit (e.g., continuous age vs. 5 year categories); these tables summarize the final specification that was selected. In addition, Tables 3.1 and 3.2 indicate whether a variable was utilized in the analysis of the NIS, the SID for California, or the SID for New York.

[3]The HCUP Clinical Classification Software assigns clusters of individual ICD-9-CM codes to clinically meaningful categories, which are intended to create higher level groupings of similar diagnoses that are easier to use. The CCS collapses more than 14,000 diagnosis codes into 285 categories.

Table 3.1: Patient-Level Explanatory Variables

Variable	Description	Database Availability		
		NIS	SID CA	SID NY
Age	21 Indicators for 5 year age intervals	Y	N	N
Sex	Male or Female	Y	Y	Y
Admission Type	Emergency, urgent, elective, newborn, trauma center, other	Y	N	Y
Admission Source	Emergency department, other health facility, court/law enforcement, routine	N[a]	Y	N[a]
Disposition	Routine, transfer to short-term hospital, other transfer, home health care, against medical advice or unknown	Y	Y	Y
Neonatal or Maternal	Indicators for maternal, neonatal, or both maternal and neonatal records	Y	Y	Y
Discharge Quarter	4 quarterly indicators	Y	Y	Y
Discharge Year	Yearly indicators for 2005-2011	N[b]	Y	Y
Weekend admission	Indicator of weekend admission	Y	Y	Y

Continued on next page

Table 3.1 – *Continued from previous page*

Variable	Description	Database Availability NIS	SID CA	SID NY
Died	Indicator of death during hospitalization	Y	Y	Y
Elixhauser Comorbidities	30 specific comorbidity indicators	Y	Y	Y
Number of Procedures	Total number of procedures coded on discharge record (up to 15: New York, NIS before 2009; up to 25: California, NIS 2009 and later)	Y	Y	Y
Number of Diagnoses	Total number of diagnoses coded on discharge record (up to 15: New York, NIS before 2009; up to 25: California, NIS 2009 and later)	Y	Y	Y
Number of Chronic Conditions	Total number of unique chronic diagnoses reported on the discharge (up to 15: New York, NIS before 2009; up to 25: California, NIS 2009 and later)	Y	Y	Y
APR DRG	316 All Patient Refined DRG indicators developed by 3M	Y	N	N

Continued on next page

Table 3.1 – *Continued from previous page*

Variable	Description	Database Availability NIS	SID CA	SID NY
APR DRG severity	4 indicators for likelihood of dying: minor, moderate, major or extreme	Y	N	N
APR DRG	4 indicators for severity of illness (loss of function): minor, moderate, major or extreme	Y	N	N
Primary Diagnosis CCS	HCUP Clinical Classification Software primary diagnosis group (285 indicators)	N[c]	Y	Y
Primary expected payer	Medicare, Medicaid, private insurance, self-pay, no charge, or other	Y	Y	Y
Zip code median income	4 quartile indicators for median household income in patient's ZIP code	Y	Y	Y

[a]Available but not used due to many missing values
[b]Available but not applicable because data were not pooled across years in the NIS
[c]Available but not used due to inferior model performance

Table 3.2: Hospital-Level Explanatory Variables

Variable	Description	Database Availability		
		NIS	SID CA	SID NY
Bed size	Small, medium or large	Y	N	N
Control/Ownership	Public, private (nonprofit), private (for profit)	Y	N	N
Region	Northeast, midwest, south, west	Y	N/A	N/A
Teaching Status & Location	Rural, urban non-teaching, urban teaching	Y	N	N
Registered Nurse Percent	Percentage of registered nurses among all licensed nurses	Y	N	N
Nurse to Patient Ratio	Number of total licensed nurse full time equivalents per 100 inpatient days	Y	N	N

3.5 Model

This section describes the theoretical model of CDI incidence and LOS that was estimated as part of this study. In order to address Objective 1, the analysis was conducted in two parts: first using the NIS and, second, using the SID for California and New York. Because the SID for California and New York lack many of the variables present in the NIS, which have been shown to be related to LOS, the NIS was used to help reinforce the results of the SID analysis. However, the NIS does not contain POA indicators for diagnoses, and as a result, an alternative estimation approach was used with this dataset to help account for CA CDI cases and, subsequently, address objective 1. This approach is described in subsection

3.5.1. To address objective 2, subsection 3.5.2 describes how CDI cases that occur post discharge were analyzed. Finally, this section concludes by describing a multilevel model, in subsection 3.5.3, that was used to estimate how much of the LOS variation between hospitals, remaining after controlling for patient and disease characteristics, was able to be captured by CDI incidence.

We begin by assuming a theoretical model where a patient's LOS is a function of individual patient/disease characteristics, along with both measured and unmeasured hospital characteristics. The following equation describes the relationship between these factors:

$$LOS_{i,j} = \beta_0 + \beta_W W_j + \beta_X X_i + \beta_Z Z_j + \epsilon_i \tag{3.8}$$

where $LOS_{i,j}$ is the LOS of patient i at hospital j, W_j are unmeasured hospital factors associated with LOS (e.g. quality or efficiency), X_i is a vector of patient/disease characteristics, Z_j is a vector of measured hospital characteristics, $\beta_0, \beta_W, \beta_X, \beta_Z$ represent coefficient values on these variables, and ϵ_i represents a patient-level error term. We next hypothesize that HA CDI incidence acts as a proxy variable for the unmeasured hospital factors W_j, because of the relationships previously described. If we assume that the relationship between CDI incidence and these unmeasured factors is fairly linear, we can express the following relationship:

$$W_j = \theta_0 + \theta_1 CDI_j^{\text{HA}}, \tag{3.9}$$

where θ_0 and θ_1 represent unknown parameters.[4] Given the relationship expressed in

[4]Note: Equation 3.9 depicts an unrealistic situation where W_j is a directly determined

equation 3.9 we can then rewrite equation 3.8 as the following:

$$LOS_{i,j} = \alpha_0 + \alpha_1 CDI_j^{\text{HA}} + \beta_X X_i + \beta_Z Z_j + \epsilon_i \tag{3.10}$$

where $\alpha_0 = \beta_0 + \beta_W \theta_0$ and $\alpha_1 = \theta_1 \beta_W$. Thus, by including CDI incidence as a proxy variable, it is possible to obtain consistent estimates even when there exist unmeasured hospital factors associated with LOS.

The model expressed in equation 3.10 is first estimated using the NIS. For the NIS analysis, CDI^{HA} is approximated using both overall CDI incidence and secondary CDI incidence, as described by equations 3.3 and 3.2. Given that secondary CDI should better reflect HA CDI, we expect secondary CDI incidence to provide a better fitting model and have a larger coefficient estimate than overall CDI incidence. However, since HA CDI cannot be directly identified in the NIS, an alternative statistical technique was used to estimate equation 3.10, in order to address Objective 1 while accounting for CA CDI. This approach is described in Subsection 3.5.1. Additionally, since each year of the NIS represents a distinct sampling of different hospitals across different states, the analysis carried out using the NIS is conducted separately for each year. Thus, for each year from 2005 to 2011 separate coefficients were estimated using the NIS.

For the state level analysis, equation 3.10 is estimated by pooling together

---

by CDI incidence. However, it is possible to write $W_j = \theta_0 + \theta_1 CDI_j^{\text{HA}} + \mu_j$, where $\theta_0 + \theta_1 CDI_j^{\text{HA}}$ is the linear projection of W_j onto 1 and CDI_j^{HA} and μ_j is an error term. It can then be shown that under certain conditions that LOS can be consistently estimated using the proxy CDI_j^{HA} (see Wooldridge, 2010, pg 67-82) [164]. Moreover, in many situations, even when an imperfect proxy variable is used, and equation 3.8 cannot be consistently estimated, the bias due to W_j being omitted may still be reduced by including the proxy CDI_j^{HA}.

the SID data from 2005-2011. However, because LOS has been trending downward over time, indicators for discharge year and quarter are added to control for trends in time. The standardized CDI incidence rates, as described in Section 3.4.2, are also calculated separately corresponding to the relative CDI incidence in each year, in order to account for possible changes in CDI incidence over time. Similar to the NIS analysis, in the SID analysis CDI^{HA} is approximated using both overall CDI incidence and secondary CDI incidence. In addition, the SID contains POA indicators so CDI^{HA} is also estimated directly using cases of CDI not POA.

In both California and New York there may exist a significant change point in CDI reporting beginning somewhere near 2009, when hospitals in both states were required to report cases of HA CDI. It is possible that the effectiveness of HA CDI as a proxy for unmeasured hospital characteristics may have changed when reporting of HA CDI became mandatory in 2009. Thus, it is conceivable that β_0, from equation 3.8, along with θ_0 and θ_1, from equation 3.9, may be different across these two reporting periods. Consequently, the value of α_1, from equation 3.10 may also vary between these time periods. In order to analyze if a change occurred in the connection between CDI incidence and LOS after the 2008 reporting requirement, the following equation is also estimated:

$$LOS_{i,j,t} = \alpha_{0,t} + \alpha_{1,PRE} CDI^{\text{HO-HA}}_{j,(t<2009)} + \alpha_{1,POST} CDI^{\text{HO-HA}}_{j,(t\geq 2009)} + \beta_X X_i + \beta_Z Z_j + \epsilon_i \quad (3.11)$$

where the values of $\alpha_{1,PRE}$ and $\alpha_{1,POST}$ are the estimated effects of CDI incidence before and after the start of 2009, respectively. If the reporting requirement led hospitals to report cases of HA CDI more accurately, or more frequently, then we

would expect $\alpha_{1,POST} > \alpha_{1,PRE}$.

3.5.1 Analyzing HA CDI from Order of Diagnosis in The NIS

The NIS does not contain POA indicators corresponding to individual diagnoses, thus, it is impossible to directly identify HA CDI cases using this data set. In our preliminary study, secondary CDI cases were used to suggest that HA CDI was the primary driver of the link between CDI incidence and LOS. However, the estimated effect of CDI incidence on LOS is still likely to be attenuated even when secondary cases are used, since not all secondary CDI cases are HA. Another way to test that the connection between LOS and CDI incidence occurs primarily via HA cases is to include both a hospital's primary and secondary CDI incidence as two separate coefficients in the regression model. Assuming that some of a hospital's CA CDI cases spillover into secondary diagnosis (e.g., patients may have more than one diagnosis when admitted), then primary CDI incidence may work to net out some of this spillover effect. To see this, consider the following model (note: the model in this section relies on the use of unstandardized CDI incidence to hold). First, the theoretical connections between CDI incidence and LOS have been described to occur primarily through HA CDI. As a result, it is reasonable to assume that a hospital's incidence of CA CDI is unassociated, or at minimum weakly associated, with LOS. Therefore, we can rewrite equation 3.8 as the following:

$$LOS_{i,j} = \alpha_0 + \mathbf{0}\, CDI_j^{\text{CA}} + \alpha_{1,t} CDI_j^{\text{HA}} + \beta_X X_i + \beta_Z Z_j + \epsilon_i \tag{3.12}$$

where CDI_j^{CA} is the incidence of CA CDI at hospital j, and $\mathbf{0}$ simply denotes that fact that the coefficient placed on CDI_j^{CA} should be, roughly, equal to 0, since CA-CDI incidence is assumed to be unrelated to LOS.

Next, let us assume that, while all HA CDI cases receive a secondary CDI diagnosis, a portion of the CA CDI cases receive a secondary CDI diagnosis as well; this was suggested by Dubberke et al. (2010).[38] Let CDI_j^1 and CDI_j^2 denote a hospital's incidence of primary and secondary CDI cases. Since secondary CDI cases may include both HA and a portion of the CA cases, hospital j's secondary CDI incidence can be expressed as the following:

$$CDI_j^2 = CDI_j^{\text{HO-HA}} + \gamma CDI_j^{\text{CA}}, \tag{3.13}$$

where γ is the fraction of HA cases that receive a secondary diagnosis. Similarly, hospital j's primary CDI incidence can be expressed as:

$$CDI_j^1 = \eta CDI_j^{\text{CA}}, \tag{3.14}$$

where η is the fraction of CA cases that receive a primary diagnosis.[5]

We can now specify the following model, which includes both primary and secondary CDI incidence:

$$LOS_{i,j} = \alpha_0 + \tilde{\alpha}_1 CDI_j^1 + \tilde{\alpha}_2 CDI_j^2 + \beta_X X_i + \beta_Z Z_j + \epsilon_i \tag{3.15}$$

[5]Note: In the interest of notational simplicity, equation 3.14 ignores the presence of CO-HA CDI; however, this could be modeled by the inclusion of an error component λ_j, such that $CDI_j^1 = \eta CDI_j^{\text{CA}} + \lambda_j$, which would appear as a hospital specific error component in equation 3.15. However, the relative size and correlation structure of λ_j is largely unknown. As part of the analysis for Objective 2, this study will also analyze the presence of CO-HA CDI cases within principal CDI diagnoses. This secondary analysis will provide some insight into how much the estimation of 3.15 may be biased by the presence of CO-HA CDI.

where $\tilde{\alpha}_1$ and $\tilde{\alpha}_2$ are coefficients on principal and secondary CDI incidence, respectively. By substituting 3.13 and 3.14 into 3.15 and then re-arranging, we can obtain the following:

$$LOS_{i,j} = \alpha_0 + (\tilde{\alpha}_1\eta + \tilde{\alpha}_2\gamma)CDI_j^{\text{CA}} + \tilde{\alpha}_2 CDI_j^{\text{HO-HA}} + \beta_X X_i + \beta_Z Z_j + \epsilon_i \qquad (3.16)$$

Comparing this to equation 3.12 we can see that the coefficient on CDI_j^{CA} should be equal to zero and, thus, we would expect that $\tilde{\alpha}_1\eta + \tilde{\alpha}_2\gamma = 0$, or $\tilde{\alpha}_1 = -\frac{\gamma}{\eta}\tilde{\alpha}_2$. In other words, the estimated coefficient on a hospital's principal CDI incidence, obtained from estimating equation 3.15, should be negative and proportional to the coefficient estimate on secondary CDI incidence. While this procedure may not be able to directly estimate the effect of HA CDI incidence on LOS, it can be used to analyze Objective 1 by providing evidence that HA CDI is the primary driver of the relationship between CDI incidence and LOS.[6]

3.5.2 Identifying and Analyzing Post-Discharge CDI

In order to address the second objective of this study, both the SID and SEDD for California and New York are used to identify cases of CDI that may have occurred post discharge. The following procedure is used to identify such cases. First, for each CDI diagnosis identified as POA, the previous records from that patient are analyzed to determine if the case represents a possible occurrence of post-discharge CDI. A

[6]In reality, equations 3.13 and 3.14 are likely oversimplified and probably should contain hospital specific parameters. For example, CA CDI cases may spillover into secondary diagnoses at different rates across hospitals. In such a setup, the values for γ and η should be hospital specific and expressed as γ_j and η_j. Therefore, the effects obtained by estimating model 3.15 may still be somewhat attenuated.

CDI case is then labeled as a CO-HA if the following conditions are satisfied: (1) the CDI diagnosis is coded POA, (2) the patient had a previous hospital discharge in the 4 weeks prior to the CDI admission, and (3) the CDI was a non-recurrent case, i.e. there were no CDI diagnoses for the patient in the previous 12 weeks. CDI cases that are identified as CO-HA are then attributed to the hospital of their previous admission, since CO-HA CDI cases may be diagnosed at a different hospital than where the CDI was acquired. This method has been used with a similar dataset to identify HA CDI cases occurring post discharge [113]. These CO-HA CDI cases are then used to compute a hospital's incidence of CO-HA and overall HA CDI, as described by equations 3.5 and 3.6. Note that while the SEDD is used to look for post discharge diagnoses of CDI, only those cases found in the SEDD that immediately follow a previous inpatient stay (i.e. the inpatient stay where the infection would be attributed) are included in this analysis.

Once each hospital's CO-HA and overall HA CDI incidence are calculated, they are used to analyze the *post-discharge effect* of CO-HA CDI cases on the relationship between HO-HA CDI incidence and LOS. As previously described, the post-discharge effect is the result that hospitals with longer average LOS are more likely to observe HA CDI cases that occur later in a patient's stay. From equation 3.6 we see that a hospital's overall HA incidence can be decomposed into its incidence of HO-HA CDI and CO-HA CDI, i.e. $CDI_j^{\text{HA}} = CDI_j^{\text{HO-HA}} + CDI_j^{\text{CO-HA}}$ or $CDI_j^{\text{HO-HA}} = CDI_j^{\text{HA}} - CDI_j^{\text{CO-HA}}$. Hence, the estimate of α_1 from equation 3.10 will be influenced by two factors: the overall rate of HA CDI (i.e. CDI_j^{HA}), and the rate at which CDI

cases occur post discharge. While the first factor is driven by the direct theoretical connections between CDI-incidence and LOS, the second factor is driven by the post-discharge effect. Thus, one way to analyze the strength of the post-discharge effect is to simply compare the rate of CO-HA CDI cases between hospitals with shorter and longer LOS. As a preliminary analysis, this study performed a bivariate comparison of CO-HA CDI across hospitals to determine if CO-HA CDI incidence is inversely correlated with a hospital's mean LOS.

The post-discharge effect may seem to imply that we would expect to see a relatively smaller rate of post-discharge CDI cases at hospitals with longer average LOS. However, this relationship may be complicated by the other theoretical links between hospital LOS and CDI incidence, since hospitals with longer LOS are hypothesized to generate more HA CDI. To see the complexity of this relationship, first consider two hospitals A and B such that $LOS_A > LOS_B$ and $CDI_A^{\text{HA}} = CDI_B^{\text{HA}}$. If the post-discharge effect were the only effect in place, hospital A would be able to observe more HA CDI cases prior to discharge, and we would expect $CDI_A^{\text{HO-HA}} > CDI_B^{\text{HO-HA}}$ while $CDI_A^{\text{CO-HA}} < CDI_B^{\text{CO-HA}}$. However, the other theoretical connections between CDI incidence and LOS, which have been described in section 2.3, suggest that hospitals with a greater average LOS will produce more HA CDI. Thus, given two hospitals A and B, such that $LOS_A > LOS_B$ we would expect $CDI_A^{\text{HA}} > CDI_B^{\text{HA}}$. This implies that even if the post-discharge effect exists, we may still have $CDI_A^{\text{CO-HA}} > CDI_B^{\text{CO-HA}}$. Because of this complex relationship, a bivariate comparison of the rates of post-discharge CDIs between hospitals may be insufficient to fully analyze the

post-discharge effect.

Another way to analyze the extent to which the estimate of α_1 in model 3.10 is being driven by the post-discharge effect, versus increased occurrence of HA CDI, is to estimate the following model:

$$LOS_{i,j} = \acute{\alpha}_0 + \acute{\alpha}_1 CDI_j^{\text{HO-HA}} + \acute{\alpha}_2 CDI_j^{\text{CO-HA}} + \beta_X X_i + \beta_Z Z_j + \epsilon_i \tag{3.17}$$

In this model, $\acute{\alpha}_1$ captures the effect of factors associated with LOS that increase HA CDI incidence, and $\acute{\alpha}_2$ captures the post-discharge effect. This model exploits the fact that HA CDI incidence can be decomposed into cases that occur prior to discharge and those which occur after. Thus, equation 3.17 can also be viewed as the result of replacing CDI^{HA} in equation 3.10 with the relationship described in equation 3.6, and then allowing for separate coefficients on $CDI_j^{\text{HO-HA}}$ and $CDI_j^{\text{CO-HA}}$. Notice that since equation 3.10 is essentially being estimated by HO-HA CDI incidence for many of the CDI incidence measures (e.g., secondary CDI or CDI not POA), estimating equation 3.17 tells us exactly how much our other estimates are being influenced by the post-discharge effect. By comparing both the magnitude and significance of the estimates obtained in model 3.17 to those estimated in model 3.10, one can better understand the role that the post-discharge effect has on the relationship between HO-HA CDI incidence and LOS.

3.5.3 Multilevel Hospital Fixed Effects Model

One final model that is analyzed as part of this study is a multilevel model to assess the amount of between-hospital variation in LOS that can be explained by

CDI incidence. If CDI incidence proves to be highly correlated with prolonged LOS in the patient-level analysis, then it is valuable to understand the amount of LOS variation between hospitals that can be captured by CDI incidence. Policy makers, or patients, wishing to compare excess LOS between hospitals may lack the information or resources to properly adjust variation in LOS due to patient characteristics. Thus, a proxy measure for excess LOS, such as CDI incidence, may serve as a marker for making such comparisons. In order for CDI incidence to serve as a useful tool for comparing hospital excess LOS, two conditions should be satisfied: (1) CDI incidence should be highly correlated with variation in patient LOS between different hospitals, (2) a significant portion of this correlation should remain after controlling for differences in patient characteristics.

To analyze the relationship between CDI incidence and between-hospital variation in LOS, a multi-level modeling approach is used. In the first level, individual patient LOS is estimated while controlling for a hospital-specific fixed effect. This model can be expressed as the following:

$$LOS_{i,j} = \beta_X X_i + \beta_{FE,j}\mathbb{I}_j + e_i \tag{3.18}$$

where $\beta_{FE,j}$ are hospital-specific fixed effects and $\mathbb{I}_j$ is a matrix of hospital indicators. In the second level of the model, the hospital-specific fixed effects estimated in (3.18) are explained using CDI incidence along with other hospital characteristics. This second-level model can be expressed as:

$$\hat{\beta}_{FE,j} = \alpha_0 + \alpha_1 CDI_j + \beta_Z Z_j + u_j \tag{3.19}$$

The estimate of α_1 from model (3.19) along with the model's goodness of fit, as measured by its R^2 value, are used to analyze how much of the LOS variation across hospitals can be captured by CDI incidence after accounting for patient characteristics. We also analyze the marginal increase/decrease in model fit that is obtained by adding/removing CDI incidence to this second-level model. This comparison allows us to determine how well CDI incidence can be used to make hospital-level comparisons of unmeasured hospital factors associated with excess LOS. This multi-level modeling approach is used in both the NIS and SID analyses.[7]

3.6 Statistical Analysis

For each year in the NIS and, across years in the two SIDs, descriptive statistics are provided for the number of CDI cases identified along with overall and secondary CDI incidence. In addition, descriptive statistics for the number of CDI cases coded as not POA and post-discharge CDI identified, along with corresponding incidence rates, are also be provided for the SID in California and New York. Bivariate analysis of LOS and each CDI incidence rate described in equations 3.1 - 3.6 are conducted, and Pearson product-moment correlation coefficients are used to analyze trends across hospitals and time. As described in section 3.5.2, a preliminary analysis of the post-discharge effect also compared the incidence of CO-HA CDI between hospitals with shorter and longer LOS.

[7]In the SID analysis, this multilevel model is estimated separately for each year of data. Estimating separate fixed effects for each hospital-year across seven years of data would be too computationally difficult. In addition, the SID does not contain other hospital-level variables so only CDI incidence is used in the second stage

The multivariate statistical analysis is carried out in a similar manner to that of the preliminary study. Because LOS is non-normally distributed and is skewed toward zero, a generalized linear modeling (GLM) approach was used to estimate LOS while comparing a variety of alternative statistical distributions. Specifically, a log link is employed along with a Gaussian, gamma, Poisson, and negative binomial distribution. The quality of these models was then compared using the Akaike information criterion (AIC), and the model with the lowest AIC value was selected as the model to be used for primary analysis. In addition, ordinary least squares (OLS) was used to estimate LOS in order to provide estimates that are "more interpretable." In the multi-level model, both stages are estimated using a standard OLS approach because of computational complexity in the first stage and to obtain easily interpretable R^2 values in the second stage.

All statistical analyses are carried out using STATA SE version 13.1. In the NIS, hospital sample weights are used to produce national level estimates and corrected standard errors. In the SID, robust standard errors are calculated to control for potential unobserved correlation between patient LOS within the same hospital.

CHAPTER 4
RESULTS

This chapter presents results for the three data sources previously described; section 4.1 summarizes results from the NIS, section 4.2 provides results from the SID in California and section 4.3 provides results from the SID in New York. From this point forward, I refer to the empirical CDI cases and corresponding incidence rates according to their coding/measurement definitions (e.g., secondary CDI, CDI not POA or post-discharge CDI) rather than the CDI class they are intended to capture (e.g., HO-HA CDI or CO-HA CDI). Because this research was conducted using observational data, and because of what appear to be systematic differences in the way various hospitals report/code CDI (described in further detail below), I have chosen to make a conscious distinction between theoretical and empirical CDI classification.

4.1 Nationwide Inpatient Sample

The relationship between CDI incidence and patient LOS was first analyzed using the NIS. Table 4.1 summarizes the number of primary and secondary CDI cases, individually and as a proportion of total CDI cases, represented in the NIS across the years 2005-2011. Table 4.1 shows a general increase over time in the number of CDI cases represented in the NIS. Total CDI cases ranged from 61,369 in 2005 to 79,633 in 2011. Both primary and secondary cases increased over this period as well. This corresponds with the general increase in CDI occurrence, which has been widely

Table 4.1: NIS - CDI Case Counts by CDI Type

	Count (% of Total)		
Year	Primary Diagnosis	Secondary Diagnosis	Total Diagnoses
2005	15,549 (25.34)	45,820 (74.66)	61,369
2006	18,337 (28.36)	46,331 (71.64)	64,668
2007	20,955 (32.07)	44,396 (67.93)	65,351
2008	23,182 (32.64)	47,844 (67.36)	71,026
2009	21,733 (32.62)	44,890 (67.38)	66,623
2010	22,294 (32.16)	47,021 (67.84)	69,315
2011	25,753 (32.34)	53,880 (67.66)	79,633

reported in other studies. In 2008 there was a relative spike in CDI cases. One note of interest is that starting in 2007 the proportion of CDI cases that are recorded as a secondary diagnosis becomes relatively stable at around 67-68%; in the two years prior, this proportion was 74.66% and 71.64%, respectively. This appears to coincide with the point in time where some states began requiring hospitals to report incidence of HA CDI [66]; however, this may also be a feature of the sampling criteria used in the NIS.

Table 4.2 reports summary statistics for the various CDI incidence measures across the seven year study period in the NIS. Consistent with the counts reported in Table 4.1, both mean and median CDI incidence is generally increasing over the study period; this appears in both primary and secondary CDI incidence as well. This increasing trend appears more stable when comparing median CDI incidence. The median CDI incidence using all CDI cases was monotonically increasing over the

seven year period, increasing from 0.590% in 2005 to 0.835% in 2011.

A bivariate analysis was conducted to compare the average LOS and CDI incidence across hospitals. Table 4.3 reports summary statistics for hospital-average LOS along with correlation coefficients comparing LOS to both overall and secondary CDI incidence. Although LOS has been decreasing over time in general [3], no clear trend in hospital-average LOS emerges in the NIS over the study period. The typical (i.e. median) hospital had an average LOS between 3.92 and 4.09 days over the study period. As can be seen from Table 4.3, both overall and secondary CDI incidence were positively and significantly correlated with hospital-average LOS. For each year, except 2009, secondary CDI incidence was more strongly correlated with LOS than overall CDI.

The first series of regression models analyzed were fit using simple OLS regression. As described in section 3.4.2 the CDI rates used in the regression analyses were standardized in order to make coefficient estimates comparable between the various measures of CDI incidence. Table 4.4 reports the coefficient estimates for yearly CDI incidence using all CDI cases and only secondary CDI cases, along with corresponding AIC values. Quarterly and yearly CDI incidence rates were compared, and for each year's worth of data, the yearly CDI rates provided a model of better quality, as measured by corresponding AIC values. The OLS results using quarterly rates can be found in Appendix B, and mirror the results reported in Table 4.4.

Table 4.4 shows that in every year CDI incidence using only secondary CDI diagnoses had both a larger coefficient estimate and provided a model of better quality.

Table 4.2: NIS - Yearly CDI Incidence Summary Statistics

CDI Cases	Mean Incidence (std dev)	Median (IQR)	Total Hospitals
2005			
All	0.962 (1.723)	0.590 (0.330-0.968)	
Secondary	0.744 (1.660)	0.401 (0.191-0.700)	946
Primary	0.215 (0.204)	0.164 (0.087-0.288)	
2006			
All	1.092 (1.854)	0.651 (0.374-1.057)	
Secondary	0.818 (1.762)	0.404 (0.213-0.730)	941
Primary	0.272 (0.286)	0.215 (0.112-0.352)	
2007			
All	1.080 (1.664)	0.706 (0.418-1.083)	
Secondary	0.761 (1.581)	0.404 (0.213-0.691)	957
Primary	0.318 (0.293)	0.254 (0.145-0.415)	
2008			
All	1.256 (2.309)	0.715 (0.417-1.150)	
Secondary	0.923 (2.236)	0.421 (0.208-0.722)	970
Primary	0.330 (0.315)	0.264 (0.150-0.417)	
2009			
All CDI	1.325 (2.702)	0.716 (0.436-1.112)	
Secondary CDI	0.987 (2.584)	0.417 (0.215-0.686)	947
Primary CDI	0.335 (0.309)	0.272(0.164-0.400)	
2010			
All	1.268 (2.026)	0.741 (0.456-1.171)	
Secondary	0.921 (1.934)	0.433 (0.244-0.730)	960
Primary	0.343 (0.385)	0.267 (0.157-0.440)	
2011			
All	1.395 (2.232)	0.835 (0.521-1.285)	
Secondary	1.002 (2.136)	0.491 (0.258-0.826)	957
Primary	0.392 (0.340)	0.306 (0.188-0.498)	

Table 4.3: NIS - Mean LOS and CDI Incidence Correlation

	Hospital Average LOS		Correlation Coefficient	
Year	Mean (Std Dev.)	Median (IQR)	Overall CDI	Secondary CDI
2005	5.15 (4.93)	4.09 (3.41-4.88)	0.688***	0.691***
2005	5.08 (4.72)	3.99 (3.36-4.76)	0.828***	0.848***
2007	5.12 (5.09)	3.97 (3.31-4.78)	0.727***	0.753***
2008	5.33 (5.34)	4.04 (3.38-4.79)	0.766***	0.773***
2009	5.56 (5.73)	4.01 (3.34-4.80)	0.759***	0.756***
2010	5.46 (5.53)	4.03 (3.36-4.79)	0.849***	0.869***
2011	5.51 (5.72)	3.92 (3.31-4.76)	0.783***	0.800***

Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Depending on the year, a theoretical patient was estimated to stay between 0.8937 and 1.9107 days longer at a hospital with a CDI incidence that was one standard deviation higher. Using only secondary CDI cases, this estimated effect was greater, ranging from 1.2629 to 2.3862. This is consistent with our hypothesis that the observed relationship occurs via HA CDI and that secondary CDI incidence is more representative of HA CDI incidence. Based on the estimated effect of CDI incidence, there appear to be two time periods where the relationship between CDI incidence and LOS was different. Prior to 2008 the estimated effect of CDI incidence ranged from 0.8937-1.1931 and 1.2629-1.5290 using all and secondary CDI cases, respectively. However, after the start of 2008 these effects were relatively greater ranging from 1.7070-1.9107, using all cases, and 2.0042 -2.3862, using secondary, cases. CDI incidence appears to have had the greatest effect in 2010 and the weakest in 2007.

Table 4.4: NIS - OLS Results By Year

CDI cases included	Coefficent (Std. Error)	AIC
2005 (N=6,655,017)		
All CDI	1.1305 (0.0062)***	40,438,298
Secondary Diagnosis	1.4297 (0.0068)***	40,427,384
2006 (N=6,752,096)		
All CDI	1.1931 (0.0057)***	40,437,348
Secondary Diagnosis	1.5290 (0.0062)***	40,420,882
2007 (N=5,971,627)		
All CDI	0.8937 (0.0058)***	35,971,388
Secondary Diagnosis	1.2629 (0.0065)***	35,958,216
2008 (N=6,272,659)		
All CDI	1.7070 (0.0066)***	37,876,432
Secondary Diagnosis	2.0042 (0.0072)***	37,866,312
2009 (N=5,781,425)		
All CDI	1.8111 (0.0075)***	34,457,666
Secondary Diagnosis	2.0878 (0.0079)***	34,446,564
2010 (N=5,926,484)		
All CDI	1.9107 (0.0067)***	35,796,210
Secondary Diagnosis	2.3537 (0.0073)***	35,772,114
2011 (N=6,097,267)		
All CDI	1.8864 (0.0066)***	36,977,966
Secondary Diagnosis	2.3862 (0.0073)***	36,952,030

Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table 4.5: NIS - GLM (Gamma) Results By Year

CDI cases included	Coefficent (Std. Error)	AIC
2005 (N=6,655,017)		
All CDI	0.0957 (0.0009)***	30,892,118
Secondary Diagnosis	0.1175 (0.0010)***	30,890,172
2006 (N=6,752,096)		
All CDI	0.0812 (0.0008)***	31,247,718
Secondary Diagnosis	0.1028 (0.0009)***	31,245,232
2007 (N=5,971,627)		
All CDI	0.0684 (0.0008)***	27,570,104
Secondary Diagnosis	0.0952 (0.0009)***	27,567,598
2008 (N=6,272,659)		
All CDI	0.1284 (0.0010)***	29,086,056
Secondary Diagnosis	0.1470 (0.0011)***	29,085,108
2009 (N=5,781,425)		
All CDI	0.1318 (0.0012)***	26,806,486
Secondary Diagnosis	0.1558 (0.0013)***	26,804,830
2010 (N=5,926,484)		
All CDI	0.1448 (0.0011)***	27,560,036
Secondary Diagnosis	0.1773 (0.0012)***	27,556,048
2011 (N=6,097,267)		
All CDI	0.1265 (0.0009)***	28,319,716
Secondary Diagnosis	0.1612 (0.0010)***	28,316,058

Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

In the regression model, containing over 400 covariates, CDI incidence was one of the strongest predictive variables in the regression. Across the years, secondary CDI incidence was between the second and fifth strongest predictor of LOS, as measured by the relative size of its test statistic; for 2008 and later, it was the second strongest variable.

For each year, the model that provided the best quality fit was the GLM model with a log link and Gamma distribution. Table 4.5 reports the results of this model using yearly CDI incidence.[1] Consistent with the general trends depicted in the OLS findings, CDI incidence had a relatively greater estimated effect, and provided better model quality, when only secondary CDI cases were included. On average, a theoretical patient was estimated to stay between 7.08 and 15.58 percent longer at a hospital with a total CDI incidence that was one standard deviation higher than another.[2] Using only secondary CDI incidence, this estimated effect was greater, ranging from 9.99 to 19.40 percent. In addition, the estimated effects appear to be relatively greater after 2008 than before. In all of the models, similar to the OLS results, secondary CDI incidence was the fifth strongest predictor of LOS, as measured by the relative size of its test statistic.

In section 3.5.1, it was hypothesized that controlling for a hospital's primary CDI incidence, in addition to secondary incidence, would allow secondary incidence to better capture HA CDI by helping to remove spillover of CA CDI cases recorded as secondary. Table 4.6 presents three different models with coefficients for (1) overall CDI incidence, (2) only secondary CDI incidence, and (3) primary and secondary CDI incidence separately. Note that in order for the relationship described in section 3.5.1 to hold, unstandardized CDI incidence must be used in the model. Thus, the results

[1]Similar to the OLS results, yearly CDI incidence performed much better that CDI incidence calculated at a quarterly level. However, the quarterly level results can also be found in Appendix B.

[2]These interpreted percentage effects are "roughly" estimated by exponentiating the GLM regression coefficients reported in Table 4.5.

Table 4.6: NIS - Using Diagnosis Order to Control for CA CDI

	Coefficent (Std. Error)		
CDI Cases	Model 1	Model 2	Model 3
2005			
Any	0.6802 (0.0037)***	-	-
Secondary	-	0.8985 (0.0043)***	1.1371 (0.0048)***
Primary	-	-	-1.9625 (0.0178)***
$\boldsymbol{R^2}$	0.4116	0.4126	0.4137
2006			
Any	0.6671 (0.0032)***	-	-
Secondary	-	0.9054 (0.0037)***	1.1727 (0.0041)***
Primary	-	-	-2.0233 (0.0146)***
$\boldsymbol{R^2}$	0.4333	0.4347	0.4363
2007			
Any	0.5546 (0.0036)***	-	-
Secondary	-	0.8320 (0.0043)***	1.1147 (0.0049)***
Primary	-	-	-1.7011 (0.0139)***
$\boldsymbol{R^2}$	0.4248	0.4261	0.4275
2008			
Any	0.7641 (0.0030)***	-	-
Secondary	-	0.9316 (0.0034)***	1.0019 (0.0036)***
Primary	-	-	-0.6664 (0.0130)***
$\boldsymbol{R^2}$	0.4181	0.4191	0.4193
2009			
Any	0.7014 (0.0029)***	-	-
Secondary	-	0.8497 (0.0032)***	0.9347 (0.0034)***
Primary	-	-	-0.9964 (0.0133)***
$\boldsymbol{R^2}$	0.4259	0.4270	0.4275
2010			
Any	0.9716 (0.0034)***	-	-
Secondary	-	1.2633 (0.0039)***	1.4317 (0.0041)***
Primary	-	-	-1.6020 (0.0136)***
$\boldsymbol{R^2}$	0.4355	0.4378	0.4391
2011			
Any	0.8738 (0.0031)***	-	-
Secondary	-	1.1648 (0.0036)***	1.3739 (0.0039)***
Primary	-	-	-1.6615 (0.0125)***
$\boldsymbol{R^2}$	0.4195	0.4220	0.4236

Note: These results use unstandardized CDI incidence rates and were described in section 3.5.1.

Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

presented Table 4.6 use raw, unstandardized CDI incidence rates. As hypothesized, for each year the model fit improves, and the estimated effect of interest increases, when moving from using all CDI cases to only secondary CDI cases (i.e. Model 1 to Model 2). Similarly, the model fit and estimated effect increased when primary CDI incidence was added to the model (Model 2 to Model 3). As was expected, the coefficient on primary CDI incidence is negative and of similar relative magnitude to the coefficient on secondary CDI incidence. These results support the hypothesis that the observed relationship between CDI incidence and increased LOS is being driven by cases of HA CDI.

The final model estimated using the NIS was the multilevel model described in section 3.5.3. In the first stage of this model, only patient characteristics along with a hospital-specific fixed effect are included. In the second level, the hospital-specific fixed effects, estimated in the first stage, are predicted using CDI incidence. Table 4.7 presents the results of two different specifications of this second-level model. First, the coefficients and R-squared values are presented for the regression of hospital fixed effects onto only CDI incidence. Using the CDI-only model, CDI incidence explained 41.31-62.96% of the variation in hospital fixed effects, when all CDI cases were used, and 41.78-65.96% of the variation when only secondary cases were used.[3] The second set of results presented in Table 4.7 is from the model where hospital fixed effects are regressed on CDI incidence and other hospital characteristics. The R^2 values

[3]These values correspond to the range in R^2 values across the various years using overall and secondary CDI incidence, respectively

reported in the second set of results correspond to the increase in the R^2 coefficient that occurs when CDI incidence is added to the model containing other hospital characteristics. When CDI incidence is added to the model the amount of variation between hospitals that could be explained increased by 31.47-47.90 percentage points, with all CDI cases, and 31.95-50.50 percentage points, when only secondary cases were used. In both models, secondary CDI incidence explains a greater amount of variation in hospital fixed effects than overall CDI incidence, with the exception of one year; in 2009 overall CDI incidence explained a slightly greater amount of variation between hospitals.

Table 4.7: NIS - Results of Multilevel Model

CDI Cases	CDI Only Coefficent (SE)	R^2	CDI + Hospital Characteristics Coefficent (SE)	R^2 Increase
2005				
All	2.2892 (0.0930)***	0.4131	2.1324 (0.0979)***	0.3147
Secondary	2.2836 (0.0919)***	0.4178	2.1282 (0.0966)***	0.3195
2006				
All	2.7334 (0.0730)***	0.6196	2.5657 (0.0766)***	0.4656
Secondary	2.7599 (0.0699)***	0.6442	2.5914 (0.0733)***	0.4874
2007				
All	2.6851 (0.0961)***	0.4726	2.4811 (0.1061)***	0.3457
Secondary	2.7265 (0.0928)***	0.4980	2.5271 (0.1026)***	0.3666
2008				
All	2.9175 (0.0839)***	0.5774	2.6231 (0.0906)***	0.3960
Secondary	2.9121 (0.0833)***	0.5804	2.6203 (0.0887)***	0.4036
2009				
All	2.8575 (0.0989)***	0.4905	2.4831 (0.1075)***	0.3152
Secondary	2.8068 (0.0982)***	0.4849	2.4326 (0.1062)***	0.3120
2010				
All	3.2721 (0.0842)***	0.6296	3.0290 (0.0891)***	0.4790
Secondary	3.3212 (0.0801)***	0.6596	3.0860 (0.0849)***	0.5050
2011				
All	3.1430 (0.0991)***	0.5329	2.8127 (0.1014)***	0.3687
Secondary	3.1957 (0.0966)***	0.5539	2.8556 (0.0986)***	0.3845

Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

4.2 State Inpatient Sample: California

The second analysis that was conducted used the SID from the state of California. The SID represents a complete population of inpatient visits in the state of California, and records from a single patient can be linked between revisits. Diagnostic codes in the California SID contain POA indicators; these were used to identify cases of CDI that were marked as HA by hospitals. As described in section 3.5.2, patient revisits in the SID, along with the SEDD, were used to identify cases of CDI that were HA but occurred post discharge. Table 4.8 summarizes the number, and proportions, of different CDI cases identified. Similar to the NIS results, the number of CDI cases per year increased over the study window from 24,448 in 2005 to 38,248 in 2011. Both primary and secondary cases grew over this period as well. One obvious feature of these data relates to the number of CDI cases that were identified by hospitals as not POA. Beginning in 2008 the number of CDI cases marked as not POA increased dramatically; from 2005-2007 slightly over 10% of all CDI cases were marked not POA and from 2008-2011 around a quarter of CDI cases were marked not POA. As previous discussed, this pattern roughly corresponds with the fact that in 2009 California began requiring hospitals to report cases of HA CDI. Although the timing does not exactly align with when the law went into full effect in 2009, hospitals likely began preparing to comply with the law in 2008. Finally, the number of post-discharge cases of CDI appeared to decrease over the study period, in both absolute terms and as a percentage of total CDI cases. Post-discharge cases made up 33.30 percent of all CDI cases in 2005 but only 20.74 percent in 2011.

Table 4.8: California SID - CDI Case Counts by CDI Type

Year	Any CDI	Count (% of Total CDI Cases) Secondary	Primary	Not POA	Not POA + Post Discharge	Post Discharge
2005	24,448	19,242 (78.71)	5,206 (21.29)	2,598 (10.63)	10,740 (43.93)	8,142 (33.30)
2006	29,861	22,796 (76.34)	7,065 (23.66)	3,009 (10.08)	12,533 (41.97)	9,524 (31.89)
2007	31,219	22,758 (72.90)	8,461 (27.10)	3,288 (10.53)	12,764 (40.89)	9,476 (30.35)
2008	34,480	25,190 (73.06)	9,290 (26.94)	8,987 (26.06)	17,173 (49.81)	8,186 (23.74)
2009	33,502	24,505 (73.14)	8,997 (26.86)	8,700 (25.97)	16,559 (49.43)	7,859 (23.46)
2010	34,473	25,141 (72.93)	9,332 (27.07)	8,392 (24.34)	16,407 (47.59)	8,015 (23.25)
2011	38,248	28,087 (73.43)	10,161 (26.57)	8,815 (23.05)	16,748 (43.79)	7,933 (20.74)
Total	226,231	167,719 (74.14)	58,512 (25.86)	43,789 (19.36)	102,924 (45.50)	59,135 (26.14)

Note: For each CDI category, the parenthesized percentages correspond to the percent of all CDI that are contained in that category (i.e. the reported count divided by the total number of CDI patients reported in the second column)

Table 4.9 reports summary statistics for the various CDI incidence measures across the seven year study period in California. Consistent with the counts reported in Table 4.8, both the mean and median CDI incidence were generally increasing over the time window; this appears in both overall CDI incidence and secondary CDI incidence. On the other hand, post-discharge CDI incidence appears relatively constant over the study period. With the exception of 2006 and 2007, mean (median) post-discharge CDI incidence ranged from 0.200% (0.187%) to 0.207% (0.191%); in 2006 and 2007 mean (median) post-discharge-CDI incidence was 0.242% (0.221%) and 0.240% (0.214%) respectively. Consistent with the trend depicted in Table 4.8, CDI labelled as not POA increased dramatically beginning in 2008. Both the mean and median incidence rates of CDI not POA increased roughly 3-fold beginning in 2008. Prior to 2008 not POA-CDI incidence was less than 0.1% and after 2008 it was greater than 0.2%.

Table 4.9: California SID - Yearly CDI Incidence Summary Statistics

CDI Incidence	Any CDI	Secondary CDI	Not POA CDI	Post Discharge CDI	Post Discharge CDI + Not POA
2005 (N=403)					
Mean (Std. Dev.)	0.623 (0.535)	0.490 (0.495)	0.066 (0.117)	0.207 (0.136)	0.273 (0.223)
Median (IQR)	0.576 (0.351-0.830)	0.429 (0.273-0.623)	0.050 (0.025-0.086)	0.191 (0.123-0.253)	0.261 (0.164-0.342)
2006 (N=400)					
Mean (Std. Dev.)	0.759 (0.650)	0.580 (0.596)	0.076 (0.115)	0.242 (0.163)	0.319 (0.243)
Median (IQR)	0.684 (0.425-0.966)	0.507 (0.308-0.734)	0.056 (0.026-0.096)	0.221 (0.137-0.318)	0.285 (0.177-0.416)
2007 (N=402)					
Mean (Std. Dev.)	0.791 (0.710)	0.576 (0.646)	0.083 (0.128)	0.240 (0.162)	0.323 (0.257)
Median (IQR)	0.740 (0.453-0.983)	0.511 (0.306-0.710)	0.064 (0.032-0.100)	0.214 (0.149-0.296)	0.304 (0.187-0.389)
2008 (N=397)					
Mean (Std. Dev.)	0.872 (0.792)	0.637 (0.727)	0.227 (0.326)	0.207 (0.128)	0.434 (0.406)
Median (IQR)	0.797 (0.499-1.082)	0.582 (0.354-0.778)	0.176 (0.101-0.296)	0.191 (0.139-0.269)	0.385 (0.264-0.548)
2009 (N=395)					
Mean (Std. Dev.)	0.855 (0.782)	0.625 (0.728)	0.222 (0.322)	0.200 (0.122)	0.422 (0.403)
Median (IQR)	0.780 (0.523-1.035)	0.564 (0.354-0.725)	0.177 (0.114-0.261)	0.187 (0.132-0.257)	0.394 (0.252-0.520)
2010 (N=387)					
Mean (Std. Dev.)	0.884 (0.834)	0.644 (0.773)	0.215 (0.329)	0.205 (0.123)	0.420 (0.400)
Median (IQR)	0.795 (0.559-1.078)	0.577 (0.358-0.778)	0.182 (0.098-0.256)	0.184 (0.136-0.258)	0.390 (0.247-0.529)
2011 (N=388)					
Mean (Std. Dev.)	0.992 (0.813)	0.728 (0.740)	0.228 (0.305)	0.205 (0.118)	0.434 (0.376)
Median (IQR)	0.941 (0.640-1.201)	0.667 (0.439-0.890)	0.184 (0.115-0.286)	0.191 (0.139-0.259)	0.393 (0.273-0.525)

Table 4.10: California SID - Mean LOS and CDI Incidence Correlation

	Hospital Mean LOS[a]		Correlation with CDI Incidence		
Year	Mean (Std Dev.)	Median (IQR)	All Cases	Secondary	Not POA
2005	4.60 (8.26)	4.38 (3.72-5.31)	0.501***	0.505***	0.491***
2006	4.58 (8.33)	4.34 (3.77-5.23)	0.434***	0.444***	0.322***
2007	4.55 (8.26)	4.33 (3.73-5.28)	0.557***	0.567***	0.468***
2008	4.56 (8.32)	4.37 (3.75-5.29)	0.568***	0.573***	0.465***
2009	4.48 (8.11)	4.36 (3.70-5.26)	0.413***	0.421***	0.379***
2010	4.45 (7.95)	4.26 (3.62-5.19)	0.389***	0.399***	0.354***
2011	4.47 (7.98)	4.22 (3.63-5.34)	0.477***	0.487***	0.408***

[a]These values correspond to the mean and median across hospitals of each hospital's mean LOS.

Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Similar to the NIS analysis, a bivariate comparison was made between the average LOS and CDI incidence across hospitals. Table 4.10 reports summary statistics for hospital-average LOS and correlation coefficients with incidence rates of overall, secondary, and CDI not POA. In general, there appears to be a slight downward trend in hospital average LOS in California over the study period. Average LOS for a typical (i.e. median) hospital decreased from 4.38 days to 4.22 days over the study period. Table 4.10 shows that CDI incidence was positively and significantly correlated with average LOS, using all cases, secondary cases or cases not POA. For every year, this correlation was strongest using secondary CDI incidence.

As described in section 3.6, a variety of models used to predict patient LOS were compared. Similar to the NIS results, the model that provided the best quality

fit was the GLM model with a log link and Gamma distribution. Table 4.11 provides the estimated effects of interest for both the best fitting gamma-GLM model along with the standard OLS model (for ease of interpretation). Models using both yearly and quarterly CDI incidence were fit and compared. As with the NIS results, yearly incidence rates provided the best quality model fit. Results presented here utilize yearly incidence; those using quarterly rates can be found in Appendix B.

As can be seen from Table 4.11, in both specifications secondary CDI incidence provided the largest estimated effect and best quality model of the various incidence rates. The OLS model suggests that a patient stayed 2.1089 days longer, on average, at a hospital with a secondary CDI incidence that was one standard deviation higher. In the chosen GLM model this effect was roughly equivalent to staying 10.24% longer.[4] The CDI incidence definition that provided the 2nd best fit, and 2nd largest effect estimate, was using all CDI cases in the OLS model but in the GLM model was using CDI cases marked not POA. In both the OLS and GLM results, CDI incidence using cases identified as not POA (with or without post-discharge cases), performed worse than secondary CDI. One result of particular interest is that post-discharge CDI incidence is positively and significantly associated with increased LOS. This suggests that the effect being captured by CDI incidence is likely outweighing any post discharge effect. However, the model quality and estimated effects diminished when post discharge cases were added to CDI cases marked not POA.

[4]These interpreted percentage effects are "roughly" estimated by exponentiating the GLM regression coefficients reported in Table 4.11.

Table 4.11: California SID - Regression Results

CDI cases included	Coefficent (Std. Error)	AIC
OLS		
All CDI	1.7917 (0.0045)***	172,476,554
Secondary Diagnosis	2.1089 (0.0047)***	172,440,732
Not POA	1.7866 (0.0047)***	172,490,130
Post Discharge	0.7640 (0.0031)***	172,567,408
Not POA + Post Discharge	1.6669 (0.0043)***	172,481,394
GLM (gamma)		
All CDI	0.0702 (0.0006)***	118,491,438
Secondary Diagnosis	0.0975 (0.0006)***	118,482,766
Not POA	0.0853 (0.0007)***	118,488,152
Post Discharge	0.0340 (0.0004)***	118,498,210
Not POA + Post Discharge	0.0801 (0.0006)***	118,485,990

Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Because a very clear change point was identified from Tables 4.8 and 4.9 beginning in 2008, and because 2009 marked the start of mandatory reporting of HA CDI in California, a second set of models were fit where CDI incidence was interacted with an indicator for time post 2008. This model was described in section 3.5 and results of this model are reported in Table 4.12.[5] In all cases, dividing the coefficient of interest into two values, before and after 2008, resulted in increased model quality. Moreover, the estimated effects showed consistent difference between these time periods. The estimated effect and relative significance level of CDI incidence increased in the period after the start of 2008 for CDI incidence measured using all CDI cases,

[5]In section 3.5 it was suggested that the model be allowed to vary before and after 2009 based on the policy implementation. However, the data very clearly indicate the change point occurred in 2008, so this was the cutoff point that was used in the final analysis.

secondary cases, and only those marked as not POA. This is most notable for CDI incidence using not POA, which increased from 1.5694 to 1.9379 in the OLS model and from 0.0780 to 0.901 in the GLM model. Although CDI incidence using only secondary CDI cases remained the best quality model, after 2008 the estimated effects using CDI cases not POA (i.e. 1.9379 and 0.0901) moved closer to the estimates using only secondary cases (i.e. 2.1597 and 0.0964). In addition, the estimated effect of CDI incidence using post-discharge cases decreased for 2008 onward.

Table 4.12: California SID - Regression Results Split By Policy Change Point

	Coefficent (Std. Error)		
CDI cases included	Pre 2008	Post 2008	AIC
OLS			
All CDI	1.7046 (0.0196)***	1.8549 (0.0164)***	172,476,270
Secondary Diagnosis	2.0364 (0.0222)***	2.1597 (0.0181)***	172,440,562
Not POA	1.5694 (0.0218)***	1.9379 (0.0187)***	172,488,636
Post Discharge	1.0478 (0.0136)***	0.5739 (0.0079)***	172,561,560
Not POA + Post Discharge	1.5520 (0.0180)***	1.7692 (0.0164)***	172,480,720
GLM (gamma)			
All CDI	0.0697 (0.0010)***	0.0705 (0.0009)***	118,491,440
Secondary Diagnosis	0.0991 (0.0011)***	0.0964 (0.0009)***	118,482,764
Not POA	0.0780 (0.0011)***	0.0901 (0.0008)***	118,488,076
Post Discharge	0.0611 (0.0011)***	0.0148 (0.0007)***	118,494,994
Not POA + Post Discharge	0.0851 (0.0012)***	0.0757 (0.0008)***	118,485,924

Note: Estimates obtained using the final model described in section 3.5.
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

As described by section 3.5.3, a multilevel model was used to estimate the amount of variation in LOS between hospitals that was explained by CDI incidence. Table 4.13 presents the results for the second stage of the multilevel model.[6] The SID did not contain the other hospital variables controlled for in the NIS results; thus, the results in Table 4.13 correspond to a second-stage model with hospital CDI incidence as the only explanatory variable. For all years, other than 2005, CDI incidence using secondary CDI explained the greatest amount of variation among hospitals. For the years after 2005, secondary CDI incidence explained between 14.88% to 28.53% of the variation in hospital fixed effects. In 2005, not-POA CDI cases were able to explain the greatest amount of variation at around 12.49%. Although post-discharge CDI incidence demonstrated a significant positive relationship with hospital fixed effects, it consistently explained the least amount of variation ranging from 3.32% to 6.24%.

[6]Models for hospital fixed effects were estimated separately for each year due to the extreme computational challenge of estimating such a large number of separate fixed effects by hospital and year.

Table 4.13: California SID - Results From Second-Level Multilevel Model

CDI cases included	All CDI	Secondary	Not POA	Post Discharge	Not POA + Post Discharge
2005 (N=399)					
Coefficent (SE)	1.4064 (.2262)***	1.7205 (.2356)***	1.9913 (.2645)***	0.6847 (.1707)***	1.4099 (.2203)***
R^2	0.0887	0.1184	**0.1249**	0.0389	0.0936
2006 (N=395)					
Coefficent (SE)	1.7056 (.2405)***	2.0774 (.2506)***	1.1345 (.2452)***	0.7326 (.1777)***	1.0754 (.2145)***
R^2	0.1134	**0.1488**	0.0517	0.0415	0.0601
2007 (N=398)					
Coefficent (SE)	1.7835 (.2147)***	2.1014 (.2219)***	1.6808 (.2305)***	0.8785 (.1712)***	1.3732 (.1991)***
R^2	0.1484	**0.1846**	0.1184	0.0624	0.1072
2008 (N=392)					
Coefficent (SE)	1.8393 (.2163)***	2.2107 (.2238)***	1.8243 (.2505)***	0.5801 (.1585)***	1.5768 (.2271)***
R^2	0.1565	**0.2002**	0.1197	0.0332	0.1100
2009 (N=389)					
Coefficent (SE)	2.1655 (.1965)***	2.4412 (.1967)***	2.3627 (.2046)***	0.8433 (.1711)***	2.1425 (.1984)***
R^2	0.2389	**0.2848**	0.2562	0.0590	0.2316
2010 (N=383)					
Coefficent (SE)	2.3639 (.2123)***	2.6399 (.2141)***	2.4651 (.2167)***	0.6373 (.1658)***	2.2286 (.2120)***
R^2	0.2455	**0.2853**	0.2536	0.0373	0.2248
2011 (N=384)					
Coefficent (SE)	1.9625 (.2065)***	2.2549 (.2110)***	2.1852 (.2131)***	0.5942 (.1548)***	1.9574 (.2055)***
R^2	0.1913	**0.2302**	0.2159	0.0371	0.1920

Notes: Standard errors in parentheses; largest R^2 in bold; significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

One of the primary goals of analyzing inpatient visits using the SID was to capture post-discharge CDI cases and attempt to control for the post-discharge effect. It was hypothesized that hospitals with longer LOS might appear to have greater rates of HO-HA CDI and secondary CDI simply because such hospitals would have less CDI occurring post discharge. Thus, we might expect to see greater rates of CO-HA CDI at hospitals with shorter LOS. This hypothesis was first analyzed at a bivariate level by comparing post-discharge CDI incidence and hospital mean LOS across years and hospitals. Contrary to this initial hypothesis, hospitals with longer LOS actually had higher incidence rates of post-discharge CDI. Hospital mean LOS and post-discharge CDI incidence rates were positively and significantly correlated (ρ=0.2495, *P-value* $<$ .0001).[7] This positive correlation existed across and within years. A simple OLS regression analysis suggests that for each day increase in a hospital's mean LOS, post-discharge CDI incidence increased by .0061 percentage points (P<.001). These results suggest that if the post-discharge effect exists, it is being masked by a much greater effect: hospitals with longer LOS generate more HO-HA CDI and CO-HA CDI than hospitals with shorter LOS.

A multivariate analysis was also conducted to test for the presence of a post-discharge effect. This analysis was described in section 3.5.2 and, essentially, adds post-discharge CDI incidence to a model with a marker for HO-HA CDI. Because secondary CDI and CDI coded not POA seem most likely to represent HO-HA CDI and tended to perform the best in the previous results, these two definitions were

[7]Using Pearson product-moment correlation coefficient.

Table 4.14: California SID - Controlling For Post-Discharge Effect

CDI Cases Used	Coefficent (Std. Error)		AIC
	CDI Incidence	Post-Discharge CDI	
OLS			
Secondary	2.5176 (0.0223)***	-0.3839 (0.0119)***	172,433,442
Not POA	1.5993 (0.0163)***	0.2577 (0.0082)***	172,485,226
GLM			
Secondary	0.1207 (0.0009)***	-0.0209 (0.0006)***	118,481,602
Not POA	0.0782 (0.0007)***	0.0090 (0.0005)***	118,487,826

Notes: Standard errors in parentheses;
significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

used to represent HO-HA CDI in the model. Table 4.14 presents the results using the either secondary CDI cases or CDI cases not POA. The multivariate analysis tells a conflicting story. When secondary CDI cases are used to calculate CDI incidence, both the OLS and GLM specifications estimate the coefficient on post-discharge CDI to be negative, while the estimated effect of secondary CDI incidence increases (from the values shown in Table 4.11). These findings suggest that the hypothesized post-discharge effect may be occurring. On the other hand, when only CDI cases coded as not POA are used to calculate CDI incidence, the coefficient estimate on post-discharge CDI is positive, while the coefficient on CDI incidence decreases. These results using CDI not POA fail to capture the post-discharge effect hypothesized. One other issue of note is that the quality of the model, as measured by AIC values, increased when post-discharge CDI was added to the model.

4.3 State Inpatient Sample: New York

The final analysis was conducted using the SID from the state of New York. Like the SID in California, the New York SID represents a complete population of inpatient visits in the state of New York, and visits from a single patient can be linked between stays. The same procedure was used to identify cases of post-discharge CDI and CDI not POA. Table 4.15 provides counts of the different CDI cases identified. Similar to the NIS and California results, the number of CDI cases per year increased over the study window, from 23,907 in 2005 to 28,926 in 2011, with spikes of 29,365 cases in 2008 and 28,385 in 2009. Both primary and secondary cases grew over this period as well. Similar to California, in New York beginning in 2008 there was a significant increase in the number of CDI cases identified by hospitals as not POA. The number of cases coded as not POA increased from around 23% of all CDI cases, prior to 2008, to around 29% of CDI cases beginning in 2008. Again, this pattern roughly corresponds with the timing in 2009 when the state of New York began requiring hospitals to report cases of HA CDI. It is worth noting that the change in the proportion of CDI cases coded as not POA was much smaller than the change in California. Additionally, in all years, New York hospitals recorded a greater percentage of CDI cases as not POA than did hospitals in California. Finally, the proportion of CDI cases that occurred post discharge appeared to decrease over the study period. Post-discharge cases made up 27.11% of all CDI cases in 2005 but only 19.96% in 2011. The proportion of CDI cases recorded as secondary was roughly equivalent in New York (74.06%) as in California (74.14%); however, considerably

more CDI cases were recorded as not POA in New York (26.79%) than in California (19.36%). Post-discharge CDI also appeared slightly less common in New York than in California.

Table 4.15: New York SID - CDI Case Counts by CDI Type

Year	Count (% of Total CDI Cases)					
	Any CDI	Secondary	Primary	Not POA	Not POA + Post Discharge	Post Discharge
2005	23,907	18,794 (78.61)	5,113 (21.39)	5,666 (23.70)	12,146 (50.81)	6,480 (27.11)
2006	24,949	18,985 (76.10)	5,964 (23.90)	5,711 (22.89)	12,281 (49.22)	6,570 (26.33)
2007	25,789	18,786 (72.85)	7,003 (27.15)	6,325 (24.53)	12,770 (49.52)	6,445 (24.99)
2008	29,365	21,348 (72.70)	8,017 (27.30)	8,757 (29.82)	15,485 (52.73)	6,728 (22.91)
2009	28,385	20,686 (72.88)	7,699 (27.12)	8,317 (29.30)	14,784 (52.08)	6,467 (22.78)
2010	27,995	20,377 (72.79)	7,618 (27.21)	7,875 (28.13)	14,034 (50.13)	6,159 (22.00)
2011	28,926	21,225 (73.38)	7,701 (26.62)	8,072 (27.91)	13,845 (47.86)	5,773 (19.96)
Total	189,316	140,201 (74.06)	49115 (25.94)	50,723 (26.79)	95,345 (50.36)	44,622 (23.57)

Notes: In 2005, post-discharge CDI cases could not be identified using the SEDD.

Table 4.16 reports summary statistics for the various CDI incidence measures across the seven year study period. Consistent with the counts reported in Table 4.15, the mean and median CDI incidence were generally increasing for both primary and secondary CDI incidence. On the other hand, post-discharge-CDI incidence remained relatively constant, with the mean varying between 0.226% and 0.258% and median varying between 0.217% and 0.239%. Consistent with the trend depicted in Table 4.15, CDI coded as not POA increased beginning in 2008. Both the mean and median incidence of CDI not POA increase by around a percentage point beginning in 2008. Before 2008 mean CDI incidence using cases not POA was slightly over 0.2%, and after 2008 it was slightly over 0.3%. All CDI incidence rates were generally greater in New York than in California.

Table 4.16: New York SID - Yearly CDI Incidence Summary Statistics

CDI Incidence	Any CDI	Secondary CDI	Not POA CDI	Post Discharge CDI	Post Discharge CDI + Not POA
2005 (N=220)					
Mean (Std. Dev.)	0.914 (0.569)	0.718 (0.487)	0.216 (0.226)	0.247 (0.144)	0.464 (0.313)
Median (IQR)	0.817 (0.553-1.221)	0.609 (0.408-0.970)	0.199 (0.050-0.295)	0.225 (0.143-0.314)	0.416 (0.234-0.622)
2006 (N=219)					
Mean (Std. Dev.)	0.949 (0.584)	0.722 (0.505)	0.217 (0.200)	0.250 (0.140)	0.467 (0.294)
Median (IQR)	0.854 (0.614-1.151)	0.622 (0.477-0.863)	0.189 (0.059-0.297)	0.239 (0.149-0.305)	0.425 (0.277-0.574)
2007 (N=223)					
Mean (Std. Dev.)	0.998 (0.581)	0.727 (0.488)	0.245 (0.182)	0.249 (0.138)	0.494 (0.280)
Median (IQR)	0.910 (0.675-1.267)	0.633 (0.474-0.949)	0.213 (0.113-0.330)	0.231 (0.164-0.313)	0.428 (0.310-0.652)
2008 (N=219)					
Mean (Std. Dev.)	1.126 (0.609)	0.818 (0.514)	0.336 (0.294)	0.258 (0.126)	0.593 (0.377)
Median (IQR)	1.068 (0.747-1.383)	0.759 (0.520-1.058)	0.327 (0.154-0.431)	0.245 (0.167-0.332)	0.550 (0.359-0.771)
2009 (N=211)					
Mean (Std. Dev.)	1.076 (0.614)	0.784 (0.524)	0.315 (0.287)	0.245 (0.123)	0.560 (0.357)
Median (IQR)	1.068 (0.699-1.340)	0.801 (0.503-1.013)	0.294 (0.166-0.442)	0.238 (0.167-0.313)	0.559 (0.330-0.737)
2010 (N=211)					
Mean (Std. Dev.)	1.079 (0.529)	0.785 (0.437)	0.303 (0.214)	0.237 (0.125)	0.540 (0.293)
Median (IQR)	1.090 (0.736-1.373)	0.817 (0.514-1.013)	0.301 (0.176-0.412)	0.217 (0.143-0.315)	0.553 (0.351-0.694)
2011 (N=207)					
Mean (Std. Dev.)	1.133 (0.556)	0.832 (0.461)	0.316 (0.226)	0.226 (0.107)	0.542 (0.295)
Median (IQR)	1.140 (0.790-1.408)	0.812 (0.523-1.054)	0.311 (0.187-0.440)	0.220 (0.150-0.298)	0.519 (0.358-0.711)

Table 4.17: New York SID - Mean LOS and CDI Incidence Correlation

	Hospital Mean LOS		Correlation with CDI Incidence		
Year	Mean (Std Dev.)	Median (IQR)	All Cases	Secondary	Not POA
2005	5.59 (9.19)	5.19 (4.62-6.13)	0.451***	0.555***	0.555***
2006	5.55 (9.18)	5.20 (4.52-6.05)	0.651***	0.704***	0.577***
2007	5.51 (9.13)	5.12 (4.45-6.15)	0.571***	0.618***	0.377***
2008	5.47 (9.01)	5.10 (4.52-5.98)	0.755***	0.810***	0.717***
2009	5.38 (8.71)	5.01 (4.45-5.94)	0.717***	0.764***	0.750***
2010	5.35 (8.72)	4.97 (4.38-5.91)	0.683***	0.407***	0.265***
2011	5.35 (8.71)	4.98 (4.37-6.02)	0.688***	0.729***	0.599***

[a]These values correspond to the mean and median across hospitals of each hospital's mean LOS.

Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

As in the NIS and California analyses, the relationship between CDI incidence and LOS was first analyzed through a bivariate comparison across hospitals. Table 4.17 reports summary statistics for hospital-average LOS and correlation coefficients using incidence rates of overall CDI, secondary CDI and CDI not POA. As in California, there appears to be a slight downward trend in hospital average LOS in New York over the study period. Average LOS for a typical (i.e. median) hospital decreased from 5.19 days to 4.98 days over the study period. Average LOS was also slightly greater in New York than in California. Table 4.17 shows that CDI incidence was positively and significantly correlated with average LOS, using all cases, secondary cases or cases not POA. For every year, except 2010, this correlation was strongest using secondary CDI incidence; it 2010 overall CDI incidence had the strongest correlation.

Similar to the findings of the NIS and California analysis, a variety of models were compared (as described in section 3.6), and the model that provided the best quality fit was the GLM model with a log link and gamma distribution. Table 4.18 provides the estimated effects of interest for both the best fitting gamma-GLM model along with the standard OLS model. As in the two previous analyses, when models using both yearly and quarterly CDI incidence were compared, yearly incidence rates provided the best quality model. Results presented here utilize yearly incidence; those using quarterly rates can be found in Appendix B.

Table 4.18 shows that, similar to previous results, in both the OLS and GLM specifications, secondary CDI incidence provided the largest estimated effect and the best quality model of all the various incidence rates. The OLS model suggests that a patient stayed approximately 1.2237 days longer, on average, at a hospital with a secondary CDI incidence that was one standard deviation higher. In the chosen GLM model this effect was roughly equivalent to staying 10.46% longer.[8] The CDI incidence definition that provided the 2nd best fit, and 2nd largest effect estimate, was using all CDI cases in both the OLS and GLM model. The 3rd "best" model in the OLS specification was using only CDI coded as not POA, while in the GLM specification was using CDI coded not POA along with post-discharge cases. Similar to the California SID results, post-discharge CDI incidence is positively and significantly associated with increased LOS, suggesting the effect captured by CDI

[8]These interpreted percentage effects are "roughly" estimated by exponentiating the GLM regression coefficients reported in Table 4.18.

Table 4.18: New York SID - Regression Results

CDI cases included	Coefficent (Std. Error)	AIC
OLS		
All CDI	0.9009 (0.0041)***	119,398,202
Secondary Diagnosis	1.2237 (0.0044)***	119,369,650
Not POA	0.4824 (0.0038)***	119,428,686
Post Discharge	0.0343 (0.0031)***	119,444,814
Not POA + Post Discharge	0.4356 (0.0037)***	119,431,192
GLM (gamma)		
All CDI	0.0785 (.0004)***	85,648,338
Secondary Diagnosis	0.0995 (.0004)***	85,635,268
Not POA	0.0423 (.0004)***	85,665,920
Post Discharge	0.0161 (.0003)***	85,674,778
Not POA + Post Discharge	0.0449 (.0004)***	85,664,910

Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

incidence likely outweighs any post-discharge effect.

As in California, in 2009 New York state began requiring hospitals to report cases of HA CDI. Because a change point was also identified in 2008 from Tables 4.15 and 4.16, a second set of models were fit where CDI incidence was interacted with an indicator for time post 2008. These results are reported in Table 4.19. In all cases, dividing the coefficient of interest into two values, before and after 2008, resulted in increased model quality. The estimated effects also showed consistent differences between these time periods. The estimated effect, and relative significance level, increased in the period after the start of 2008 for CDI incidence measured using all CDI cases, secondary cases, those marked as not POA, and those marked as not POA or post discharge. The change in estimates was most apparent in CDI

incidence using not POA. After the change point in 2008, the effect of not-POA-CDI incidence increased from 0.2404 to 1.2401 in the OLS model and from 0.0244 to 0.0786 in the GLM model. Similar to the findings in California, CDI incidence using only secondary CDI cases remained the best quality model; however, after 2008 the estimated effects using not-POA-CDI cases (i.e. 1.2401 and 0.0786) moved much closer to the estimates using only secondary cases (i.e. 1.2880 and 0.1041). In addition, the estimated effect of CDI incidence using post-discharge cases decreased for 2008 onward, and was negative after 2008 in the OLS model.

Table 4.19: New York SID - Regression Results Split By Policy Change Point

	Coefficent (Std. Error)		
CDI cases included	Pre 2008	Post 2008	AIC
OLS			
All CDI	0.8844 (0.0203)***	0.9218 (0.0250)***	119,398,182
Secondary Diagnosis	1.1704 (0.0248)***	1.2880 (0.0309)***	119,369,466
Not POA	0.2404 (0.0116)***	1.2401 (0.0438)***	119,415,068
Post Discharge	0.2295 (0.0070)***	-0.0608 (0.0049)***	119,442,696
Not POA + Post Discharge	0.2803 (0.0101)***	0.7677 (0.0317)***	119,427,158
GLM (gamma)			
All CDI	0.0750 (0.0011)***	0.0825 (0.0012)***	85,648,276
Secondary Diagnosis	0.0949 (0.0011)***	0.1041 (0.0012)***	85,635,186
Not POA	0.0244 (0.0009)***	0.0786 (0.0015)***	85,662,368
Post Discharge	0.0377 (0.0007)***	0.0055 (0.0006)***	85,673,210
Not POA + Post Discharge	0.0339 (0.0008)***	0.0611 (0.0014)***	85,663,908

Note: Estimates obtained by the final model described in section 3.5
Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

As described in section 3.5.3, a multilevel model was used to estimate the amount of variation in LOS between hospitals explained by CDI incidence. Table 4.20 presents the results for the second stage of this model.[9] The amount of variation in hospital CDI incidence that could be potentially explained by CDI incidence appears to have systematically increased from 2005-2009 and then decreased from 2010-2011. From 2005-2008, secondary CDI incidence explained the greatest amount of variation in hospital fixed effect, ranging from 4.93% in 2005 to 20.49% in 2008. Beginning in 2009, CDI not POA explained the greatest amount of variation, ranging from 20.55% in 2009 to 12.58% in 2011. Prior to 2008, CDI not-POA explained significantly less variation than secondary CDI, ranging from 0.39% to 3.37%; however, from 2008 onward CDI not-POA explained roughly an equivalent amount of variation as secondary CDI. Unlike the California results, in New York post discharge CDI incidence was not significantly related to hospital fixed effects.

[9] As in California, models for hospital fixed effects were estimated separately for each year due to the extreme computational challenge of estimating such a large number of separate fixed effects by hospital and year.

Table 4.20: New York SID - Results From Second-Level Multilevel Model

CDI cases included	All CDI	Secondary	Not POA	Post Discharge	Not POA + Post Discharge
2005 (N=399)					
Coefficent (SE)	0.9106 (.3378)***	1.2175 (.3466)***	0.6352 (.3137)**	0.2786 (.4282)	0.5828 (.3315)*
R^2	0.0296	**0.0493**	0.0169	0.0018	0.0128
2006 (N=395)					
Coefficent (SE)	2.3001 (.4239)***	3.0129 (.4569)***	1.1339 (.3959)***	0.4399 (.3930)	0.9226 (.3728)**
R^2	0.1114	**0.1561**	0.0337	0.0053	0.0254
2007 (N=398)					
Coefficent (SE)	2.3172 (.3814)***	3.1202 (.4137)***	0.2496 (.2590)	0.5652 (.3266)*	0.3989 (.2702)
R^2	0.1352	**0.1942**	0.0039	0.0125	0.0091
2008 (N=392)					
Coefficent (SE)	1.9267 (.3205)***	2.6913 (.3480)***	3.0279 (.3928)***	0.1855 (.2832)	2.1975 (.3660)***
R^2	0.1348	**0.2049**	0.2039	0.0018	0.1345
2009 (N=389)					
Coefficent (SE)	2.0409 (.3502)***	2.9091 (.3817)***	3.2310 (.4235)**	-0.0473 (.2182)	2.2159 (.2836)***
R^2	0.1311	0.2051	**0.2055**	0.0002	0.1292
2010 (N=383)					
Coefficent (SE)	3.2105 (.6764)***	2.8231 (.4326)***	3.4781 (.4822)***	0.1524 (.2699)	2.2487 (.4282)***
R^2	0.0918	0.1604	**0.1892**	0.0014	0.1101
2011 (N=384)					
Coefficent (SE)	1.4707 (.3563)***	2.2101 (.3979)***	2.6596 (.4696)***	0.0613 (.2512)	1.7832 (.4141)***
R^2	0.0710	0.1216	**0.1258**	0.0003	0.0768

Notes: Standard errors in parentheses; largest R^2 in bold; significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

As previously described, a primary goal of this work was to analyze the existence and impact of a post-discharge effect. Similar to the findings in California, New York hospitals with longer LOS actually had higher incidence rates of post-discharge CDI. Hospital mean LOS and post-discharge CDI incidence were positively and significantly correlated (ρ=0.1315, $P < .0001$)[10] This positive correlation existed across and within years, and was statistically significant ($P < .001$) in all years except for 2005 and 2009. Simple linear regression analysis suggests that each day increase in a hospital's mean LOS corresponded to an increase in post-discharge CDI incidence of .0039 percentage points (P<.001). These results are in contrast to the hypothesized relationship of the post-discharge effect and suggest that if the post-discharge effect exists, it is being masked by a much greater effect: hospitals with longer LOS generate more HO-HA CDI and CO-HA CDI than hospitals with shorter LOS.

A multivariate analysis was also conducted to test for the presence of a post-discharge effect. The results using secondary CDI cases and CDI cases not POA to represent HO-HA CDI are presented in Table 4.21. The results in New York, consistently suggest the presence of a post discharge effect. In both the OLS and GLM models, the estimated coefficient on post discharge CDI was negative and statistically significant, with the exception of in the GLM model when CDI not POA was used. When post-discharge CDI was added to the model, the estimated effect of CDI incidence increased. Moreover, in the models where post-discharge CDI was significant, the model quality (as measured by AIC values) improved when post-discharge

[10]Using the Pearson product-moment correlation coefficient.

Table 4.21: New York SID - Controlling For Post-Discharge Effect

CDI cases used	Coefficent (Std. Error)		AIC
	CDI Incidence	Post Discharge CDI	
OLS			
Secondary	2.0899 (0.0307)***	-0.9077 (0.0127)***	119,319,492
Not POA	0.5629 (0.0163)***	-0.1563 (0.0050)***	119,426,542
GLM (gamma)			
Secondary	0.1401 (0.0006)***	-0.0502 (0.0004)***	85,625,226
Not POA	0.0426 (0.0004)***	-0.0007 (0.0004)*	85,665,920

Significance levels: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

CDI was added. These findings suggest that a post-discharge effect may be occurring and that controlling for it may also allow the model to better estimate the effect of HA-CDI incidence.

CHAPTER 5
DISCUSSION

The primary purpose of this research was to further establish the connection between HA-CDI incidence and LOS in patients without CDI. While this relationship had been suggested by our preliminary study, further research was needed to provide evidence that this relationship was (1) driven by cases of HA CDI and (2) was not simply a spurious correlation resulting from hospitals with shorter LOS capturing more HA CDI prior to discharge. The first objective set forth at the beginning of this research was to estimate the effect of HA-CDI incidence on LOS across hospitals and determine if this relationship has varied over time. Because CDI incidence and policies regarding the reporting of HAIs have been changing, the time window used in the preliminary research was expanded to cover the years 2005-2011. The NIS and SIDs for California and New York were selected for this analysis because they offered discharge records for all patients within a given hospital, spanned a range of hospitals, contained a wide variety of patient characteristics for which to control and, in the cases of the two SIDs, included indicators for HA status of diagnoses.

The NIS was used to address the first objective because it offered a wider sampling of hospitals around the US and larger number of patient and hospital characteristics than the two SIDs. In each year from 2005-2011, after controlling for a wide range of patient and hospital characteristics, secondary CDI incidence was found to be strongly associated with increased LOS in patients without CDI. Estimates of the increased patient LOS associated with hospital's having a one standard deviation

greater secondary-CDI incidence ranged from 10.0% to 19.4%[1] or 1.3 to 2.4 days, depending on the model used. After controlling for differences in patient and disease characteristics, CDI incidence helped to explain an additional 31.2% to 50.5% of the variation in LOS among hospitals, beyond the other hospital characteristics available in the NIS. Moreover, in each year CDI incidence was consistently one of the top 5 predictive variables for inpatient LOS, in regression models containing over 400 different covariates. Although the NIS contains no markers of HA CDI, all evidence from the NIS results suggests that the effect being captured by CDI incidence is driven primarily by cases of HA CDI. Not only were effect estimates greater when only secondary cases were used, but model quality systematically improved when primary cases were removed from CDI incidence. Furthermore, the analysis outlined in Table 4.6, which used our hypothesized statistical approach to remove CA-CDI cases that spillover into secondary diagnoses, consistently suggested that HA-CDI incidence was the primary driver of our results.

The NIS results depict an increasing trend in both the incidence and number of CDI cases, which has been well established in existing literature [11, 5, 84, 45]. Because the NIS represents a sampling of hospitals, it may be difficult to observe changes in the connection between CDI incidence and LOS over this time. Sample weights, intended to create a representative patient population using the NIS, do not necessarily provide a hospital population that is representative of hospital CDI

[1]These interpreted percentage effects are obtained by exponentiating the GLM regression coefficients reported in table 4.5.

incidence and its relationship to LOS. There does, however, appear to be a slight change that may have occurred in 2008. Stating in 2008 the effect estimates in both the OLS and GLM regression models, presented in Tables 4.4 and 4.5, appear to systematically increase. However, this 2008 change point does not appear in the estimates obtained in the multilevel model presented in Table 4.7. In either case, CDI incidence was a strong predictor of inpatient LOS variation both before and after 2008.

The SIDs in both California and New York were selected to address the first study objective because they offered the ability to assign CDI cases as HA using POA indicators. It has been previously suggested that POA indicators, as opposed to secondary diagnosis, should be used to assign HA status to CDI cases [74, 93]. In both California and New York, CDI incidence calculated using cases not POA was a strong predictor of inpatient LOS. In California, estimates of the increased LOS associated with a standard deviation increase in not-POA-CDI incidence were roughly 8.9%[2] and 1.8 days using the best-fitting-GLM and OLS models, respectively. In New York, these same estimates were 4.3%[3] and 0.5 days. However, in nearly all models secondary CDI incidence worked as a stronger predictor of inpatient LOS than CDI not POA in terms of model fit, coefficient significance and estimated effect size. In California, a standard deviation increase in secondary CDI incidence was associated with a 10.2%

[2]These interpreted percentage effects are obtained by exponentiating the GLM regression coefficients reported in table 4.11.

[3]These interpreted percentage effects are obtained by exponentiating the GLM regression coefficients reported in table 4.18.

or 2.1 day increase in LOS, using the GLM and OLS models respectively. And in New York these same estimates were 10.5% and 1.2 days. Despite these differences, the SID results reinforce the strong connection between CDI incidence and inpatient LOS variations among hospitals. In California, the multilevel model suggested that CDI incidence can explain between 12.5%-28.5% of the between-hospital variation in LOS. In New York, CDI incidence explained 4.9%-20.6% of the between-hospital variation in LOS.

One clear trend that was observed from the SID analysis was a change that occurred beginning in 2008. In both states, there was a dramatic increase in the incidence and number of CDI cases coded as not POA. In California the proportion of CDI cases coded as not POA increased from around 10%, before 2008, to around 25%, beginning in 2008. Similarly, in New York the proportion of CDI cases coded as not POA increased from around 23%, before, to around 29% beginning in 2008. These changes roughly correspond to the point when regulations in both states required hospitals to report HA CDI. Although occurring the year before regulations went into full effect in 2009, it seems likely that hospitals may have altered their recording practices in anticipation of the regulations. While some of this may also be related to changes in the types of tests used to identify CDI cases (e.g., PCR testing) it seems most likely to be driven by the reporting mandate.

Because of the apparent 2008 change point, models were also fit where the estimated effect of CDI incidence was allowed to vary before and after 2008. Modeling this change point had some impact on all the measures of CDI incidence (e.g., using

all or secondary CDI cases), but it had the most dramatic effect on CDI incidence measured using cases not POA. In both California and New York, following the 2008 mandate there was a dramatic increase in the estimated effect of CDI incidence using cases not POA. In California, the effect estimates increased from 1.57 to 1.94 days and 8.11% to 9.43% in the OLS and GLM models, respectively. In New York, these same effect estimates increased from 0.24 to 1.24 days and 2.47% to 8.18%. Similarly, in both states beginning in 2008 there was a dramatic increase in the amount of between-hospital variation captured in the multilevel model by CDI incidence using cases not POA. However, the finding that is, perhaps, of greatest note is beginning in 2008 the results obtained using CDI not POA appear to converge to those obtained using secondary CDI; this occurs in both the coefficient estimates and the amount of variation captured in the multilevel model. This finding seems to suggest that the effect being captured by secondary CDI, before and after 2008, is also being captured by CDI cases coded as not POA after 2008.

Secondary CDI incidence generally outperformed CDI not POA in the SID results; however, there is strong evidence to suggest that secondary CDI may actually better reflect a hospital's true HA-CDI incidence than CDI coded as not POA. If variation between hospitals in the utilization of POA indicators exceeds the amount of variation between hospitals in HA-CDI incidence, then secondary CDI may better reflect true HA-CDI variation even if inadvertently capturing some CA-CDI cases. Three pieces of evidence suggest this may be the case. First, the SID analysis revealed what appears to be systemic underutilization of POA indicators in both states. Rates

of CDI not POA increased sharply in 2008, just prior to regulations requiring hospitals to repot HA CDI. Yet even after this increase, the rates of HA CDI marked not POA appear significantly below incidence rates reported in existing literature. Previous research has reported HO-HA CDI to be the most common class of CDI, with over 50% of HO cases being HO-HA CDI [46, 83, 87]. However, incidence of CDI not POA in California and New York appear much lower than these previous estimates. Given that POA indicators were still utilized to some extent prior to the mandate, there is clearly some degree of hospital discretion in the use of such indicators.

Second, variation in the proportion of CDI cases coded as not POA, across both years and the two states, was not reflected by corresponding changes in the proportion of CDI cases coded as a secondary diagnosis. Because nearly all HO-HA CDI should be coded as secondary, any increase or decrease in HA CDI incidence should be reflected by corresponding increases or decreases in secondary CDI. In both California and New York, changes across years in the proportion of CDI cases marked not POA were not accompanied by similar changes in the proportion of secondary CDI cases. Moreover, the proportion of CDI not POA was much greater in New York than in California (i.e. 26.79% vs. 19.36%), while the proportion of primary and secondary CDI was nearly identical between the these two states (i.e. 74.06% vs 74.14%). This would suggest that New York hospitals may be more apt to utilize POA indicators than hospitals in California.

Finally, as previously discussed, after 2008 when POA indicators became more heavily utilized, the estimated effect of CDI incidence seemed to converge to the

estimates obtained using secondary CDI. As shown from Tables 4.12 and 4.19, the estimates using secondary CDI incidence remained relatively stable before and after 2008 while those of CDI not POA increased substantially, becoming similar to the estimates for secondary CDI. Furthermore, in the multilevel model for New York, where reporting of CDI not POA was higher, CDI incidence using not POA became the best predictor of between-hospital LOS variation beginning in 2009.

The second objective of this research was to determine the extent to which the relationship captured by CDI incidence was attributable to unobserved cases of CO-HA CDI. It was initially hypothesized that the observed relationship between CDI incidence and LOS may be inflated by the fact that HA-CDI cases are more likely to be hospital onset, and thus captured in discharge records at hospitals with longer LOS. This effect, which we have referred to as the post-discharge effect, would make a hospital with longer LOS appear to generate more HA CDI, if such a hospital simply recorded a greater number of HA-CDI cases prior to discharge. The SID in California and New York were used in order to address this second objective because in both states inpatient and emergency department visits are able to be linked in time between multiple visits for a given patient. Both the SID and SEDD in California and New York were used to identify cases of HA CDI occurring post discharge. Cases were labeled as post-discharge CDI using the standard definition within the literature for CO-HA CDI (i.e. occurring with 4 weeks of a previous discharge).

We first conducted a univariate analysis comparing mean LOS and post-discharge-CDI incidence across hospitals. Contrary to our initial hypothesis, hospitals

with longer LOS did not have lesser incidence of post-discharge CDI; in fact, the opposite was true. In both California and New York, hospital mean LOS was positively and significantly correlated with increased incidence of post-discharge CDI. We also include post-discharge incidence as one of the CDI-incidence measures used in the regression models estimating patient LOS. In both California and New York, and in both the OLS and GLM models, a significant and positive coefficient estimate was obtained when post-discharge incidence was used to predict LOS in patients without CDI. Moreover, in California, when post-discharge-CDI incidence was used in the multilevel model it was able to significantly explain 3.3-6.2% of the between-hospital variation in LOS. These findings suggest that hospitals with longer LOS tended to generate both more HO-HA CDI and CO-HA CDI. Thus, any post-discharge effect that might be occurring is potentially being masked by hospitals with longer LOS generating more HA CDI overall.

In order to further search for the existence of a post-discharge effect, we estimated a model that included separate incidence measures for both HO-HA CDI and CO-HA CDI. Because our previous findings suggested that secondary CDI may better capture variation in HA-CDI incidence, we used both secondary CDI and CDI not POA as separate estimates for HO-HA CDI. Tables 4.14 and 4.21 report the results of the model controlling for both HO-HA-CDI incidence and the post discharge effect. As hypothesized, we obtained evidence for the existence of a post-discharge effect (i.e. a negative coefficient estimate on post-discharge CDI incidence) in both California and New York when secondary CDI incidence was used to represent HO-HA CDI. We

also obtained evidence of this effect in New York using CDI not POA to represent HO-HA CDI. However, contrary to our initial expectations, in the models where the post-discharge effect was observed, the coefficient estimate on CDI incidence actually increased when we controlled for the post-discharge effect. This is, however, consistent with the finding that hospitals with longer LOS generated more HA CDI, both HO and CO. It appears that controlling for the post-discharge effect allowed CDI incidence to better capture CDI variation between hospitals. Therefore, our results strongly suggest that the observed correlation between CDI incidence and LOS is not a spurious relationship being driven by the post-discharge effect; if anything, the post-discharge effect makes the relationship between CDI incidence and LOS appear much weaker than it actually is.

5.1 Limitations

While the results of this study bolster the hypothesized connection between HA-CDI incidence and LOS variations across hospitals, this study is not without limitations. This research was conducted using observational data sources and was, thus, constrained by the amount of information available in these data. One of the primary limitations of this study was an inability to directly confirm CDI diagnoses or the HA status of these diagnoses using clinical microbiological test results. Most research involving HA CDI has relied on microbiologic test results along with the timing of diagnosis to identify and classify CDI cases as HA. However, such research often involves only a single or small number of hospitals. Because this study's primary

objective was to make LOS comparisons across a wide variety of hospitals, it was conducted using administrative discharge data which could span the greatest number of hospitals. Consequently, we did not have access to CDI test results, the day of CDI diagnosis, the day symptoms first occurred or the types of tests that hospitals used to identify CDI. Each of these variables may be necessary to conclusively determine whether a diagnosis of CDI was a true case and if a case were HA. While it is likely possible that some of the CDI cases identified in our data may have been incorrectly diagnosed, the use of ICD-9-CM codes for surveillance of overall hospital CDI incidence has been thoroughly validated elsewhere [38, 41, 140]. However, it has also been suggested that the type of test hospitals use to identify CDI cases can alter observed CDI incidence, and it is possible that testing differences between hospitals could be partially affecting our results [94, 80].

Another limitation that remains is establishing CDI cases as HA based on secondary diagnosis. Previous research has called into question the accuracy of a using a secondary diagnosis as a marker for HA status, and POA indicators have been suggested as a means of ensuring accuracy of HA CDI [38]. However, as previously described, our results suggest that POA indicators may be widely underutilized and that secondary CDI incidence may work better for capturing HA-CDI incidence. Future research, with more detailed clinical data on CDI testing and diagnosis, would be needed to throughly validate this result. While such research may require a large number of resources to acquire such granular patient data spanning a large number of hospitals, the current study may provide motivation for such future work.

In addition to the lack of detailed clinical diagnostic data, another limitation of this study was the lack of detailed hospital characteristics and unreported patient characteristics. It remains a possibility that CDI incidence may not be acting as a proxy for quality and/or efficiency, as has been theorized, and that some other set of unobserved hospital or patient characteristics is driving this connection. The NIS contains only a handful of the most basic hospital characteristics while the SID contains virtually none. In addition, a number of patient characteristics go unreported by states or hospitals, and many are not collected by our data sources. For example, no information is available regarding medications or other types of treatments received. Thus, we cannot conclusively rule out other hospital or patient level factors diving the observed relationship between CDI incidence and LOS. However, given the large number of patient characteristics that were controlled for, and the consistency of results across time and three separate data sources, its seems more likely that a missing hospital characteristic would be responsible for confounding the observed relationship. Future research would be needed to rule out other hospital factors not contained in these data.

Another limitation of this study was the inability to observe CDI diagnoses that occurred outside of inpatient or emergency department settings when calculating post-discharge-CDI incidence. Although we used the entire population of inpatient and emergency department records to identify post-discharge cases of CDI, it is highly likely that a number of post-discharge diagnoses occurred in other ambulatory settings. Patients may have also received diagnoses, or moved, out of state. Thus, the

post discharge incidence rates that were calculated likely represent an underestimate of the true CO-HA-CDI incidence. As a result, we may not have been able to observe the full post-discharge effect, and it is possible that post-discharge CDI had a greater impact on our findings than was estimated.

As a final note, it is important that the estimates of this study be interpreted and used with caution. This research posited that CDI incidence may operate as a "proxy variable" for unobserved hospital heterogeneity that is potentially associated with hospital quality and/or efficiency. However, the purpose of this research was to build further evidence for the connection between HA-CDI incidence and LOS, while accounting for potential confounding associated with CO-HA-CDI cases occurring post discharge. The objective here was not to establish a direct connection to hospital quality or efficiency, and analyzing this potential connection remains an area for future research. Thus, the author advises caution in using CDI incidence alone to directly assess hospital quality, or efficiency, related to LOS. In addition, it would be incorrect to interpret the coefficient estimates obtained herein to imply that by decreasing LOS at a hospital level, one can necessarily expect to see a drop in CDI incidence, or vice versa. Because we believe CDI incidence to be a proxy variable for other hospital characteristics associated with excess LOS, the exact factors driving this connection would need to be targeted instead.

5.2 Areas for Future Research

This study identified a number of areas for future research. First and foremost, the strength and consistency of the estimated relationship between HA-CDI incidence and inpatient LOS creates a strong motivation to further study the exact sources of this connection. Future research should attempt to validate or refute the hypothesized connection between hospital quality and efficiency. Such work will need to compare previously validated measures of hospital quality and efficiency to both LOS and measures of HA CDI. In addition, a wide range of hospital-specific factors, such as staff training and experience, hospital protocols, or hospital resources, should also be explored as potential confounding factors. Future research aimed at identifying the primary factors driving the observed relationship will require data sources that not only provide a large number patient and disease characteristics and span a wide range of hospitals, but also provide a rich source of hospital-specific information. Although such data may require a large number of resources to obtain and are likely to require multiple sources of information, the current study provides a strong basis for such continued work.

Another area for future work is to explore the factors that influence hospitals' utilization of POA indicators. Differences in the results from California and New York suggest that hospitals may inconsistently use POA indicators to indicate HA CDI. As previously highlighted, over the course of the study period, New York hospitals consistently recorded a greater percentage of CDI cases as not POA than did hospitals in California, yet both states had a nearly identical portion of CDI cases recorded

as secondary. In addition, both states had significant increases in the proportion of CDI cases coded as not POA following regulatory changes. These findings suggest that hospitals have a fair degree of flexibility in determining how to assign POA indicators. Future research should explore the sources that drive variation in the use of POA indicators. Moreover, the 2008 change point suggests that state policies can be highly influential in how hospitals report HAIs. It is likely incentives that influence the use of POA indicators differ between hospitals and states. Future research should take advantage of the variation in state policies to study how such policies affect the use of POA indicators.

A final area for future work is the need to evaluate secondary CDI and CDI not POA as markers for HA-CDI incidence. The prevailing wisdom in existing research has held that secondary CDI should not be used as a marker for HA CDI because CA-CDI cases are often assigned a secondary diagnosis [38]. Thus, secondary CDI incidence would tend to overestimate the true incidence of HA CDI. This was one of the primary motivators for the selection of the two SIDs used in the current research, which both contained POA indicators. However, our results strongly suggest that POA indicators may by widely underutilized and as a result the use of such indicators to assign HA CDI may vastly underestimate the true incidence of CDI. Not only were our findings stronger and more consistent using secondary CDI, when reporting became mandatory, the results using CDI not POA tended to converge to those using secondary CDI. It is strongly suspected that, while CDI coded as not POA may offer a better marker of HA status for individual cases, CDI incidence calculated

using secondary diagnosis may offer a better marker for HA-CDI variations across hospitals. Indeed, the hospital incentives for using POA indicators, just discussed, offer some justification as to why secondary CDI may provide a more accurate estimate of variations in HA CDI. Future research should analyze the relative accuracy of using either secondary diagnoses or POA indicators to compute HA-CDI incidence.

5.3 Conclusions

Inpatient LOS is a major contributor to hospital costs and overall healthcare spending. Consequently, excess LOS has also become a frequently used outcome measure to assess costs and burden associated with many diseases. In addition, systematic variation in LOS among hospitals has been used as a measure of hospital efficiency and quality. Prior to this work, our preliminary study found a significant correlation between increased LOS and hospital incidence of CDI. We hypothesized that this connection was occurring because of increased HA CDI, which may be due to hospital quality and/or efficiency. One of the limitations of the preliminary work was the inability to distinguish between HA- and CA-CDI cases; only secondary diagnosis was used to suggest the link was occurring via HA CDI. In addition, the preliminary work was unable to account for the occurrence of CDI post discharge. One alternative explanation for the observed relationship between CDI incidence and LOS is that hospitals with shorter LOS simply have more HA CDIs that occur post discharge. If this were the case, the observed relationship could be overestimated or considered spurious. The primary purpose of this research was to further establish

the relationship between hospital incidence of HA CDI and excess LOS in patients without CDI by addressing these limitations. In order to do so, the current research expanded on the NIS data used in the preliminary work and added data from the SIDs in California and New York. This allowed for HA-CDI cases to be assigned using POA indicators and for post-discharge CDI to be identified through patient revisits.

Consistent with the findings of our preliminary research, this study found a strong relationship between CDI incidence and inpatient LOS. CDI incidence was consistently found to be one of the strongest predictive variables for inpatient LOS, and this result held across time, locations and nearly all of the model specifications that were analyzed. Consistent with the hypothesis that CDI incidence acts as a proxy variable for unobserved hospital heterogeneity, the results of the multilevel models suggest that CDI incidence is largely working to explain between-hospital variation in LOS. In addition, HA-CDI incidence, as measured using POA indicators, was consistently associated with increased LOS; such incidence became an even stronger predictor following mandatory reporting of HA CDI in both California and New York. This suggests that HA CDI is the factor driving this relationship. While it was not expected that secondary CDI incidence would outperform CDI not POA, as was frequently the case, this study also found compelling evidence that POA indicators are systematically underutilized. It is the author's firm belief that the evidence presented here strongly suggests secondary CDI incidence to be a better marker for variation in HA-CDI incidence than CDI not POA. Future research may be need to validate this finding; however, secondary CDI incidence remains is an incredibly strong predictor

of inpatient LOS.

This study's second primary objective was to evaluate the potential that the observed relationship between CDI incidence and LOS was being driven by a post-discharge effect, where hospitals with shorter LOS had a greater proportion of HA-CDI cases occur post discharge. The results of this study emphatically find this not to be the case. Although evidence of a post-discharge effect was somewhat inconsistent between the models and states that were analyzed, if any post-discharge effect exists at all, it appears to be outweighed by the fact that hospitals with longer LOS actually tended to generate both more HO-HA CDI and more CO-HA CDI. In fact, our univariate results found hospitals with longer LOS to have greater incidence of post-discharge CDI. Contrary to our initial hypothesis, the post-discharge effect seems to be weakening rather than increasing the observed effect of CDI incidence on LOS. In each of the models where significant evidence for a post-discharge effect could be observed, the effect of HA-CDI incidence improved when secondary CDI incidence was controlled for. While this study was limited by the number of post-discharge CDI cases that could be observed using inpatient and emergency-department data, there was no evidence to suggest the post-discharge effect was inflating the estimated connection between CDI incidence and LOS.

CDI incidence is relatively easy to measure and record, it can be readily identified using administrative data, and many states have begun mandating hospitals to report incidence of HA CDI. In addition, CDI incidence has a straightforward interpretation and is easy to compare across hospitals. The current study provides strong

evidence that CDI incidence can be used to capture variations in excess LOS between hospitals. Thus, CDI incidence may represent an ideal tool for researchers, along with patients and policy makers, wishing to make hospital comparisons. This study also provides a strong motivational foundation for future research into the nature of this observed relationship. Our findings provide strong evidence that this relationship is driven by HA CDI, and suggest the connection is not being driven by a post-discharge effect. If future research can validate a connection with CDI incidence and hospital efficiency or quality, policy makers may wish to consider CDI incidence as a powerful measure to rank and compare hospitals. The findings of this research also provide motivation to further study the connection between POA indicators and HA-CDI incidence, as well as the policy factors that influence hospital reporting of HAIs. Moreover, given the large amount of research that has sought to explain LOS variations between patients and hospitals, the strong relationship between CDI incidence and LOS should be taken into consideration by future researchers wishing to analyze variation of inpatient LOS.

APPENDIX A
ABBREVIATIONS

Table A.1 provides a list of the various abbreviations used throughout this thesis.

Table A.1: Table of Abbreviations

Abbreviation	Full Name
CDI	*Clostridium difficile* Infection
LOS	Length Of Stay
HAI	Hospital Acquired Infection
HA	Hospital/Healthcare Acquired
CA	Community Acquired
HO-HA	Hospital Onset, Hospital Acquired
CO-HA	Community Onset, Hospital Acquired
POA	Present On Admission
AHRQ	Agency for Healthcare Research and Quality
HCUP	Healthcare Cost and Utilization Project
NIS	Nationwide Inpatient Sample
SID	State Inpatient Databases
MRSA	Methicillin-Resistant *Staphylococcus aureus*
CMS	Centers for Medicare & Medicaid Services
DRG	Diagnosis-Related Group
APR	All Parient Refined

APPENDIX B
QUARTERLY RESULTS

This appendix contains regression results using quarterly CDI incidence rates for the NIS, and SID analyses in California and New York. Tables B.1 and B.2 provide the OLS and GLM (gamma) results, respectively, for the NIS. Table B.3 summarizes regression results for the SID in California. Table B.4 summarizes regression results for the SID in New York. One issue of note is that the AIC values reported in these tables, in many instances, are lower than those reported in the corresponding results from Chapter 4. This is simply due to the fact that, in each case, the number of observations was reduced when quarterly rates were used. In some quarters there were hospitals that did not report any cases of CDI and, as a result, observations from such hospitals were excluded. However, when yearly rates were used with this restricted population of patients, the yearly incidence measures unanimously resulted in lower AIC values.

Table B.1: NIS OLS Quarterly Results By Year

CDI cases included	Coefficent (Std. Error)	AIC
2005 (N=6,505,438)		
All CDI	0.9530 (0.0058)	39,582,178
Secondary Diagnosis	1.2095 (0.0063)	39,572,808
2006 (N=6,621,299)		
All CDI	1.2644 (0.0064)	39,627,416
Secondary Diagnosis	1.5998 (0.0069)	39,612,956
2007 (N=5,856,932)		
All CDI	0.7980 (0.0056)	35,044,032
Secondary Diagnosis	1.1033 (0.0062)	35,032,498
2008 (N=6,160,889)		
All CDI	1.6492 (0.0069)	37,239,918
Secondary Diagnosis	1.9599 (0.0076)	37,230,312
2009 (N=5,697,525)		
All CDI	1.6524 (0.0072)	33,828,230
Secondary Diagnosis	1.9009 (0.0075)	33,817,106
2010 (N=5,824,128)		
All CDI	1.8928 (0.0073)	35,183,232
Secondary Diagnosis	2.1904 (0.0074)	35,162,482
2011 (N=6,020,364)		
All CDI	1.6143 (0.0063)	36,542,708
Secondary Diagnosis	1.9977 (0.0068)	36,522,580

Table B.2: NIS GLM (Gamma) Quarterly Results By Year

CDI cases included	Coefficent (Std. Error)	AIC
2005 (N=6,505,438)		
All CDI	0.0800 (0.0008)	30,233,770
Secondary Diagnosis	0.0985 (0.0009)	30,232,068
2006 (N=6,621,299)		
All CDI	0.0891 (0.0009)	30,685,448
Secondary Diagnosis	0.1112 (0.0009)	30,683,224
2007 (N=5,856,932)		
All CDI	0.0619 (0.0008)	27,075,966
Secondary Diagnosis	0.0841 (0.0009)	27,073,844
2008 (N=6,160,889)		
All CDI	0.1245 (0.0010)	28,605,736
Secondary Diagnosis	0.1444 (0.0012)	28,604,716
2009 (N=5,697,525)		
All CDI	0.1203 (0.0011)	26,432,464
Secondary Diagnosis	0.1410 (0.0012)	26,430,864
2010 (N=5,824,128)		
All CDI	0.1480 (0.0011)	27,111,018
Secondary Diagnosis	0.1696 (0.0012)	27,107,484
2011 (N=6,020,364)		
All CDI	0.1038 (0.0009)	27,979,032
Secondary Diagnosis	0.1295 (0.0010)	27,976,464

Table B.3: SID Quarterly Regression Results in California

CDI cases included	Coefficent (Std. Error)	AIC
OLS Results		
All CDI	1.8017 (.0044)	165,908,405
Secondary Diagnosis	2.1012 (.0046)	165,872,129
Not POA	1.5991 (.0045)	165,950,492
Post Discharge	0.5338 (.0031)	166,046,617
Not POA + Post Discharge	1.4475 (.0041)	165,952,396
GLM Results		
All CDI	0.0843 (0.0006)	115,503,294
Secondary Diagnosis	0.1100 (0.0006)	115,494,539
Not POA	0.0861 (0.0006)	115,504,462
Post Discharge	0.0238 (0.0004)	115,518,987
Not POA + Post Discharge	0.0738 (0.0005)	115,505,716

Notes: Total observations N=25,109,686.

Table B.4: SID Quarterly Regression Results in New York

CDI cases included	Coefficent (Std. Error)	AIC
OLS Results		
All CDI	0.7780 (0.0039)	242,480,031
Secondary Diagnosis	1.1043 (0.0043)	242,479,135
Not POA	0.3608 (0.0037)	242,479,459
Post Discharge	0.0507 (0.0033)	242,469,768
Not POA + Post Discharge	0.3699 (0.0037)	242,472,902
GLM (gamma) Results		
All CDI	0.0703 (.0004)	85,161,030
Secondary Diagnosis	0.0922 (.0004)	85,150,332
Not POA	0.0326 (.0004)	85,179,698
Post Discharge	0.0152 (.0004)	85,184,690
Not POA + Post Discharge	0.0388 (.0004)	85,177,516

Notes: Total observations N=17,242,489.

REFERENCES

[1] 3M^{TM} APR DRG Classification System (Version 26.1):Methodology Overview. 3M Health Information Systems. Wallingford, CT. 2008. http://www.hcup-us.ahrq.gov/db/nation/nis/v261_aprdrg_meth_ovrview.pdf.

[2] Committee to Reduce Infection Deaths. State Legislation & Iniatives on Healthcare-Associated Infections. http://www.hospitalinfection.org/legislation.shtml. Accessed: 2014-11-08.

[3] National Center for Health Statistics. Health, United States, 2013: With special feature on prescription drugs. Hyattsville, MD. 2014. http://www.cdc.gov/nchs/data/hus/hus13.pdf#095.

[4] Overview of hcup. http://www.hcup-us.ahrq.gov/overview.jsp. Accessed: 2014-11-10.

[5] A. N. Ananthakrishnan. Clostridium difficile infection: epidemiology, risk factors and management. *Nature Reviews. Gastroenterology & Hepatology*, 8(1):17–26, Jan 2011.

[6] A. N. Ananthakrishnan, E. L. McGinley, and D. G. Binion. Excess hospitalization burden associated with Clostridium difficile in patients with inflammatory bowel disease. *Gut*, 57(2):205, 2008.

[7] D. C. Angus, W. Linde-Zwirble, C. A. Sirio, A. J. Rotondi, L. Chelluri, R. C. Newbold, J. R. Lave, and M.R. Pinsky. The effect of managed care on icu length of stay: implications for medicare. *JAMA*, 276(13):1075, 1996.

[8] E. Apostolopoulou, V. Raftopoulos, K. Terzis, and I. Elefsiniotis. Infection probability score: a predictor of Clostridium difficile-associated disease onset in patients with haematological malignancy. *European Journal of Oncology Nursing*, 15(5):404–409, Dec 2011.

[9] J. Ashby, S. Guterman, and T. Greene. An analysis of hospital productivity and product change. *Health Affairs*, 19(5):197, 2000.

[10] R. A. Bankowitz, B. Doyle, M. Duan, E. Kroch, and J. Martin. Identifying hospital-wide harm: a set of icd-9-cm-coded conditions associated with increased cost, length of stay, and risk of mortality. *American Journal of Medical Quality*, 29(5):373, 2014.

[11] J. G. Bartlett. Narrative review: the new epidemic of clostridium difficile-associated enteric disease. *Annals of Internal Medicine*, 145(10):758, 2006.

[12] S. E. Berki, M. L. Ashcraft, and W. C. Newbrander. Length-of-stay variations within icda-8 diagnosis- related groups. *Medical Care*, 22(2):126, 1984.

[13] G. E. Bignardi. Risk factors for Clostridium difficile infection. *The Journal of Hospital Infection*, 40(1):1, 1998.

[14] M. A. Blegen, R. C. Reiter, C. J. Goode, and R. R. Murphy. Outcomes of hospital-based managed care: a multivariate analysis of cost and quality. *Obstetrics and Gynecology*, 86(5):809, 1995.

[15] D. Z. Bliss, S. Johnson, K. Savik, C. R. Clabots, K. Willard, and D. N. Gerding. Acquisition of Clostridium difficile and Clostridium difficile- associated diarrhea in hospitalized patients receiving tube feeding. *Annals of Internal Medicine*, 129(12):1012, 1998.

[16] E. Brown, G. H. Talbot, P. Axelrod, M. Provencher, and C. Hoegg. Risk factors for clostridium difficile toxin- associated diarrhea. *Infection Control and Hospital Epidemiology*, 11(6):283, 1990.

[17] M. D. Brownell and N. P. Roos. Variation in length of stay as a measure of efficiency in Manitoba hospitals. *CMAJ*, 152(5):675, 1995.

[18] L. R. Burns, J. A. Chlingerian, and D. R. Wholey. The effect of physician practice organization on efficient utilization of hospital resources. *Health Services Research*, 29(5):583, 1994.

[19] L. R. Burns and D. R. Wholey. The effects of patient, hospital, and physician characteristics on length of stay and mortality. *Medical Care*, 29(3):251, 1991.

[20] N. Y. Busaba and D. A. Schaumberg. Predictors of prolonged length of stay after major elective head and neck surgery. *Laryngoscope*, 117(10):1756–1763, 2007.

[21] S. M. Campbell, M. O. Roland, and S. A. Buetow. Defining quality of care. *Social Science & Medicine*, 51(11):1611–1625, 2000.

[22] K. Carey and T. Stefos. Controlling for quality in the hospital cost function. *Health Care Management Science*, 14(2):125–134, 2011.

[23] H. T. Chang, D. Krezolek, S. Johnson, J. P. Parada, C. T. Evans, and D. N. Gerding. Onset of symptoms and time to diagnosis of Clostridium difficile-associated disease following discharge from an acute care hospital. *Infection Control and Hospital Epidemiology*, 28(8):926, 2007.

[24] G. M. Chertow, E. Burdick, M. Honour, J. V. Bonventre, and D. W. Bates. Acute kidney injury, mortality, length of stay, and costs in hospitalized patients. *Journal of the American Society of Nephrology*, 16(11):3365, 2005.

[25] D.C. Classen, S. L. Pestotnik, R. S. Evans, J. F. Lloyd, and J. P. Burke. Adverse drug events in hospitalized patients: excess length of stay, extra costs, and attributable mortality. *JAMA*, 277(4):301, 1997.

[26] R. Coello, H. Glenister, J. Fereres, C. Bartlett, D. Leigh, J. Sedgwick, and E. M. Cooke. The cost of infection in surgical patients: a case-control study. *Journal of Hospital Infection*, 25(4):239–250, 1993.

[27] T. C. Collins, J. Daley, W. H. Henderson, and S. F. Khuri. Risk factors for prolonged length of stay after major elective surgery. *Annals of Surgery*, 230(2):251, 1999.

[28] M. I. Correia and D. L. Waitzberg. The impact of malnutrition on morbidity, mortality, length of hospital stay and costs evaluated through a multivariate model analysis. *Clinical Nutrition*, 22(3):235–239, 2003.

[29] S. E. Cosgrove, Y. Qi, K. S. Kaye, S. Harbarth, A. W. Karchmer, and Y. Carmeli. The impact of methicillin resistance in staphylococcus aureus bacteremia on patient outcomes: mortality, length of stay, and hospital charges. *Infection Control and Hospital Epidemiology*, 26(2):166, 2005.

[30] R. Datta, N. N. Kazerouni, J. Rosenberg, V. Q. Nguyen, M. Phelan, J. Billimek, C. Cao, P. McLendon, K. Cummings, and S. S. Huang. Substantial variation in hospital rankings after adjusting for hospital-level predictors of publicly-reported hospital-associated Clostridium difficile infection rates. *Infection Control and Hospital Epidemiology*, 36(4):464–466, April 2015.

[31] I de Glind van, S. de Roode, and A. Goossensen. Do patients in hospitals benefit from single rooms? a literature review. *Health Policy*, 84(2-3):153, 2007.

[32] J. B. Dimick, P. J. Pronovost, R. F. Heitmiller, and P. A. Lipsett. Intensive care unit physician staffing is associated with decreased length of stay, hospital cost, and complications after esophageal resection. *Critical Care Medicine*, 29(4):753, 2001.

[33] A. Donabedian. Evaluating the quality of medical care. *The Milbank Quarterly*, 83(4):691–729, 2005.

[34] R. H. Drew, R. White, C. Macdougall, E. D. Hermsen, and R. C. Owens. Insights from the society of infectious diseases pharmacists on antimicrobial stewardship guidelines from the Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America. *Pharmacotherapy*, 29(5):593–607, 2009.

[35] E. R. Dubberke, A. M. Butler, B. Hota, Y. M. Khan, J. E. Mangino, J. Mayer, K. J. Popovich, K. B. Stevenson, D. S. Yokoe, L. C. McDonald, J. Jernigan, and V. J. Fraser. Multicenter study of the impact of community onset infection on surveillance for infection. *Infection Control and Hospital Epidemiology*, 30(6):518–525, 2009.

[36] E. R. Dubberke, A. M. Butler, K. A. Reske, D. Agniel, M. A. Olsen, G. D'Angelo, L. C. McDonald, and V. J. Fraser. Attributable outcomes of endemic Clostridium difficile-associated disease in nonsurgical patients. *Emerging Infectious Diseases*, 14(7):1031, 2008.

[37] E. R. Dubberke, A. M. Butler, D. S. Yokoe, J. Mayer, B. H., J. E. Mangino, Y. M. Khan, K. J. Popovich, and V. J. Fraser. Multicenter study of Clostridium difficile infection rates from 2000 to 2006. *Infection Control and Hospital Epidemiology*, 31(10):1030–1037, 2010.

[38] E. R. Dubberke, A. M. Butler, D. S. Yokoe, J. Mayer, B. Hota, J. E. Mangino, Y. M. Khan, K. J. Popovich, K. B. Stevenson, L. C. McDonald, M. A. Olsen, and V. J. Fraser. Multicenter study of surveillance for hospital onset infection by the use of diagnosis codes. *Infection Control and Hospital Epidemiology*, 31(3):262–268, 2010.

[39] E. R. Dubberke, D. N. Gerding, D. Classen, K. M. Arias, K. Podgorny, D. J. Anderson, H. Burstin, et al. Strategies to prevent Clostridium difficile infections in acute care hospitals. *Infection Control and Hospital Epidemiology*, 29:S81–S92, 2008.

[40] E. R. Dubberke, K. M. McMullen, J. L. Mayfield, K. A. Reske, P. Georgantopoulos, D. K. Warren, and V. J. Fraser. Hospital-associated Clostridium difficile infection: is it necessary to track community-onset disease? *Infection Control and Hospital Epidemiology*, 30(4):332, 2009.

[41] E. R. Dubberke, K. A. Reske, L. C. McDonald, and V. J. Fraser. ICD-9 codes and surveillance for Clostridium difficile-associated disease. *Emerging Infectious Diseases*, 12(10):1576, 2006.

[42] E. R. Dubberke, K. A. Reske, M. A. Olsen, L. C. McDonald, and V. J. Fraser. Short- and long-term attributable costs of Clostridium difficile-associated disease in nonsurgical inpatients. *Clinical Infectious Diseases*, 46(4):497, 2008.

[43] E. R. Dubberke, K. A. Reske, M. A. Olsen, K. M. McMullen, J. L. Mayfield, L. C. McDonald, and V. J. Fraser. Evaluation of Clostridium difficile-associated disease pressure as a risk factor for C difficile-associated disease. *Archives of Internal Medicine*, 167(10):1092, 2007.

[44] E. R. Dubberke, K. A. Reske, Y. Yan, M. A. Olsen, L. C. McDonald, and V. J. Fraser. Clostridium difficile-associated disease in a setting of endemicity: identification of novel risk factors. *Clinical Infectious Diseases*, 45(12):1543, 2007.

[45] E. R. Dubberke and A. Wertheimer. Review of current literature on the economic burden of infection. *Infection Control and Hospital Epidemiology*, 30(1):57–66, 2008.

[46] G. Dumyati, V. Stevens, G. E. Hannett, A. D. Thompson, C. Long, D. MacCannell, and B. Limbago. Community-associated Clostridium difficile infections, Monroe County, New York, USA. *Emerging Infectious Diseases*, 18(3):392–400, 2012.

[47] T. J. Dunne, I. Gaboury, and M. C. Ashe. Falls in hospital increase length of stay regardless of degree of harm. *Journal of Evaluation in Clinical Practice*, 20(4):396, 2014.

[48] M. J. Durkin, A. W. Baker, K. V. Dicks, S. S. Lewis, L. F. Chen, D. J. Anderson, D. J. Sexton, and R. W. Moehring. A comparison between National Healthcare Safety Network laboratory-identified event reporting versus traditional surveillance for Clostridium difficile infection. *Infection Control and Hospital Epidemiology*, 36(2):125–131, 2015.

[49] W. H. Edwards, J. A. Morris, J. M. Jenkins, S. M. Bass, and E. J. MacKenzie. Evaluating quality, cost-effective health care. vascular database predicated on hospital discharge abstracts. *Annals of Surgery*, 213(5):433, 1991.

[50] A. Elixhauser, C. Steiner, D. R. Harris, and R. M. Coffey. Comorbidity measures for use with administrative data. *Medical Care*, 36(1):8, 1998.

[51] D. W. Eyre, M. L. Cule, D. J. Wilson, D. Griffiths, A. Vaughan, L. O'Connor, C. L. Ip, T. Golubchik, E. M. Batty, et al. Diverse sources of C. difficile infection identified on whole- genome sequencing. *The New England Journal of Medicine*, 369(13):1195, 2013.

[52] J. Feinglass and J. J. Holloway. The initial impact of the medicare prospective payment system on U.S. health care: A review of the literature. *Medical Care Research and Review*, 48(1):91–115, 1991.

[53] R. Fekety and A. B. Shah. Diagnosis and treatment of Clostridium difficile colitis. *JAMA*, 269(1):71, 1993.

[54] M. J. Fine, H. M. Pratt, D. S. Obrosky, J. R. Lave, L. J. McIntosh, D. E. Singer, C. M. Coley, and W. N. Kapoor. Relation between length of hospital stay and costs of care for patients with community-acquired pneumonia. *The American Journal of Medicine*, 109(5):378, 2000.

[55] A. J. Forster, M. Taljaard, N. Oake, K. Wilson, V. Roth, and C. van Walraven. The effect of hospital-acquired infection with Clostridium difficile on length of stay in hospital. *CMAJ*, 184(1):37, 2012.

[56] The Henry J. Kaiser Family Foundation. Hospital adjusted expenses per inpatient day. `http://kff.org/other/state-indicator/expenses-per-inpatient-day/`, 2014.

[57] A Freitas, T. Silva-Costa, F. Lopes, I. Garcia-Lema, A. Teixeira-Pinto, P. Brazdil, and A. Costa-Pereira. Factors influencing hospital high length of stay outliers. *BMC Health Services Research*, 12(1):265, 2012.

[58] S. S. Ghantoji, K. Sail, D. R. Lairson, H. L. DuPont, and K. W. Garey. Economic healthcare costs of Clostridium difficile infection: a systematic review. *Journal of Hospital Infection*, 74(4):309–318, 2010.

[59] L. G. Glance, P. W. Stone, D. B. Mukamel, and A. W. Dick. Increases in mortality, length of stay, and cost associated with hospital-acquired infections in trauma patients. *Archives of Surgery*, 146(7):794, 2011.

[60] E. M. Graffunder and R. A. Venezia. Risk factors associated with nosocomial methicillin-resistant staphylococcus aureus (MRSA) infection including previous use of antimicrobials. *The Journal of Antimicrobial Chemotherapy*, 49(6):999, 2002.

[61] N. Graves, F. Birrell, and M. Whitby. Effect of pressure ulcers on length of hospital stay. *Infection Control and Hospital Epidemiology*, 26(3):293–297, 2005.

[62] N. Graves, S. Harbarth, J. Beyersmann, A. Barnett, K. Halton, and B. Cooper. Estimating the cost of health care-associated infections: mind your p's and q's. *Clinical Infectious Diseases*, 50(7):1017, 2010.

[63] N. Graves, D. Weinhold, E. Tong, F. Birrell, S. Doidge, P. Ramritu, K. Halton, D. Lairson, and M. Whitby. Effect of healthcare-acquired infection on length of hospital stay and cost. *Infection Control and Hospital Epidemiology*, 28(3):280, 2007.

[64] V. B. Haley, A. G. DiRienzo, E. C. Lutterloh, and R. L. Stricof. Quantifying sources of bias in national healthcare safety network laboratory-identified Clostridium difficile infection rates. *Infection Control and Hospital Epidemiology*, 35(1):1–7, Jan 2014.

[65] K. Hauck and X. Zhao. How dangerous is a day in hospital? a model of adverse events and length of stay for medical inpatients. *Medical Care*, 49(12):1068, 2011.

[66] C. T. A. Herzig, J. Reagan, M. Pogorzelska-Maziarz, D. Srinath, and P. W. Stone. State-mandated reporting of health care-associated infections in the united states: Trends over time. *American Journal of Medical Quality*, June 2014.

[67] B. L. Hoh, Y. Y. Chi, M. F. Waters, J. Mocco, and F. G. Barker. Effect of weekend compared with weekday stroke admission on thrombolytic use, in-hospital mortality, discharge disposition, hospital charges, and length of stay in the Nationwide Inpatient Sample database, 2002 to 2007. *Stroke*, 41(10):2323, 2010.

[68] B. Hollingsworth. The measurement of efficiency and productivity of health care delivery. *Health Economics*, 17(10):1107–1128, 2008.

[69] L. H. F. Hoonhout, M. de Bruijne C., C. Wagner, M. Zegers, R. Waaijman, P. Spreeuwenberg, H. Asscheman, G. der Wal van, and M. van Tulder W. Direct medical costs of adverse events in Dutch hospitals. *BMC Health Services Research*, 9:27, 2009.

[70] S. D. Horn, P. D. Sharkey, J. M. Buckle, J. E. Backofen, R. F. Averill, and R. A. Horn. The relationship between severity of illness and hospital length of stay and mortality. *Medical Care*, 29(4):305, 1991.

[71] J. Janka and N. P. O'Grady. Clostridium difficile infection: current perspectives. *Current Opinion in Critical Care*, 15(2):149, 2009.

[72] W. R. Jarvis. Infection control and changing health- care delivery systems. *Emerging Infectious Diseases*, 7(2):170, 2001.

[73] M. H. Jen, S. Saxena, A. Bottle, P. Aylin, and R. C. G. Pollok. Increased health burden associated with Clostridium difficile diarrhoea in patients with inflammatory bowel disease. *Alimentary Pharmacology & Therapeutics*, 33(12):1322–1331, June 2011.

[74] Y. Jiang, S. Viner-Brown, and R. Baier. Burden of hospital-onset Clostridium difficile infection in patients discharged from Rhode Island hospitals, 2010-2011: application of present on admission indicators. *Infection Control and Hospital Epidemiology*, 34(7):700, 2013.

[75] V. Monge Jodra, L. Sainz de Los Terreros Soler, C. Diaz-Agero Perez, C. M. Saa Requejo, and N. Plana Farras. Excess length of stay attributable to surgical site infection following hip replacement: a nested case-control study. *Infection Control and Hospital Epidemiology*, 27(12):1299, 2006.

[76] T. Johnson, J. M. Kane, R. Odwazny, and R. McNutt. Association of the position of a hospital-acquired condition diagnosis code with changes in medicare severity diagnosis-related group assignment. *Journal of Hospital Medicine*, 9(11):707–713, November 2014.

[77] K. R. Jones. Predicting hospital charge and stay variation. the role of patient teaching status, controlling for diagnosis-related group, demographic characteristics, and severity of illness. *Medical Care*, 23(3):220, 1985.

[78] B. Jönsson and B. Lindgren. Five common fallacies in estimating the economic gains of early discharge. *Social Science and Medicine. Part C Medical Economics*, 14(1):27–33, 1980.

[79] R. L. Kalish, J. Daley, C. C. Duncan, R. B. Davis, G. A. Coffman, and L. I. Iezzoni. Costs of potential complications of care for major surgery patients. *American Journal of Medical Quality*, 10(1):48, 1995.

[80] A. Kaltsas, M. Simon, L. H. Unruh, C. Son, D. Wroblewski, K. A. Musser, K. Sepkowitz, N. E. Babady, and M. Kamboj. Clinical and laboratory characteristics of Clostridium difficile infection in patients with discordant diagnostic test results. *Journal of Clinical Microbiology*, 50(4):1303, 2012.

[81] M. Kamboj, N. E. Babady, J. W. Marsh, J. L. Schlackman, C. Son, J. Sun, J. Eagan, Y. Tang, and K. Sepkowitz. Estimating risk of C. difficile transmission from PCR positive but cytotoxin negative cases. *PloS One*, 9(2):e88262, 2014.

[82] A. G. Kamthan, H. W. Bruckner, S. Z. Hirschman, and S. G. Agus. Clostridium difficile diarrhea induced by cancer chemotherapy. *Archives of Internal Medicine*, 152(8):1715, 1992.

[83] S. V. Kazakova, K. Ware, B. Baughman, O. Bilukha, A. Paradis, S. Sears, A. Thompson, B. Jensen, L. Wiggs, et al. A hospital outbreak of diarrhea due to an emerging epidemic strain of Clostridium difficile. *Archives of Internal Medicine*, 166(22):2518, 2006.

[84] V. R. Kee. Clostridium difficile infection in older adults: A review and update on its management. *American Journal of Geriatric Pharmacotherapy*, 10(1):14–24, 2012.

[85] C. Kenneally, J. M. Rosini, L. P. Skrupky, J. A. Doherty, J. M. Hollands, E. Martinez, W. McKenzie, T. Murphy, J. R. Smith, et al. Analysis of 30-day mortality for Clostridium difficile-associated disease in the ICU setting. *Chest*, 132(2):418, 2007.

[86] J. Kim, S. A. Smathers, P. Prasad, K. H. Leckerman, S. Coffin, and T. Zaoutis. Epidemiological features of clostridium difficile- associated disease among inpatients at children's hospitals in the United States, 2001-2006.(report). *Pediatrics*, 122(6):1266, 2008.

[87] P. K. Kutty, S. R. Benoit, C. W. Woods, A. C. Sena, S. Naggie, J. Frederick, J. Engemann, S. Evans, B. C. Pien, et al. Assessment of Clostridium difficile-associated disease surveillance definitions, North Carolina, 2005. *Infection Control and Hospital Epidemiology*, 29(3):197, 2008.

[88] L. Kyne, M. B. Hamel, R. Polavaram, and C. P. Kelly. Health care costs and mortality associated with nosocomial diarrhea due to Clostridium difficile. *Clinical Infectious Diseases*, 34(3):346, 2002.

[89] S. J. Lawrence, L. A. Puzniak, B. N. Shadel, K. N. Gillespie, M. H. Kollef, and L. M. Mundy. Clostridium difficile in the intensive care unit: epidemiology, costs, and colonization pressure. *Infection Control and Hospital Epidemiology*, 28(2):123, 2007.

[90] K. Lesperance, M. W. Causey, M. Spencer, and S. R. Steele. The morbidity of Clostridium difficile infection after elective colonic resection results from a national population database. *The American Journal of Surgery*, 201(2):141–148, 2011.

[91] D. R. Levinson. Adverse events in hospitals: National incidence among medicare beneficiaries. Technical report, Washington, DC: Office of the Inspector General, Department of Health and Human Services, 2010.

[92] J. Librero, S. Peiro, and R. Ordiana. Chronic comorbidity and outcomes of hospital care: length of stay, mortality, and readmission at 30 and 365 days. *Journal of Clinical Epidemiology*, 52(3):171, 1999.

[93] M. J. Lipp, D. C. Nero, and M. A. Callahan. Impact of hospital acquired Clostridium difficile. *Journal of Gastroenterology and Hepatology*, 27(11):1733–1737, 2012.

[94] Y. Longtin, J. Longtin, D. Goulet, C. Beaulieu, V. Loungnarath, C. Garenc, B. Paquet-Bolduc, G. Brochu, and S. Trottier. Impact of the type of diagnostic assay on clostridium difficile infection and complication rates in a mandatory reporting program. *Clinical Infectious Diseases*, 56(1):67, 2013.

[95] V. G. Loo, A. Bourgault, L. Poirier, F. Lamothe, S. Michaud, N. Turgeon, B. Toye, A. Beaudoin, E. H. Frost, et al. Host and pathogen factors for Clostridium difficile infection and colonization. *The New England Journal of Medicine*, 365(18):1693, 2011.

[96] M. Macedo-Vinas, G. De Angelis, P. Rohner, E. Safran, A. Stewardson, C. Fankhauser, J. Schrenzel, D. Pittet, and S. Harbarth. Burden of meticillin-resistant staphylococcus aureus infections at a swiss university hospital: excess length of stay and costs. *Journal of Hospital Infection*, 84(2):132, 2013.

[97] A. P. Macgowan, I. Brown, R. Feeney, A. Lovering, S. Y. McCulloch, D. S. Reeves, M. G. Cheesman, H. G. Shetty, M. H. Wilcox, and J. G. Cunnliffe. Clostridium difficile-associated diarrhoea and length of hospital stay. *The Journal of hospital infection*, 31(3):241, 1995.

[98] P. A. Maguire, I. C. Taylor, and R. W. Stout. Elderly patients in acute medical wards: factors predicting length of stay in hospital. *British medical journal (Clinical research ed.)*, 292(6530):1251, 1986.

[99] C. McDermott and G. N. Stock. Hospital operations and length of stay performance. *International Journal of Operations & Production Management*, 27(9):1020–1042, 2007.

[100] M. F. McDermott, D. G. Murphy, R.J. Zalenski, R. J. Rydman, M. McCarren, D. Marder, B. Jovanovic, K. Kaur, R. R. Roberts, et al. A comparison between emergency diagnostic and treatment unit and inpatient care in the management of acute asthma. *Archives of Internal Medicine*, 157(18):2055, 1997.

[101] L. C. McDonald, G. E. Killgore, A. Thompson, R. C. Owens, S. V. Kazakova, S. P. Sambol, S. Johnson, and D. N. Gerding. An epidemic, toxin gene-variant strain of Clostridium difficile. *The New England Journal of Medicine*, 353(23):2433, 2005.

[102] L. V. McFarland, M. E. Mulligan, R. Y. Kwok, and W. E. Stamm. Nosocomial acquisition of Clostridium difficile infection. *The New England Journal of Medicine*, 320(4):204, 1989.

[103] S. M. Mcglone, R. R. Bailey, S. M. Zimmer, M. J. Popovich, Y. Tian, P. Ufberg, R. R. Muder, and B. Y. Lee. The economic burden of Clostridium difficile. *Clinical Microbiology and Infection*, 18(3):282–289, 2012.

[104] E. A. McGlynn, S. M. Asch, J. Adams, J. Keesey, J. Hicks, A. DeCristofaro, and E. A. Kerr. The quality of health care delivered to adults in the United States. *The New England Journal of Medicine*, 348(26):2635, 2003.

[105] R. McNutt, T. J. Johnson, R. Odwazny, Z. Remmich, K. Skarupski, S. Meurer, S. Hohmann, and B. Harting. Change in MS-DRG assignment and hospital reimbursement as a result of Centers for Medicare & Medicaid changes in payment for hospital-acquired conditions: is it coding or quality? *Quality Management in Health Care*, 19(1):17–24, January 2010.

[106] A. C Miller, L. A. Polgreen, and P. M. Polgreen. Practical limitations of case-control exact matching. Working paper, 2014.

[107] A. C. Miller, L. A. Polgreen, P. M. Polgreen, and J. E. Cavanaugh. Hospital incidence of Clostridium difficile infections as a novel marker for prolonged length of stay in uninfected patients. Working paper, 2014.

[108] R. H. Miller and H. S. Luft. Managed care plan performance since 1980: a literature analysis. *JAMA*, 271(19):1512, 1994.

[109] B. G. Mitchell and A. Gardner. Prolongation of length of stay and Clostridium difficile infection: a review of the methods used to examine length of stay due to healthcare associated infections. *Antimicrobial Resistance and Infection Control*, 1(1):14, 2012.

[110] B. G. Mitchell, A. Gardner, A. G. Barnett, J. E. Hiller, and N. Graves. The prolongation of length of stay because of Clostridium difficile infection. *American Journal of Infection Control*, 42(2):164, 2014.

[111] M. C. Munin, T. E. Rudy, N. W. Glynn, L. S. Crossett, and H. E. Rubash. Early inpatient rehabilitation after elective hip and knee arthroplasty. *JAMA*, 279(11):847–852, March 1998.

[112] E. Munoz, K. Mulloy, J. Goldstein, N. Tenenbaum, and L. Wise. Costs, quality, and the volume of surgical oncology procedures. *Archives of Surgery*, 125(3):360, 1990.

[113] C. R. Murphy, T. R. Avery, E. R. Dubberke, and S. S. Huang. Frequent hospital readmissions for Clostridium difficile infection and the impact on estimates of hospital-associated C. difficile burden. *Infection Control and Hospital Epidemiology*, 33(1):20–28, 2012.

[114] D. M. Needham, G. Anderson, G. H. Pink, I. McKillop, G. A. Tomlinson, and A. S. Detsky. A province-wide study of the association between hospital resource allocation and length of stay. *Health Services Management Research*, 16(3):155, 2003.

[115] L. D. Nordgren, T. Johnson, M. Kirschbaum, and M. L. Peterson. Medical errors: excess hospital costs and lengths of stay. *Journal for Healthcare Quality*, 26(2):42, 2004.

[116] C. M. Nylund, A. Goudie, J. M. Garza, G. Fairbrother, and M. B. Cohen. Clostridium difficile infection in hospitalized children in the United States. *Archives of Pediatrics & Adolescent Medicine*, 165(5):451, 2011.

[117] J. A. O'Brien, B. J. Lahue, J. J. Caro, and D. M. Davidson. The emerging infectious challenge of Clostridium difficile-associated disease in Massachusetts hospitals: clinical and economic consequences. *Infection Control and Hospital Epidemiology*, 28(11):1219, 2007.

[118] A. Pakyz, N. V. Carroll, S. E. Harpe, M. Oinonen, and R. E. Polk. Economic impact of clostridium difficile infection in a multihospital cohort of academic health centers. *Pharmacotherapy*, 31(6):546–551, 2011.

[119] C. Pant, M. P. Anderson, J. A. O' Connor, C. M. Marshall, A. Deshpande, and T. J. Sferra. Association of Clostridium difficile infection with outcomes of hospitalized solid organ transplant recipients: results from the 2009 Nationwide Inpatient Sample database. *Transplant Infectious Disease*, 14(5):540–547, 2012.

[120] P. Pavese, N. Saurel, J. Labarere, C. Decouchon, J. P. Vittoz, L. Foroni, M. Maurin, and P. Francois. Does an educational session with an infectious diseases physician reduce the use of inappropriate antibiotic therapy for inpatients with positive urine culture results? a controlled before-and-after study. *Infection Control and Hospital Epidemiology*, 30(6):596, 2009.

[121] N. R. Payne, J. H. Carpenter, G. J. Badger, J. D. Horbar, and J. Rogowski. Marginal increase in cost and excess length of stay associated with nosocomial bloodstream infections in surviving very low birth weight infants. *Pediatrics*, 114(2):348, 2004.

[122] J. Pepin, L. Valiquette, and B. Cossette. Mortality attributable to nosocomial clostridium difficile- associated disease during an epidemic caused by a hypervirulent strain in quebec. *CMAJ*, 173(9):1037, 2005.

[123] D. Pittet, D. Tarara, and R. P. Wenzel. Nosocomial bloodstream infection in critically ill patients: excess length of stay, extra costs, and attributable mortality. *JAMA*, 271(20):1598, 1994.

[124] R. Plowman, N. Graves, M. A. Griffin, J. A. Roberts, A. V. Swan, B. Cookson, and L. Taylor. The rate and cost of hospital-acquired infections occurring in patients admitted to selected specialties of a district general hospital in England and the national burden imposed. *The Journal of Hospital Infection*, 47(3):198–209, Mar 2001.

[125] E. Polverejan, J. C. Gardiner, C. J. Bradley, M. Holmes-Rovner, and D. Rovner. Estimating mean hospital cost as a function of length of stay and patient characteristics. *Health Economics*, 12(11):935–947, 2003.

[126] A. I. Qureshi, M. M. Adil, H. Zacharatos, and M. F. Suri. Factors associated with length of hospitalization in patients admitted with transient ischemic attack in United States. *Stroke*, 44(6):1601, 2013.

[127] J. Rapoport, D. Teres, S. Lemeshow, J. S. Avrunin, and R. Haber. Explaining variability of cost using a severity-of-illness measure for ICU patients. *Medical Care*, 28(4):338, 1990.

[128] J. Reagan and C. Hacker. Laws pertaining to healthcare-associated infections: a review of 3 legal requirements. *Infection Control and Hospital Epidemiology*, 33(1):75–80, January 2012.

[129] J. Reagan, C. T. A. Herzig, M. Pogorzelska-Maziarz, A. W. Dick, P. W. Stone, and J. Divya Srinath. State law mandates for reporting of healthcare-associated clostridium difficile infections in hospitals. *Infection Control and Hospital Epidemiology*, 36(3):350–352, March 2015.

[130] R. Ricciardi, K. Harriman, N. N. Baxter, L. K. Hartman, R. J. Town, and B. A. Virnig. Predictors of Clostridium difficile colitis infections in hospitals. *Epidemiology and Infection*, 136(7):913–921, July 2008.

[131] M. M. Riggs, A. K. Sethi, T. F. Zabarsky, E. C. Eckstein, R. L. P. Jump, and C. J. Donskey. Asymptomatic carriers are a potential source for transmission of epidemic and nonepidemic clostridium difficile strains among long-term care facility residents. *Clinical Infectious Diseases*, 45(8):992, 2007.

[132] T. V. Riley, J. P. Codde, and I. L. Rouse. Increased length of hospital stay due to Clostridium difficile associated diarrhoea. *Lancet*, 345(8947):455, 1995.

[133] P. E. Rivard, S. L. Luther, C. L. Christiansen, S. S. Zhao, A. Loveland, P. S. Elixhauser, A. K. Romano, and A. K. Rosen. Using patient safety indicators to estimate the impact of potential adverse events on outcomes. *Medical Care Research and Review*, 65(1):67, 2008.

[134] R. R. Roberts, P. W. Frutos, G. C. Ciavarella, L. M. Gussow, E. K. Mensah, L. M. Kampe, H. E. Straus, G. Joseph, and R. J. Rydman. Distribution of variable vs fixed costs of hospital care. *JAMA*, 281(7):644, 1999.

[135] R. R. Roberts, R. J. Zalenski, E. K. Mensah, R. J. Rydman, G. Ciavarella, L. Gussow, K. Das, L. M. Kampe, B. Dickover, et al. Costs of an emergency department- based accelerated diagnostic protocol vs hospitalization in patients with chest pain: a randomized controlled trial. *JAMA*, 278(20):1670, 1997.

[136] J. C. Robinson. Decline in hospital utilization and cost inflation under managed care in California. *JAMA*, 276(13):1060–1064, 1996.

[137] A. K. Rosen, P. Rivard, S. Zhao, S. Loveland, D. Tsilimingras, C. L. Christiansen, A. Elixhauser, and P. S. Romano. Evaluating the patient safety indicators: how well do they perform on Veterans Health Administration data? *Medical Care*, 43(9):873, 2005.

[138] R. J. Rydman, M. L. Isola, R. R. Roberts, R. J. Zalenski, M. F. McDermott, D. G. Murphy, M. M. McCarren, and L. M. Kampe. Emergency department observation unit versus hospital inpatient care for a chronic asthmatic population: a randomized trial of health status outcome and cost. *Medical Care*, 36(4):599, 1998.

[139] T. Micek S., A. Schramm G., R. Morrow L., H. Frazee E., H. Personett H., H. Doherty J., H. Hampton N., H. Hoban A., H. Lieu A., H. Mckenzie M., H. Dubberke E., and H. Kollef M. Clostridium difficile infection: A multicenter study of epidemiology and outcomes in mechanically ventilated patients. *Critical Care Medicine*, 41(8):1968–1975, 2013.

[140] D. B. Scheurer, L. S. Hicks, E. F. Cook, and J. L. Schnipper. Accuracy of ICD-9 coding for Clostridium difficile infections: a retrospective cohort. *Epidemiology and Infection*, 135(6):1010, 2007.

[141] M. A. Schuster, E. A. McGlynn, and R. H. Brook. How good is the quality of health care in the United States? *The Milbank Quarterly*, 83(4):843–895, 2005.

[142] W. B. Schwartz and D. N. Mendelson. Hospital cost containment in the 1980s: hard lessons learned and prospects for the 1990s. *The New England Journal of Medicine*, 324(15):1037, 1991.

[143] I. Scott, D. Youlden, and M. Coory. Are diagnosis specific outcome indicators based on administrative data useful in assessing quality of hospital care? *Quality & Safety in Health Care*, 13(1):32, 2004.

[144] E. M. Skobeloff, W. H. Spivey, S. S. St. Clair, and J. M. Schoffstall. The influence of age and sex on asthma admissions. *JAMA*, 268(24):3437, 1992.

[145] X. Song, J. G. Bartlett, K. Speck, A. Naegeli, K. Carroll, and T. M. Perl. Rising economic impact of clostridium difficile-associated disease in adult hospitalized patient population. *Infection Control and Hospital Epidemiology*, 2008.

[146] W. N. Southern, E. Y. Bellin, and J. H. Arnsten. Longer lengths of stay and higher risk of mortality among inpatients of physicians with more years in practice. *American Journal of Medicine*, 124(9):868, 2011.

[147] A. E. Stephen and D. L. Berger. Shortened length of stay and hospital cost reduction with implementation of an accelerated clinical care pathway after elective colon resection. *Surgery*, 133(3):277–282, March 2003.

[148] C. M. Surawicz, L. J. Brandt, D. G. Binion, A. N. Ananthakrishnan, S. R. Curry, P. H. Gilligan, L. V. McFarland, M. Mellow, and B. S. Zuckerbraun. Guidelines for diagnosis, treatment, and prevention of Clostridium difficile infections. *The American Journal of Gastroenterology*, 108(4):478, 2013.

[149] P. A. Taheri, D. A. Butz, and L. J. Greenfield. Length of stay has minimal impact on the cost of hospital admission. *Journal of the American College of Surgeons*, 191(2):123, 2000.

[150] D. M. Tehrani, M. J. Phelan, C. Cao, J. Billimek, R. Datta, H. Nguyen, H. Kwark, and S. S. Huang. Substantial shifts in ranking of california hospitals by hospital- associated methicillin-resistant staphylococcus aureus infection following adjustment for hospital characteristics and case mix. *Infection control and hospital epidemiology*, 35(10):1263, 2014.

[151] D. M. Tehrani, M. J. Phelan, C. Cao, J. Billimek, R. Datta, Ho. Nguyen, H. Kwark, and S. S. Huang. Substantial shifts in ranking of California hospitals by hospital-associated methicillin-resistant Staphylococcus aureus infection following adjustment for hospital characteristics and case mix. *Infection Control and Hospital Epidemiology*, 35(10):1263–1270, October 2014.

[152] M. Than, S. Aldous, S. J. Lord, S. Goodacre, C. M. A. Frampton, R. Troughton, P. George, C. M. Florkowski, M. Ardagh, et al. A 2-hour diagnostic protocol for possible cardiac chest pain in the emergency department: A randomized clinical trial. *JAMA Internal Medicine*, 2013.

[153] J. W. Thomas, K. E. Guire, and G. G. Horvat. Is patient length of stay related to quality of care? *Hospital & Health Services Administration*, 42(4):489–507, Winter 1997.

[154] P. Thungjaroenkul, G. G. Cummings, and A. Embleton. The impact of nurse staffing on hospital costs and patient length of stay: a systematic review. *Nursing Economic$*, 25(5):255, 2007.

[155] N. G. Tornieporth, R. B. Roberts, J. John, A. Hafner, and L. W. Riley. Risk factors associated with vancomycin-resistant enterococcus faecium infection or colonization in 145 matched case patients and control patients. *Clinical Infectious Diseases*, 23(4):767, 1996.

[156] R. P. Vonberg, C. Reichardt, M. Behnke, F. Schwab, S. Zindler, and P. Gastmeier. Costs of nosocomial Clostridium difficile-associated diarrhoea. *The Journal of Hospital Infection*, 70(1):15, 2008.

[157] F. Vrijens, F. Hulstaert, de Sande Van, S. Devriese, I. Morales, and Y. Parmentier. Hospital-acquired, laboratory-confirmed bloodstream infections: linking national surveillance data to clinical and financial hospital data to estimate increased length of stay and healthcare costs. *Journal of Hospital Infection*, 75(3):158, 2010.

[158] R. M. Wachter and L. Goldman. The hospitalist movement 5 years later. *JAMA*, 287(4):487, 2002.

[159] A. S. Walker, D. W. Eyre, D. H. Wyllie, K. E. Dingle, R. M. Harding, L. O'Connor, D. Griffiths, A. Vaughan, J. Finney, et al. Characterisation of Clostridium difficile hospital ward-based transmission using extensive epidemiological data and molecular typing. *PLoS Medicine*, 9(2):e1001172, 2012.

[160] T. Wen, S. He, F. Attenello, S. Y. Cen, M. Kim-Tenser, P. Adamczyk, A. P. Amar, N. Sanossian, and W. J. Mack. The impact of patient age and comorbidities on the occurrence of never events? in cerebrovascular surgery: an analysis of the nationwide inpatient sample: Clinical article. *Journal of neurosurgery*, 121(3):580–586, 2014.

[161] J. D. Whitehouse, N. D. Friedman, K. B. Kirkland, W. J. Richardson, and D. J. Sexton. The impact of surgical-site infections following orthopedic surgery at a community hospital and a university hospital: adverse quality of life, excess length of stay, and extra cost. *Infection Control and Hospital Epidemiology*, 23(4):183, 2002.

[162] M. H. Wilcox, J. G. Cunniffe, C. Trundle, and C. Redpath. Financial burden of hospital-acquired Clostridium difficile infection. *The Journal of hospital infection*, 34(1):23, 1996.

[163] C. A. Wong, A. J. Recktenwald, M. L. Jones, B. M. Waterman, M. L. Bollini, and W. C. Dunagan. The cost of serious fall-related injuries at three midwestern hospitals. *Joint Commission Journal on Quality and Patient Safety*, 37(2):81, 2011.

[164] J. M. Wooldridge. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, 2nd edition, 2010.

[165] A. Z. Yaldo, J. L. Sullivan, and Z. Li. Factors influencing physicians' decision to discharge hospitalized patients infected with methicillin-resistant staphylococcus aureus. *American Journal of Health-System Pharmacy*, 58(18):1756, 2001.

[166] J. Yang, C. Peek-Asa, V. Allareddy, G. Phillips, Y. Zhang, and G. Cheng. Patient and hospital characteristics associated with length of stay and hospital charges for pediatric sports-related injury hospitalizations in the United States, 2000-2003. *Pediatrics*, 119(4):e813, 2007.

[167] H. Yasunaga, H. Horiguchi, H. Hashimoto, S. Matsuda, and K. Fushimi. The burden of Clostridium difficile-associated disease following digestive tract surgery in japan. *Journal of Hospital Infection*, 82(3):175–180, 2012.

[168] C. Zhan and M. R. Miller. Excess length of stay, charges, and mortality attributable to medical injuries during hospitalization. *JAMA*, 290(14):1868, 2003.

[169] M. D. Zilberberg, G. S. Tillotson, and C. McDonald. Clostridium difficile infections among hospitalized children, United States, 1997- 2006. *Emerging Infectious Diseases*, 16(4):604, 2010.