COST ATTRIBUTABLE TO HOSPITAL-ACQUIRED CLOSTRIDIUM DIFFICILE INFECTION (CDI)

by

Kelly Baekyung Choi

Thesis submitted to the

Faculty of Graduate and Postdoctoral Studies

in partial fulfillment of the requirements for the degree of

M.Sc.in Epidemiology

Epidemiology and Community Medicine

Faculty of Medicine

University of Ottawa

Abstract

Introduction: *Clostridium difficile* infection (CDI) is a common hospital-acquired infection and a financial burden on the healthcare system. There is a need to reduce its impact on patients and the entire health system. More accurate estimates of the financial impact of CDI will assist hospitals in creating better CDI reduction strategies with limited resources. Previous research has not sufficiently accounted for the skewed nature of hospital cost data, baseline patient mortality risk, and the time-varying nature of CDI.

Objective: We conducted a retrospective cohort study to estimate the cost impact of hospital-acquired CDI from the hospital perspective, using a number of analytical approaches.

Method: We used clinical and administrative data for inpatients treated at The Ottawa Hospital to construct an analytical data set. Our primary outcome was direct costs and our primary exposure was hospital-acquired CDI. We performed the following analyses: Ordinary least square regression and generalized linear regression as time-fixed methods, and Kaplan-Meier survival curve and Cox regression models as time-varying methods.

Results: A total of 49,888 admissions were included in this study (mean (SD) age of 64.6 ± 17.8 years, median (IQR) baseline mortality risk of 0.04 (0.01-0.14)). 360 (0.73%) patients developed CDI. Estimates of incremental cost due to CDI were substantially higher when using time-fixed methods than time-varying methods. Using methods that appropriately account for the time-varying nature of the exposure, the estimated incremental cost due to CDI was $8,997 per patient. In contrast, estimates from time-fixed methods ranged from $49,150 to $55,962: about a six fold difference.

Conclusion: Estimates of hospital costs are strongly influenced by the time-varying nature of CDI as well as baseline mortality risk. If studies do not account for these factors, it is likely that the impact of hospital-acquired CDI will be overestimated.

Acknowledgements

I would like to express my sincere gratitude to my thesis supervisors Drs Alan Forster and Kathryn Suh for providing me with all the necessary facilities. Dr. Forster continually and persuasively conveyed a spirit of adventure in regard to my project. He was always available for every question in order to move forward with my thesis. There were also times when he challenged me with various questions and tasks. At that time, it was very difficult to address them but now that I look back, those challenges made my thesis more meaningful.

I have been extremely lucky to have Dr. Suh as my supervisor. She was the person who opened up my eyes to healthcare epidemiology. I am very grateful for Dr. Suh's invaluable advice and support on both an academic and personal level. She provided the scientific and clinical guidance throughout the process of completing this thesis. I will never forget her ongoing encouragements of "You can DO IT!" And finally, I DID IT!

Besides my thesis supervisors, I would like to thank the rest of my committee: Drs Monica Taljaard and Karen Lee. Despite Dr.Taljaard's many other academic and professional commitments, she always provided me all the necessary guidance and insight throughout my thesis. Her wisdom, knowledge and commitment inspired and motivated me. Dr. Lee's advice on healthcare costs was greatly appreciated.

Dr. Virginia Roth, for providing the funding to qualify for the entrance scholarship was greatly appreciated.

I would like to thank Natalie Oake for her kindness in always responding to my questions about SAS code and problems. I also would like to thank Deanna Rothwell for her valuable time through our regular meetings to discuss my thesis. I remember all her generosity and encouragement. I would like to thank Nadine, Li, Eric, Ciera, Laura and Jocelyn for all their encouragement and the many discussions we had regarding SAS programs. Ilze and Naghmeh's endless moral support will forever be remembered as a wonderful part of my life. I would also like to give sincere thanks to Nadea Saikaley for all her understanding and support.

I would like to thank Denise Gravel who provided me the flexibility of fine tuning my thesis while working. If I had not been given this flexibility, completing my thesis would have taken much longer than anticipated.

This thesis would not have been possible without the help, support and great patience of my amazing and lovely husband Andrew. I will never forget our numerous late nights of reading my thesis draft and figuring out statistical formulas. I am so glad that I am married to someone like him who is willing to spend our QUALITY TIME discussing statistical formulas.

My parents and my sister Erin Choi who have always supported, encouraged and believed in me, in all my endeavours and who so lovingly and unselfishly cared for our girls, Alyssa and Jane. Their endless support allowed me to focus on my thesis free of worry.

Lastly, I would like to thank God for leading me to where I am supposed to be. I always believed that this thesis was part of God's plan for me. And because I believe that "I can do everything through him who gives me strength." (Philippians 4:13), I close this chapter of my life and move on to the next.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

1.1 *Hospital-acquired Clostridium difficile infection (CDI)*

Clostridium difficile is a spore forming gram-positive bacillus that causes significant diseases in humans[1]. It exerts its virulence through the production of toxins that bind to human colonic epithelial cells, resulting in inflammation[2]. The manifestations of *Clostridium difficile* infection (CDI) include diarrhea, pseudo-membranous colitis, toxic megacolon, sepsis, and even death[3].

Clostridium difficile is the most frequent cause of healthcare associated diarrhea[4]. In addition, there is a growing body of evidence that the incidence and severity of hospital-acquired CDI is increasing in the United States, Canada and many other countries worldwide[4-6].

Firstly, the Canadian Nosocomial Infection Surveillance Program (CNISP) estimated that the incidence of hospital-acquired CDI in Canada was 4.45 cases per 1,000 admissions in 2005 and reported that the mortality rate was 5.7%, which was 4 times higher than the mortality rate reported in 1997[7]. Secondly, hospitals in Quebec experienced a dramatic increase in the incidence of hospital-acquired CDI between 2002 to late 2006 [8-10]. Finally, the incidence of hospital-acquired CDI all across the US doubled between 2001 and 2003 [11]. There are multiple reasons for this increase, but the frequent use of antibiotics on hospitalized patients is likely a major factor. This increase in rate is especially alarming because of the emergence of a more virulent strain of *C. difficile* (B1/NAP1) that can produce more toxin than the historic strain [12, 13].

As a result of this evolving epidemiology, there a greater urgency to more clearly understand CDI's impact and thereby develop better policies and practices to reduce its incidence rate and lower its associated health care cost. A number of studies have evaluated the epidemiology of hospital-acquired CDI and have determined consistently that the most common risk factors for CDI are older age, increased severity of underlying disease, increased length of stay and use of antibiotics[5]. For example, one study reported that the risk of acquiring CDI is about 1.04 times (95% CI 1.00-1.07) higher as each year of age increases[5]. In another study by Kyne et al., they reported that patients with extremely severe underlying disease were 17.6 (95% CI 5.8-54.5) times more likely to acquire CDI versus those with mild disease[14]. As these factors are also associated with health care costs, defining the cost specifically attributable to CDI remains a challenge.

1.2 *Financial burden of hospital-acquired CDI*

Several studies have analyzed the financial burden associated with hospital-acquired CDI. As expected, their results suggest a higher healthcare cost for CDI patients. The reproducibility in these results varies, as there were differences in the study populations, settings, or the statistical methods that were used in the different studies[15, 16].

For example, in a prevalence study that was conducted by Miller et al. of 19 acute care Canadian hospitals, variables such as the average length of stay for hospital-acquired CDI readmission, minimum cost per day per bed, and annual cost of antibiotic therapy were used to estimate the cost for treating CDI readmission cases. They reported that hospital-acquired CDI cases cost a minimum of $128,200 (Canadian dollars) per year per facility[17].

Similarly, studies in the US show a trend of increasing healthcare expenditure for CDI patients. In one prospective cohort study conducted by Kyne et al., they reported that the adjusted hospital cost of patients whose stay was complicated by hospital-acquired CDI was 54% (95% CI: 17% - 103%) higher than those whose stay was not complicated by hospital-acquired CDI[16]. After controlling for potential confounding variables using a linear regression model, the cost associated with hospital-acquired CDI was reported to be $3,669 US (95% CI, $1,126-$7,024)[16]. By extrapolation, the investigators predict that the total annual cost for treating CDI in all US hospitals would exceed $1.1 billion.

Pakyz et al. conducted a multi-center cohort study in the US and reported that the mean cost for CDI patients was $55,796 (95%CI $55,369-$56,168) versus $28,609 (95% CI $28,330-$28,888) for the control group, a two-fold difference[18].

A notable deficiency in the aforementioned studies is the failure to consider the onset of CDI as a time-varying variable. Given that CDI typically develops after hospital admission, it is inappropriate to attribute the entire hospital cost of a CDI admission to the infection. The reasons for not accounting for CDI as a time-varying variable may be due to: the limited source of data (i.e. lack of information available about micro-cost for each patient), the inability to accurately determine the date of CDI acquisition, and most importantly, the statistical difficulty of properly handling time-varying covariates. For these reasons, prior studies most likely overestimated the cost impact of CDI.

1.3 *Analytical challenges and approaches for estimating cost*

The following factors make analyzing the cost associated with hospital-acquired CDI challenging: accurately identifying hospital-acquired CDI cases, needing to control for time-

dependent exposure, accounting for confounding factors, handling skewness of cost data, determining the incremental cost impact on an additive scale, and accounting for mortality. The following sections will discuss these challenges in more detail and will elaborate on possible approaches to handle them.

Firstly, hospital-acquired CDI cases must be identified accurately. While in theory, one could simply use coding within the separation record (which takes advantage of the International Classification of Diseases (ICD)), this would likely lead to an underestimation of the true incidence of CDI as the sensitivity of this approach is only 70%[19, 20].

In order to improve hospital-acquired CDI case detection, it is possible to use electronic laboratory data. For example, investigators at The Ottawa Hospital developed an algorithm based on a text-searching technique applied to the electronic *C.difficile* laboratory reports[21]. By linking this information to admission and discharge data the investigators were able to classify CDI cases as either *hospital* or *community acquired*. This algorithm was shown to have a 100% sensitivity[21].

Secondly, the time-dependent nature of hospital-acquired CDI must be accounted for in the analysis. It is important to take into account for the time at which CDI was acquired when trying to estimate the cost attributable to hospital-acquired CDI. However, previous studies classified patients as either *CDI* or *non-CDI* cases at the time of their admission even though the CDI acquisition may have occurred during the patient's hospital stay. In these approaches, CDI acquisition was modeled as a time-fixed variable. As stated above, the probable consequence of this was overestimating the cost of CDI (i.e. Pakyz et. al. showed a two fold increase versus those that did not develop CDI [18]).

In order to control for the time-dependent nature of CDI, two pieces of information are required: the date of CDI acquisition and the daily cost of each encounter. From the CDI acquisition date, hospital costs can be categorized as either *before CDI acquisition* or *after CDI acquisition*.

Thirdly, the models must make adjustments for potential confounding factors that may be associated with hospital cost and CDI acquisition (such as the disease severity and patient demographic factors such as gender and age). For example, sicker patients or aged patients are likely to cost more due to their slow recovery process or their need for more assistance for daily activities such as eating or bathing. These factors contribute to higher hospital costs ***and*** increased risk of CDI acquisition. It is necessary, therefore, to include an analytical technique that takes into account for the potential confounding.

Fourthly, the skewed cost data must be handled appropriately. Because hospital costs are not normally distributed (they are usually positively skewed due to a small portion of patients whose cost of care is extremely high), standard linear regression cannot be applied directly. One commonly used method to improve the normality of the data is to log transform the outcome variable prior to the analysis. However, this method produces estimates on the multiplicative scale, which does not give inferences on the original scale (in dollars). Therefore, additional steps of retransforming the estimated values are required in order to produce cost estimates on the scale of interest (mean cost in dollars). However, this simple retransformation can produce biased cost estimates when the error terms are not normally distributed. In such case, it requires a retransformation bias correction. One approach to correct such bias is Duan's smearing factor adjustment, which is a well-known non-parametric method[22]. Other studies used a generalized linear regression model (GLM) as a more sophisticated approach for analyzing highly skewed

outcome variables. GLM handles the cost data skewness by predetermining the link function and the family type. More details about GLM are presented in section 1.4.1.2

Finally, we must account for competing risks such as death. Patients who die in the hospital have higher hospital costs than those who do not[23]. If one were to simply exclude these patients, then the approach may produce a biased result. Previous studies used survival models (the Kaplan-Meier (KM) survival curve and the Cox proportional hazard (Cox PH) regression model [24, 25] to censor patients who were discharged dead. However, a biased estimate can still be produced if the fundamental assumption of non-informative censoring is not satisfied[26]. De Angelis et al. suggested more sophisticated statistical approaches be considered such as the multistate model, which incorporates the discharge status as a competing risk when analyzing healthcare cost[27].

1.4 *Statistical methods for cost analysis*

The following statistical methods can be used to analyze the association between an exposure and its related hospital cost: ordinary least squares regression (CDI time-fixed), generalized linear regression (CDI time-fixed), Kaplan-Meier survival curve (CDI time-varying), and Cox PH regression (CDI time-varying). This section will discuss the advantages and disadvantages of each method.

1.4.1 Time-fixed method

1.4.1.1 *Ordinary least squares regression (OLS)*

Performing cost analysis using the ordinary least squares model, in which hospital costs are the dependent variable, is straightforward. The regression model can be written as:

$$C_i = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k + \epsilon \quad \text{, where } \epsilon \sim N(0.\sigma^2) \qquad \text{(Equation 1.4.1)}$$

Under the assumption of a random error that has a mean of zero and a constant variance, the advantage of this model is that the effects from the covariates are estimated directly on the additive scale. Direct insights about incremental costs between different groups can be made since the predicted mean cost is easy to interpret in this form. However, since hospital cost is positively skewed, this model approach typically fails to meet a basic assumption of the standard linear regression analysis (homoscedasticity).

There are several analytical methods available to handle the skewed nature of cost data: the OLS on log transformed cost and the generalized linear regression methods. The OLS on log transformed cost model is designed to remove the cost data skewness. The regression model can be written as:

$$\ln(C_i) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots \beta_k X_k + \epsilon \quad \text{, where } \epsilon \sim N(0.\sigma^2) \qquad \text{(Equation 1.4.2)}$$

Since it reduces the cost skewness by taking the logarithm of cost, it is more likely to satisfy the normal distribution assumption versus the OLS model without any transformation. As a result, predicting the mean cost from this approach is more reliable. On the other hand, if the regression coefficients are given on the multiplicative scale, then it is difficult to directly interpret the results on the original scale. For example, a regression coefficient of 2 is interpreted as the logarithm of the proportional change in the median cost with a one-unit change of the predictor variable. To provide estimates on the original scale of the response (cost in dollars), the regression coefficients can be converted via exponentiation, which is referred to as a geometric mean. However, this geometric mean can be biased if there is a large variability between the observations. For example, Group 1 presented in Table 1.1 shows a large variability between the

observations. The difference between the predicted mean (geometric mean of 73.986) from the log transformed data and the mean on the original scale (arithmetic mean of 90) is large. On the other hand, observations in Group 2 show less variability. The predicted mean on the log scale (geometric mean of 88.493) is very close to the arithmetic mean of 90. Therefore, the example presented in Table1.1 illustrates that the simple retransformation of the regression coefficients could produce biased estimates in the presence of a large variability between the observations. In such cases, a bias correction is required to provide more accurate predictions.

Table 1.1 Arithmetic mean versus geometric mean

	No transformation		**Log transformation**	
Observations	Group 1	Group 2	Group 1	Group 2
1	30	70	3.4012	4.2485
2	90	90	4.4998	4.4998
3	150	110	5.0106	4.7005
Arithmetic mean	**90**	**90**	--	--
Log (arithmetic mean)	--	--	4.4998	4.4998
Arithmetic mean of log	--	--	4.3039	4.4829
Geometric mean*	--	--	**73.986**	**88.493**
Standard deviation of logs	--	--	**0.8224**	**0.2265**

-- represents not applicable, * Geometric mean is calculated via exponentiation of the arithmetic mean of log

To correct any bias, the smearing factor (nonparametric bias-correction factor) can be applied when the retransformation via the exponentiation is performed [22]. The following formula is used to calculate the smearing factor:

$$Smearing\ factor = \frac{1}{N} \sum_{i=1}^{N} e^{(Zi - Z\imath)}, \qquad \text{(Equation 1.4.3)}$$

where Zi is the predicted value for each observation and $Z\imath$ is the predicted mean.

Table 1.2 Smearing factor adjustment

	Group 1			Group 2		
Observation	ln	$Zi - Z\imath$	$e^{(Zi-Z\imath)}$	ln	$Zi - Z\imath$	$e^{(Zi-Z\imath)}$
1	3.4012	-0.9027	0.4055	4.2485	-0.2344	0.7910
2	4.4998	0.1959	1.2164	4.4998	0.0169	1.0170
3	5.0106	0.7067	2.0274	4.7005	0.2176	1.2430
Z	**4.3039**	--	--	**4.4829**	--	--
Smearing factor*	--	--	**1.2164**	--	--	**1.0170**
Geometric mean	**73.986**	--	--	**88.493**		--
Geometric mean x smearing factor	**90**	--	--	**90**	--	--

*smearing factor is the average of $e^{(Zi-Z\imath)}$ in each group

When the error terms are heteroscedastic (variance of the error terms is not constant), it is recommended to calculate the specific smearing factor[28] for each group and then multiply it by the geometric mean. For example, a group specific smearing factor presented in Table 1.2 is multiplied by the geometric mean for each group producing '90', which is the same as the arithmetic mean of 90 shown in Table 1.1.

Overall, the OLS log transformed model with smearing factor adjustment can theoretically produce more accurate estimates versus the general OLS model with untransformed cost. However, the smearing factor adjustment should be applied to obtain unbiased results on the original scale.

1.4.1.2 *Generalized linear regression (GLM)*

The generalized linear regression (GLM) model is an extension of the simple OLS that can model non-normal response families. The GLM predetermines the following mean-variance relationship: Gaussian (for a constant variance), Poisson (for a variance that is proportional to the mean) or Gamma (for a variance that is proportional to the square of mean). However, in econometric studies of healthcare cost, the gamma family is widely used and the most recommended. The GLM also predetermines the link function for connecting the mean of the

response variable with the linear combination of predictors. The choices for the link function are the following: 1. identity link used for linear regression, 2. logit link used for logistic regression, or 3. log link used for count data (i.e.poisson distribution). From previous econometric studies, the log link function is appropriate for healthcare cost data as it incorporates the logarithmic scale of cost measurement directly into the model[29-31]. With this approach, the link function acts as a transformation for the response variable. As a result, unlike OLS, it does not suffer from any retransformation bias when converting parameter estimates on the original scale via exponentiation of estimates. However, this approach does not account for the time-dependency of the covariates unless it is implemented as a longitudinal model (e.g. using daily cost data). Even though the time of CDI acquisition varies among patients, the GLM treats CDI as a time-fixed variable, producing time-biased estimates. To account for this time dependency, more sophisticated methods can be used. These models will be discussed in the following sections.

1.4.2 Time-varying method

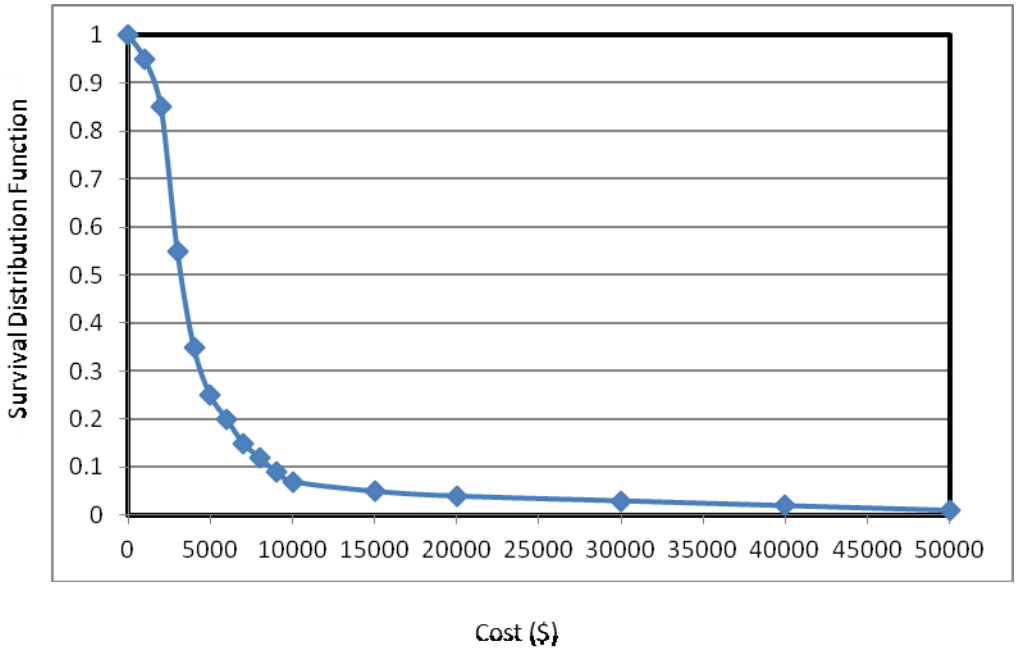
Several survival analysis methods have been used in econometric studies[24, 32]. The Kaplan-Meier survival curve and Cox proportional hazards regression are two such methods to analyze healthcare expenditures by treating hospital cost as "survival time". To illustrate this concept, a hypothetical dataset of 100 patients is presented in Table 1.3. In Figure 1.1, the Kaplan-Meier survival curve is constructed using the data from this table. From this figure, the probability of spending $3,000 or more to discharge is 0.55.

Table 1.3 Calculation for survival rate replacing survival time with hospital costs for hypothetical patients

Costs to discharge	Number of patients discharged	Number of patients remain	Survival Probability of discharge within interval	Probability of remain within interval	Survival Distribution Function
0	0	100	0/100	100/100=1	1
1000	5	95	5/100	95/100=0.95	1*0.95=0.95
2000	10	85	10/95	85/95=0.89	0.95*0.89=0.85
3000	30	55	30/85	55/85=0.65	0.85*0.65=0.55
4000	20	35	20/55	35/55=0.64	0.55*0.64=0.35
5000	10	25	10/35	25/35=0.71	0.35*0.71=0.25
6000	5	20	5/25	20/25=0.80	0.25*0.80=0.20
7000	5	15	5/20	15/20=0.75	0.20*0.75=0.15
8000	3	12	3/15	12/15=0.80	0.15*0.80=0.12
9000	3	9	3/12	9/12=0.75	0.12*0.75=0.09
10000	2	7	2/9	7/9=0.78	0.09*0.78=0.07
15000	2	5	2/7	5/7=0.71	0.07*0.71=0.05
20000	1	4	1/5	4/5=0.80	0.05*0.80=0.04
30000	1	3	1/4	3/4=0.75	0.04*0.75=0.03
40000	1	2	1/3	2/3=0.67	0.03*0.67=0.02
50000	1	1	1/2	1/2=0.50	0.02*0.50=0.01

Note : Total number of patients = 100

Figure 1.1 Hypothetical Kaplan-Meier survival curve on hospital cost



1.4.2.1 *Kaplan-Meier survival curve (KM survival curve)*

The KM survival curve is a non-parametric method that can account for the time-varying variables, and it can also be used in the presence of censoring. The KM estimator for hospital costs, which estimates the probability of cost to discharge, is given by:

$$KM\ \ hopsital\ costs\ \ = \ \ _{j\ c_{j<c}}\ \ a - \frac{d_j}{n_j} \qquad \text{(Equation 1.4.4)}$$

Where d_j is the number of events at hospital cost C_j and n_j is the number remaining at cost C_j. The KM survival curve can be used to account for the time-dependency of the disease acquisition as long as the data are arranged in the "counting process format". The counting process format is created by treating every admitted patient as a **control** subject before hospital-acquired CDI (HA-CDI) acquisition and a **case** after HA-CDI acquisition. A patient only becomes a case patient once HA-CDI is acquired. A patient who acquired HA-CDI would have two rows

representing the initial status as 'before HA-CDI (also referred to as pre HA-CDI)' and the following status as 'after HA-CDI (also referred to as post HA-CDI)'. Table 1.4 presents how the counting process format is applied to a hypothetical *patient 2* with HA-CDI. For *patient 2,* the hospital costs are split into the following categories: 1) cost before HA-CDI acquisition (control) and 2) cost after HA-CDI acquisition (case). However, patient 1 has only one row of data because his/her status has not changed (he/she did not acquire CDI) during his/her hospital stay.

Table 1.4 Example of arranging data in the counting process format

	HA-CDI patient	**Event (CDI status)**	**Control or Case**	Admission date	HA-CDI acquisition date	Discharge date	Initial hospital cost in the current status	Final hospital cost in the current status
Patient 1	No	**No**	**Control**	Jan 1, 2010	N/A	Jan 10, 2010	0	$10,000
Patient 2	Yes	**No**	**Control**	Jan 1, 2010	Jan 5, 2010	N/A	0	$30,000
Patient 2	Yes	**Yes**	**Case**	N/A	Jan 5, 2010	Jan 10, 2010	**$30,000**	$50,000

When applying the KM survival curve, an additional issue to consider is how to deal with those patients who die. If one were to simply censor the observation, this MAY violate the *non-informative censoring* assumption required for survival analysis. Non-informative censoring refers to where censoring is assumed to be random and independent of survival time (healthcare cost). However, this is difficult to be satisfied when analyzing healthcare cost because patients who are discharged dead tend to have higher overall hospital costs versus those who are discharged alive[33]. Hence the non-informative assumption is violated, producing an underestimated true hazard because the remaining uncensored patients were used to produce the estimates.

A disadvantage of this approach is that the KM survival curve cannot control for continuous confounding variables such as baseline mortality risk (a score that predicts a patient risk of dying in hospital). Nevertheless, the KM survival curve is still appealing because it can account for CDI as a time-varying covariate.

1.4.2.2 *Cox proportional hazard model (Cox PH model)*

The Cox PH model is atypical for cost analysis, yet it has been suggested in some healthcare econometric studies over the last two decades[24, 34]. There are several characteristics that make it appealing for analyzing cost data. Firstly, it makes no assumption about the distribution of the outcome variable. As discussed above, healthcare costs are typically highly skewed. Secondly, the Cox PH can be set up to account for time-varying covariates. Thirdly, the Cox PH model can include multiple continuous and/or categorical covariates. Therefore, a model can be developed to adjust for one or multiple confounding factors. Lastly, the Cox PH model can handle incomplete data. A patient is said to have 'incomplete data' when the observation is terminated by death. This is managed through censoring (which, as in the KM method, assumes *non-informative* censoring).

Like other methods, there are similar limitations from using the Cox PH model approach. When the PH assumption is not satisfied, it may produce biased estimates. In practice, one assesses whether covariates of interest meet the proportional hazard assumption over time, and if they do not, then the covariates must either be stratified or treated as time-varying covariates. In addition, since the model produces estimates on the multiplicative scale (i.e. a hazard ratio) the interpretation of the model may not be easily translated into a dollar amount. Furthermore, the Cox PH model also requires an assumption of non-informative censoring.

1.5 *Objective of study*

This study will estimate the hospital cost attributable to hospital-acquired CDI using the following statistical approaches: ordinary least square regression, generalized linear regression, Kaplan-Meier survival curve and Cox proportional hazard regression models.

Chapter 2 Methods

2.1 *Study design*

We conducted this retrospective observational study from the perspective of the hospital. The unit of analysis for the study was the "hospital inpatient encounter", which we will simply refer to as "encounter" (herein). Each encounter refers to an admission to the hospital. We used clinical and administrative databases to identify patients, exposures and outcomes.

2.2 *Study setting*

The study was conducted at The Ottawa Hospital (TOH). TOH is a tertiary-care teaching hospital comprised of three inpatient acute care campuses - the General, Civic and University of Ottawa Heart Institute. TOH is located in Ottawa, Ontario, Canada and serves a population of about 1.5 million. TOH also has outpatient facilities at the Riverside Campus, a regional cancer program, and a rehabilitation hospital. Across all campuses, TOH has approximately 1,050 beds.

2.3 *Patient population*

All inpatients treated at the General and Civic campuses were eligible if they were admitted after April 1, 2008 and discharged before March 31, 2011. We excluded admissions with hospital stay less than 72 hours, those with an age less than 15 years, and those with child-related care (i.e. admissions to obstetric service).

2.4 *Data sources*

We gathered data from two data sources: the Ottawa Hospital Data Warehouse (termed the OHDW) and the Ottawa Hospital Case Costing system. We used the OHDW to extract data describing encounters, CDI exposure, and relevant covariates; and, we used the Ontario Case Costing system to extract data describing hospital costs.

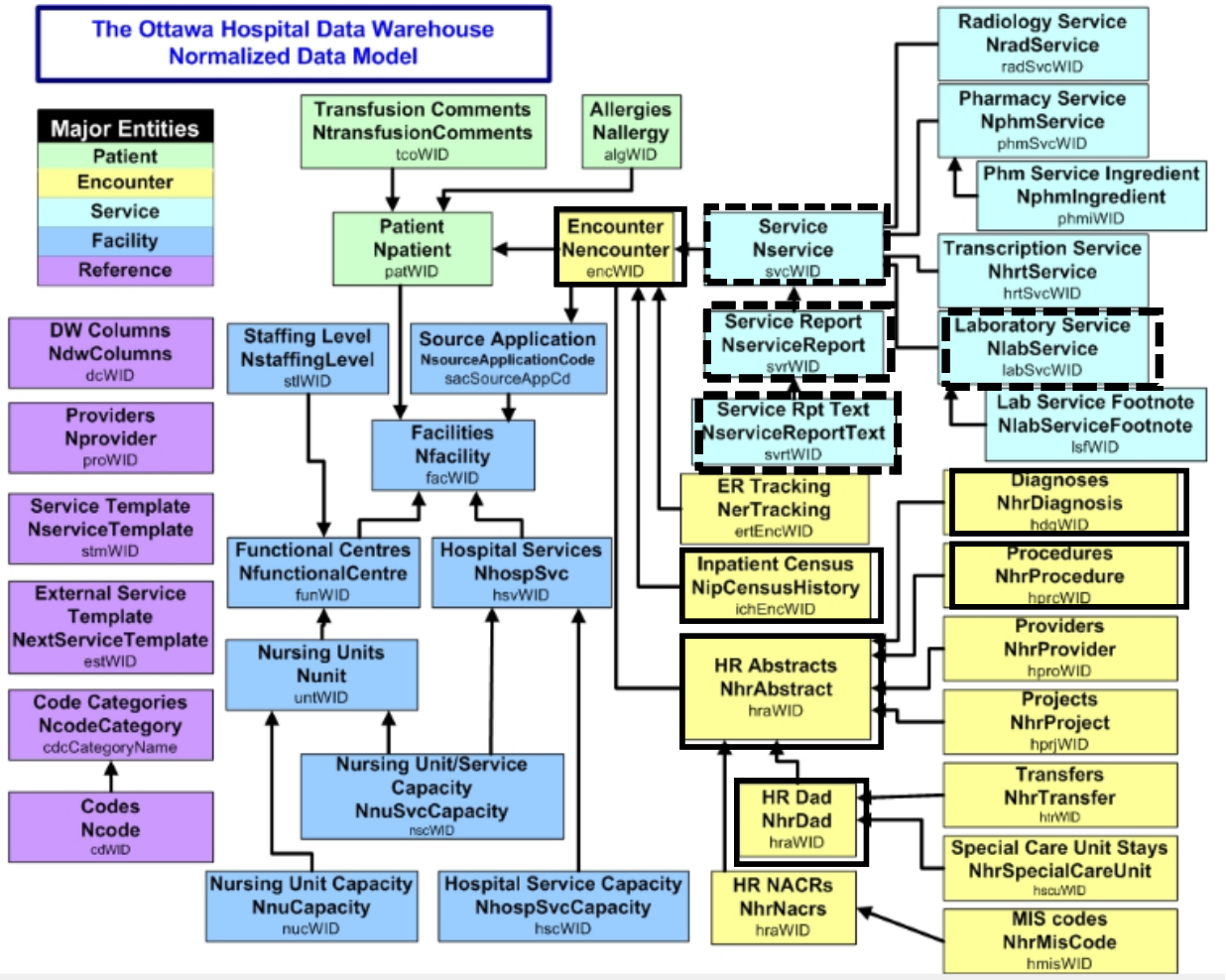
2.4.1 Ottawa Hospital Data Warehouse

The Ottawa Hospital Data Warehouse (OHDW) is TOH's information system and stores all of its enterprise data. It comprises data derived from clinical and administrative information systems. Three important data sources for this project are: the patient registration system (called SMS), the electronic health record system (called vOACIS), and the health records abstracting system (called WinRx).

The SMS has demographic information as well as data regarding dates and times of all patient movement (admissions, discharges or transfers). Clerical staff enters this information into the SMS. The vOACIS has information about laboratory tests, medications ordered, and diagnostic imaging. These data are entered directly into the system from departmental information systems. The WinRx has information on the hospital diagnosis and procedures for all inpatients. After a patient is discharged, a health record analyst reviews the chart and assigns the appropriate International Classification of Diseases, 10^{th} revision, Clinical Modification (ICD–10-CM) codes for the patient's diagnoses, and the Canadian Classification of Health Interventions (CCI) codes for the patient's procedures.

The OHDW stores data according to a data model consisting of five major entities: patient, encounter, service, facility and reference (Figure 2.1). Each entity is composed of a series of connected tables that has a unique warehouse identifier (WID) to relate the data across each of the entities. These unique WIDs are patient identifiers to identify each individual.

Figure 2.1 The Ottawa Hospital Data Warehouse Normalized Data Model



Note: Encounter Table and sub-tables are highlighted in solid line. Service Table and sub-tables are highlighted in dotted line

For this study, we obtained information from two main entities, the encounter and service tables, along with their sub-entities. Each row in the encounter table reflects a single encounter. The encounter table includes the following: patient demographics, the type of encounter, and details about the encounter (e.g. start/end dates). We used the following sub-entities of the encounter table: Health Record (HR) Abstracts table, Census History table, HR Discharge Abstract Database (DAD) table, Diagnoses table and Encounter Cost Detail table. Each record in the HR

Abstract table reflects single inpatient encounter, each record in the Census History table reflects a single transfer between beds, services, or physicians; each record in the HR DAD table reflects a single in-patient and overnight same-day surgery, and each record in the Diagnoses table reflects a single diagnosis classified using ICD codes. The diagnostic type (e.g. admission, most responsible diagnosis, and complication) is also recorded for each diagnosis. Each record in the encounter cost detail table reflects a single intermediate product received and includes its associated cost. Each record in the service table reflects a single service, and each service has a code. The service table includes the following types of data: lab test, radiology test, or pharmacy order. We used the following sub-entities of the service table: service report, service report text table and laboratory service table. The service report and service report text tables are used together and contain the text of all patient reports. Each record in the laboratory service table reflects a single test received. It includes the date and test results (if the test result is numeric). All the variables and the table from which they were derived are listed in Table 2.1.

Table 2.1 OHDW tables and variables used in the study

Encounter table			
Variable	***Sub-tables***	***Table description***	***Information obtained***
Age	*HR Abstract table*	Describes inpatient encounters with basic demographics and other >50 variables	Age at admission
Gender	*HR Abstract table*	Same as above	Female or male
Admitting services	*Census history table*	Contains information about patient's physical location.	
Acuity of admission	*HR DAD table*	Contains information on in-patient and overnight same-day surgery admissions	Elective or emergent
Charlson index	*Diagnoses table*	Contains all diagnostic information. Each row in this table represents one diagnosis. Diagnoses are	• Myocardial Infarction • Congestive Heart Failure • Peripheral Vascular Disease

		classified using International Classification of Diseases (ICD) codes. Diagnostic type (e.g. admission, most responsible diagnosis, complication) is also recorded for each diagnosis.	• Cerebrovascular Disease • Dementia Chronic Obstructive • Pulmonary Disease • Connective tissue / rheumatic disease • Peptic Ulcer Disease • Mild Liver Disease • Moderate or Severe Liver Disease • Diabetes without chronic complications • Diabetes with chronic complications • Hemiplegia or Paraplegia • Renal Disease • Primary Cancer (includes lymphoma and leukemia and excludes malignant neoplasm of skin) • Metastatic Cancer • HIV/AIDS
Costs	*Encounter Ccost Detail table*	Contains all cost information. Each row in this table represents one intermediate product and associated costs. Cost for the each intermediate product is broken down to variable direct, variable indirect, fixed direct and fixed indirect costs	Variable direct and fixed direct costs
Service table			
Variable	***Sub-tables***	***Table description***	***Information obtained***
Severity of acute disease	*Laboratory service tables*	Contains the specific details of laboratory tests received during encounters. Specifically, it includes the date and results of tests (if the test result is numeric).	• Serum albumin • Serum chloride • Arterial pH • PaCO2, and PaO2 • Bicarbonate • Total serum bilirubin • Blood urea nitrogen • Serum creatinine • Serum glucose • Serum sodium • Serum troponin I • Hematocrit • Total white blood cell count

CDI status	*Service report and service report text table*	The *service report* and *service report text* tables are used together and contain the text of all patient reports. Examples of reports found in these tables include discharge abstracts, pathology reports, and toxin assay results.	*Clostridium difficile* toxin assay result as 'positive' or 'negative'

2.4.2 Ottawa Hospital Case Costing system

We also used data derived from TOH's case costing system. These data are collected at TOH using methods established and approved by the Ontario Case Costing Initiative.

2.4.2.1 *Ontario Case Costing Initiative*

The Ontario Case Costing Initiative was launched with the goal of developing a standard approach for attributing costs to healthcare services, which in turn would allow better management decisions and more valid comparisons between institutions. Ontario hospitals were invited to join the Ontario Case Costing Initiative in 2005, and currently there are 49 hospitals participating in the program[35]. This initiative is considered vital by some as data derived from it will be used to guide the current Health Services Funding Reform activities in Ontario which will fundamentally alter how acute care hospitals are funded.

Further, the Ontario Case Costing Initiative data are a valuable resource for analysis and research initiatives. One example is the Werb et al. study that estimated the cost associated to treat hepatitis C infection among injection drug users in Canada[36]. Another example is the cost-effectiveness study of using transdermal nitroglycerin for preterm labour that was conducted

based on TOH data through the Ontario Case Costing Initiative [37]. There are many other examples, like those mentioned above, in the literature.

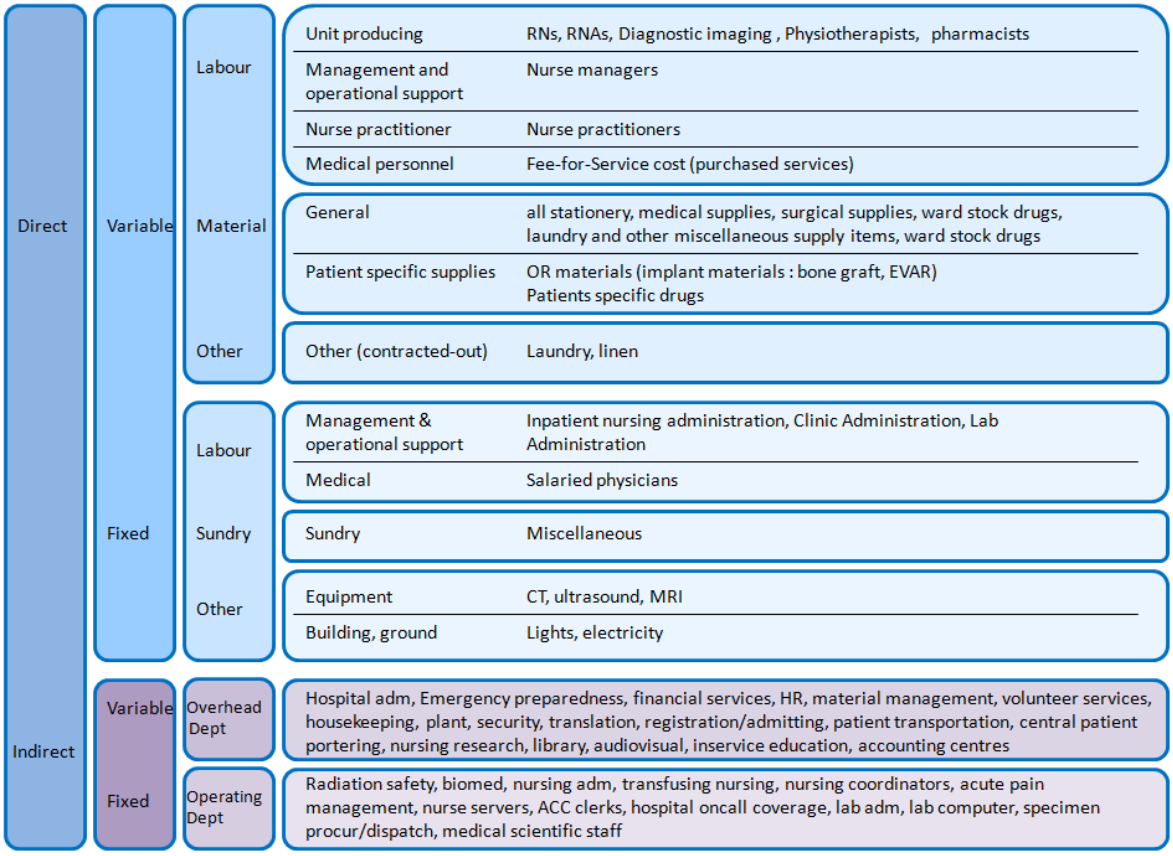
2.4.2.2 *Case Costing Methodology*

The goal of the case costing system is to collect patient-specific costs. Figure 2.2 is a conceptual diagram illustrating various cost components and their respective cost categories. We have included examples of intermediate products within each cost category. Each intermediate product is mapped onto these cost categories. For example, some labour and material intermediate products are considered direct variable costs (e.g. unit producing and medical fee-for-service labor and patient-specific supplies); while other labour and material intermediate products are considered direct fixed costs (e.g. management/operational support and other medical costs such as salaried physicians, sundries and buildings, equipment and grounds). Indirect costs are generated from overhead expenses (variable) and operating departments (fixed).

A detailed review of the case costing methodology is beyond the scope of this thesis. There are excellent descriptions publically available if the reader is interested[38]. In short, the case costing system links financial, clinical and patient activity information stored within the existing information system to define so-called 'intermediate products'. Examples of intermediate products are nursing time, medications, lab tests, surgical materials, and imaging exams. The cost for each intermediate product is calculated by capturing the patient specific quantity of direct resources (for example, labour and supplies) and indirect resources (for instance, overhead expenses such as administration support, housekeeping, and finance) from functional cost centers. To attribute indirect expenses at the patient level, it is necessary to allocate expenditures

from non-patient-care functional cost centers. Patient care functional centers are referred to as the absorbing functional centers (AFCs), because in case costing methods they 'receive' allocation of indirect costs generated from departments associated with the administrative and support (overhead functional) centers, which are in turn termed transient functional centers (TFCs). Indirect costs generated from TFCs are allocated to the AFCs. To allocate indirect costs, the costing methods use the Simultaneous Equation Allocation Method (SEAM) to determine relative units received by each AFCs from each TFCs. The total indirect costs allocated are further distributed to each patient according to that specific patient's utilization. Both direct and indirect costs are further categorized into variable and fixed components.

Figure 2.2 Cost components of Case costing



2.5 *Creation of the analytical dataset*

The OHDW and case costing data were used to construct the analytical data set. This process consisted of the following steps:

1) Study Cohort–identification of all in-patient encounters meeting the study inclusion criteria
2) Independent Variable–identification of *Clostridium difficile* positive tests
3) Dependent Variables–calculation of direct and total hospital costs
4) Covariates

2.5.1 Study Cohort

We used *the encounter table* to identify all inpatient encounters at the Civic and General campuses if they were admitted after April 1, 2008 and discharged before March 31, 2011 and their overall length of stay was less than 72 hours. We excluded patients who were younger than 15 years of age and who were admitted to the obstetrics service. For each encounter, we defined the patient's encounter identifier and the admission and discharge dates.

2.5.2 Independent variable

We used the *service table* to identify the *Clostridium difficile* toxin test results and classify them as positive or negative. All tests that were performed between April 1, 2008 and March 31, 2011 were selected. We retrieved the laboratory reports for these toxin assays by relating the *service report table* to the *service report text table.* We then used text-searching techniques available in SAS to identify tests whose toxin assay results were "positive", "negative", "detected", or "not detected". For each test, we defined the test date as the CDI acquisition date to categorize the CDI patient's hospitalization as either before CDI acquisition or after CDI acquisition.

Next, we linked the *Clostridium difficile* positive tests to the encounters to determine the hospital-acquired CDI status for all encounters as directed by the Ontario Provincial Infectious Disease Advisory Committee guideline[39]. We followed previously established logical rules to define each encounter based on whether or not the patient acquired the *Clostridium difficile* infection (CDI) on that encounter[21]. CDI cases were classified as hospital-acquired if the toxin assay was positive for *Clostridium difficile* more than 72 hours after admission. Patients who were discharged within the previous 8 weeks with CDI tests that were positive less than 72 hours

after admission were excluded. Those were excluded due to a lack of information (e.g. admission to other facilities) from the 8 weeks between the previous discharge and readmission, because we were not able to determine whether those cases were attributable to TOH or not. In addition, when there were multiple positive *Clostridium difficile* test results during the same admission, only the first *Clostridium difficile* test was considered for this study.

2.5.3 Study outcome

The primary outcome was direct costs and the secondary outcome was total hospital cost. We derived total cost using the direct costs rather than calculating it directly from the indirect costs because we noted that a very small number of observations appeared to be outliers due to extremely high indirect costs. This seemed to be occurring more commonly after 2010 when the ministry modified the methods of assigning indirect costs. We were worried that these few observations would skew our results and we were uncertain whether these costs were actually attributable to true increases in costs or rather due to errors in attributing overhead costs fairly. Because direct costs did not exhibit this extreme variation in any observations, we decided to focus our analysis on direct costs for our primary analysis. Furthermore, we decided to extrapolate total costs from the direct costs (rather than use the measured indirect costs) for our analysis at the patient level. While on an aggregate level, the use of indirect costs would likely be acceptable, when assessing individual cases, it is possible the accounting for errors would lead to spurious results. Therefore, as other investigators have done, we calculated indirect costs at an individual patient level using the overall costs for the entire study period. During the study period at TOH, the ratio of direct to indirect costs was 7:3. We therefore multiplied direct costs by a factor of 1.428 to derive total costs.

We used the *Nencouter Cost Detail* table to obtain cost data that determines each patient's adjusted direct costs and total cost. From this table, we extracted details about individual intermediate products such as date of utilization, functional center and cost information. We did not include any costs that were incurred before the admission date and after the discharge date. Table 2.2 presents the steps and formula deriving outcome variables. To obtain total direct costs for each intermediate product, we summed the variable and fixed direct costs to obtain the INTERMEDIATE PRODUCT total direct cost. Next, we added the total direct cost for all INTERMEDIATE PRODUCTs used on a calendar date to calculate the DAILY direct costs. Finally, we summed DAILY direct costs to define the total DIRECT costs for the encounter. We derived the total HOSPITAL costs by multiplying total direct costs by the factor of 1.428. Finally, we adjusted all costs to 2010 Canadian dollars by using the healthcare component from the Canadian consumer price index[40].

Table 2.2 Steps and formula deriving outcome variables

Calculated variables	Formula
INTERMEDIATE PRODUCT total direct cost	Intermediate fixed direct cost + Intermediate variable direct cost
DAILY direct costs	$\sum_{i=1}^{n} intermediate\ total\ direct\ cost$, n represents intermediate products received a given day
Total DIRECT costs	$\sum_{j=1}^{m} daily\ direct\ cost$, m represents hospital days
Total HOSPTITAL costs for each admission	Total DIRECT costs X 1.428

*Each cost component is from individual admission.

2.5.4 Covariates

Risk of in-hospital death (herein referred to as baseline mortality risk) is an important factor associated with both hospital costs [41] and CDI acquisition[21]. Therefore, it is important to account

for such a potential confounding effect when investigating the association between hospital-acquired CDI and hospital cost.

Baseline mortality risk for each admission is a score that predicts the likelihood of dying during a hospital stay and is calculated via a regression model that was previously validated at TOH[42]. This model utilizes the following covariates: age at admission, gender, admitting hospital service, acuity of admission, co-morbidity, and severity of acute disease. Co-morbidity is represented by a Charlson index, which is a weighted scoring system that accounts for a patient's chronic health conditions at the time of admission. Table 2.1 lists the 17 types of co-morbidity that were identified in this study. In addition, the severity of the acute disease was represented using the Labotatory-based Acute Physiology Score (LAPS), which is a continuous variable that was calculated by summarizing the results of 14 lab tests performed within the first 24 hours of admission.

We used the *abstract table* to obtain the age and gender for each encounter. Admitting hospital service type (e.g. medical, surgical, or intensive unit care) was determined by using the *census history table.* To determine the acuity of admission, we used the *HR DAD table* to categorize the status as 'elective' or 'emergent'. To identify patient co-morbidity, we used the *diagnoses table*. Diagnoses were classified according to the ICD-10-CM code for the 17 diagnoses listed in Table 2.1. The tests and results necessary for LAPS were obtained from the *service* and *laboratory service tables*.

2.6 *Statistical analysis*

We used SAS Version 9.2 (Cary, NC) for all statistical analyses performed in this study.

2.6.1 Exploratory analyses

2.6.1.1 *Cohort characteristics*

Basic descriptive statistics were generated for the study cohort stratified by CDI status. For continuous variables, the mean and standard deviation are presented for variables with a normal distribution (age) and medians and inter-quartile ranges (IQR) for variables with a skewed distribution (baseline mortality risk). For the categorical variables, we report frequency distributions.

2.6.1.2 *Cost outcomes*

We summarized cost data for the primary outcome (i.e. the inflation-adjusted direct costs) and the secondary outcome (i.e. the derived inflation adjusted total hospital costs) for the following sub-groups of patients: by gender, age group, campus, fiscal year, Charlson index and CDI status. For each sub-group, we reported the mean with standard deviation and the median with IQR.

2.6.2 Multivariable analysis

To account for confounding related to patient factors, we decided a priori to include baseline mortality risk in all models: OLS, GLM and Cox PH regression analysis. For each patient, we calculated their estimated baseline mortality risk (risk of dying in hospital) using an established risk scoring algorithm. The risk score is calculated using patient's age, gender, admitting services, acuity of admission, a Charlson index and severity of acute disease. We created a histogram from baseline mortality risk to visualize the distribution. To investigate the functional form of the association with baseline mortality risk, we plotted mean cost against mean baseline

mortality risk within 10 groups, defined based on percentiles. We decided to address the skewed distribution by log-transforming baseline mortality risk score. We further tested for non-linearity by including the square-log transformed baseline mortality risk in each model and retaining this term in the model if its p-value was less than 0.05.

2.6.2.1 *Time-fixed model*

We analyzed the effect of CDI on the direct costs using the following two methods in which CDI was treated as a time-fixed variable. In both models below, we adjusted for baseline mortality risk to account for any potential confounding related to patient factors.

2.6.2.1.1 Ordinary least squares (OLS) regression

We used the OLS regression method to investigate the effect of the hospital-acquired CDI on the direct costs. Diehr et al. [43] and others suggested that the OLS regression model is the simplest approach for the cost analysis [32]. We used log-transformed costs as outcome variable to account for costs with highly skewed distribution.

The OLS regression coefficient was used to determine the mean log-cost for patients with hospital-acquired CDI and those without CDI, while holding baseline mortality risk constant at its median. To determine the cost difference in dollars, we retransformed the mean of log cost for each group by exponentiation of the parameter estimates, which then provided the geometric mean of costs (Table 1.1). To eliminate the possibility of retransformation bias, we considered a smearing factor adjustment[22]. If the error terms are not log normal, a simple retransformation would likely produce biased estimates (retransformation bias). Table 1.1 presents an example of retransformation bias resulting from a simple retransformation.

We calculated a group specific smearing factor for those with HA-CDI (1.56509) and without HA-CDI (1.53991) and multiplied them by the geometric mean cost to generate the estimated direct costs for those with and without HA-CDI. This was done for the following reasons: 1) the smearing factor adjustment can be applied as a method to correct for retransformation bias and 2) Manning et al.[28] recommended using group specific smearing factors to account for any possible differences in variances between the groups.

2.6.2.1.2 Generalized linear regression model (GLM)

We utilized a GLM to analyze the effect of hospital-acquired CDI on the direct costs after adjusting for a potential confounder (baseline mortality risk). Blough et al.[29]and others recommended a GLM with gamma family and log link function to model healthcare cost data[31, 44]. Previous studies stated that a GLM is suitable for healthcare cost data because it allows one to fit models whose outcome variable is non-normally distributed (i.e. belonging to the exponential family). It is a more sophisticated method than the traditional OLS regression method for modeling an outcome variable with non-normal distribution and it avoids the smearing factor adjustment. It does so by modeling the outcome variable with a particular distribution in the exponential family by predetermining the link function. Therefore, the model produces the log of the mean rather than the mean of the log-transformed data. More details on GLM are already described in section 1.4.1.2. Therefore, we modeled the direct costs of patient hospital-acquired CDI status with the recommended family type (gamma) and link function (log) controlling for baseline mortality risk as a potential confounding factor. To obtain the estimated mean costs for the group with CDI and the group without CDI in dollars, we retransformed the regression coefficients via exponentiation by holding baseline mortality risk constant at its median. We assessed each model by plotting a scaled deviance residual against the fitted value.

We took four random samples with 500 (10% of the study population) observations to have a better visual presentation rather than having the data points for the entire cohort of 49,888.

2.6.2.2 *Time-varying model*

We used the Kaplan-Meier survival curve and the Cox Proportional Hazard regression model as methods to account for CDI as a time-varying variable.

2.6.2.2.1 Kaplan-Meier (KM) survival curve

We used the Kaplan-Meier survival curve to describe the costs to discharge while accounting for the time-varying nature of CDI[45]. This required the dataset to be rearranged according to the 'counting process format' of the cost data. In this way, the group with CDI received two data lines. The first line represented cost between the time of admission up to the time of CDI acquisition (herein referred as cost before CDI). The second line represented cost between the time of CDI acquisition up to the time of discharge (herein referred to as cost after CDI). To calculate the cost attributable to CDI after accounting for CDI as a time-varying variable, the cost before CDI was attributed to the group without CDI and the cost after CDI was attributed to the group with CDI. We used the KM curve macro provided by the Mayo Clinic [46] to generate the survival curve for patients with and without CDI. The KM survival curve provided the median costs to discharge where the survival rate was at 0.5.To estimate the mean cost, we calculated the area under the estimated survival curve (AUC) using the Riemann sum method[47]. The formula for Riemann sum is given below;

Mean costs $=\sum_1^z(\text{cost}i_i - \text{cost}i_{i-1})$ X (survival rate_i+survival $\text{rate}_{i-1)}$/2, (Equation 2.6.1)

where i represents a distinct cost value presented in our study dataset. In order to calculate AUC, we used the values from the dataset produced by surviveTD macro where it contained the survival rate at each cost for those with CDI and without CDI. We subtracted the costs between two sequential points and multiplied the difference by the average of the survival rate to calculate the area for each rectangle. To control for baseline mortality risk in the KM curve, we stratified the cohort by quartile of baseline mortality risk. Therefore, we were able to estimate the mean costs for each stratum by calculating AUC using the Riemann sum estimation method. We took 1000 bootstrap samples from the original cohort to report 2.5^{th} and 97.5^{th} percentiles of the estimated mean. To compare the results for the method that treats CDI as a time-varying versus time-fixed variable, we repeated the above steps on the dataset where patients with CDI received only one line treating CDI as a time-fixed variable instead of two lines (i.e. before CDI and after CDI). We coded patients with CDI as 1 and patients without CDI as 0.

2.6.2.2.2 Cox Proportional Hazard regression model

We utilized the Cox proportional hazard (Cox PH) regression model to simultaneously account for the time-varying nature of CDI and baseline risk of mortality. We used the dataset formatted as 'counting process' that was described in the KM curve section above. Hence, the cost before CDI was attributed to the group without CDI, and the cost after CDI was attributed to the group with CDI. Pre CDI was coded as 0 and post CDI was coded as 1. We also utilized the Cox PH regression model treating CDI as a time-fixed variable for the purpose of a comparison between treating CDI as time-varying versus time-fixed variable. In order to assess the model, we calculated the c-statistic using the survcstd macro provided by the Mayo Clinic[46].

We tested the PH assumption with a log-negative-log plot creating strata of the quartile baseline risk of mortality. Furthermore, we plotted the Schoenfeld residual for log transformed baseline risk of mortality and square of log transformed baseline risk of mortality for testing the PH assumption.

2.6.3 Sensitivity analysis

We performed a stratified Cox PH regression analysis to accommodate any non-proportional hazards. We used baseline mortality risk score to stratify each group (CDI absent and CDI present) by quartiles.

2.6.4 Model assessment

Plots of residuals against fitted values were obtained from the OLS and GLM regression models. Since the proportion of patients without CDI was so much larger than that of patients with CDI, a random sample was used to generate the residual plots to better observe the pattern. Otherwise, it would be hard to draw any conclusions from the residual plots containing the entire cohort (n=49,888). In our study, we took several random samples of 360 patients without CDI (same number as CDI patients), and plots of residuals against fitted values were obtained from the OLS and GLM regression models.

For the Cox PH regression models, the *c*-statistic values were observed for both the time-fixed time-varying models.

Chapter 3 Results

3.1 *Cohort characteristics*

Table 3.1 summarizes the cohort characteristics of the encounters in this study. Between April 1, 2008 and March 31, 2011, there were 49,888 eligible encounters whose length of stay was more than 3 days (72 hours) at the Civic and General campuses of The Ottawa Hospital. The mean (SD) age was 64.6 ± 17.8 years old; 51.2% were female (at both campuses) and 43.4% were admitted to the Civic campus. In general, 30.2% had a Charlson index of 0, 27.3% had a Charlson index of 1-2, 16.2% had a Charlson index of 3-4, and 26.2% had a Charlson index of 5+. The most common chronic condition was primary cancer (27.6%), followed by diabetes without complications (17.8%), and diabetes with complications (15.5%). The median (IQR) predicted baseline mortality was 0.04 (0.01 - 0.14), and the median (IQR) length of stay for the cohort was 8.00 (5.00-15.00) days. Overall, 6.9% of all admitted patients died.

Table 3.1 Characteristics of the study cohort

Characteristics	N = 49,888
Age (years), mean (SD*)	64.6 ± 17.8
Female, n* (%)	25,525 (51.2%)
Campus, n (%)	
Civic	21,661 (43.4%)
General	28,227 (56.6%)
Charlson index, n (%)	
0	15,087 (30.2%)
1 – 2	13,641 (27.3%)
3 – 4	8,067 (16.2%)
5+	13,093 (26.2%)
Most common chronic conditions, n (%)	
Myocardial Infarction	2,525 (5.1%)
Congestive Heart Failure	5,285 (10.6%)
Peripheral Vascular Disease	3,345 (6.7%)

Cerebrovascular Disease	4,738	(9.5%)
Dementia	2,620	(5.3%)
Chronic Obstructive Pulmonary Disease	5,531	(11.1%)
Connective tissue / rheumatic disease	614	(1.2%)
Peptic Ulcer Disease	1,014	(2.0%)
Mild Liver Disease	1,291	(2.6%)
Diabetes without complications	8,865	(17.8%)
Diabetes with complications	7,740	(15.5%)
Hemiplegia or Paraplegia	1,504	(3.0%)
Renal Disease	4,111	(8.2%)
Primary Cancer	13,763	(27.6%)
Moderate or Severe Liver Disease	752	(1.5%)
Metastatic Cancer	6,013	(12.1%)
HIV/AIDS	266	(0.5%)
Predicted baseline mortality risk, median (IQR*)	0.04	(0.01 - 0.14)
Length of stay in hospital (days), median (IQR)	8.00	(5.00 - 15.00)
In-hospital deaths, n (%)	3,454	(6.9%)

SD = standard deviation, n = number, IRQ = interquartile range

Table 3.2 provides a description of the cohort characteristics by hospital-acquired CDI status. From the study cohort, 360 encounters were identified with CDI, resulting in an overall infection risk of 0.73%. In addition, the campus and gender distributions were similar for admissions with and without CDI. The median number of days from admission to CDI detection was 13 (IQR of 7-29 days). When compared to non-CDI patients, CDI patients were older (mean age in years - 71.3±16.1 versus 64.5±17.8), had more chronic illnesses (Charlson score 3 or more - 54.5% versus 42.3%), and had a higher baseline mortality risk (median (IQR) baseline mortality risk - 0.12 versus 0.04). Although the prevalence of each chronic condition was similar for patients with and without CDI, CDI patients had a much higher risk of congestive heart failure, dementia, mild liver disease and diabetes – all of which are very important factors associated with death (Table 3.2). In addition, the median length of stay (LOS) was significantly higher for patients with CDI (median LOS - 36 days, IQR of 20-62 days) than patients without CDI (median LOS -

8 days, IQR of 5-15). Finally, the in-hospital mortality rate was approximately 4 times higher for patients with CDI than those without CDI (24.2% versus 6.8%).

Table 3.2 Characteristics of the study cohort by hospital-acquired CDI status

Characteristics	Hospital-acquired CDI	
	No	**Yes**
	N = 49,528	N = 360
Age (years), mean (SD)	64.5 ± 17.8	71.3 ± 16.1
Female, n (%)	25,345 (51.2%)	180 (50.0%)
Campus, n (%)		
Civic	21,477 (43.4%)	184 (51.1%)
General	28,051 (56.6%)	176 (48.9%)
Charlson index, n (%)		
0	15,016 (30.3%)	71 (19.7%)
1 - 2	13,548 (27.4%)	93 (25.8%)
3 - 4	7,989 (16.1%)	78 (21.7%)
5+	12,975 (26.2%)	118 (32.8%)
Most common chronic conditions, n (%)		
Myocardial Infarction	2,507 (5.1%)	18 (5.0%)
Congestive Heart Failure	5,218 (10.5%)	67 (18.6%)
Peripheral Vascular Disease	3,310 (6.7%)	35 (9.7%)
Cerebrovascular Disease	4,702 (9.5%)	36 (10.0%)
Dementia	2,585 (5.2%)	35 (9.7%)
Chronic Obstructive Pulmonary Disease	5,476 (11.1%)	55 (15.3%)
Connective tissue / rheumatic disease	609 (1.2%)	5 (1.4%)
Peptic Ulcer Disease	1,005 (2.0%)	9 (2.5%)
Mild Liver Disease	1,273 (2.6%)	18 (5.0%)
Diabetes without complications	8,804 (17.8%)	61 (16.9%)
Diabetes with complications	7,657 (15.5%)	83 (23.1%)
Hemiplegia or Paraplegia	1,488 (3.0%)	16 (4.4%)
Renal Disease	4,065 (8.2%)	46 (12.8%)
Primary Cancer	13,660 (27.6%)	103 (28.6%)
Moderate or Severe Liver Disease	740 (1.5%)	12 (3.3%)
Metastatic Cancer	5,984 (12.1%)	29 (8.1%)
HIV/AIDS	266 (0.5%)	0 (0.0%)
Predicted baseline mortality risk, median (IQR)	0.04 (0.01-0.14)	0.12 (0.04-0.27)
Length of stay in hospital (days), median (IQR)	8.00 (5.00-15.00)	36.00 (20.00-62.00)
CDI detection from admission (days), median (IQR)	N/A	13 (7-26)
In-hospital deaths, n (%)	3,367 (6.8%)	87 (24.2%)

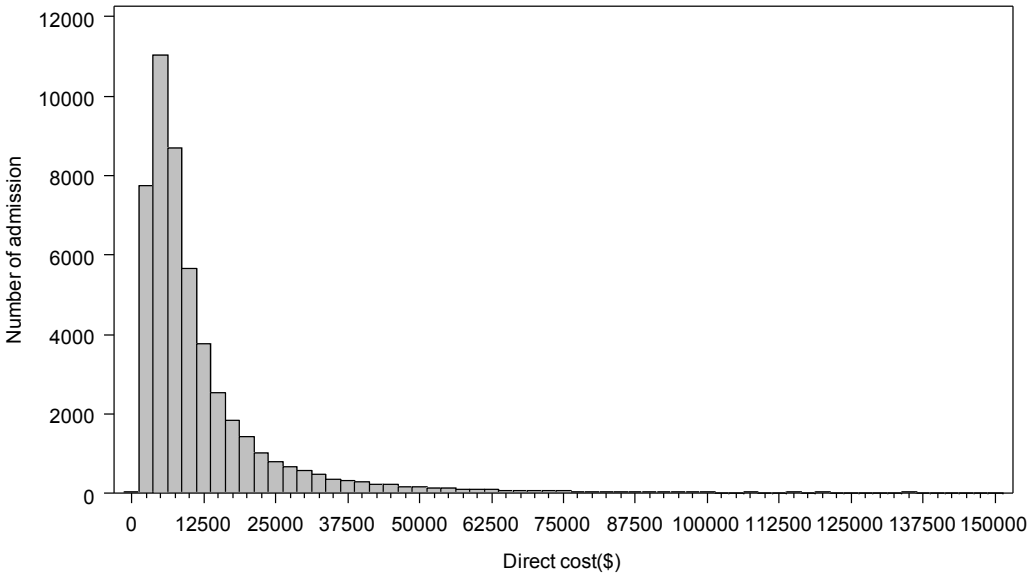
SD = standard deviation, n = number, IRQ = interquartile range

3.2 *Crude costs analyses for the main cohort*

Figure 3.1 displays the distribution of the direct costs for the entire study cohort. The costs were heavily skewed to the right with a mean of $13,580 and a standard deviation of $20,214. The median direct costs were $7,871 with interquartile range of $4,765 and $14,171.

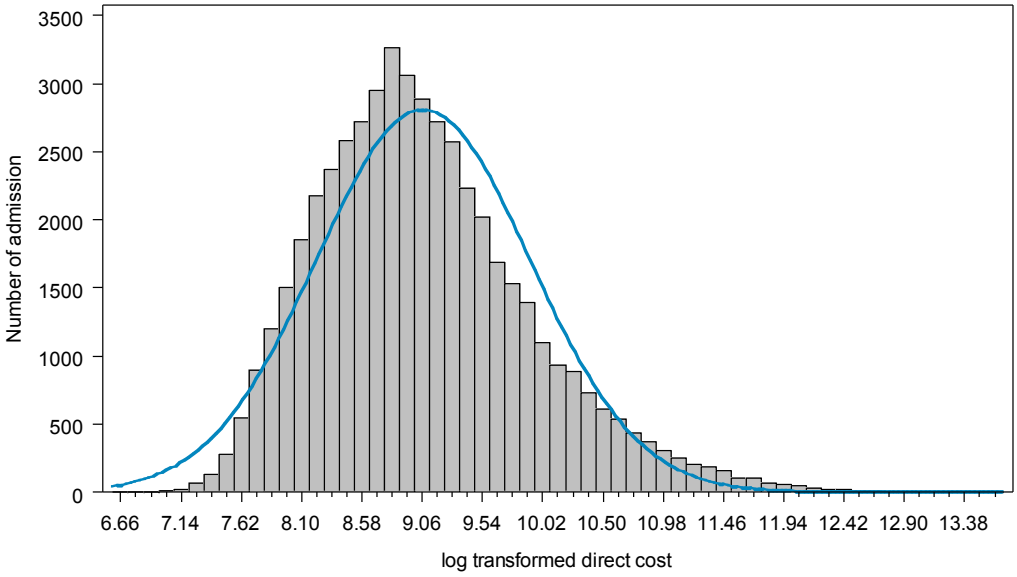
The histogram of the log-transformed direct costs in Figure 3.2 is less skewed than the histogram of untransformed cost in Figure 3.1. The untransformed and log transformed skewness coefficients for direct costs were 7.6 and 0.7, respectively. This provides evidence that the log transformation of the direct costs improved normality. This interpretation is based on the larger value of the skewness coefficients, implying an increased in skewness. Thus, the log-transformation has resulted in a considerable improvement in the cost distribution.

Figure 3.1Histogram of the direct costs



*There are 179 admissions with cost that exceeded $150,000 not illustrated in this figure.

Figure 3.2 Histogram of log transformed direct costs



3.2.1 Direct and total hospital costs by patient characteristics

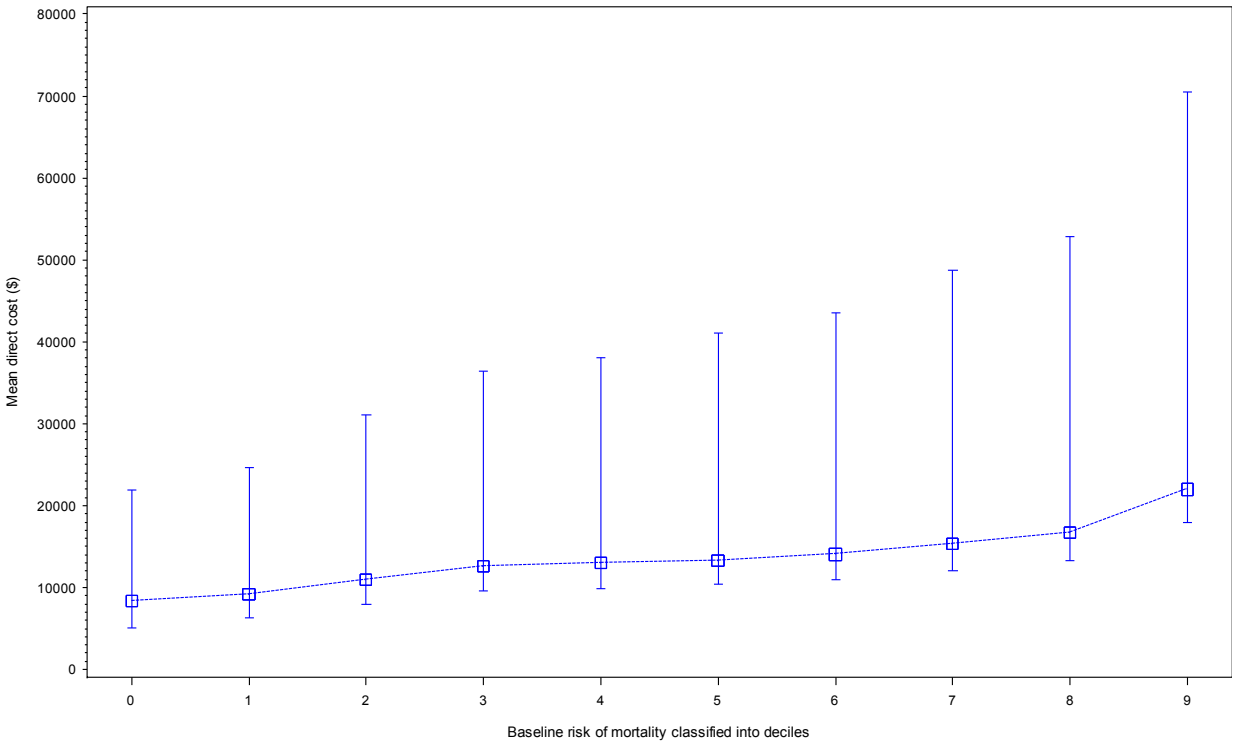
Table 3.3 displays the crude direct and estimated total (direct costs multiplied by a factor of 1.482) hospital costs stratified by patient characteristics. The direct and total costs increased with age even though the differences between the age strata were small. For instance, the mean difference between the direct costs of the youngest group of patients (age 15-19 years) and the oldest group of patients (age 60 years and above) was about $1,000. Also, patients admitted to the Civic campus had a slightly higher cost versus those admitted to the General campus. When patients were stratified by co-morbidity as indicated by the Charlson index score, it was apparent that costs increased as the co-morbidity increased (Table 3.3 and Figure 3.3). For example, the mean (±SD) direct costs of patients who had a Charlson index score of 0 was $10,864 ± $17,181, but the mean direct costs for patients who had a Charlson index of 5 or higher was $15,019 ± $20,011. In addition, the cost also increased with the baseline mortality risk. For patients in the 1st quartile of baseline death risk, the mean direct costs was $9,109± $13,867; while for patients in the 4th quartile, the mean direct costs was $18,551 ± $25,747. Figure 3.3 displays the mean direct costs stratified by deciles of baseline mortality risk with 10th to 90th percentile range. As can be seen, the direct costs increased as baseline mortality risk increased. Finally, the costs increased over the study period.

Table 3.3 Direct and total hospital costs (in thousands) by patient characteristics

		Direct costs ($, in thousands)		Total costs ($, in thousands)*	
Variable	**N**	**Mean ± SD**	**Median (IQR)**	**Mean ± SD**	**Median (IQR)**
Age					
15 - 19	466	12.8 ± 20.3	6.8 (4.0 - 12.3)	18.3 ± 28.9	9.7 (5.8 - 17.6)
20 - 39	4,472	12.4 ± 24.0	6.6 (4.0 - 11.9)	17.7 ± 34.2	9.4 (5.5 - 17.0)
40 - 59	12,972	13.4 ± 22.1	7.5 (4.5 - 13.1)	19.2 ± 31.6	10.7 (6.5 - 18.7)
60 above	31,978	13.8 ± 18.8	8.3 (5.0 - 15.2)	19.7 ± 26.8	11.8 (7.2 - 21.4)
Gender					
F	25,525	13.0 ± 18.6	7.8 (4.8 - 13.6)	18.5 ± 26.5	11.1 (6.9 - 19.4)
M	24,363	14.2 ± 21.8	8.0 (4.7 - 14.8)	20.3 ± 31.1	11.4 (6.7 - 21.2)
Campus					
Civic	21,661	14.2 ± 20.6	8.2 (4.8 - 15.2)	20.3 ± 29.4	11.6 (6.9 - 21.7)
General	28,227	13.1 ± 19.9	7.7 (4.7 - 13.4)	18.7 ± 28.4	11.1 (6.8 - 19.1)
Charlson Index					
0	15,087	10.9 ± 17.2	6.9 (4.4 - 10.9)	15.5 ± 24.5	9.8 (6.3 - 15.6)
1 - 2	13,641	14.1 ± 21.7	8.1 (4.8 - 14.7)	20.1 ± 31.0	11.6 (6.8 - 20.9)
3 - 4	8,067	15.5 ± 22.5	8.8 (4.9 - 16.6)	22.2 ± 32.1	12.5 (7.0 - 23.7)
5+	13,093	15.0 ± 20.0	9.0 (5.2 - 16.6)	21.5 ± 28.6	12.8 (7.4 - 23.7)
Baseline mortality risk					
1st quartile	12,479	9.1 ± 13.9	6.8 (4.5 - 9.8)	13.0 ± 19.9	9.7 (6.5 - 14.0)
2nd quartile	12,465	12.6 ± 17.9	7.8 (4.6 - 13.5)	18.0 ± 25.5	11.1 (6.6 - 19.3)
3rd quartile	12,472	14.1 ± 20.3	8.1 (4.6 - 15.3)	20.1 ± 29.1	11.6 (6.5 - 21.9)
4th quartile	12,472	18.6 ± 25.7	10.4 (5.6 - 20.6)	26.5 ± 36.8	14.9 (7.9 - 29.5)
Study year					
FY2008	15,988	12.6 ± 17.0	7.6 (4.6 - 13.3)	17.9 ± 24.3	10.9 (6.1 - 19.0)
FY2009	16,981	13.9 ± 21.9	7.9 (4.7 - 14.2)	19.8 ± 31.3	11.3 (6.7 - 20.3)
FY2010	16,919	14.3 ± 21.2	8.1 (4.9 - 14.9)	20.3 ± 30.3	11.5 (7.0 - 21.3)

* Total costs refers to total hospital costs, SD =standard deviation, n = number, IQR = interquartile range
Costs were adjusted to 2010 Canadian dollars by using the healthcare component from the Canadian consumer price index and were rounded to thousands with one decimal place.

Figure 3.3 Mean direct costs stratified by the deciles of the baseline risk mortality with 10th and 90th percentile range



3.2.2 Comparison of crude cost analyses by hospital-acquired CDI status

Table 3.4 compares the costs for encounters with and without CDI status. The results show that the direct and total hospital costs were significantly higher for patients with CDI. For patients without CDI, the mean direct cost was $13,290 and the median direct cost was $7,816. For patients with CDI, the mean direct cost was $53,582 and the median direct cost was $32,954.The mean total hospital costs difference between patients with and without CDI was about $40,000, which was about four-fold increase for those with CDI. Similarly, the median cost difference between the two groups was about $25,000.

Table 3.4 Crude cost analyses by hospital-acquired *C.difficlie* infection status

Cost type	Hospital-acquired CDI	
	No	**Yes**
	N = 49,528	N = 360
Direct costs ($)		
Mean ± SD	13,290 ± 19,181	53,582 ± 66,388
Median (IQR)	7,816 (4,735-13,992)	32,954 (17,068 - 63,735)
Total hospital costs ($)		
Mean ± SD	18,984 ± 27,400	76,542 ± 94,835
Median (IQR)	11,165 (6,764-19,987)	47,075 (24,382 - 91,045)

SD = standard deviation, N = number, IQR = interquartile range

3.3 *Result from the multivariable regression analysis*

3.3.1 Functional forms of covariate

To account for patient-related confounding factors, we included baseline mortality risk in the following models: OLS, GLM and Cox PH regression analysis. The baseline mortality risk is calculated using the following patient characteristics: age, gender, admitting services, acuity of admission, a Charlson index and the severity of acute disease. Figure 3.3 illustrates the relation between the log transformed baseline mortality risk score and the outcome (direct costs) variable. As there was some evidence of non-linearity, we included the square of log-transformed baseline mortality risk in each model if this term was statistically significant if it had a p-value less than 0.05.

3.3.2 Ordinary least square and generalized regression models

We utilized the ordinary least square regression (OLS) and the generalized linear regression (GLM) models to estimate the incremental impact of CDI on direct costs, while controlling for the effect of baseline mortality risk. Table 3.5 presents parameter estimates for variables used for

the OLS and GLM regression models. Both models showed significant effects for CDI status confirming that CDI is an important determinant of hospital cost. Furthermore, both linear and quadratic terms for baseline mortality risk were significant. Therefore, they remained in both models. Table 3.6 presents a summary of the mean estimates of total direct costs for admissions with CDI and without CDI that were derived from the OLS and GLM regression models. Based on the OLS model, the estimated mean direct costs per CDI admission was $47,505 (95% CI$43,628 to $51,728). The estimated mean difference between admission with CDI and without CDI was $34,419 (95% CI, $30,648 to $38,841). The increase is about 263% for an "average" admission, i.e., an admission with baseline mortality risk fixed at the median.

GLM is a more sophisticated method than the OLS model in predicting healthcare expenditures because the estimated mean (i.e. cost) does not require any further smearing adjustment to correct retransformation bias as OLS. Table 3.6 shows that the estimated mean cost obtained directly from the GLM model was very similar to the OLS model. The estimated mean direct costs for patients with CDI from the GLM was $48,025 (95% CI, $44,006 to $52,411). The estimated mean difference of direct costs between an admission with CDI and one without CDI was about $35,138 (95% CI, $31,238 to $39,426), an increase of about 272% for an "average" admission, i.e., an admission with a median baseline risk of mortality.

Table 3.5 Parameter estimates from OLS and GLM regression models

Variable		Estimated Regression Coefficient (p-value)	
Model type		OLS	GLM
Intercept		9.0497 (<0.0001)	9.4632 (<0.0001)
CDI status	CDI No	Reference	Reference
	CDI Yes	1.2730	1.3163 (<0.0001)
Log baseline risk mortality		0.1085 (<0.0001)	0.1432 (<0.0001)
Square of log baseline risk mortality		0.0114 (<0.0001)	0.0111 (<0.0001)

Table 3.6 Estimated mean direct costs from OLS and GLM regression models

		OLS (95% CI)	GLM (95% CI)
Cost ($)	CDI Yes	47,505 (43,628-51,728)	48,025 (44,006-52,411)
	CDI No	13,086 (12,980-13,194)	12,887 (12,768-12,985)
Mean cost difference ($)*		34,419 (30,648-38,841)	35,138 (31,238-39,426)
CDI Increase in costs (%)**		263	272

P-value <0.0001. Mean estimates for patients with the median baseline risk of mortality. *mean cost difference = total direct costs of patients with CDI – total direct costs of patients without CDI. **CDI increase in costs (%) = costs for CDI No / CDI attributable costs * 100

3.3.3 Kaplan-Meier survival curve

Figures 3.4-3.6 display the Kaplan-Meier survival curves for admissions with CDI versus admissions without CDI. These curves differ from the standard Kaplan-Meier survival curve in that direct cost is plotted on the X-axis, instead of time, as is found in a typical survival model. Figure 3.4 displays the KM cost curve with CDI as a time-fixed variable. Based on the Kaplan-Meier survival curves, the median cost to discharge was $32,954 for patients with CDI versus $7,816 for patients without CDI, which translates to a median incremental effect of $25,138 (Incremental effect $= CDI present $ - CDI absent $). To provide the mean estimates, the area under the curve was calculated. As shown in Table 3.7, the estimated mean cost to discharge was

$52,471 for patients with CDI and $13,282 for patients without CDI yielding an incremental effect estimate of $39,189.

Table 3.7 Summary of the direct costs estimates using Kaplan-Meier survival curves, treating CDI as time-fixed versus time-varying variable

CDI		CDI present	CDI absent	Incremental effect of CDI
Time-fixed	Median($)	32,954	7,816	25,138
	Mean($)	52,471	13,282	39,189
Time-varying	Median($)	11,520	7,846	3,674
	Mean($)	19,670 (13,930, 24,399)*	13,370 (13,199, 13,533)*	6,300

*95% bootstrap confidence interval constructed from 1000 samples

Figure 3.5 displays the KM cost curve with CDI as a time-varying variable. Compared with Figure 3.4, this curve shows that the incremental effect of CDI is substantially less when properly accounting for CDI as a time-varying variable. The median cost to discharge was $11,520 for patients with CDI versus $7,846 for patients without CDI which yields an incremental effect estimate of $3,674. In addition, the estimated mean cost to discharge was $19,670 for patients with CDI and $13,370 for patients without CDI: an incremental effect of $6,300 (Table 3.7).

The preceding analyses did not control for baseline mortality risk, which has been shown to be strongly associated with both the risk of CDI and hospital cost. In order to assess the impact of CDI on direct costs, while controlling for the effect of baseline mortality risk, we conducted a stratified analysis, using quartiles of baseline risk (Figure 3.6). The mean estimates for each stratum are displayed in Table 3.8. In general, the difference in mean estimated direct costs increased as the risk of mortality increased, with the exception that patients with CDI in the 2nd quartile showed higher cost than the patients in the 3rd quartile (Table 3.8). Patients with CDI in

the 2nd quartile of the baseline risk of mortality on average were observed to have direct costs equaling $22,405 whereas those in the 3rdquartile had direct costs equaling $21,345 (Table 3.8). This may be due to more variability of the cost among patients with CDI in the 2nd quartile versus the cost for the other strata. 95% bootstrap confidence interval estimates for patients with CDI in the 2nd quartile were wide (between $1,906 and $34,660) suggesting higher variability in costs among the admission within the 2nd quartile (Table 3.8). Overall, the direct costs attributable to CDI (Incremental effect $= CDI present $ - CDI absent $) increased regardless of stratum and the estimated incremental cost increase was similar across the four strata ranging between $6,562 and $9,969 after controlling the time-varying nature of CDI.

Table 3.8 Estimated mean direct costs estimates from Kaplan-Meier survival curve stratified by the baseline risk of mortality accounting for time-varying nature of CDI

Baseline mortality risk quartile	CDI present, $ (95% CI)*	CDI absent, $ (95% CI)*	Incremental effect, $
1st quartile	15,572 (10,560, 20,385)	9,010 (8,804-9,231)	6,562
2nd quartile	22,405 (1,906, 34,660)	12,436 (12,120, 12,731)	9,969
3rd quartile	21,348 (15,143, 27,920)	13,823 (13,496, 14,167)	7,525
4th quartile	25,432 (20,183, 32,897)	18,182 (17,709, 18,568)	7,270

*95% bootstrap confidence interval constructed from 1000 samples

Figure 3.4 Kaplan-Meier survival curves with CDI as a time-fixed variable

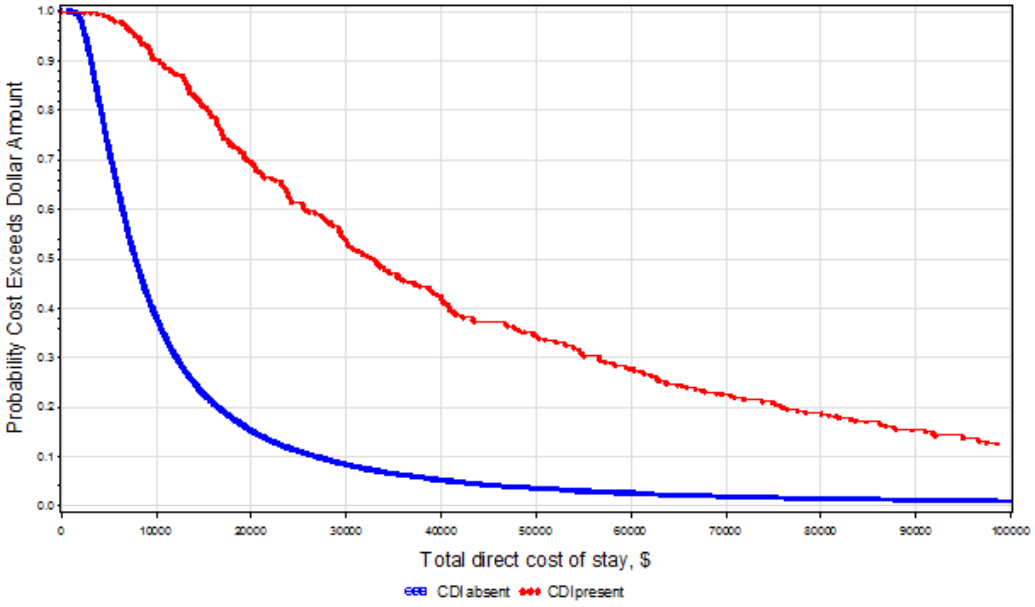


Figure 3.5 Kaplan-Meier survival curves with CDI as a time-varying variable

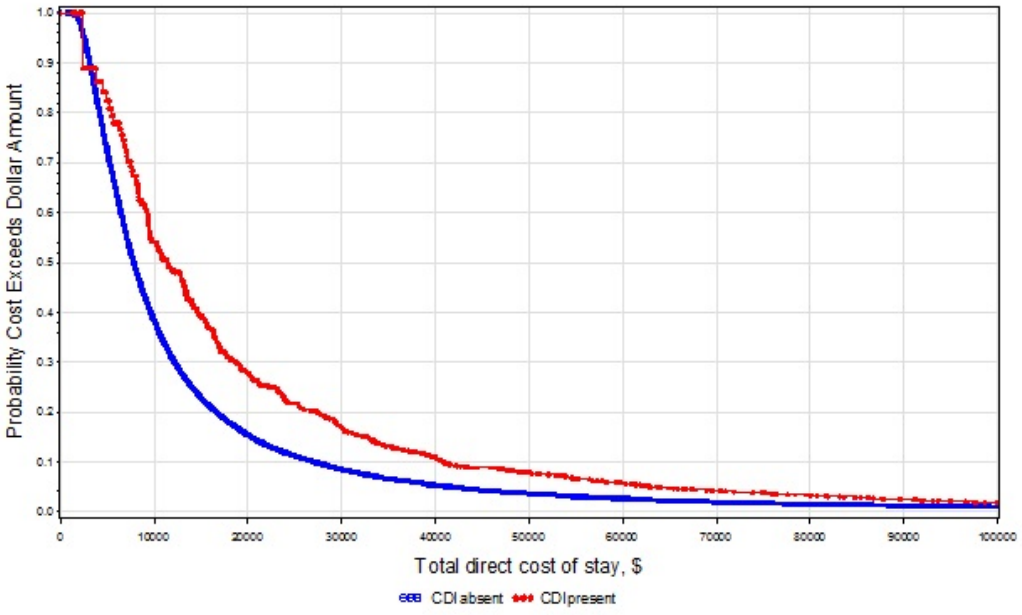
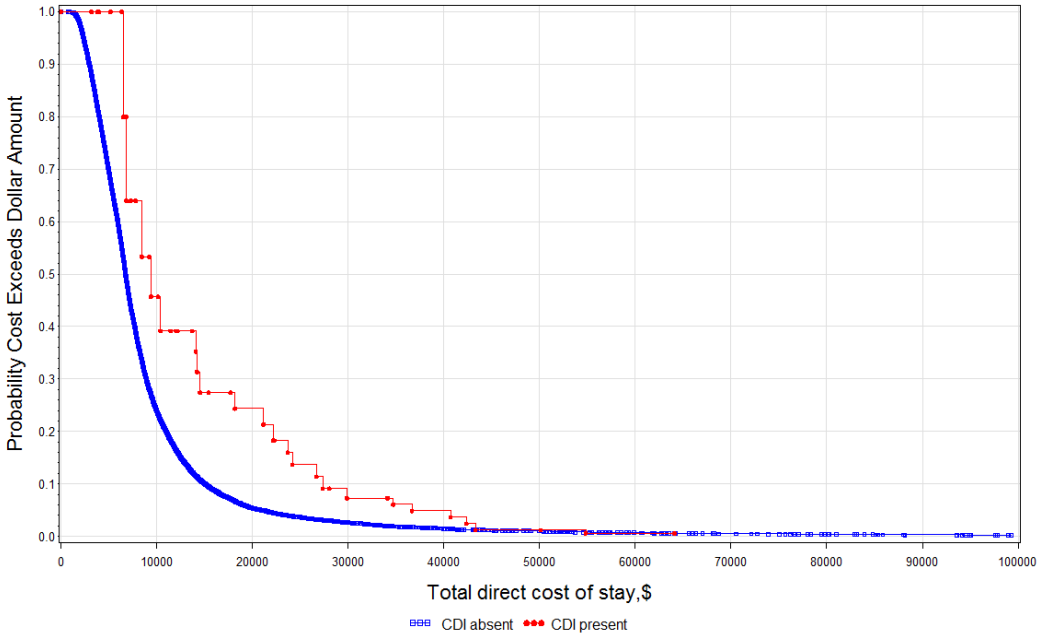
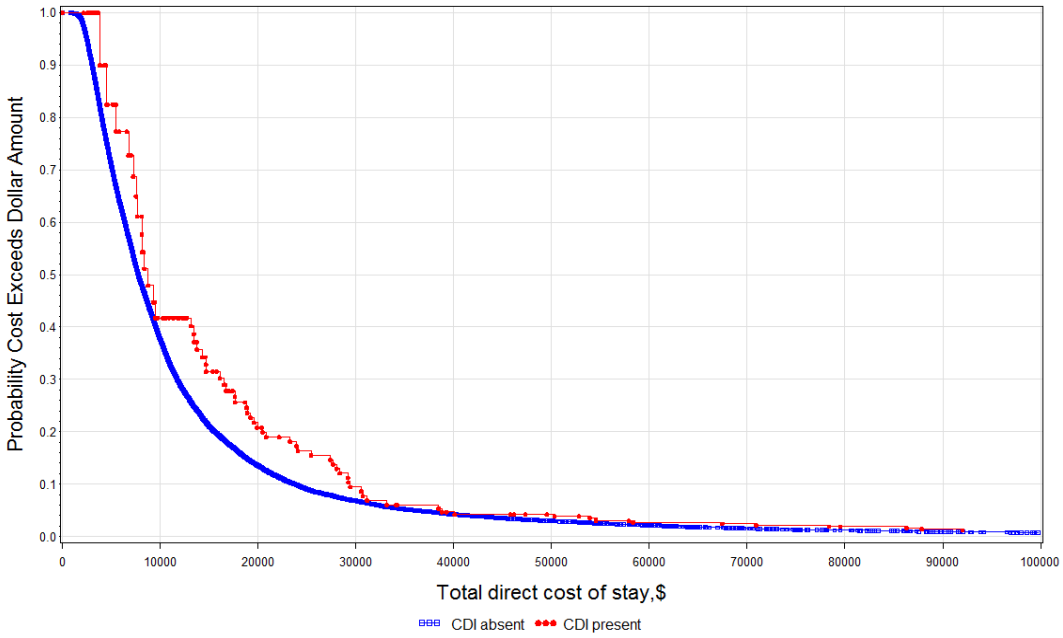


Figure 3.6 Kaplan-Meier survival curves with CDI as a time-varying variable stratified by the baseline risk of mortality

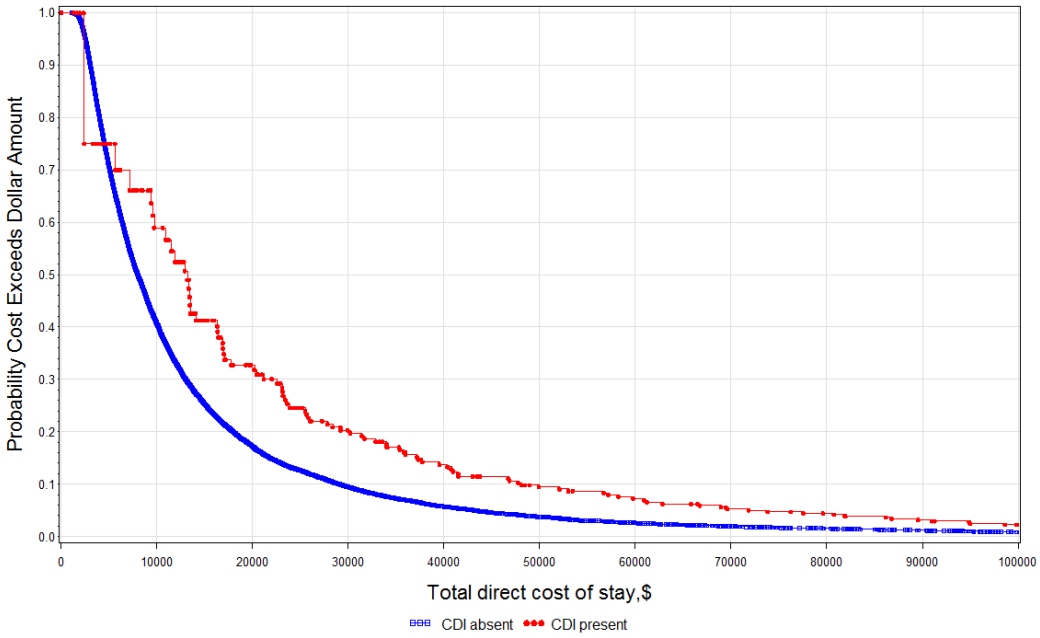
(a) 1st quartile



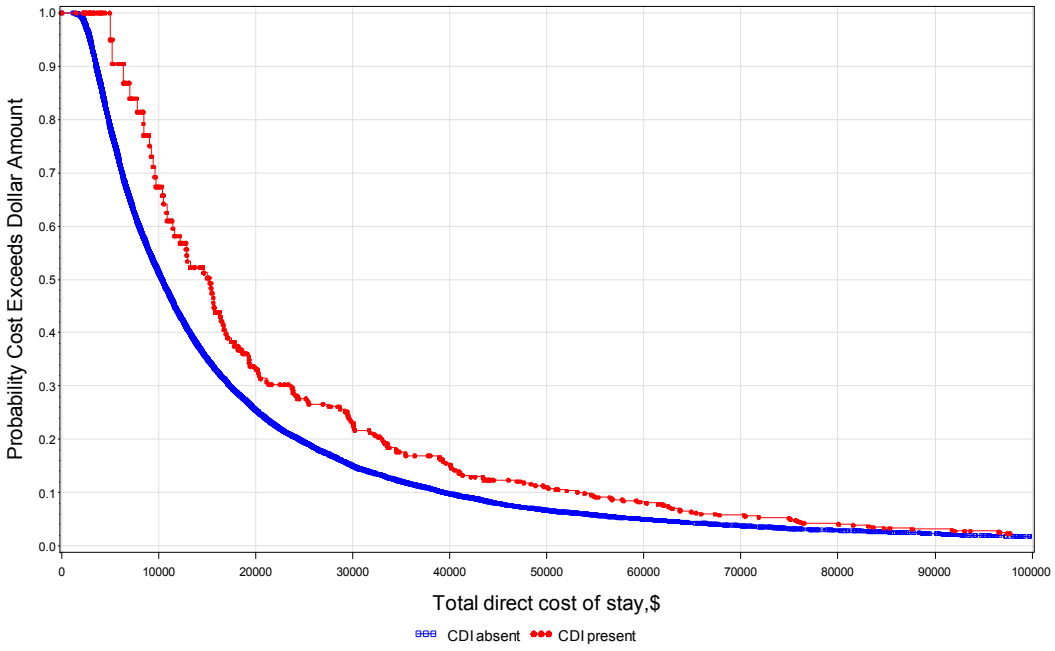
(b) 2nd quartile



(c) 3rd quartile



(d) 4th quartile



3.3.4 Cox proportional hazard regression analyses

We used the Cox proportional hazard (Cox PH) regression model to simultaneously account for the time-varying nature of CDI and the baseline risk of mortality. Table 3.9 presents the regression coefficient estimates and corresponding p-values for the time-fixed versus time-varying models. Both time-fixed and time-varying models showed that the coefficients for CDI and two functional forms of baseline mortality risk were significant.

Table 3.10 presents the hazard ratio for treating CDI as a time-fixed variable and a time-varying variable. As mentioned earlier, the effect estimates from the Cox PH regression model are on the multiplicative scale, which is not intuitive for interpreting cost data. The hazard ratio represents the relative change in the hazard of attaining the final cost or the hazard of not spending another dollar. Thus, hazard ratios less than 1 would be associated with harmful exposures. To aid interpretation of the results, we inverted the hazard ratios associated with each covariate. Thus, we are essentially modeling the hazard of spending another dollar. For the time-fixed Cox PH model, the multiplicative effect of CDI was 2.77 (95% CI, 2.49 to 3.07), i.e., acquiring CDI increased the hazard of spending another dollar almost three fold. When CDI was analyzed as a time-varying variable, the effect was much smaller: 1.15 (95% CI, 1.04 to 1.28). Thus, patients with CDI have a 15% increase in the hazard of spending another dollar to discharge versus those without CDI, even after accounting for baseline mortality risk.

Table 3.9 Model coefficients from Cox proportional hazard regression model

Model type	Covariate	Coefficient (p-value)
Time-fixed model	CDI Yes	-1.0179 (<0.0001)
	Log baseline risk mortality	-0.1315 (<0.0001)
	Square of log baseline risk mortality	-0.0108 (<0.0001)
Time-varying model	CDI Yes	-0.1434 (<0.0001)
	Log baseline risk mortality	-0.1314 (<0.0001)
	Square of log baseline risk mortality	-0.0107 (<0.0001)

Table 3.10 Hazard ratio from Cox proportional hazard regression model

Model type	Hazard ratio (95%CI)	P-value	*c*-statistic (95% CI)
Time-fixed model	2.77 (2.49-3.07)	<0.0001	0.5686 (0.5658- 0.5713)
Time-varying model	1.15 (1.04-1.28)	<0.0001	0.5650 (0.5622- 0.5677)

3.3.5 Sensitivity analysis – stratifying Cox PH regression model for baseline mortality risk

We tested whether the proportional hazard assumption is satisfied for the log of baseline mortality risk and square of log baseline mortality risk. Firstly, we plotted the log-negative-log plot for log transformed baseline risk mortality. Figure 3.7 indicates that the proportional hazard assumption is not satisfied because lines of quartile groups cross each other. Secondly, we plotted Schoenfeld residuals for log transformed baseline risk of mortality and the square of log transformed baseline risk of mortality. Figure 3.8 presents Schoenfeld residuals for log transformed baseline risk of mortality. The regression line suggests that the violation of the proportional hazard assumption is not serious because the slope is -0.00521 (p-value of 0.0015) which is almost close to zero. Figure 3.9 presents Schoenfeld residuals for square of log transformed baseline risk of mortality. The regression line suggests that the violation of the

proportional hazard assumption is not serious either because the slope is 0.00421 (p-value<0.0001) which is almost close to zero.

Nevertheless, to accommodate possible non-proportional hazards, we performed a stratified Cox PH regression analysis. Table 3.11 presents the regression coefficient estimate and the hazard ratios for a time-fixed model and a time-varying model. We found that the hazard ratios (95% CI) for CDI Yes was 1.15 (1.04-1.28) which was almost identical compared to that obtained using a non-stratified Cox PH model (Table 3.10). This confirms the effect of CDI as an important time-varying predictor of cost in both non-stratified and stratified Cox PH models, and that the results are robust against violation of the proportional hazard assumption.

Figure 3.7 Log-negative-log plot for log transformed baseline risk mortality

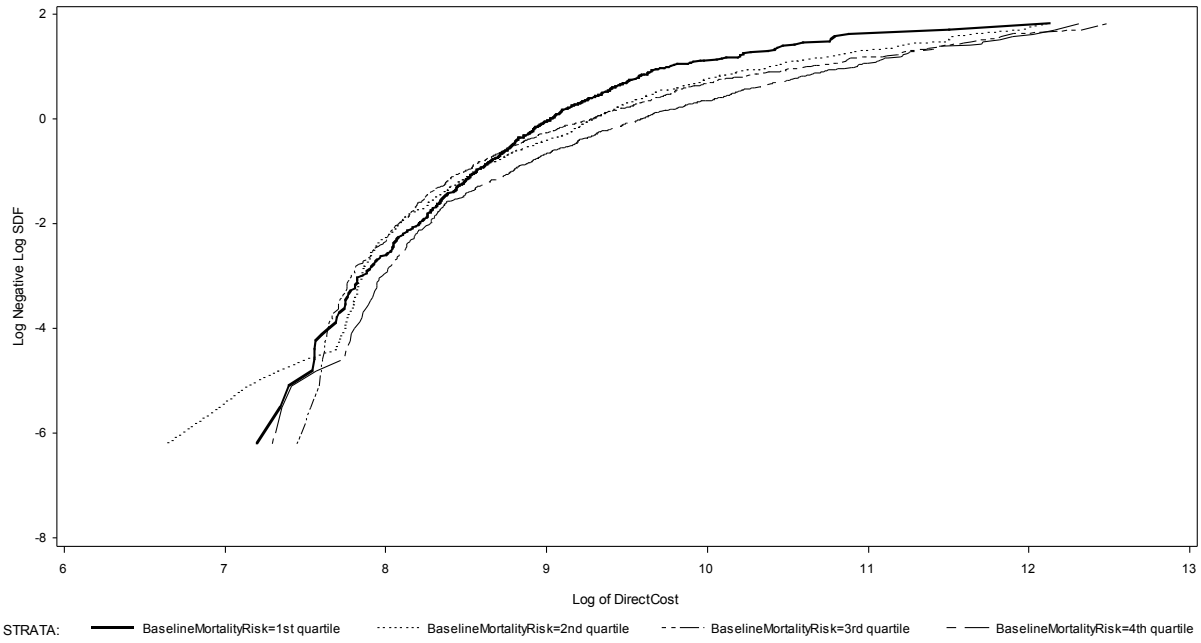
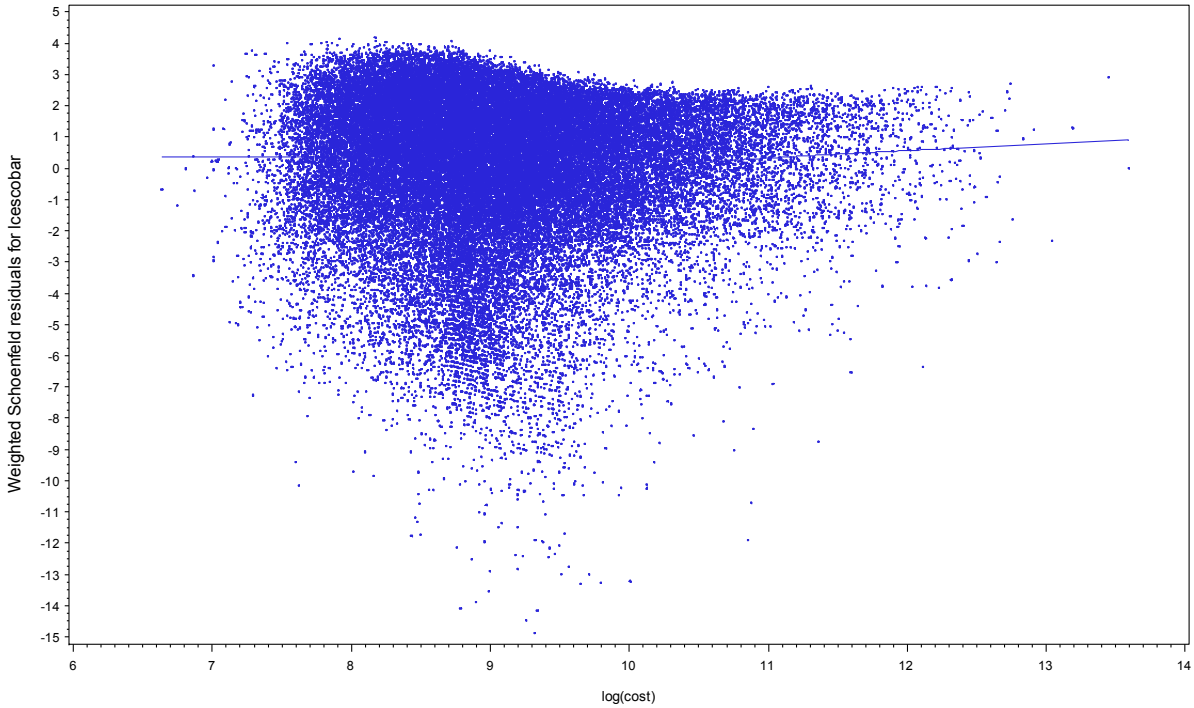
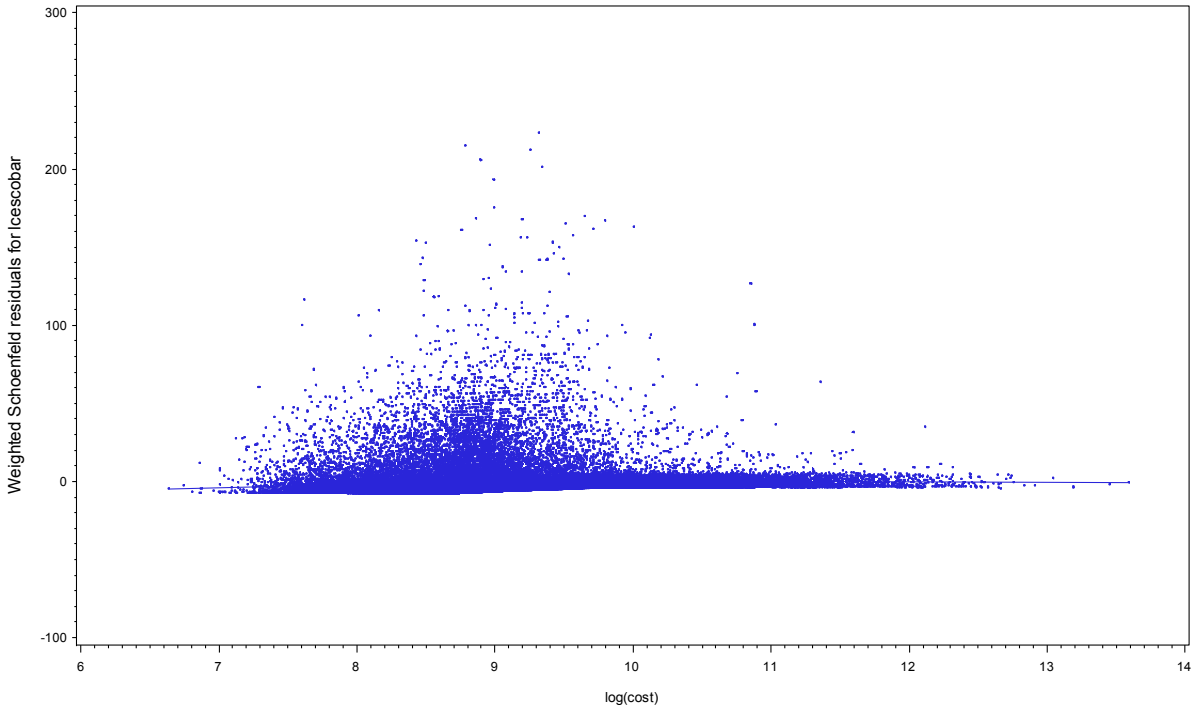


Figure 3.8 Schoenfeld residuals plots for log transformed baseline risk of mortality



Schoenfeld residuals for log transformed baseline risk mortality has a slope of -0.00521 (p-value=0.0015). This suggests that the violation of the proportional hazard assumption is not serious because the slope is very close to zero.

Figure 3.9 Schoenfeld residuals plots for squared of log transformed baseline risk of mortality



Schoenfeld residuals for squared of log transformed baseline risk mortality has a slope of 0.00421 (p-value<0.0001) with the p-value of <0.0001. This suggests that the violation of the proportional hazard assumption is not serious because the slope is very close to zero.

Table 3.11 Stratified Cox proportional hazard regression model coefficients and hazard ratio

Model type	Coefficient	Hazard ratio (95%CI)	*C*-statistic (95% CI)
Time-fixed model	-1.0752*	2.93 (2.64-3.26)*	0.5052 (0.5047- 0.5058)
Time-varying model	-0.1451*	1.15 (1.04-1.28)*	0.5004 (0.5002- 0.5006)

Note: stratified by the baseline risk of mortality, *p-value <0.0001

3.4 *Summary of results*

Overall, the CDI status was significantly associated with increased hospital costs. The cost attributable to CDI was estimated to be higher when CDI was modeled as a time-fixed variable

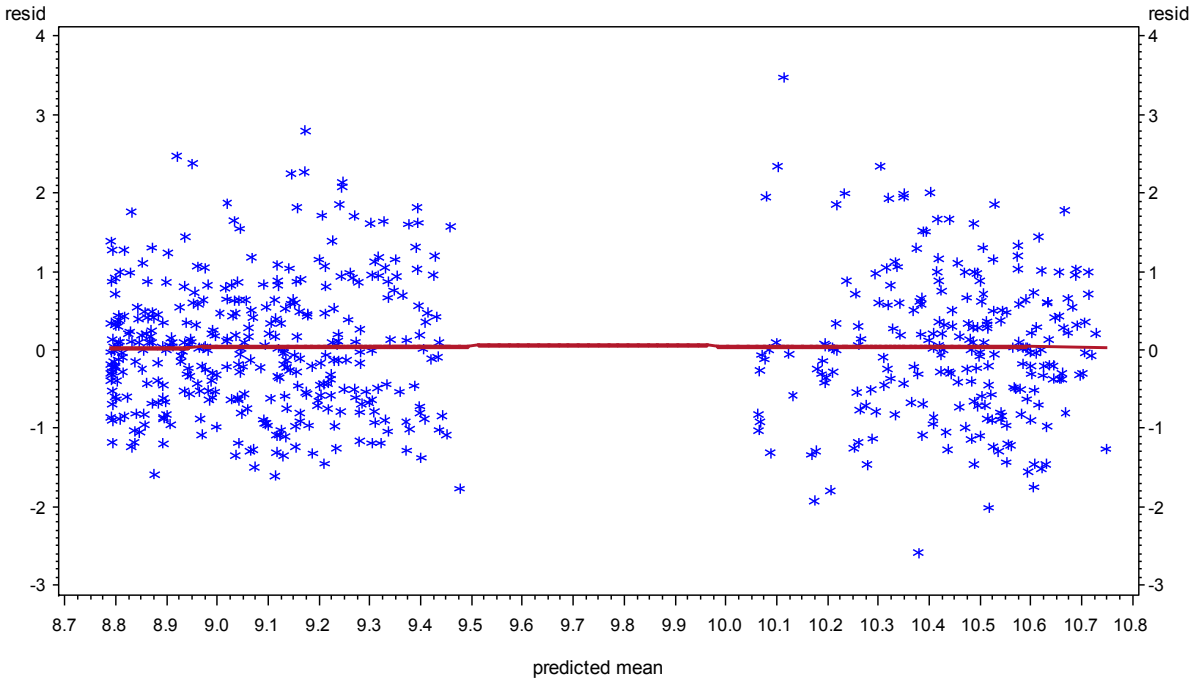
versus a time-varying variable. The cost attributable to CDI was similarly higher when using crude versus multivariable regression analyses in which the baseline mortality risk was included.

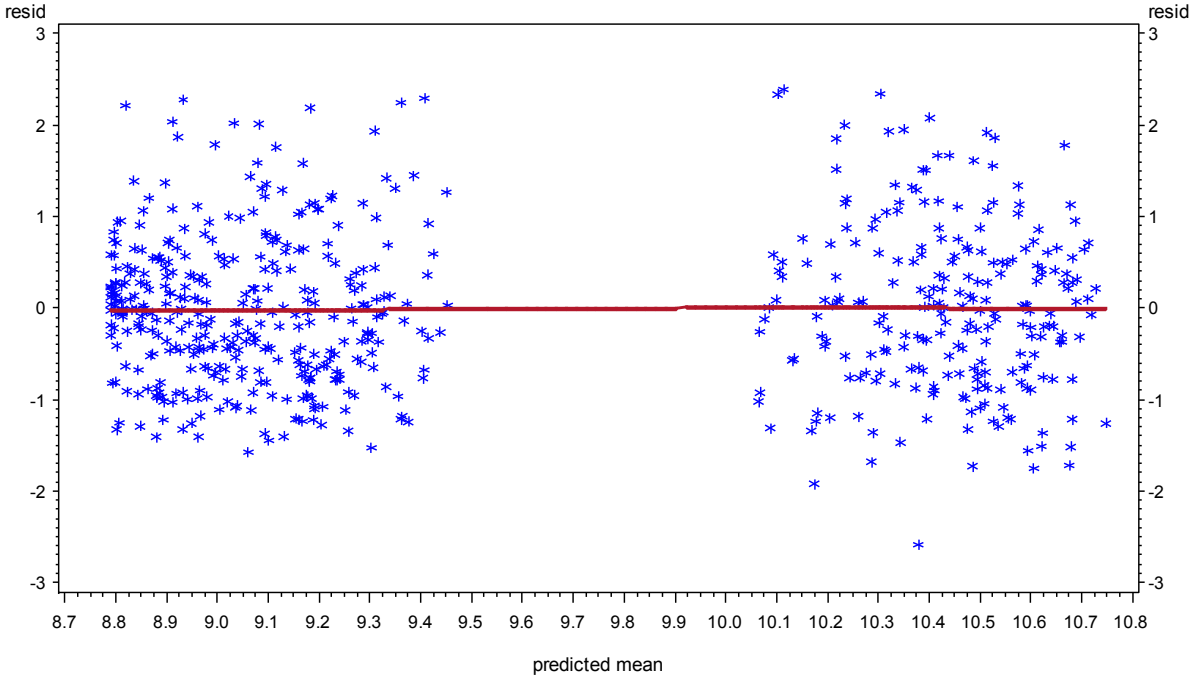
3.5 *Model assessment*

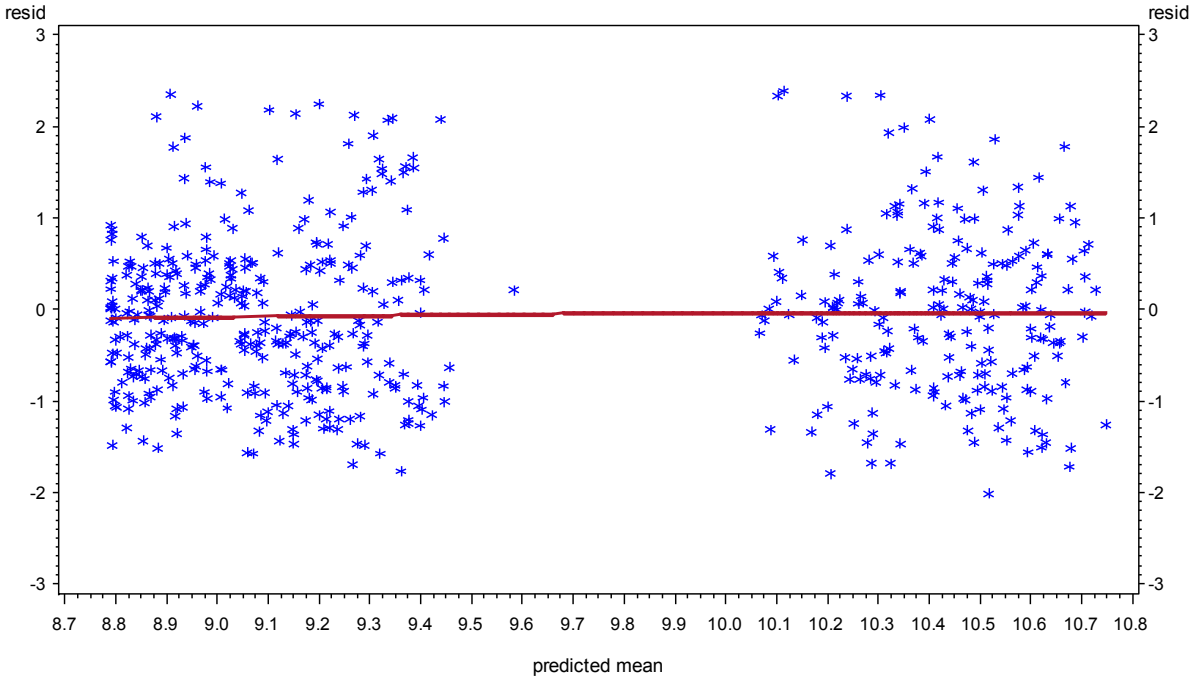
Figure 3.10 and 3.11 present the regression diagnostics for the time-fixed models, the OLS and GLS regression, respectively. From the plots of residuals versus predicted values obtained from the OLS, it was observed that the residuals were randomly distributed with no obvious trends (Figure 3.10). Figure 3.11 depicts plots of the deviance residuals against the predicted direct costs obtained from the GLM model. In each of the four plots, the deviance residuals are randomly distributed with no obvious trends (Figure 3.11).

For the Cox PH regression models, the *c*-statistic values were observed to be very similar for both the time-varying and time-fixed models (Table 3.10). The *c*-statistic for the time-fixed model was 0.5686 (95% CI, 0.5658- 0.5713) and for the time-varying model it was 0.5650 (95% CI, 0.5622- 0.5677). After controlling for the time-varying nature of CDI, the Cox PH regression model still exhibited moderate discrimination compared to the time-fixed model. Therefore, the time dependent model did not improve the discrimination when compared to the time-fixed Cox PH regression model.

Figure 3.10 Raw residual plots against fitted values from OLS regression model with random samples






resid
resid
3
2
1
0
-1
-2
-3
8.7 8.8 8.9 9.0 9.1 9.2 9.3 9.4 9.5 9.6 9.7 9.8 9.9 10.0 10.1 10.2 10.3 10.4 10.5 10.6 10.7 10.8
predicted mean

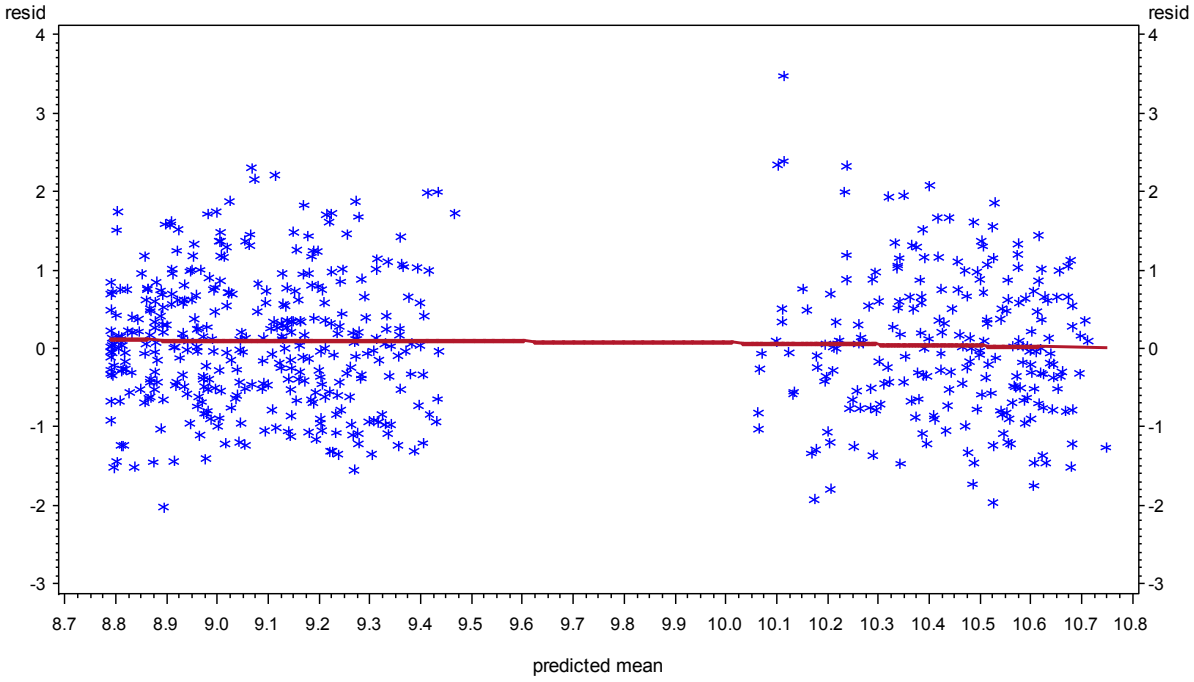
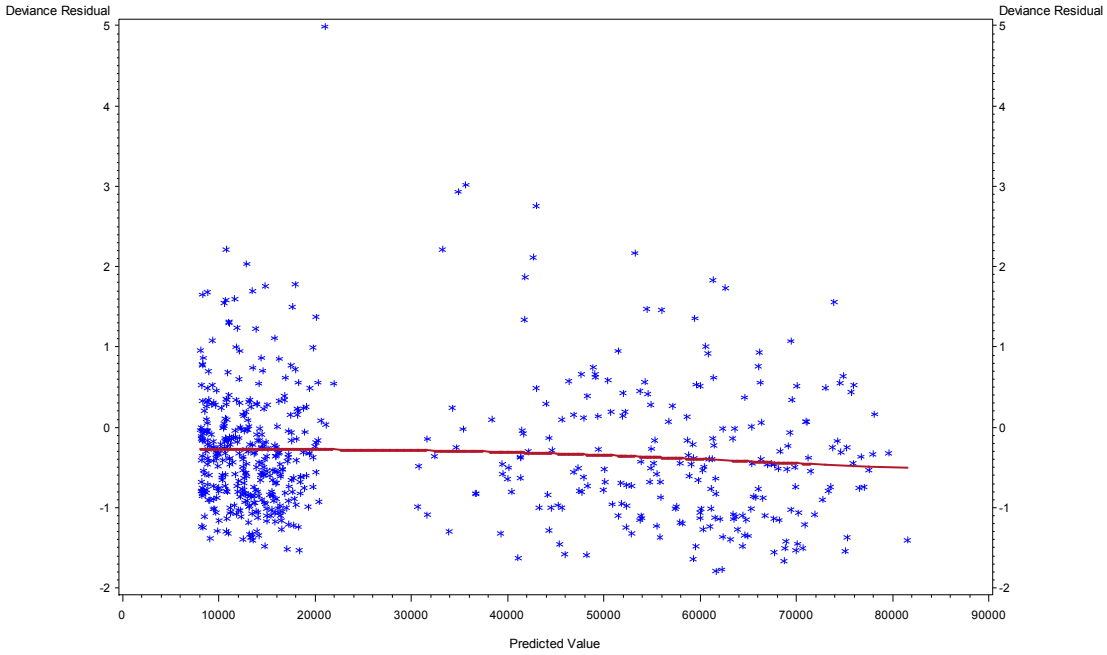
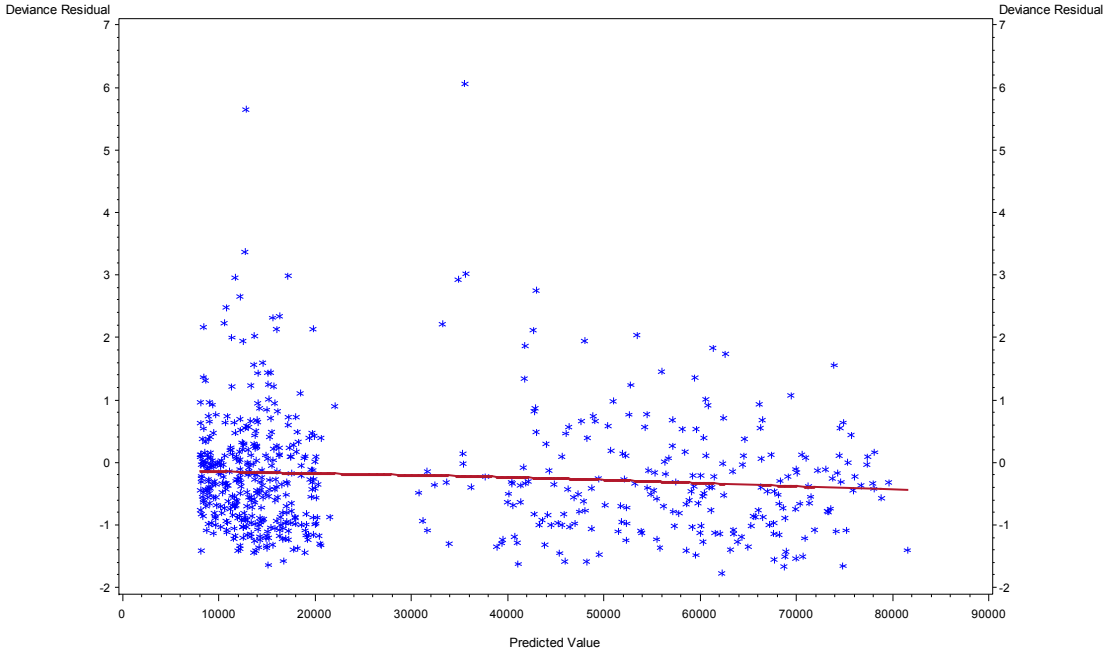
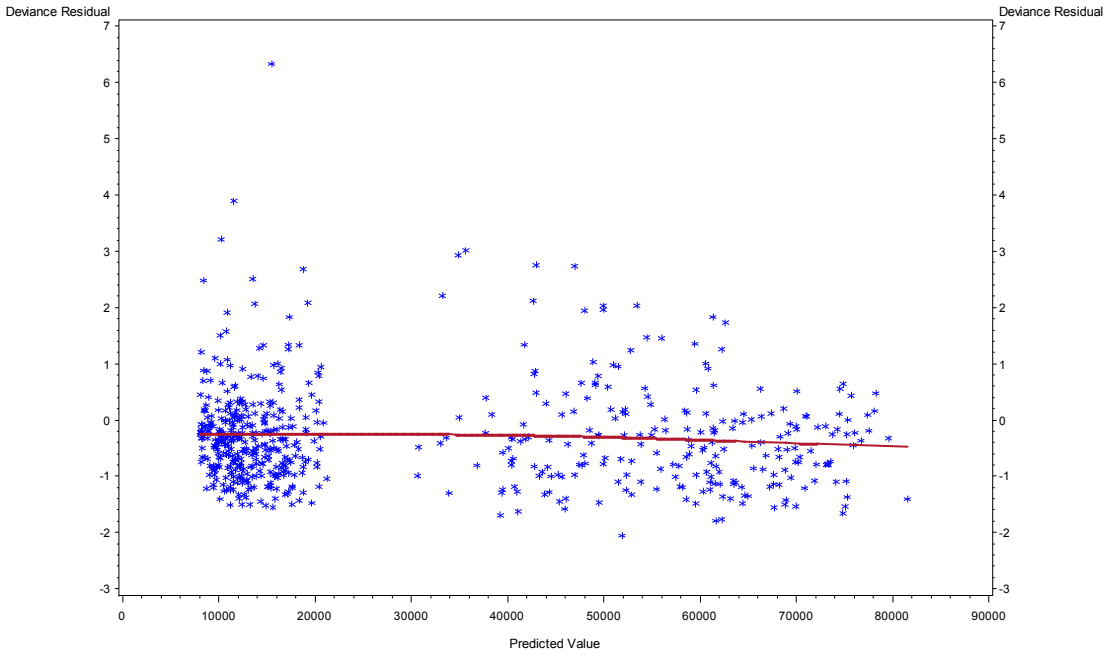
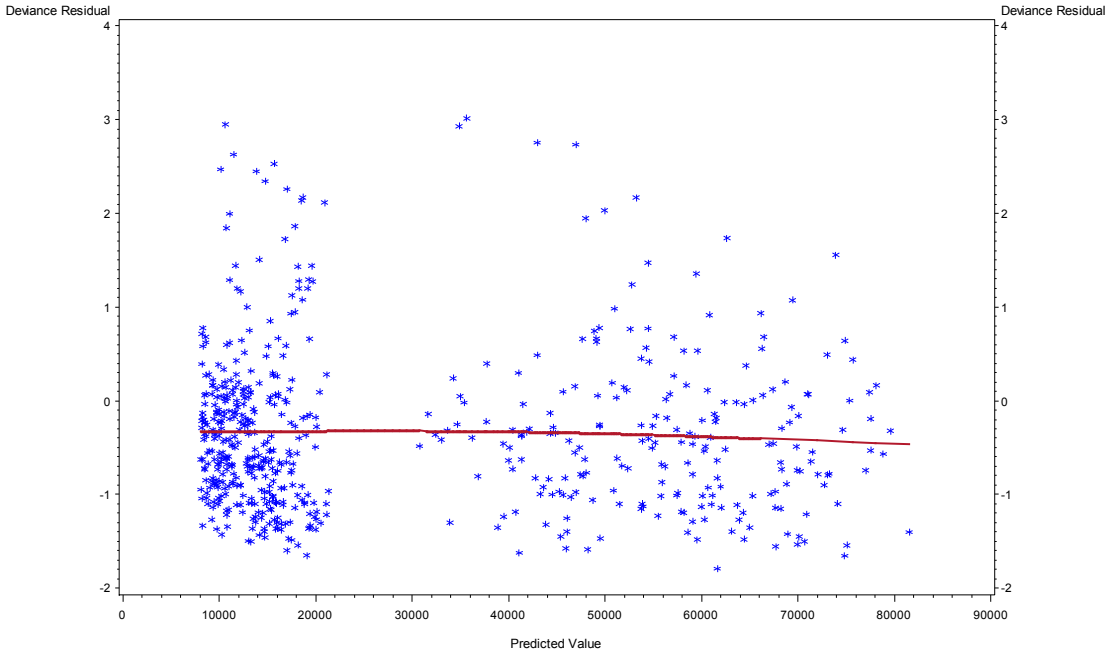

resid
resid
4
3
2
1
0
-1
-2
-3
8.7 8.8 8.9 9.0 9.1 9.2 9.3 9.4 9.5 9.6 9.7 9.8 9.9 10.0 10.1 10.2 10.3 10.4 10.5 10.6 10.7 10.8
predicted mean

Figure 3.11Deviance residual plots against fitted values from GLM regression model with random samples






Deviance Residual
Deviance Residual
7
6
5
4
3
2
1
0
-1
-2
-3
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
Predicted Value


Deviance Residual
Deviance Residual
4
3
2
1
0
-1
-2
0
10000
20000
30000
40000
50000
60000
70000
80000
90000
Predicted Value

Chapter 4 Discussion

4.1 *Overview*

This study applied different analytical approaches to determine the financial impact of CDI (an important hospital-acquired infection) in acute care hospitals, in particular, The Ottawa Hospital. The methods that were used varied with regards to accounting for baseline mortality risk and for the timing of CDI acquisition. We compared the results from statistical models that did not account for either of these factors with those that did account for both factors.

We found that failing to account for baseline mortality risk as well as the time-varying nature of CDI led to substantially over-estimated cost impact. The overall cost impact derived from the models that accounted for only the baseline risk were more similar to the overall cost impact in unadjusted models than those which only included the impact of timing. The models adjusting for the time-varying nature of CDI acquisition were the most conservative.

From a conceptual point-of-view these results make sense for the following reasons. Firstly, the relationship between CDI and cost is influenced by a patient's health status at hospital admission. As sicker patients cost more to care for *and* are at higher risk for CDI, not accounting for their health status will lead to inflated estimates of the impact of CDI on cost. Similarly, patients who stay in the hospital for long durations (which may occur due to factors other than health status) will also have a higher risk of CDI. As length of stay is the primary driver of hospital costs, then not accounting for CDI timing will lead to inflated estimates of cost. Secondly, for obvious reasons, it is not sensible to attribute pre-CDI acquisition costs to CDI infection.

We conclude that the method of assessing the financial impact of CDI matters: studies not accounting for both the time of CDI acquisition and baseline risk factors almost certainly overestimate financial impact. Any evaluation of costs not accounting for these factors should be interpreted cautiously.

In the sections that follow, we will highlight the most salient results of this study, place these in the context of other studies, and address the implications of this work. We will then describe the strengths and weaknesses of this study. Finally, we will conclude by making recommendations to policy makers and researchers.

4.2 *Summary of the most important findings*

In Table 4.1, we present the projected annual cost and the impact for *Clostridium difficile* infection at The Ottawa Hospital based on the results of this study. We compare projected annual cost based on inappropriate methods, with the projected annual cost on the recommend method. We are currently experiencing an average of 15 hospital-acquired CDI cases per month at The Ottawa Hospital (180 cases per year). Although all of the models estimate care costs to be similar in the patients who *do not* acquire CDI, the inpatient cost of care for those patients acquired CDI during their hospital stay was very different, leading to vastly different estimates of institutional impact. For example, projections based on the OLS and GLM models lead one to estimate that the annual cost due to CDI to be approximately $8.85 and $9.03 million, respectively. These models accounted for baseline mortality risk but did not account for the time-varying nature of CDI acquisition and as such, they provide over-estimates of cost. Likewise, our analyses using simple KM survival curves, which were not adjusted for CDI timing, led to a projected overall institutional cost due to hospital-acquired CDI of $10.07

million annually - also an overestimate. When we adjusted for the time-varying nature of CDI in the KM survival curve, the projected annual cost due to hospital-acquired CDI dropped dramatically to $1.62 million compared with other time models treating CDI as time-fixed variable. Thus, a more accurate estimate of the annual cost due to hospital-acquired CDI at The Ottawa Hospital is approximately $1.6 million.

Table 4.1 Annual inpatient costs at The Ottawa Hospital

Analytical method	Adjust for BMR*	Adjust for time-varying nature of CDI	Hospital cost in non-CDI per patient	Hospital cost in CDI per patient	Cost due to CDI per patient	Projected annual cost due to CDI¥($, in million)
Crude mean	No	No	18,984	76,542	57,558	10.36
OLS	Yes	No	18,687	67,837	49,150	8.85
GLM	Yes	No	18,388	68,580	50,192	9.03
KM – time-fixed	No	No	18,975	74,929	55,954	10.07
KM –time-varying	No	Yes	19,092	28,089	8,997	1.62
KM –time-varying†	Yes†	Yes	12,866-25,963φ	22,237-36,347φ	9,671 – 12,384φ	1.69-1.87φ
Cox PH	Yes	Yes	Unknown	Unknown		Unknown

* Baseline mortality risk, ¥ Annual cost due to CDI= (overall cost for CDI-overall cost for non-CDI) X annual CDI cases at TOH (180), †Baseline risk mortality adjusted by the stratification, φ Lower value represents the cost for those patients in the 1st quartile baseline risk mortality and higher value represents the cost for those patients in the 4th quartile baseline risk mortality.

4.2.1 Impact of accounting for baseline mortality risk

Compared to cost estimates in which baseline mortality risk was ignored, those models that adjusted for baseline mortality risk yielded lower estimated costs for CDI. This finding reinforces our supposition that baseline risk of mortality has an impact on the hospital cost, and therefore, it needs to be properly controlled while performing cost analyses. The importance of accounting for baseline mortality risk is well demonstrated in the KM survival curve analysis. The estimated overall cost for CDI patients in the highest quartile was $37,690, which was 63%

higher than those CDI patients in the lowest quartile ($23,078) (Table 4.1). As expected, the estimated annual TOH hospital cost due to CDI was greatest among the patients in the highest risk quartile (Table 4.1). This suggests that models estimating cost attributable to CDI should account for baseline risk of death.

4.2.2 Impact of accounting for CDI as a time-varying variable

In all models, the estimates for the cost of hospital care for non-CDI patients were similar (the range was between $18,833 and $19,092). However, the estimates for the cost of hospital care for *patients with CDI* varied from $28,089 and $76,542 with the lower values obtained when accounting for CDI as a time-varying attribute.

When accounting for the time-varying nature of CDI, the estimated total hospital cost due to hospital-acquired CDI from the KM curve was $8,997. In other words, there is an $8,997 difference in estimated total hospital cost between those with CDI and without CDI (Table 4.1). In comparison, the estimates from the methods that did not control for the time-varying nature of CDI had a 6-fold higher hospital cost than models that adjusted for the time-varying nature of CDI. ($8,997 x 6 is roughly the same as the cost attributable to CDI from the models treating CDI as time-fixed variable, which was between $51,009 and $52,089). Hence, the cost estimates by methods not adjusting for the CDI time-varying nature are overestimated. Based on the data, we conclude that estimates of hospital cost are strongly influenced by the time-varying nature of CDI and this must be factored in when performing hospital cost analyses.

4.3 *Implications of this study*

This work explores the use of time-varying models to predict the cost attributable to hospital-acquired CDI, thus filling a gap in the current literature. Because exposure to CDI varies between patients and can occur at any point during hospitalization, the estimated cost attributable to CDI in models that do not properly account for its time-varying nature are usually overestimated. Since more accurate models are greatly needed, the time-varying model represents an advance in methods of estimating the cost relating to hospital-acquired CDI. Accurate cost estimates for CDI can help policy makers and funders devise more cost effective prevention strategies because hospitals can use these data to help them allocate limited health resources and evaluate any interventions or infection control practices for potential cost savings. In addition, it will be possible to apply them to other hospital-based complications for which one can define a specific onset date. For example, one will be able to apply this time-varying method to other types of hospital-acquired infections, procedure complications, and adverse drug events.

4.4 *Comparison to previous studies*

4.4.1 For models treating CDI as a fixed-variable

Although earlier studies that are similar to this one seem to agree that the overall healthcare cost for patients with CDI is higher than patients without CDI (i.e. the increase in costs for patient with CDI ranged around 33% - 41% by Dubberke et al. and 46% by O'Brien et al.), comparing estimate costs between all these studies is quite challenging[15].

There is marked variability between each study that makes direct comparison between them difficult, such as the study population (entire inpatient population, limited inpatient population

such as including only those admitted to the medicine units), study design (retrospective cohort study, prospective observational study, matched cohort study versus unmatched study), adjustment for confounders (types and number varies), statistical analysis (simple linear regression, multiple linear regression) or cost type (micro-costs or charges). Therefore, when comparing cost estimates between different studies, one should acknowledge these listed differences.

According to our models treating CDI as a time-fixed variable, patients with hospital-acquired CDI cost between 3.6 (as predicted by OLS model) and 3.7 (as predicted by the GLM model) times more than those without hospital-acquired CDI. These differences are much higher than reported by Dubberke et al., which reported that patients with CDI cost 1.4 times more than the control. The difference between this study and Dubberke et al. may be due to different criteria that were used for the study population. For example, in Dubberke et al. only patients admitted to the medical unit were included, which may have excluded all of the surgical population. This may have created a downward bias since the surgical population may cost more than non-surgical population[15]. Furthermore, cost results from Dubberke et al. included both hospital-acquired and community-acquired CDI patients. This might explain the lower difference of cost between CDI and non-CDI populations in their study because hospital-acquired CDI - an adverse event during hospital stay while treating another disease – typically requires more cost to treat than community acquired CDI.

In another study by O'Brien et al., a study in the literature that is most comparable to the current one, it was reported that the mean cost for patients with hospital-acquired CDI was $29,946 +/- $46,489[48]. Their study was comparable to the current one because they used the same study population (included the entire inpatient population), CDI cases (included only the hospital-

acquired CDI cases), and study design (non-matching retrospective cohort). However, their estimated mean cost for CDI of$29,946 +/- $46,489 was much lower than the current study of $68,580 (derived from the direct costs of $48,025, (95% CI $44,006-$52,411) multiplied by 1.428 to calculate the total hospital costs). This discrepancy from O'Brien et al. can be explained by the fact the length of stay (LOS), which is a major component of hospital costs[49], is significantly higher for this study. The mean LOS for the current study was 45.6 days (95% CI, 42.0-49.5 days - data not shown) versus 15.7 +/-16.9 days for O'Brien et al[48]. In order to factor in the difference of LOS between these two studies, we had to calculate the average daily costs for each one. The average daily cost for the current study was found to be $1,504 versus $1,907 for O'Brien et al. Because the O'Brien study was multi-centered (they included 2656 hospital-acquired CDI cases from 77 hospitals in Massachusetts) versus the current study (360 CDI cases in a single institute) which is single-centered, they had a broader spectrum of patient groups. This may have contributed to the higher daily hospital cost observed in their estimates[48] when comparing against the current study. Also, the O'Brien et al. study did not explicitly mentioned that relapsed CDI cases were excluded, which may have also contributed to the higher daily cost that was observed since relapse CDI cases are known to cost more to treat than non-relapse cases [48]. Overall *the total hospital cost* estimate from the current study was significantly higher than the study by O'Brien. This may be associated with much greater LOS found in this study as opposed to LOS from O'Brien study. Despite the higher *total hospital cost* found in this study, average daily cost was lower by $400 compared to the study by O'Brien et al. which may be due to the broader spectrum of patients and inclusion of relapsed cases in O'Brien's study.

4.4.2 For time-varying model

We believe that the present study provides a comprehensive estimate of cost attributable to CDI by controlling its time-varying nature, whereas previous studies did not explore this area. Although it was not possible to compare cost estimates from the time-varying model with the previous studies directly, we were able to compare the hazard ratio from the Cox PH regression model that treated CDI as time-fixed versus one that treated CDI as time-varying (Table 4.10). It was clear that the hazard ratio was substantially lower when treating CDI as a time-varying (1.15) versus a time-fixed (2.77) variable, indicating evidence that the time-fixed model produced overestimated result. From this observation, we can conclude that methods treating hospital-acquired CDI as a time-varying factor is more accurate at predicting cost estimates because it can properly control the time-dependency nature of CDI.

4.5 *Strengths*

The major strength of the study is that it provides a comprehensive estimate of cost attributable to CDI by controlling for its time-varying nature simultaneously with baseline risk of mortality. Since the estimates are more accurate, hospitals can use them if they need to justify or evaluate intervention strategies. The other strength is that because this study used an electronic algorithm for detecting CDI cases that has a sensitivity and specificity of 100%, it is less likely to face any impact of a misclassification bias.

4.6 *Limitations*

The biggest limitation of the study comes from the fact that it does not include all possible CDI cases and sources of costs relating to hospital-acquired CDI. Firstly, the study's time horizon was

the encounter. As a result, costs incurred by other parts of the health system and other societal costs such as the lost productivity associated with hospital-acquired CDI were not assessed. In addition, we excluded cases in which the infection was likely truly hospital-acquired and attributable to a preceding encounter (i.e. tests positive within 72 hours with a prior hospitalization within 8 weeks)[39]. We excluded these cases because we could not reliably ensure all such cases were captured, and also because we could not determine with certainty that CDI was not acquired elsewhere. Patients may get CDI following an encounter and not return to TOH (if the symptoms are self-limiting, if they are treated by an outpatient physician, or if they are treated at another hospital in our region). For this reason we focused on the encounter as the unit of analysis. Thus, our results should be considered conservative estimates of the overall costs to the health system.

Lastly, it was not possible to derive the estimated mean cost for each different patient population from the Cox PH regression model. Although the hazard of spending another dollar or more for those with hospital-acquired CDI was estimated to be 15% higher than for those without, this estimate cannot be directly translated to actual dollars that can be saved by the hospital, which is the main interest for decision makers. Therefore, the Cox PH regression model alone may not be sufficient for predicting hospital costs, but it may be useful when comparing increased costs due to hospital-acquired CDI among different nursing units or patient population.

4.7 *Recommendations*

To estimate the cost attributable to CDI accurately, analysts should account for the time-varying nature of CDI and baseline risk of death simultaneously. Our results indicate that hospitals

should consider prioritizing patients according to the baseline risk of mortality when implementing intervention strategies to reduce hospital-acquired CDI.

4.8 *Conclusion*

From the hospital perspective, obtaining cost estimates that factor in the time-varying nature of CDI is essential because it provides greater accuracy when predicting the amount of dollars that can be saved when each hospital-acquired CDI case is prevented. However, none of the studies in the literature we have investigated the cost attributable to CDI with the adjustment for time-dependency, making the current study the first one to explore it.

Chapter 5 References

[1] Sunenshine RH, McDonald C. *Clostridium difficile*-associated disease: New challenges from an established pathogen. Clevel Clin J Med 2006;73(2):187-97.

[2] Riegler M, Sedivy R, Pothoulakis C, Hamilton G, Zacherl J, Bischof G. *Clostridium difficile* toxin B is more potent than toxin A in damaging human colonic epithelium in vitro. J Clin Invest 1995;95(5):2004-11.

[3] Bartlett JG. Antibiotic-associated diarrhea. New Engl J Med 2002;346(5):334-9.

[4] Pouteen SM, Simor AE. *Clostridium difficile*-associated diarrhea in adults. Canadian Medical Association Journal 2004;171(1):51-58.

[5] McFarland LV, Surawicz CM, Stamm WE, Risk factors for *Clostridium difficile* carriage and *C.difficile*-associated diarrhea in a cohort of hospitalized patients. J Infect Dis 1990; 162(3):678-84.

[6] Kelly CP, Pothoulakis C, LaMont JT, *Clostridium difficile* colitis. N Engl J Med 1994;330:257-262.

[7] Gravel D, Miller M, Simor A, Taylor G, Gardam M. McGeer A, Hutchinson J, Moore D, Kelly S, Boyd D, Mulvey M, Bryce E, Conly J, Dow G, Embil J, Embree J, Forgie S, Frenette C, Henderson E, John M, Johnston L, Kibsey P, Langley J, L. Health care-associated *Clostridium difficile* infection in adults admitted to acute care hospitals in Canada: A Canadian nosocomial infection surveillance program study. Clinical Infectious

Diseases 2009;(48)5:568-76.

[8] Pépin J, Valiquette L, Cossette B. Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. CMAJ 2005;173(9):1037-41.

[9] Valiquette L, Low DE, Pépin J, McGeer A. *Clostridium difficile* infection in hospitals: a brewing storm. CMAJ 2004;171(1):27-9.

[10] Karas JA, Enoch DA, Aliyu S. A review of mortality due to *Clostridium difficile* infection. J Infect 2010; 61:1-8.

[11] McDonald LC, Owings M, Jernigan DB. *Clostridium difficile* infection in patients discharged from US short-stay hospitals, 1996-2003. Emerg Infect Dis 2006;12(3):409-15.

[12] Khanna S, Pardi DS. The growing incidence and severity of *Clostridium difficile* infection in inpatient and outpatient settings. Expert Rev Gastroenterol Hepatol 2010; 4(4):109-16.

[13] Rupnik M, Wilcox MH, Gerding DN. *Clostridium difficile* infection: New developments in epidemiology and pathogenesis. Nat Rev Microbiol 2009;7:526-36.

[14] Kyne L, Hamel MB, Polavaram R, Kelly CP. Health care costs and mortality associated with nosocomial diarrhea due to *Clostridium difficile*. Clin Infect Dis 2002;34(3):346-53.

[15] Dubberke ER, Reske KA, Olsen MA, McDonald LC, Fraser VJ. Short- and long-term attributable costs of *Clostridium difficile*-associated disease in nonsurgical inpatients. Clin Infect Dis 2008;46:497-504.

[16] Kyne L, Hamel MB, Polavaram R, Kelly CP. Health care costs and mortality associated with nosocomial diarrhea due to *Clostridium difficile*. Clin Infect Dis 2002:34(3):346-53.

[17] Miller MA, Hyland M, Ofner-Agostini M, Gourdeau M, Ishak M. Morbidity, mortality, and healthcare burden of nosocomial *Clostridium difficile*-associated diarrhea in canadian hospitals. Infect Control Hosp Epidemiol 2002;23:137-40.

[18] Pakyz A, Carroll NV, Harpe SE, Oinonen M, Polk RE. Economic impact of *Clostridium difficile* infection in a multihospital cohort of academic health centers. Pharmacotherapy 2001;31:546-51.

[19] Scheurer DB, Hicks LS, Cook EF, Schnipper JL. Accuracy of ICD-9 coding for *Clostridium difficile* infections: a retrospective cohort. Epidemiol Infect 2007;135:1010-3.

[20] Dubberke ER, Reske KA, McDonald LC, Fraser VJ. ICD-9 codes and surveillance for *Clostridium difficile*-associated disease. Emerg Infect Dis 2006;12:1576–9.

[21] Oake N, Taljaard M, Van Walraven C, Wilson K, Roth V, Forster AJ. The effect of hospital-acquired *Clostridium difficile* infection on in-hospital mortality.Arch Intern Med 2010;170:1804-10.

[22] Duan N. Smearing estimate: a nonparametric retransformation method. J Am Stat Assoc 1983;78:605-10.

[23] Zhao Y, Encinosa W. The Costs of End-of-Life Hospitalizations, 2007. Agency for Healthcare Research and Quality, Rockville, 2009.

[24] Dudley RA, Harrell FE, Jr, Smith LR, Mark DB, Califf RM, Pryor DB, Glower D, Lipscomb J, Hlatky M. Comparison of analytic models for estimating the effect of clinical factors on the cost of coronary artery bypass graft surgery. J Clin Epidemiol 1993;46(3) 261-71.

[25] Fenn P, McGuire A, Phillips V, Backhouse M, Jones D. The analysis of censored treatment cost data in economic evaluation. J Am Stat Assoc 1995;33:851-63.

[26] Lin D, Feuer E, Etzioni R, Wax Y. Estimating Medical Costs from Incomplete Follow-Up Data. Biometrics 1997;53:419-34.

[27] Angelis GD, Murthy A, Beyersmann J, Harbarth S. Estimating the impact of healthcare-associated infections on length of stay and costs.Clin Microbiol Infect 2010;16:1729–1735.

[28] Manning W. THe logged dependent variable, heteroscedasticity, and the retrasnformation problem. J Health Econ 1998;17:283-95.

[29] Blough DK, Madden CW, Hornbrook MC. Modeling risk using generalized linear models. J Health Econ 1999;18:153-71.

[30] Buntin MB and Zaslavsky AM. Too much ado about two-part models and transformation?: comparing methods of modeling Medicare expenditures. J Health Econ 2004;23:525-42.

[31] Manning WG, Basu A and Mullahy J. Generalized modeling approaches to risk adjustment of skewed outcomes data. J Health Econ 2005;24:465-88.

[32] Austin PC, Ghali WA, Tu JV. A comparison of several regression models for analysing cost

of CABG surgery. Stat Med 2003;22(17):2799-2815.

[33] Scitovsky AA. The High Cost of Dying" Revisited. The Milbank Quarterly 1994;72(4):561-591.

[34] Lipscomb J, Ancukiewicz M, Parmigiani G, Hasselblad V, Samsa G, Matchar D. Predicting the Cost of Illness: A comparison of Alternative Models Applied to Stroke. Med Decis Making 1998;18:S39-S56.

[35] The Ontario Case Costing Initiative, [Online]. Available: http://www.occp.com/mainPage.htm.

[36] Werb D, Wood E, Kerr T, Hershfield N, Palmer RWH, Remis RS. Treatement costs of hepatitis C infection among infection drug users in Canada, 2006-2026. Int J Drug Policy 2011;22(1):70-6.

[37] Guo Y, Longo CH, Xie R, Wen SW, Walker MC, Smith GN. Cost-effectiveness of transdermal nitroglycerin use for preterm labor. Value Health 2011;14(2):240-6, 2011.

[38] Ministry of Health and Long-Term Care. Ontario Case Costing Guide. 2010.

[39] Provincial Infectious Diseases Advisory Committee. Routine Practices and Additional Precautions. 2009.

[40] Statistics Canada. Consumer Price Index, health and personal care, by province. [Online]. Available: http://www.statcan.gc.ca/tables-tableaux/sum-som/l01/cst01/econ161a-eng.htm.

[41] Iezzoni LI, Shwartz M, Moskowitz MA, Ash AS, Sawitz E, Burnside S. The ability of MedisGroups and its clinical variables to predict cost and in-hospital death. Report prepared for the Health Care Financing Administration, Cooperative Agreement No. 18-C-98526/1-04. Boston: Health Care Research Unit, Boston University Medical Center, Washington, DC, 1988.

[42] Van Walraven C, Escobar GJ, Greene JD, Forster AJ. The Kaiser Permanente inpatient risk adjustment methodology is valid in an external patient population. J Clin Epidemiol 2010; 63(2):798-803.

[43] Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for analyzing healthcare utilization and costs. Annu. Rev. Public Health 1999;20:125-44.

[44] Manning WG, Mullahy J. Estimating log models: to transform or not to transform?. J Health Econ 2001;20(4):461-94.

[45] Forster AJ, Taljaard M, Oake N, Wilson K, Roth V, Van Walraven C. The effect of hospital-acquired infection with *Clostridium difficile* on length of stay in hospital. CMAJ 2012;184(1):37-42.

[46] Division of Biomedical Statistics and Informatics. Locally Written SAS Macros. 2010.

[47] Zigon G, Rosato R, Gregori D, Bo S. A Comparison of Analytic Models for the Costs of the Hospitalized Diabetic Patients. Metodoloski zvezki 2005;2(1):135-45.

[48] O'Brien JA, Lahue BJ, Caro JJ, Davidson DM. The emerging infectious challenge of

Clostridium difficile-associated disease in Massachusetts hospitals: clinical and economic consequences. Infect Control Hosp Epidemiol 2007;28(11):1219-27.

[49] Young DS, Sachais BS, Jefferies LC. The Costs of Disease. Clinical Chemistry 2000;46(7):955-66.