

A Longitudinal Study of the  
Effect of Renal Failure on  
Readmission Rates of Patients  
with *Clostridium Difficile*



Brian Detweiler  
College of Arts and Sciences  
University of Nebraska, Omaha

A thesis submitted for  
*Master of Science*

May 13, 2017

## Abstract

*Clostridium difficile* Infection (C. *diff*, or simply CDI) is a highly contagious endospore forming bacterium that is transferred through physical contact with an infected surface. Symptoms range from diarrhea to life-threatening colitis and is most commonly acquired in a hospital setting where antimicrobials have been administered. Increased mortality in CDI patients with renal failure comorbidities has appeared in the literature as early as 1998 [1]. In this study, we use the *Nationwide Readmissions Database* to assess the risk of 30, 60, and 90 day readmissions in patients with comorbid CDI and renal failure conditions. We also discuss general CDI trends from 2001-2014, using the *Nationwide Inpatient Sample*.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Overview of Clostridium <i>difficile</i> . . . . .	1
1.2	Overview of Renal Failure . . . . .	2
1.2.1	The MDRD Equation vs. the CKD-EPI Equation . . . . .	2
1.2.2	Classifying Chronic Kidney Disease . . . . .	3
1.2.3	End stage renal disease . . . . .	3
1.3	Readmissions and the ACA . . . . .	3
1.3.1	Readmission measures . . . . .	4
1.4	Overview of the data . . . . .	4
1.4.1	The Nationwide Inpatient Sample (NIS) . . . . .	5
1.4.2	The Nationwide Readmissions Database (NRD) . . . . .	5
1.4.3	Limitations . . . . .	5
<b>2</b>	<b>Methods</b>	<b>8</b>
2.1	Data source . . . . .	8
2.1.1	Persistence . . . . .	9
2.1.2	Diagnosis and procedure codes . . . . .	10
2.1.3	Determining index admissions and readmissions . . . . .	11
2.1.4	Choosing features . . . . .	12
2.2	Statistical analysis . . . . .	13
<b>3</b>	<b>Results</b>	<b>15</b>
3.1	Trends . . . . .	15
3.2	Readmission risk modeling . . . . .	15
<b>4</b>	<b>Discussion</b>	<b>22</b>
4.1	Trends in CDI . . . . .	22
4.1.1	Increasing infection rates . . . . .	22
4.1.2	Infections at a younger age . . . . .	23

4.2	Trends in renal failure . . . . .	23
4.3	Renal failure as a CDI readmission risk factor . . . . .	23
<b>5</b>	<b>Conclusion and future work</b>	<b>24</b>
5.1	Future work . . . . .	24

# List of Figures

2.1	State diagram for determining what constitutes an index admission and a subsequent readmission. Additional rules include cutting off index events by October, November, or December, depending on whether we are looking for 90, 60, 30 day readmissions, respectively; filtering out deaths on index events; lengths of stay that included transfers and same-day stays; and infants less than 1 year of age, where <i>C. diff</i> is common but shows no symptoms.	14
3.1	CDI trends show a somewhat linearly increasing trend . . . . .	16
3.2	Comparatively, CDI occurrences are not increasing as rapidly as renal diseases . . . . .	16
3.3	Age distribution for ESRD has remained consistent since ICD-9-CM coding standards changed in October, 2005 . . . . .	17
3.4	C . . . . .	17
3.5	All age groups below 75 show a steady increase in CDI infections	18
3.6	Females are almost 1.4 times as likely to contract CDI than males	19
3.7	The median age for inpatient CDI admissions has been sharply declined since 2006 . . . . .	19
3.8	Coefficient estimates for the 30-day readmission model . . . . .	20
3.9	Coefficient estimates for the 60-day readmission model . . . . .	20
3.10	Coefficient estimates for the 90-day readmission model . . . . .	21

# List of Tables

1.1	GFR classifications . . . . .	3
2.1	ICD-9-CM renal failure codes . . . . .	11

# Chapter 1

## Introduction

### 1.1 Overview of *Clostridium difficile*

*Clostridium difficile* Infection - also referred to by its shortened name, *C. diff*, or simply CDI - has been an increasing concern in the last two decades among healthcare providers. The organism itself is a resilient endospore-forming bacterium, resistant to heat, acid, and antibiotics, and can survive on surfaces for up to 5 months, if proper sanitation is not carried out. [2]

In past years, the most common CDI cases occurred in elderly patients, 65 years or older, who were admitted as inpatients in a hospital or nursing home setting, and given antimicrobial therapy. In fact, it has become the most frequent nosocomial (hospital-acquired) disease, surpassing methicillin-resistant *Staphylococcus aureus* (MRSA). [3]

Antimicrobials deplete the healthy gut flora in the intestines which protect against harmful organisms like *C. diff*. [4]

Adding to the complexity of the situation, CDI carriers can remain asymptomatic, making them stealth transporters and allowing the disease to propagate undetected until it is too late.

In 2013, the CDC estimated that around 250,000 Americans contracted *C. diff* in a single year, causing 14,000 deaths. That estimate was later updated to half a million in 2015, causing 15,000 deaths. [5] [6] Another study puts that number even higher, at 29,000 deaths in 2011.

CDI is also costly. The CDC estimates that in 2008, it cost acute healthcare facilities alone more than \$4.8 billion. The mean cost of an incident of CDI was found to be \$11,498 (inflation adjusted to 2008 dollars) and as high as \$15,397 when CDI was hospital acquired. [7]

## 1.2 Overview of Renal Failure

While CDI is a singular diagnosis category, renal failure falls into one of two umbrella categories, acute kidney injury (AKI), and chronic kidney disease (CKD). AKI is further broken into subcategories. Acute tubular necrosis is the most common form of AKI. Other subcategories include renal cortical necrosis, renal medullary necrosis, lesions, and a category for unspecified AKI.

Chronic kidney disease is broken into categories based on stages that are calculated using one of the estimated Glomerular Filtration Rate equations. The equations model kidney health as a function of age, sex, race, and blood creatinine, a waste product that is produced from normal muscle use.

### 1.2.1 The MDRD Equation vs. the CKD-EPI Equation

The MDRD equation encodes sex and race (African American or not) and does not rely on height or weight due to using  $1.73m^2$  surface area, the generally accepted mean human adult body surface area.

$$GFR = 175 \times S_{cr} - 1.154 \times \text{Age}^{-0.203} \times 0.742 \cdot I(F) \times 1.212 \cdot I(AA) \quad (1.1)$$

The CKD-EPI equation makes use of a 2-slope spline. It was shown to outperform the MDRD, with lower bias and increased precision. [8, 9]

$$GFR = 141 \times \min\left(\frac{S_{cr}}{\kappa}, 1\right)^{\alpha} \times \max\left(\frac{S_{cr}}{\kappa}, 1\right)^{-1.209} \times 0.993^{\text{Age}} \times 1.018 \cdot I(F) \times 1.159 \cdot I(AA) \quad (1.2)$$

where:

- F is female sex
- AA is African American race
- I is an indicator function that returns 1 if true, the reciprocal of the preceding term if false (thereby making the preceding term 1)
- $S_{cr}$  is serum creatinine in mg/dL
- $\kappa$  is 0.7 for females and 0.9 for males
- $\alpha$  is -0.329 for females and -0.411 for males



CKD Stage	Description	GFR	Kidney Function	ICD-9-CM
1	Normal function	90+	90-100%	585.1
2	Mild loss	60-89	60-89%	585.2
3	Mild to severe	30-59	30-59%	585.3
4	Severe	15-29	15-29%	585.4
5	Kidney failure	15 or less	15% or less	585.5

Table 1.1: GFR classifications

## 1.2.2 Classifying Chronic Kidney Disease

Once the GFR is calculated, patients can be placed into one of five categories. Table 1.1 shows the GFR rating along with the stage of CKD and the level of kidney function. **585** is used in the ICD-9-CM coding system to indicate CKD. If the level is known, a more specific coding is used. **585.6** is used for end stage renal disease, and **585.9** is used if the level is unspecified.

## 1.2.3 End stage renal disease

End stage renal disease (ESRD) is diagnosed when CKD reaches its most severe point and dialysis or a kidney transplant is needed to stay alive. The most common causes of ESRD are diabetes and high blood pressure. Risk of ESRD also increases with age.

## 1.3 Readmissions and the ACA

A 2014 study done by the Agency for Healthcare Research and Quality (AHRQ) under the Healthcare Cost and Utilization Project (HCUP) found hospital readmissions accounted for about \$41.3 billion in hospital costs. [10]

Under the Readmission Reduction Program, a provision of the Affordable Care Act, Hospitals face penalties on Medicare payments if they exceed certain 30-day readmission standards. While the American Hospital Association strongly opposes the measure, citing a lack of control over the chain of events that can lead to readmission [11, 12], the Affordable Care Act is still the rule of law, and hospitals must seek to reduce readmissions in order to avoid penalties.

For this reason, readmission statistics are an important key metric for hospitals interested in optimizing their operations. Using large surveys, researchers are able to determine trends and end results, but not necessarily causes, of read-

missions. Still, high level trends can point healthcare providers in a direction where they can more efficiently focus their attention.

### 1.3.1 Readmission measures

The Centers for Medicare & Medicaid Services (CMS) sets guidelines that hospitals must follow to avoid penalization on Medicare payments. The CMS measures "excess readmissions" as a ratio of predicted-to-expected readmissions and each hospital's relative performance, based on a 30-day risk standardized measure. All-cause unplanned readmissions to the same or another applicable acute care hospital, occurring within 30 days - for any reason, regardless of principal diagnosis - from the index admission are counted in this measure. Some planned readmissions are not counted. [13]

For fiscal years 2013 to 2018, the following formula is used to calculate the Payment Readjustment Factor (PRF):

$$PRF = 1 - \min\left(0.03, \sum_{dx} \frac{\text{Payment}(dx) \cdot \max((ERR(dx) - 1.0), 0)}{\text{All payments}}\right) \quad (1.3)$$

Where  $dx$  is one of six measure cohorts:

- acute myocardial infarction (AMI)
- heart failure (HF)
- pneumonia
- chronic obstructive pulmonary disease (COPD)
- coronary artery bypass graft (CABG) surgeries
- elective primary total hip and/or total knee arthroplasty (THA/TKA)

ERR is a hospital's performance measure  $dx$ , and payment refers to base operating DRG payments. [14, 15]

## 1.4 Overview of the data

The Agency for Healthcare Research and Quality (AHRQ), under the Department of Health and Human Services (DHHS), sponsors the Healthcare Cost and Utilization Project (HCUP), a collection of databases including the Nationwide Inpatient Sample (NIS) and the Nationwide Readmissions Database (NRD). [16]

### 1.4.1 The Nationwide Inpatient Sample (NIS)

The Agency for Healthcare Research (AHRQ) has been conducting the National (later renamed to "Nationwide") Inpatient Sample since 1988, as part of the Healthcare Cost and Utilization Project (HCUP). It estimates a weighted 35 million hospitalizations per calendar year using around 7-8 million unweighted discharges per year. It is the largest database of its kind in the United States. [17]

### 1.4.2 The Nationwide Readmissions Database (NRD)

Similar to the NIS, the NRD tracks hospitalizations. In addition, it tracks patients across admissions, using an ID key, an admission reference date, and a length of stay for each admission. This allows analysts to track anonymized readmission cases. The NRD tracks around 14 million unweighted patients, when weighted, estimates about 36 million weighted patients across admissions per calendar year. [18]

### 1.4.3 Limitations

Working on such a rich dataset does not come without limits. Prior to obtaining the NIS or NRD, analysts must take the HCUP Data Usage Agreement (DUA). At a high level, HCUP requires that researchers protect individual identities. Cell sizes where  $n \leq 10$  may not be reported. Attempting to identify individual patients or health care providers through vulnerability or penetration testing, or any other means, is prohibited. Publication of any methodology that could identify individuals is prohibited.

Furthermore, HCUP data may only be used for research, not for commercial or competitive purposes. Institutions may not be contacted to verify any of the data within the HCUP datasets either.

For the above reasons, the data may also not be posted online, and anyone wishing to work with or even see the data must take the DUA class and sign the agreement. [19]

In a couple of recent publications, Khera and Krumholz expanded on these base requirements and offered a checklist [20] for analysts to follow as they work with the NIS. In a followup study, they found only 10.5% (95% CI, 4.7%-16.4%) of published research projects based on NIS data followed all of the guidelines. [21]

Listed below are the guidelines from Khera and Krumholz, and notes on how we have conformed to them.

- **Section A: Research Design**

- ✓ Does the study consider that it can only detect disease conditions, procedures, and diagnostic tests in hospital settings?

**Yes, we make no assumptions about events occurring outside of the hospital setting.**

- ✓ Does the study acknowledge that it includes encounters, not individual patients?

**Yes, all of our assumptions are made upon the basis of inpatient discharges and readmissions, not individuals.**

- ✓ Does the study avoid diagnosis/procedure-specific volume assessments for units that are not part of the sampling frame of the NIS, and are therefore not representatively sampled, including

- geographic units, like U.S. states
- healthcare facilities (after 2011)
- individual healthcare providers?

**Yes, we only make assessments at the national level.**

- **Section B: Data Interpretation**

- ✓ Does the study attempt to identify disease conditions or procedures of interest using administrative codes or their combinations that have been previously validated?

**Yes, when checking for renal failure, the comorbidity indicators (`renlfail`, `cm_renlfail`) were used before assessing ICD-9-CM codes.**

- ✓ Does the study limit its assessments to only in-hospital outcomes, rather than those occurring after discharge?

**Yes, the only outcome assessments were readmission status and mortality (died), both of which are in-hospital events.**

- ✓ Does the study distinguish complications from comorbidities or clearly note where it cannot?

**Yes, renal failure comorbidities were distinguished using the the comorbidity indicators (`renlfail`, `cm_renlfail`). CDI while most often a complication, cannot usually be distinguished between complication and comorbidity however.**

- **Section C: Data Analysis**

- ✓ Does the study clearly account for the survey design of the NIS and its components -clustering, stratification, and weighting?

**Yes, the R survey package was used to account for survey design.**

- ✓ Does the study adequately address changes in data structure over time (for trend analyses)?

**Yes, since we are only doing national-level assessments, and we are not using ICD-10-CM codes in the 2015 datasets, we don't need to worry about the changes in the survey design.**

# Chapter 2

## Methods

All analyses were done in R version 3.4.3 (2017-11-30) on an x86\_64-pc-linux-gnu (64-bit) running Ubuntu 16.04.4 LTS. Complex survey designs of the NIS and NRD were accounted for using the **survey** package, version 3.33-2. [22] Data were stored and retrieved in MonetDB using MonetDBLite version 0.5.1. [23]

### 2.1 Data source

Years 2001-2014 of the NIS, as well as years 2010-2014 of the NRD were provided courtesy of Creighton University School of Medicine. Both datasets originate in comma-separated variables (CSV) format.

The NIS covers between 7-8 million unweighted patients per calendar year, resulting in file sizes around 3 GB on average, totaling around 43 GB of raw ASCII text.

```
$ 1 -ha NIS* | awk '{print $5, $9}'  
2.7G NIS2001.csv  
2.9G NIS2002.csv  
3.0G NIS2003.csv  
3.1G NIS2004.csv  
3.1G NIS2005.csv  
3.1G NIS2006.csv  
3.4G NIS2007.csv  
3.4G NIS2008.csv  
3.5G NIS2009.csv  
3.6G NIS2010.csv  
3.7G NIS2011.csv  
2.6G NIS2012.csv  
2.6G NIS2013.csv  
2.8G NIS2014.csv
```

The NRD is normalized into 4 different files, 3 discharge-level files and a hospital-level file.

The **Core** file contains data elements necessary for readmission analysis. The **Severity** files contain data related to the severity of the patients' conditions, including, for our purposes, comorbidity flags. The **DX.PR** file (diagnoses and procedures) file contains ICD-9-CM codes and other fields related to diagnoses and procedures. Finally, the hospital file contains information on the hospital characteristics.

The NRD is a record of around 14 million unweighted admissions per calendar year with identifiers that allow analysts to track readmissions from a particular index event.

```
$ 1 -ha NRD*/*.CSV | awk '{print $5, $9}' | sed -e 's/ /\t/'
5.0G      NRD2010/NRD_2010_Core_V2.CSV
3.4G      NRD2010/NRD_2010_DX_PR_GRPS_V2.CSV
88K       NRD2010/NRD_2010_Hospital_V2.CSV
1.2G      NRD2010/NRD_2010_Severity_V2.CSV
5.1G      NRD2011/NRD_2011_Core_V2.CSV
3.4G      NRD2011/NRD_2011_DX_PR_GRPS_V2.CSV
87K       NRD2011/NRD_2011_Hospital_V2.CSV
1.2G      NRD2011/NRD_2011_Severity_V2.CSV
4.9G      NRD2012/NRD_2012_Core_V2.CSV
3.3G      NRD2012/NRD_2012_DX_PR_GRPS_V2.CSV
82K       NRD2012/NRD_2012_Hospital_V2.CSV
1.1G      NRD2012/NRD_2012_Severity_V2.CSV
5.2G      NRD2013/NRD_2013_Core.CSV
3.4G      NRD2013/NRD_2013_DX_PR_GRPS.CSV
92K       NRD2013/NRD_2013_Hospital.CSV
1.1G      NRD2013/NRD_2013_Severity.CSV
6.6G      NRD2014/NRD_2014_Core.CSV
4.1G      NRD2014/NRD_2014_DX_PR_GRPS.CSV
98K       NRD2014/NRD_2014_Hospital.CSV
1.2G      NRD2014/NRD_2014_Severity.CSV
```

This totals to over 50 GB of raw ASCII text.

## 2.1.1 Persistence

There are numerous ways to handle a dataset of this size. If we know exactly what unique codes we want, we could simply **grep** for them. However, this isn't scalable and we would not be able to calculate population proportions.

For this project, we imported the data into MonetDBLite, an in-process version of MonetDB. We chose MonetDB, because it fits well in the academic space, being open source with strong R integration and good community support. It also fits well in the data warehousing space, being a pioneer in column-store technologies.

Column-store databases partition each column as an array, making data retrieval extremely fast when only a subset of the columns need to be loaded into memory. [24]

### 2.1.2 Diagnosis and procedure codes

The International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) is based on the World Health Organization’s Ninth Revision, International Classification of Diseases (ICD-9). It is the coding standard for diseases and procedures used in the NIS and NRD up to October, 2015, when HCUP upgraded to ICD-10-CM.

Although we had access to the 2015 NRD data, it was not used due to the added complexity of accounting for ICD-10-CM changes, as well as not having 2015 data for NIS, which would have caused inconsistencies in trend analysis.

With the acknowledgement that ICD-9-CM codes are not perfect [25], they are still the best thing we have for longitudinal epidemiological studies on a large scale.

All diagnosis fields, `dx1`, `dx2`, . . . `dx30` were queried for code **00845** (*Intestinal infection due to Clostridium difficile*). The decision not to look exclusively at `dx1`, the principal diagnosis, was deliberate, due to the nature of *C. diff*. Patients rarely contract CDI independently. It is typically contracted while in a hospital setting while being treated for a separate disease.

To complicate matters, ICD-9-CM codings are not an exact science, and are often done based on cost and seriousness of comorbid conditions. Coders must use their judgement for determining a principal diagnosis when comorbid conditions are present. [26] For this reason, we simply queried for the presence of the condition on any diagnosis field.

The same was done for renal failure codes.

For trend analysis, the CDI and renal failure selections were joined. This provided full samples of CDI and renal patients, as well as patients with both.



Table 2.1: ICD-9-CM renal failure codes

Code	Description
584	Acute kidney failure
584.5	Acute kidney failure with lesion of tubular necrosis convert
584.6	Acute kidney failure with lesion of renal cortical necrosis convert
584.7	Acute kidney failure with lesion of renal medullary [papillary] necrosis
584.8	Acute kidney failure with lesion of with other specified pathological lesion in kidney
584.9	Acute kidney failure, unspecified
585	Chronic kidney disease (ckd)
585.1	Chronic kidney disease, Stage I
585.2	Chronic kidney disease, Stage II (mild)
585.3	Chronic kidney disease, Stage III (moderate)
585.4	Chronic kidney disease, Stage IV (severe)
585.5	Chronic kidney disease, Stage V (mild)
585.6	End stage renal disease
585.9	Chronic kidney disease, unspecified
586	Renal failure, unspecified

### 2.1.3 Determining index admissions and readmissions

Unlike the NIS, the NRD allows tracking patients across hospital visits within a given calendar year. The field `nrd_visitlink` provides a key that identifies a single patient across multiple visits. To determine temporality, a length of stay field (`los`) is provided for each visit, as well as a reference date, `nrd_daystoevent`. To ensure anonymity, a randomly selected date is chosen for each patient. `nrd_daystoevent` then references the random date and lists the days from the epoch date. This way, no precise date can be determined, thereby protecting patient privacy, while providing the researcher with the data he or she needs.

The NRD leaves readmission determination up to the analyst. First an index event must be chosen. We selected all cases of CDI (ICD-9-CM code **00845**) and retrieved all unique `nrd_visitlink` identifiers. Then a second query was performed retrieving all events for the `nrd_visitlink` identifiers.

We then grouped the cases first by `nrd_visitlink`, then chronologically. Then, we scanned for the first occurrence of a CDI event (ICD-9-CM code **00845**) and marked it as the index event. All information from the index event was stored in a "patient profile" object. The remaining patient admissions, if any, were scanned. If the event contained a CDI identifier and fell within the given readmission day window (30, 60, and 90 days, separately), the event was considered a readmission and the number of readmissions were stored. If the

patient died on a readmission, that was also stored. If the secondary event fell outside of the readmission window, it was considered another index event, and the process started over.

The following additional rules were applied for determining index events for  $d$ -day readmissions, where  $d = \{30, 60, 90\}$ :

1. For years 2010-2014:  $(1 \leq \text{DMONTH} \leq 12 - \text{ceil}(d/30))$

We needed to cut off index events with enough time to determine if there was a readmission, since only calendar years can be analyzed.

2.  $\text{DIED} \neq 0$

A death on index does not allow for readmission.

3.  $\text{LOS} > 0$

A length of stay equal to zero represents transfers and same-day stays that were combined which represents a more complex type of care. [27]

4.  $\text{AGE} > 0$

About 70% of infants under one year of age carry C. *diff* without showing signs or symptoms. [4]

### 2.1.4 Choosing features

To determine renal failure comorbidities, the `cm_renlfail` flag was first used, and then more specific ICD-9-CM codes were identified. Acute kidney failure, or acute kidney injury (AKI) were grouped by all sub-category codes into a single AKI category. This included codes **584**, **584.5**, **584.6**, **584.7**, **584.8**, and **584.9**.

Chronic kidney disease (CKD) stages were individually analyzed, but unspecified or unknown CKD cases were grouped, (ICD-9-CM codes **585** and **585.9**).

Additionally, we considered hospital characteristics as independent factors, including hospital control (Government, nonfederal; Private, non-profit; Private, invest-own), urban/rural designation (9 categories from smallest to largest), teaching designation, and bedsize.

Hospital urban/rural designations contained 9 categories 1 being the largest and 9 being the smallest. These were reversed in order to have a meaningful effect in the regression.

Sex was also included in the regression.

Patients' age is included in the eGFR formula, and as such, would be a confounding variable, so it was not included in the regression.

## 2.2 Statistical analysis

Descriptive and inferential statistics were done using the NIS and NRD complex survey design, supplying `hospid` as the clusters, `nis_stratum` as the strata, and `discwt` as the weighting. Lonely primary sampling units (PSU) - in our case, hospitals - were excluded using `options(survey.lonely.psu="remove")`. [28]

The primary readmission analysis was done with multivariable logistic regression to determine the effect of the covariates and confounding variables on the likelihood of being readmitted with CDI under the three readmission windows, 30, 60, and 90 days.

The model was fitted independently on each year's data. We could have included the years as independent variables and attempted to fit the entire dataset, but we chose not to. The primary reason being, the dataset is very large, so a single all-year fit was somewhat impractical due to hardware resource limitations. More academically, however, we are able to capture more nuance in modelling individual years, given each year is a separate independent data set, and medical trends do change over time. Fitting 5 years of data all at once, is likely to miss smaller subtleties, such as increases or decreases in a particular coefficient over time.

Variance estimation was done using Jackknife Repeated Replication (JRR), specifically the JK<sub>kn</sub> approach. Replication methods have shown better precision and reduced bias compared to Taylor Series Linearization. [29, 30]

All charts were done with the `ggplot2` package.

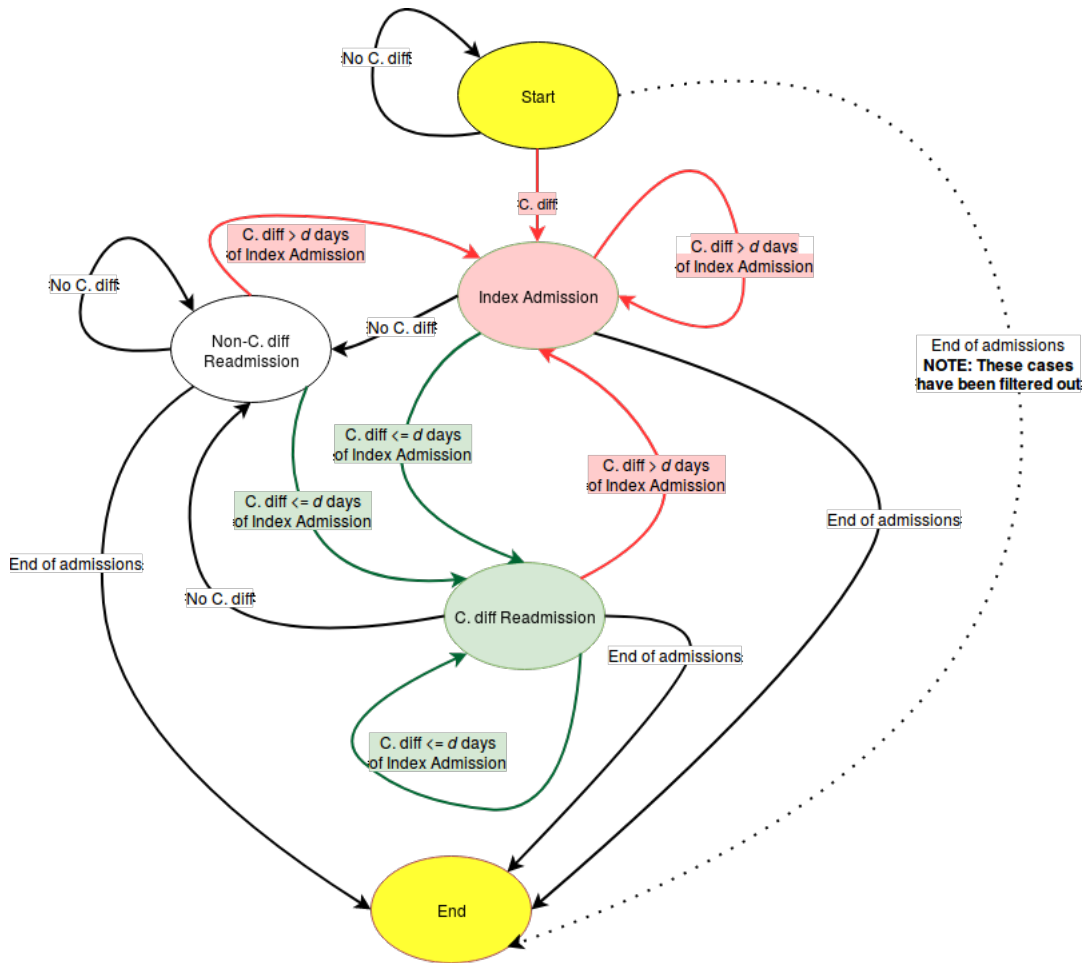


Figure 2.1: State diagram for determining what constitutes an index admission and a subsequent readmission. Additional rules include cutting off index events by October, November, or December, depending on whether we are looking for 90, 60, 30 day readmissions, respectively; filtering out deaths on index events; lengths of stay that included transfers and same-day stays; and infants less than 1 year of age, where *C. diff* is common but shows no symptoms.

# Chapter 3

## Results

### 3.1 Trends

Over the 14 year period, from 2010 - 2014, the proportion of inpatients with CDI increased from 0.003998292% (95% CI, 0.003773284%-0.004236661%) in 2001, roughly 141,540 (95% CI, 133,574-149,978) people, to 0.01023635% (95% CI, 0.01004522%-0.01043108%) in 2014, or 362,367 (95% CI, 355,601-369,260) people - a difference of about 0.006238055%, or 220,827 people.

In Figures 3.1 and 3.4, we can see the median age for inpatient CDI has been lowering over the years.

### 3.2 Readmission risk modeling

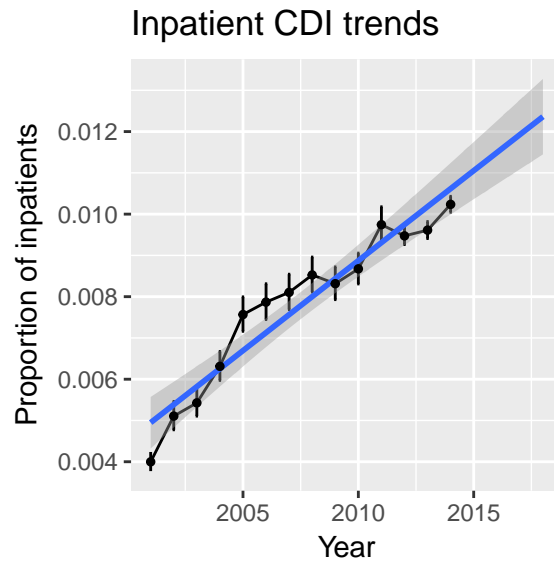


Figure 3.1: CDI trends show a somewhat linearly increasing trend. If we extrapolate to 2018, we have a rough idea of the proportion of the inpatient population we can expect to be diagnosed with CDI.

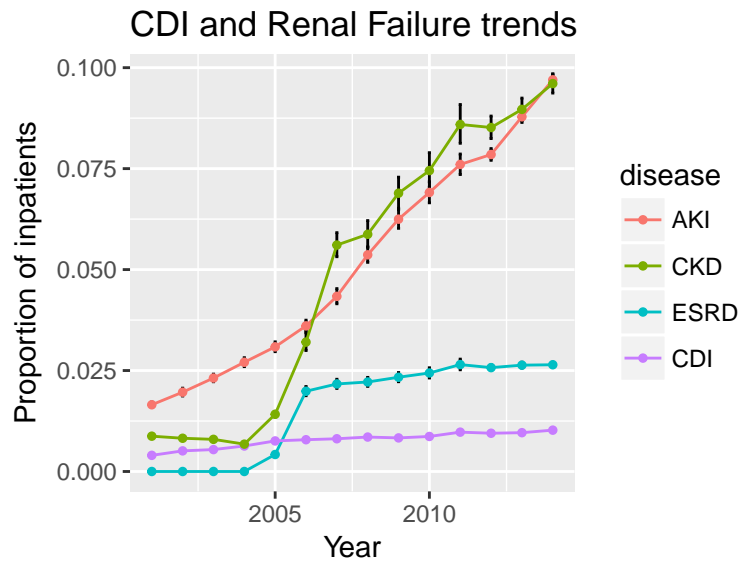


Figure 3.2: Comparitively, CDI occurrences are not increasing as rapidly as renal diseases. The sharp spikes from 2004-2006 were likely due to ICD-9-CM coding requirements implemented in October of 2005.

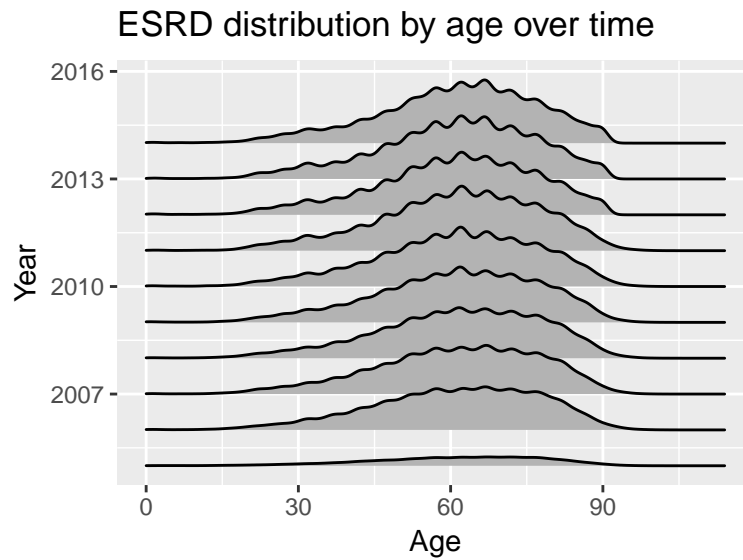


Figure 3.3: Age distribution for ESRD has remained consistent since ICD-9-CM coding standards changed in October, 2005.

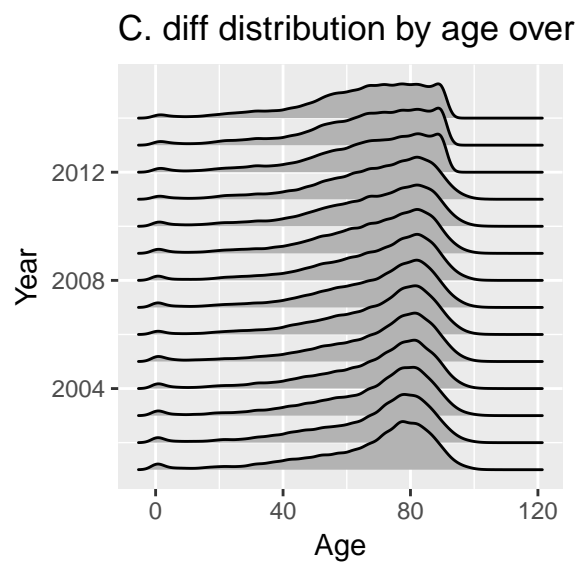


Figure 3.4: C. extitdiff is trending into younger generations in recent years.

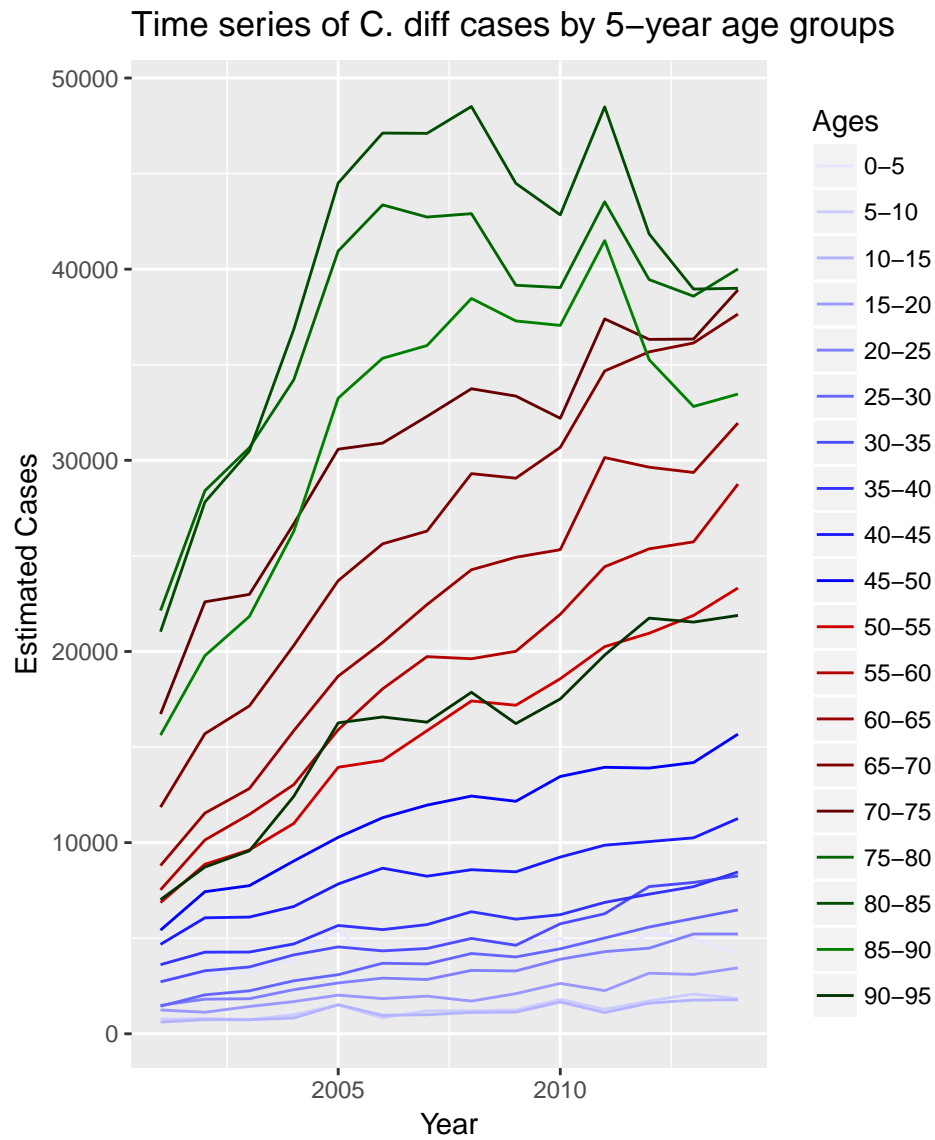


Figure 3.5: All age groups below 75 show a steady increase in CDI infections. 75 and older are down from their peaks in the mid 2000s.



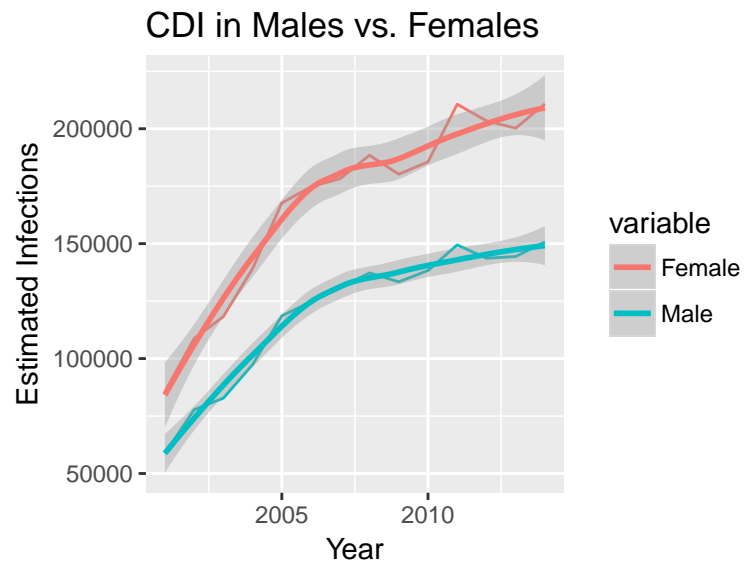


Figure 3.6: Females are almost 1.4 times as likely to contract CDI than males.

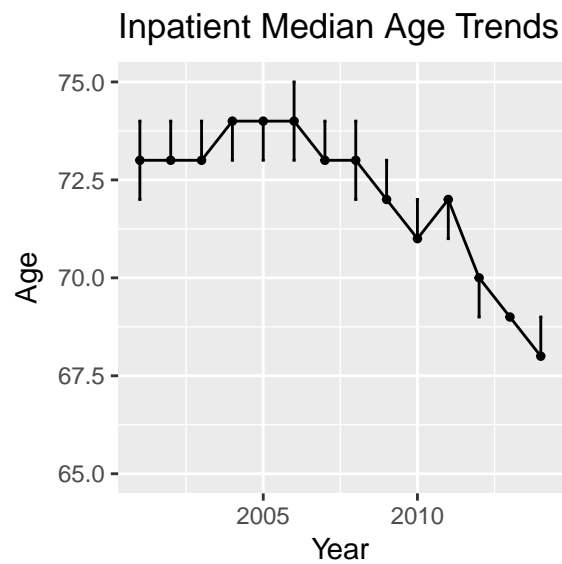


Figure 3.7: The median age for inpatient CDI admissions has been sharply declined since 2006. The vertical bars indicate a 95% confidence interval.

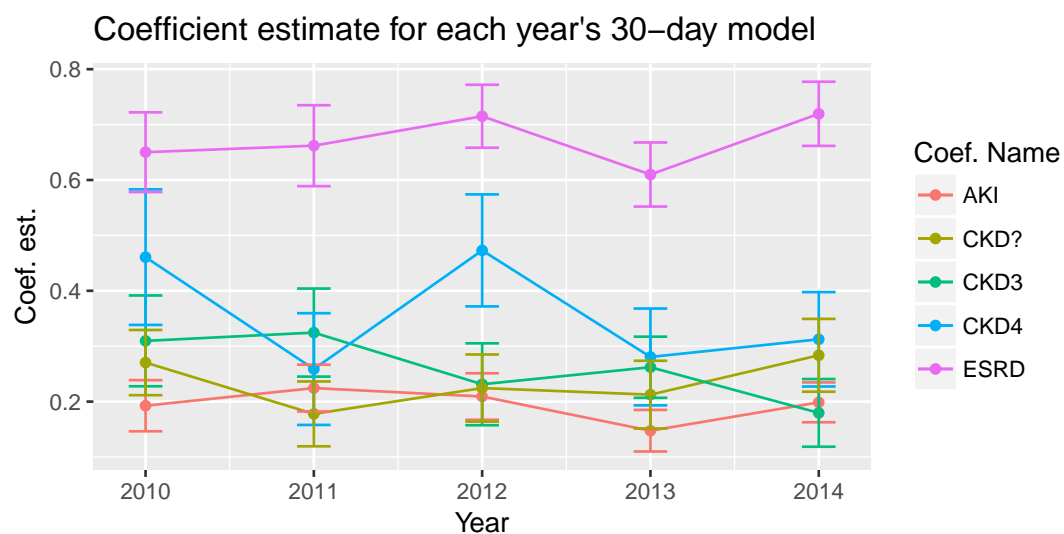


Figure 3.8: Coefficient estimates for the 30-day readmission model. Here, we only show coefficients that were statistically significant across all years. ESRD is highly significant and a strong predictor of readmission.

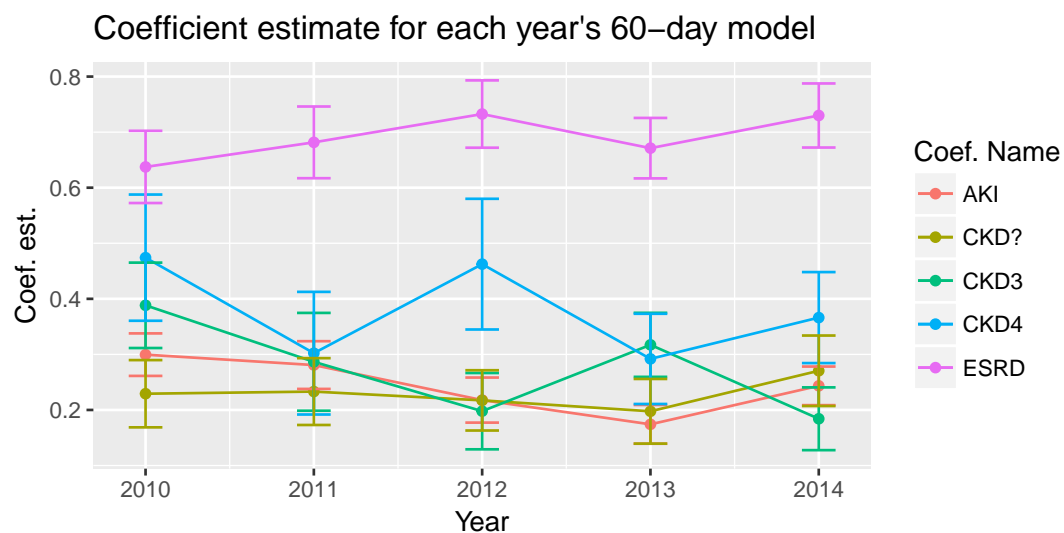


Figure 3.9: Coefficient estimates for the 60-day readmission model. Here, we only show coefficients that were statistically significant across all years. ESRD is highly significant and a strong predictor of readmission.

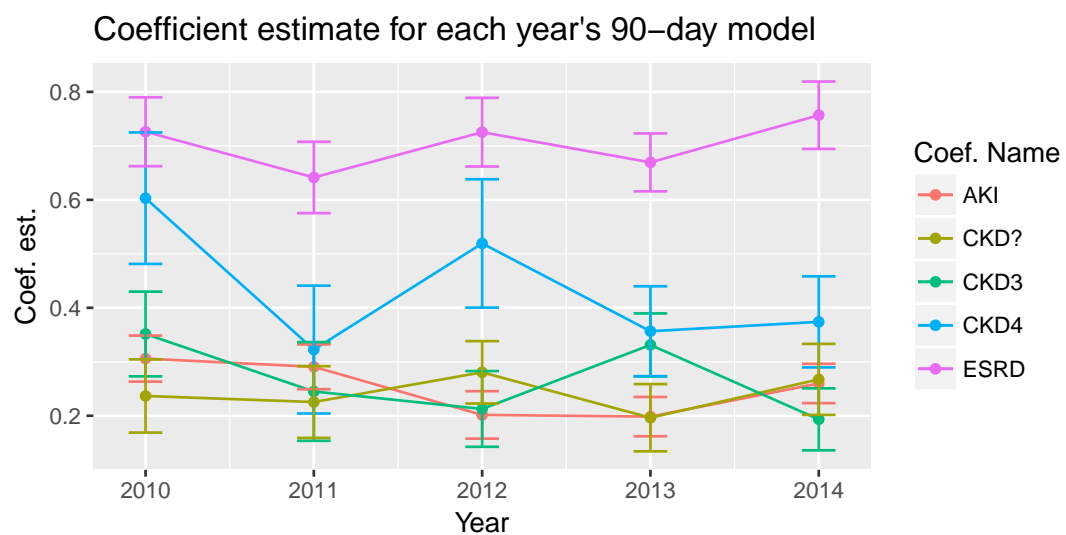


Figure 3.10: Coefficient estimates for the 90-day readmission model. Here, we only show coefficients that were statistically significant across all years. ESRD is highly significant and a strong predictor of readmission.

# Chapter 4

## Discussion

### 4.1 Trends in CDI

#### 4.1.1 Increasing infection rates

*C. diff* has been on the rise since the first reported major outbreak of ribotype 027, a hypervirulent strain, in 2004 [33]. Figure 3.1 shows the trend for CDI using the data from 2001-2014, extrapolated through 2018. Over 1% of the inpatient population in 2014 was diagnosed with CDI. If this trend continues linearly, we can expect that number to increase to nearly 1.2% by 2018.

While the CDC found that around half a million people contracted CDI in 2015, about 150,000 of those were not documented in inpatient records, making their inpatient estimate around 350,000. [5] This is somewhat consistent with our CDI linear model in Figure 3.1, which estimates about 391,293.4 (95% CI, 381,061.4-401,802.9) people in 2015, without having the data.

Using this model, we can expect CDI infections to increase to 437,605.1 (95% CI, 427,984.2-447,380.8) hospital inpatients in 2018, and 468,479.6 (95% CI, 459,266.1-477,766.1) in 2020, if the trend continues.

This proportion pales in comparison to renal failure, however, which has also been on the rise, with nearly 10% of patients coded with some form of acute kidney injury (ICD-9-CM codes **584**, and **584.5-584.9**) or chronic kidney disease (ICD-9-CM codes **585** and **585.1-585.5**, as well as **585.9**) in 2014. AKI has risen by 0.6189396% per year on average.

While renal failure is a broad general category, even more specific codings, including the most serious, End-Stage Renal Disease, showed much higher inpatient rates than CDI. This suggests that CDI is not yet a national epidemic. The trend should continue to be monitored, however, for drastic increases. If

data for future years show the process breaking the previously linear trend, it could be an indication that something has changed, such as a mutation in the bacterium. Conversely, if infections make a downward turn, research should be done to determine what methods have been effective in fighting the disease.

#### **4.1.2 Infections at a younger age**

We showed a trend in CDI moving into younger age groups. This is consistent with findings by Gupta and Khanna [3].

Increased age as a risk factor may be confounded by increased risk of other acquired comorbidities such as renal failure. [31, 32]

Figure 3.1 shows the distribution of *C. diff* patients by age for all data from 2001-2014.

In 2001, the median age for CDI was 73 (95% CI, 72-74). In 2014, it had dropped to 68 (95% CI, 68-69).

### **4.2 Trends in renal failure**

The distribution of ESRD has remained consistent since the coding changes in 2005.

### **4.3 Renal failure as a CDI readmission risk factor**

In their 2015 meta-analysis, Phatharacharukul, et. al. found the pooled relative risk of *C. diff*-associated diarrhea (CDAD) in patients with CKD were 1.95 (95% CI, 1.81-2.10) and 2.63 (95% CI, 2.04-3.38), in ESRD patients. [35] Recurrent CDAD is the primary reason for CDI readmissions.

# Chapter 5

## Conclusion and future work

### 5.1 Future work

Due to time constraints, we were unable to explore some of the larger questions that arose from this study. Is the trend in CDI expanding into younger members of the population due to community-acquired CDI, as shown by Gupta and Khanna? [3] Are antibiotics still being used liberally, causing *C. diff* to spread into a broader population? More work needs to be done to investigate these trends.

On average, females were 1.397 times as likely to contract CDI. We did not investigate the reasons for this. One possible contributing factor was that women giving birth may be more susceptible to CDI, but verification of this line of thinking is left to future work.

Another topic we were unable to cover was mortality. If ESRD patients contracted CDI, what is their mortality rate? Findings at the National Kidney Foundation showed mortality to be at 3.8% among patients with ESRD, compared to 1.46% for CDI patients without ESRD. [36]

These findings should be verified and updated with newer data.

# References

- [1] Robert J. Cunney et al. “Clostridium difficile colitis associated with chronic renal failure”. In: *Nephrology Dialysis Transplantation* 13.11 (1998), pp. 2842–2846. DOI: 10.1093/ndt/13.11.2842. eprint: /oup/backfile/content\_public/journal/ndt/13/11/10.1093\_ndt\_13.11.2842/1/132842.pdf. URL: <http://dx.doi.org/10.1093/ndt/13.11.2842>.
- [2] Dale N. Gerding, Carlene A. Muto, and Jr. Robert C. Owens. “Measures to Control and Prevent Clostridium difficile Infection”. In: *Clinical Infectious Diseases* 46 (Supplement 1 Jan. 2008), S43–S49. DOI: 10.1086/521861.
- [3] Arjun Gupta and Sahil Khanna. “Community-acquired Clostridium difficile infection: an increasing public health threat”. In: *Infection and Drug Resistance* 7 (), 6372. DOI: 10.2147/IDR.S46780.
- [4] J. Thomas Lamont. *C. Diff in 30 Minutes: A Guide to Clostridium Difficile for Patients and Families*. i30 Media Corporation, 2013-2017. First.
- [5] *Antibiotic/Antimicrobial Resistance: Biggest Threats*. [https://www.cdc.gov/drugresistance/biggest\\_threats.html](https://www.cdc.gov/drugresistance/biggest_threats.html). 2018. (Visited on 04/14/2018).
- [6] *Nearly half a million Americans suffered from Clostridium difficile infections in a single year*. <https://www.cdc.gov/media/releases/2015/p0225-clostridium-difficile.html>. Feb. 2015. (Visited on 04/14/2018).
- [7] Erik R. Dubberke and Margaret A. Olsen. “Burden of Clostridium difficile on the Healthcare System”. In: *Clinical Infectious Diseases: An Official Publication of the Infectious Diseases Society of America* 55 (Supplement 2 2012), S88–S92. DOI: 10.1093/cid/cis335.
- [8] Andrew S. Levey et al. “A New Equation to Estimate Glomerular Filtration Rate”. In: *Annals of Internal Medicine* 150 (9), 604612.
- [9] *Estimating Glomerular Filtration Rate*. <https://www.niddk.nih.gov/health-information/communication-programs/nkdep/laboratory-evaluation/glomerular-filtration-rate/estimating>. (Visited on 04/14/2018).
- [10] Anika L. Hines et al. *Statistical Brief 172: Conditions With the Largest Number of Adult Hospital Readmissions by Payer, 2011*. <https://www.hcup-us.ahrq.gov/reports/statbriefs/sb172-Conditions-Readmissions-Payer.pdf>. Apr. 2014. (Visited on 04/14/2018).

- [11] Sabriya Rice. *Most hospitals face 30-day readmissions penalty in fiscal 2016*. <http://www.modernhealthcare.com/article/20150803/NEWS/150809981>. Aug. 2015. (Visited on 04/14/2018).
- [12] *Hospital Readmissions Reduction Program Factsheet*. <https://www.aha.org/system/files/2018-01/fs-readmissions.pdf>. 2018. (Visited on 04/14/2018).
- [13] *Hospital Readmissions Reduction Program (HRRP)*. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/Value-Based-Programs/HRRP/Hospital-Readmission-Reduction-Program.html>. (Visited on 04/14/2018).
- [14] *Hospital Readmissions Reduction Program (HRRP)*. <https://www.cms.gov/medicare/medicare-fee-for-service-payment/acuteinpatientpps/readmissions-reduction-program.html>. (Visited on 04/14/2018).
- [15] Fernanda C. Lessa et al. “Burden of Clostridium difficile infection in the United States”. In: *New England Journal of Medicine* (372 Feb. 2015), pp. 825–834. DOI: 10.1056/NEJMoa1408913.
- [16] Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. *Overview of HCUP*. <https://www.hcup-us.ahrq.gov/overview.jsp>. (Visited on 04/14/2018).
- [17] Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. *Overview of the National (Nationwide) Inpatient Sample (NIS)*. <https://www.hcup-us.ahrq.gov/nisoverview.jsp>. (Visited on 04/14/2018).
- [18] Healthcare Cost and Utilization Project, Agency for Healthcare Research and Quality. *Overview of the Nationwide Inpatient Sample (NRD)*. <https://www.hcup-us.ahrq.gov/nrdoverview.jsp>. (Visited on 04/14/2018).
- [19] *DUA Training - Accessible Version*. [https://www.hcup-us.ahrq.gov/DUA/dua\\_508/DUA508version.jsp](https://www.hcup-us.ahrq.gov/DUA/dua_508/DUA508version.jsp). (Visited on 04/14/2018).
- [20] Rohan Khera and Harlan M. Krumholz. “With Great Power Comes Great Responsibility: Big Data Research From the National Inpatient Sample”. In: *Circulation: Cardiovascular Quality and Outcomes* 10 (2017). DOI: 10.1161/CIRCOUTCOMES.117.003846.
- [21] Rohan Khera et al. “Adherence to Methodological Standards in Research Using the National Inpatient Sample”. In: *JAMA* 318.20 (2017), pp. 2011–2018. DOI: 10.1001/jama.2017.17653.
- [22] Thomas Lumley. *Analysis of Complex Survey Samples*. <https://cran.r-project.org/web/packages/survey/survey.pdf>. Mar. 2018.
- [23] Hannes Muehleisen. *In-Process Version of 'MonetDB'*. <https://cran.r-project.org/web/packages/MonetDBLite/MonetDBLite.pdf>. Jan. 2018.
- [24] *Column store features*. <https://www.monetdb.org/content/column-store-features>.



- [25] Eduardo Uchiyama et al. “Accuracy of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) As a Research Tool for Identification of Patients with Uveitis and Scleritis”. In: *Ophthalmic Epidemiology* 22.2 (2015), pp. 139–141. DOI: 10.3109/09286586.2015.1012274.
- [26] *Accuracy of the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) As a Research Tool for Identification of Patients with Uveitis and Scleritis*. <http://www.hcpro.com/content/271142.pdf>. 2011.
- [27] *Introduction To The HCUP Nationwide Readmissions Database (NRD) 2013*. [https://www.hcup-us.ahrq.gov/db/nation/nrd/NRD\\_Introduction\\_2013.jsp](https://www.hcup-us.ahrq.gov/db/nation/nrd/NRD_Introduction_2013.jsp). Nov. 2015. (Visited on 04/14/2018).
- [28] Thomas Lumley. *Lonely PSUs*. <http://r-survey.r-forge.r-project.org/survey/exmample-lonely.html>. (Visited on 04/14/2018).
- [29] Sadeq Chodhury. “A Comparison of Taylor Linearization and Balanced Repeated Replication Methods for Variance Estimation in Medical Expenditure Panel Survey”. In: (July 2013). (Visited on 04/14/2018).
- [30] Philip J. Smith, K.P. Srinath, and Michael P. Battaglia. “Issues Relating to the Use of Jackknife Methods in the National Immunization Survey”. In: (2000). (Visited on 04/14/2018).
- [31] Greta Krapohl et al. “Preoperative Risk Factors for Postoperative Clostridium difficile Infection in Colectomy Patients”. In: *American Journal of Surgery* 205 (3 Mar. 2013). DOI: 10.1016/j.amjsurg.2012.10.028.
- [32] Aikaterini Masgala, Spiros G. Delis, and Christos Dervenis. “Clostridium Difficile Infection: An Increasing Postsurgical Complication”. In: *Infectious Diseases & Therapy* 2 (6 2014). DOI: 10.4172/2332-0877.1000176.
- [33] Jacques Ppin et al. “Clostridium difficile-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity”. In: *Canadian Medical Association Journal* 171 (5 Aug. 2004), pp. 466–472. DOI: 10.1503/cmaj.1041104.
- [34] Karen C. Carroll and John G. Bartlett. “Biology of Clostridium difficile: Implications for Epidemiology and Diagnosis”. In: *Annual Review of Microbiology* 65 (2011), pp. 501–521. DOI: 10.1146/annurev-micro-090110-102824.
- [35] Parkpoom Phatharacharukul et al. “The Risks of Incident and Recurrent Clostridium difficile-Associated Diarrhea in Chronic Kidney Disease and End-Stage Kidney Disease Patients: A Systematic Review and Meta-Analysis”. In: *Digestive Diseases and Sciences* 60 (2015). DOI: 10.1007/s10620-015-3714-9.
- [36] Ed Susman. *C. Diff Dangerous in ESRD*. <https://www.medpagetoday.com/meetingcoverage/nkf/38255>. Apr. 2013. (Visited on 04/14/2018).