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



## Brian Detweiler

Under the advisory supervision of
Dora Matache, Ph.D., Renuga Vivekanandan, M.D., and Ryan Walters, Ph.D.

College of Arts and Sciences

University of Nebraska, Omaha

A project submitted for

*Master of Science*

May 4, 2017

# Abstract

Clostridium *difficile*(C. *diff*) 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 C. *diff* infected patients with renal failure comorbidities has appeared in the literature as early as 1998 [1]. In this study, we analyze C. *diff* trends from 2001-2014 using the *National Inpatient Sample*. We also assess the risk of 30, 60, and 90 day readmissions in patients with comorbid C. *diff* infection and renal failure conditions using the *Nationwide Readmissions Database* from 2010-2014.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Overview of Clostridium *difficile*

Clostridium *difficile*, or C. *diff* Infection (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 of the intestines which protect against harmful organisms like C. *diff* [4].

Adding to the complexity of the situation, C. *diff* 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 CDI 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.

When evaluating the effect of renal failure on outcome variables, we first need to understand how renal failure is classified.

Chronic kidney disease is broken into categories based on stages that are calculated using one of the estimated Glomerular Filtration Rate (GFR) 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. One of two formulas may be used, the Modification of Diet in Renal Disease (MDRD) formula, or the newer Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) formula [8]. The models are included below to show the roles that age, sex, and race play in the diagnosis of CKD.

### 1.2.1    The MDRD and CKD-EPI Equations

Patients are classified as having a particular CKD stage by testing their Glomerular Filtration Rate (GFR) using one of two equations. Both equations measure the creatinine in a patient's blood with a serum creatinine test. Creatinine is a waste product produced by normal muscle wear and tear.

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.

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

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)

| 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 for stages of Chronic Kidney Disease

- $S_{cr}$ is serum creatinine in mg/dL

[9]

The CKD Epidemiology (CKD-EPI) was a single equation selected out of a large number of candidate equations, that uses transformations of continuous variables and additional variables and interactions. Serum creatinine is modeled as a 2-slope spline with sex-specific knots at 0.7 mg/dL for women and 0.9 mg/dL for men. It was shown to outperform the MDRD, with lower bias and increased precision. [8, 10]

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

where:

- $\kappa$ is 0.7 for females and 0.9 for males

- $\alpha$ is -0.329 for females and -0.411 for males

[10]

## 1.2.2 Classifying Chronic Kidney Disease

Once the GFR is calculated using either the MDRD or the CKD-EPI formulas, 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. **585.1-585.5** indicate CKD Stage 1 through 5. **585.6** is used for end stage renal disease (dialysis or transplant), and **585.9** is used if the level is unspecified.

### 1.2.3  End stage renal disease

When CKD reaches Stage 5, it is considered end stage renal disease (ESRD). 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  Index Admissions and Readmissions

An inpatient admission begins on the first day a patient is admitted to a hospital under a doctor's order. The last day before discharge is the last inpatient day [11].

Readmissions are subsequent admissions from a given *index* admission within a specified time interval. Methods and inclusion/exclusion criteria for determining an index admission and readmissions vary.

### 1.3.1  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 [12].

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 [13, 14], 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 readmissions. Still, high level trends can point healthcare providers in a direction where they can more efficiently focus their attention. Studies like this one focus on narrow cases where a better understanding can contribute to reduced readmissions and an overall reduction in penalties.

### 1.3.2 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 [15].

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

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

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 against $dx$, and payment refers to base operating DRG payments [16, 17].

## 1.4 Overview of the data

The data obtained for this study comes from the Agency for Healthcare Research and Quality (AHRQ), under the Department of Health and Human Services (DHHS). ARHQ sponsors the Healthcare Cost and Utilization Project (HCUP), a collection of databases including the Nationwide Inpatient Sample (NIS) and

the Nationwide Readmissions Database (NRD) [18]. For this study, we obtained years 2001-2014 of the NIS, and 2010-2014 of the NRD.

Both datasets are based on complex survey designs. The Primary Sampling Units (PSUs) are hospitals, stratified by region, teaching status, and bedsize. Weights are calculated for each discharge which are used to "map" the sample back to an unbiased representation of the survey population [19].

### 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 [20].

### 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 [21].

### 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 (groups of people) 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. For instance, if an analyst noticed that a certain hospital used a particular coding scheme that was unique to that hospital and could be used to identify aspects of a patient, the analyst may not use this information. Furthermore, probing for such vulnerabilities is prohibited as well. 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 [22].

In a couple of recent publications, Khera and Krumholz expanded on these base requirements and offered a checklist [23] 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 [24].

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.**

## 1.4.4 Goals

The goals of this study are to evaluate trends in CDI and renal failure over the period of 2001-2014 using the NIS, and to determine risk factors of readmission for CDI patients over the period of 2010-2014 using the NRD.

# 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 [25]. Data were stored and retrieved in MonetDB using MonetDBLite version 0.5.1 [26].

## 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.

```
$ l −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
```
Listing 2.1: Showing file sizes for the NIS dataset from 2001-2014

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.

```
$ l -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
```
Listing 2.2: Showing file sizes for the NRD dataset from 2010-2014

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 [27].

### 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 [28], they are still the best thing we have for longitudinal epidemiological studies on a large scale. The ICD-9-CM diagnosis codes are stored in fields `dx1, dx2, ...` `dx30`. `dx1` is the principal diagnosis, or the diagnosis primarily responsible for the patient being admitted. This is not always clear cut, as patients often have multiple diseases, and any one could be responsible for their admission, so in these cases, they are often coded with consideration to cost and reimbursement.

These 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 CDI. Patients rarely contract CDI independently. It is typically contracted while in a hospital setting while being treated for a separate disease.

| ICD-9-CM 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 |
| 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 (end stage) |
| 585.6 | End stage renal disease |
| 585.9 | Chronic kidney disease, unspecified |
| 586 | Renal failure, unspecified |

Table 2.1: ICD-9-CM renal failure codes

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 [29]. For this reason, we simply queried for the presence of the condition on any diagnosis field.

The same was done for renal failure codes, shown in table 2.1.

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.

### 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 (ICDM-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 state diagram is shown in Figure .

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

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

   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. DIED $\neq 0$

   A death on index does not allow for readmission.

3. 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 [30].

4. $AGE > 0$

   About 70% of infants under one year of age carry C. *diff* without showing signs or symptoms of infection [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** (see Table 2.1).

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")` [31].

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. Logistic regression is used throughout the medical literature to explain effects on a binary outcome variable. At its core is a linear model that is both effective and easy to explain.

The odds are given by

$$Pr(\text{readmitted}) = \frac{e^{\beta \mathbf{X}}}{1 - e^{\beta \mathbf{X}}} \tag{2.1}$$

where $\mathbf{X}$ is a matrix consisting of a constant (slope), and the feature variables shown in Table 2.2. The full list of regression coefficient estimates are provided in Appendix A (Tables A.1, A.2, and A.3).

Note the following:

- `hosp_hcontrl_govt` and `hosp_hcontrl_priv_np` were compared against a baseline of `hosp_hcontrl_priv_invest_own`, which represents a hospital's control/ownership of private, investor owned (proprietary)

14

- `hosp_urcat4` is a discrete variable from 1 to 4, representing the smallest to largest areas. In this case, an increasing integer size corresponds to an increasing metropolitan area size. The original variable used a descending order, where 1 was the largest and 4 was the smallest, so we reversed the order to make the variable naturally meaningful without the need for dummy variables

- `hosp_ur_teach_metro` and `hosp_ur_teach_metro_teaching` were compared against a baseline of `hosp_ur_nonmetro`, which did not distinguish between teaching and non-teaching because rural hospitals were rare

- `hosp_bedsize` is a discrete variable from 1 to 3, indicating small, medium, and large hospital bedsizes

- `female` is a binary variable, 0 is male, 1 is female

- `acute_kidney_failure` grouped all forms of AKI. The baseline is not having any form of AKI

- `chronic_kidney_disease2-6` and `chronic_kidney_disease_unk` are the various stages of CKD, compared against a baseline of CKD Stage 1 (mostly healthy)

- `renal_failure_unspecified` was compared against not having any unspecified renal failure

Age was not included, because it is factored in to the CKD stage classification in the GFR equations, 1.1 and 1.2.

The model was fitted independently on each year's data. Another approach would be to include the years as independent variables and attempt to fit the entire dataset. Because the dataset is very large, a single all-year fit was somewhat impractical due to hardware resource limitations. From a practical aspect, we are able to capture more nuance in modeling 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 particular coefficient estimates over time.

Variance estimation was done using the non-parametric method of Jackknife Repeated Replication (JRR), specifically the JKn approach, which is the default

| NRD variable | Description |
|---|---|
| hosp_hcontrl_govt | Hospital's ownership/control - Government, nonfederal |
| hosp_hcontrl_priv_np | Hospital's ownership/control - Private, not-profit |
| | |
| hosp_urcat4 | Urban-rural categorization (1-4, smallest to largest) |
| hosp_ur_teach_metro | Hospital teaching status - metro, non-teaching |
| hosp_ur_teach_metro_teaching | Hospital teaching status - metro, teaching |
| hosp_bedsize | Hospital bedsize (1-3, small to large) |
| female | Sex, binary, female |
| acute_kidney_failure | AKI, all types |
| chronic_kidney_disease2 | CKD Stage 2 |
| chronic_kidney_disease3 | CKD Stage 3 |
| chronic_kidney_disease4 | CKD Stage 4 |
| chronic_kidney_disease5 | CKD Stage 5 |
| chronic_kidney_disease6 | ESRD |
| chronic_kidney_disease_unk | CKD, unknown |
| renal_failure_unspecified | Renal failure, unspecified |

Table 2.2: ICD-9-CM renal failure codes

in the `survey` package. Details on JKn and other approaches to variance estimation can be found in [19]. Replication methods have shown better precision and reduced bias compared to Taylor Series Linearization [32, 33].

All charts were done with the `ggplot2` package [34], and the coefficient tables in Appendix A were done with the `stargazer` package [**Hlavac2018**].

All code for this project can be found on the author's GitHub account at [35]. A journal has also been kept, documenting the research process, which can be found on the author's website at [36].
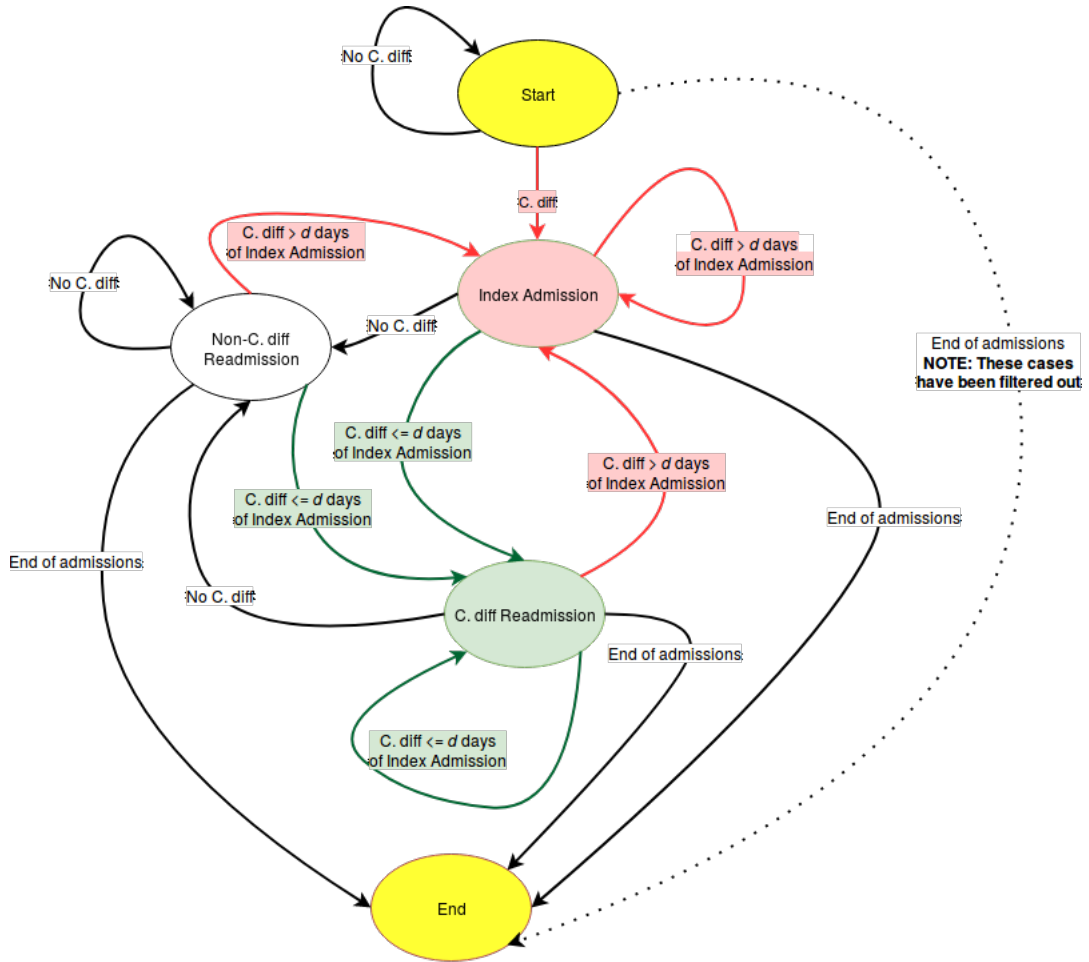
Figure 2.1: State diagram for determining what constitutes an index admission and a subsequent readmission. Additional rules include cutting off index events by October, November, or December, depending on whether we are looking for 90, 60, 30 day readmissions, respectively; filtering out deaths on index events; lengths of stay that included transfers and same-day stays; and infants less than 1 year of age, where C. *diff* bacteria are common but the patient 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) cases, to 0.01023635% (95% CI, 0.01004522%-0.01043108%) in 2014, or 362,367 (95% CI, 355,601-369,260) cases - a difference of about 0.006238055%, or 220,827 people.

In Figures 3.1 and 3.2, we can see the median age for inpatient CDI has been lowering over the years. In fact, the entire distribution of CDI cases by age is becoming less left-skewed and more mesokurtic indicating a trend toward a more normal distribution (Figure 3.3).

Figure 3.4 shows the median age of CDI lowering from 73 years old in the mid-2000s, to about 68 years old in 2014.

We can also separate the trends over time by age groups, broken down into 5-year buckets. Figure 3.7 shows more interesting trends. Notably, nearly every age group is increasing in the number of CDI cases, except the 75-90 year groups, which appear to be on something of a downward trend. The 50-75 year group are showing steady increases and are poised to take the lead on the number of CDI occurrences.

Lastly, we look at the differences in males and females. Figure 3.8 shows that women are, on average, 1.397 times more likely to contract CDI than males.

## 3.2  Readmission risk modeling

Dealing with 12 individual model fittings is cumbersome, particularly when trying to convey results. But by modeling each readmission window and year in-
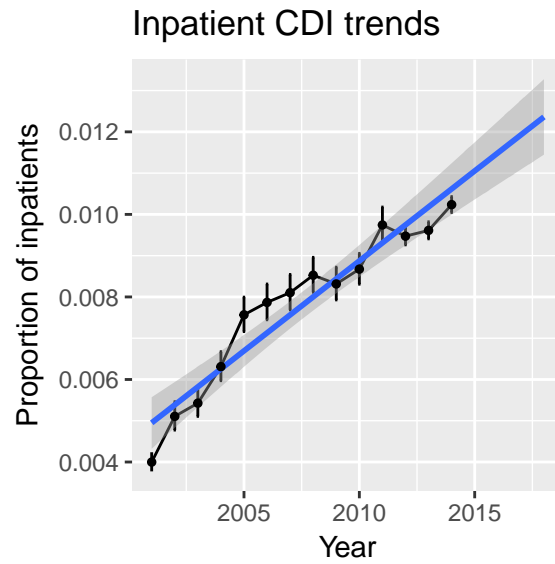
Figure 3.1: CDI trends show a somewhat linearly increasing trend. If we extrapolate to 2018, we have a rough idea of the proportion of the inpatient population we can expect to be diagnosed with CDI.
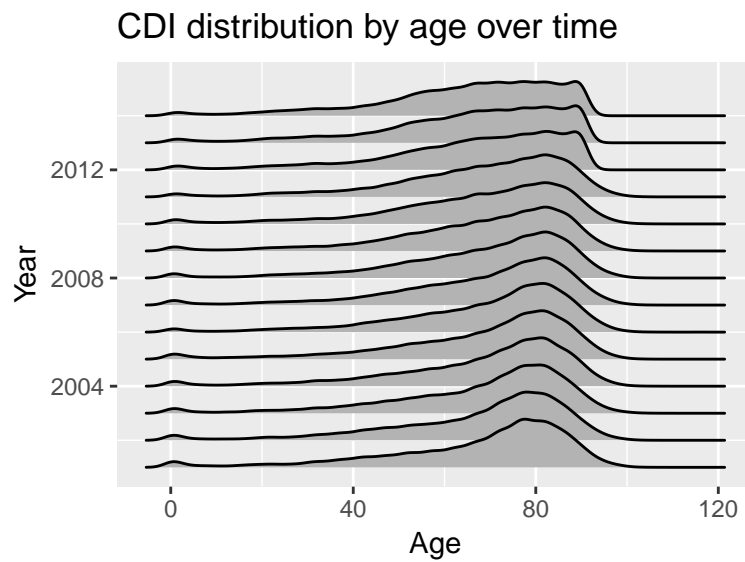


Figure 3.2: CDI is trending into younger generations in recent years.
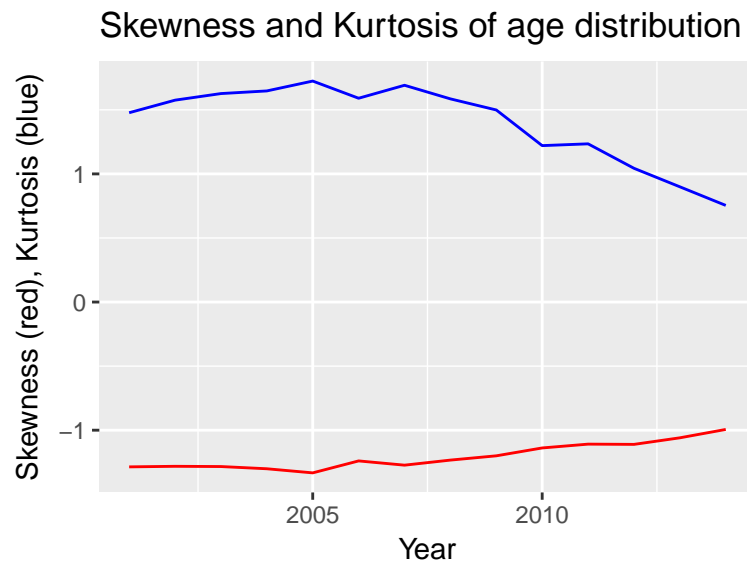
Figure 3.3: The age distrubution of CDI is becoming more platykurtic and less left-skewed, indicating that the CDI is starting to infect younger people in recent years.
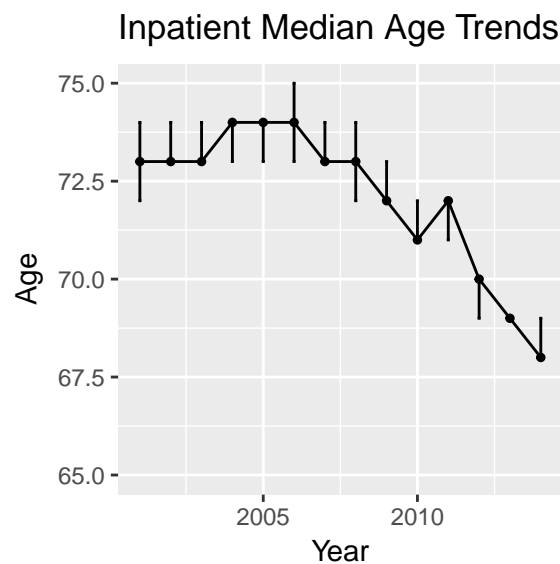


Figure 3.4: The median age for inpatient CDI admissions has been sharply declined since 2006. The vertical bars indicate a 95% confidence interaval.
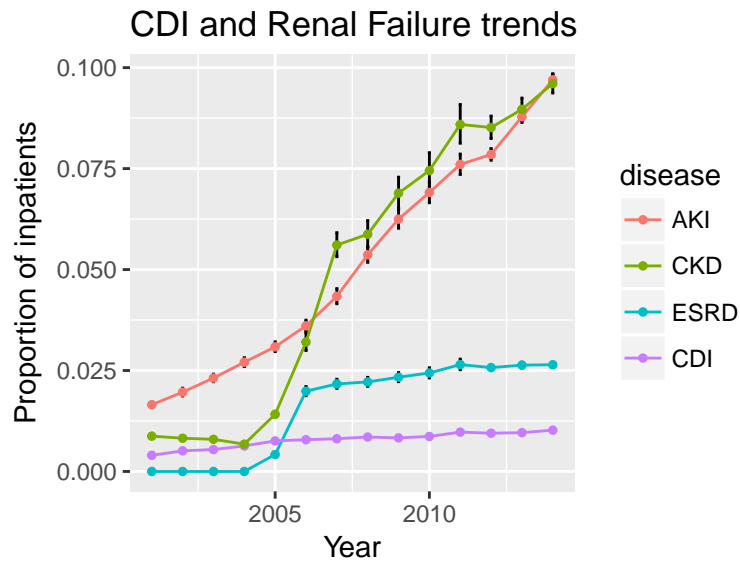
Figure 3.5: Comparitively, CDI occurrences are not increasing as rapidly as renal diseases. The sharp spikes from 2004-2006 were likely due to ICD-9-CM coding requirements implemented in October of 2005.
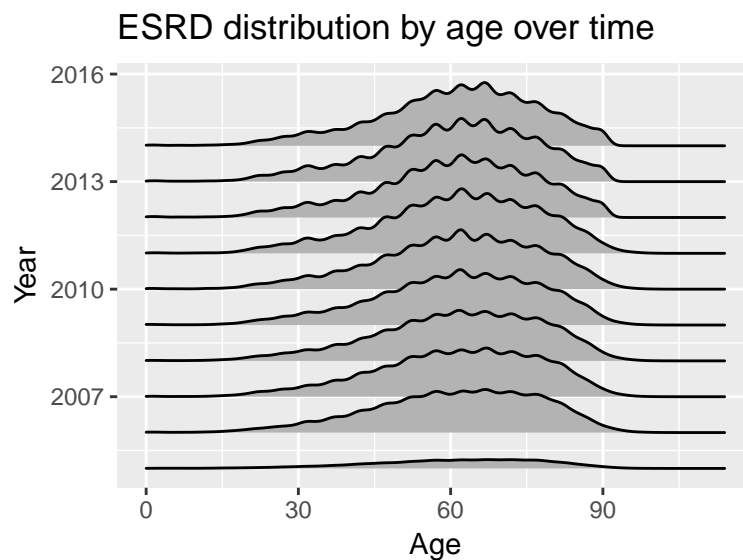


Figure 3.6: Age distribution for ESRD has remained consistent since ICD-9-CM coding standards changed in October, 2005.
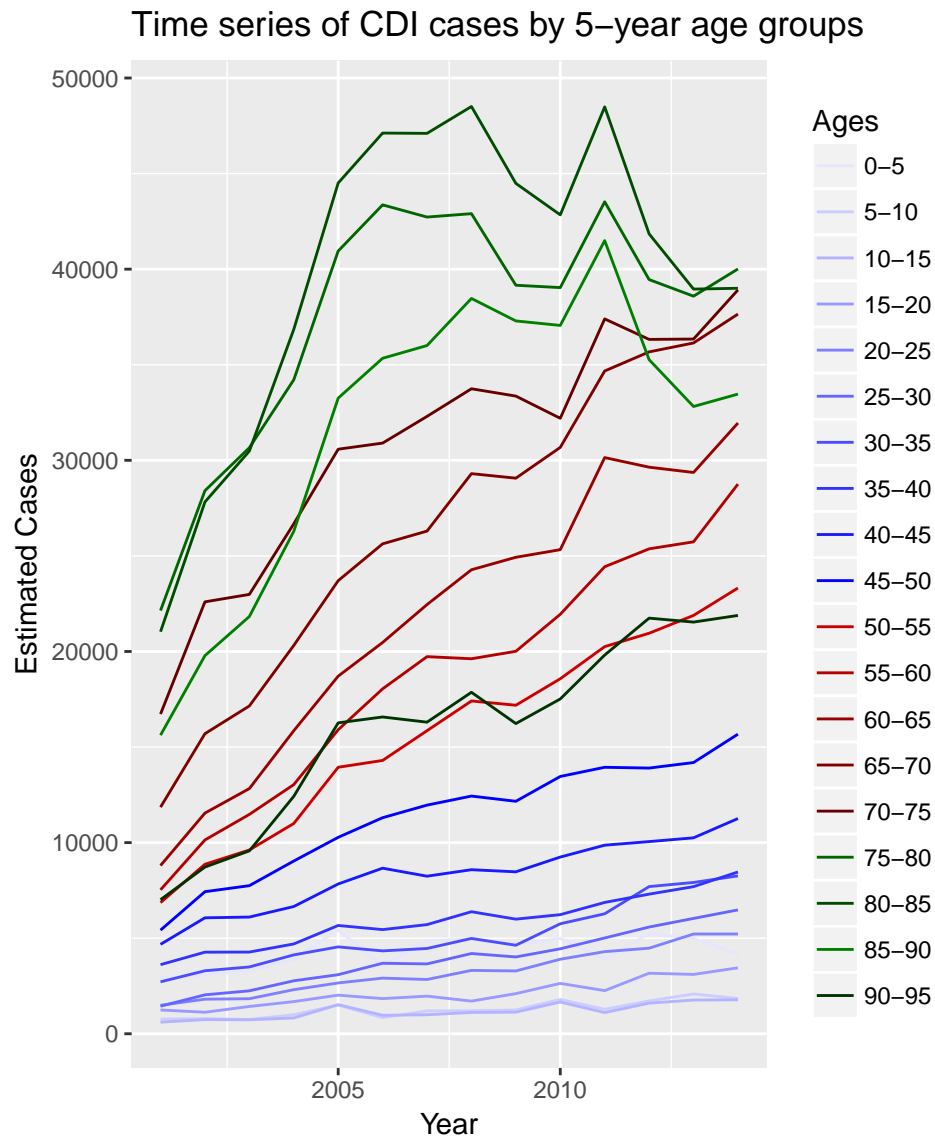
Figure 3.7: All age groups below 75 show a steady increase in CDI infections. 75 and older are down from their peaks in the mid 2000s.
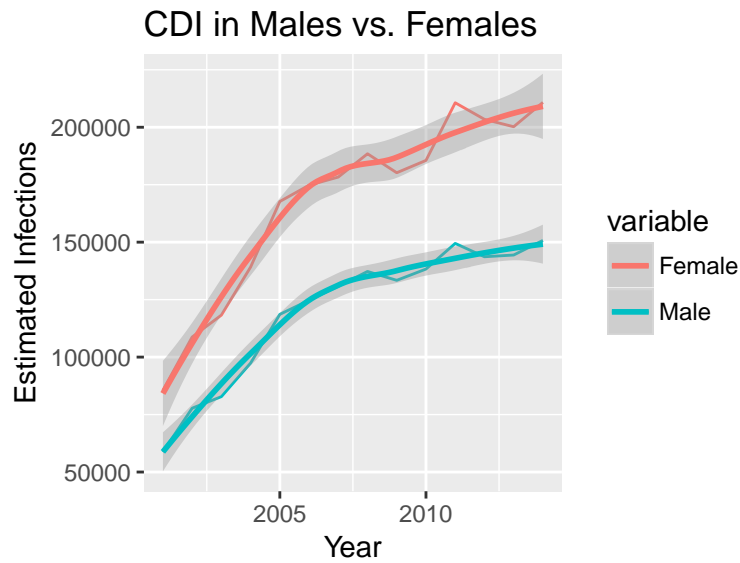
Figure 3.8: Females are almost 1.4 times as likely to contract CDI than males.
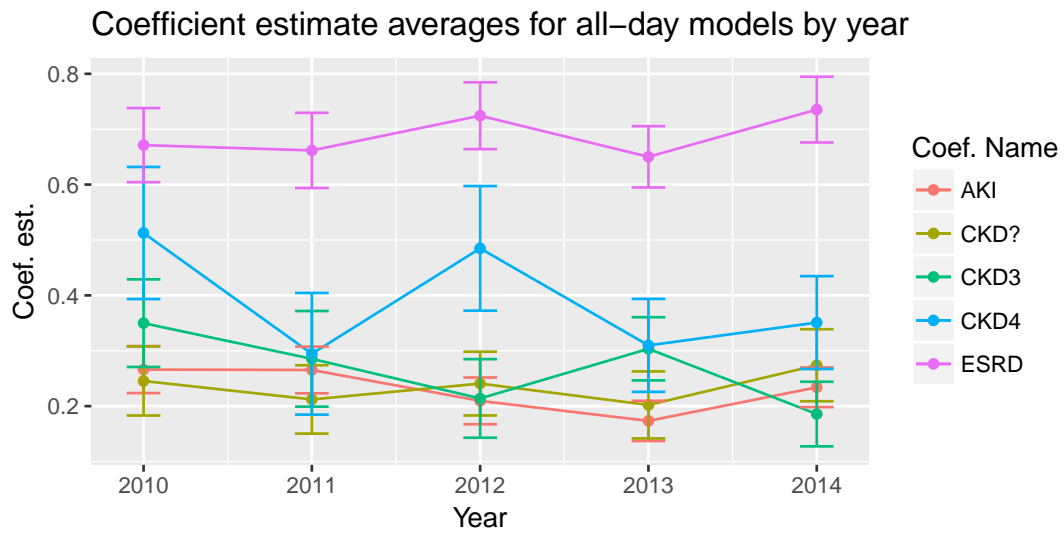


Figure 3.9: Mean coefficient estimates for the 30-, 60-, and 90-day readmission models. Here, we capture the overall effects by averaging the 30-, 60-, and 90-day readmission coefficients to get a more general idea of how the coefficients are behaving across all readmissions. We only show only coefficients that were statistically significant across all years. ESRD is highly significant and a strong predictor of readmission.

dividually, we can capture nuances exclusive to those subsets. For instance, metropolitan and metropolitan teaching hospitals, and hospitals with larger bedsizes showed virtually no effect on readmissions for years 2010-2013. However, in 2014 the coefficients were quite large and highly significant (see Tables A.1, A.2, and A.3). If we had fit across all years, these may have showed up as somewhat significant and with less of an overall effect, and we would have missed the point: Something happened in 2014 causing larger metropolitan and metropolitan teaching hospitals to have a large, significant effect on CDI readmissions.

Another benefit is clarity. We can more easily detect the signal from the noise. Coefficients that show up on some years as statistically significant and not at all on other years could just be noise. But a variable that shows up as significant across all years is probably a signal.

The only coefficients that showed significance across all years at $p < 0.01$, were `AKI`, `CKD3`, `CKD4` (CKD Stages 3 and 4), `CDK?` (CKD unknown), and `ESRD`. Furthermore, `AKI` and `ESRD` showed high significance at $p < 0.001$ across all years. This indicates that there is a strong, consistent effect on readmission for patients with CDI (see Appendix A, figures A.1, A.2, and A.3).

Because the the 30-, 60-, and 90-day readmission model coefficients were all fairly similar, we can average the coefficients and get condensed estimates for a general readmission model (Figure 3.9).

# 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 [37]. 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 [6]. 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 C. *diff* 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.

The proportion of C. *diff* infected individuals pales in comparison to those with 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.

We can also separate the trends over time by age groups, broken down into 5-year buckets. Figure 3.7 shows more interesting trends. Notably, nearly every age group is increasing in the number of CDI cases, except the 75-90 year groups, which appear to be on something of a downward trend. The 50-75 year group are showing steady increases and are poised to take the lead on the number of CDI occurrences.

### 4.1.2   Infections at a younger age

We showed a trend in CDI moving into younger age groups. In 2001, the median age for CDI was 73 (95% CI, 72-74). In 2014, it had dropped to 68 (95% CI, 68-69). The age distribution of CDI becoming less left-skewed and more mesokurtic. Whether it ever reaches the conditions of a normal distribution is doubtful, but the trend is troubling. This means that where CDI once flourished only in the elderly, it is beginning to move into younger demographics. This is consistent with findings by Gupta and Khanna [3], and could partly account for the rise in CDI numbers. If the bacterium can infect younger, healthier individuals, there will be more opportunities for the disease to spread.

When breaking down the distribution into age buckets of 5 year intervals, we see that the previously dominant 75+ age group where CDI once thrived, has been on something of a decline since 2011, while all other age groups continue to rise in CDI instances. The 70-75 age group is poised to become the most frequently infected group, overtaking the 75-85 groups, which had vastly outpaced all other groups in the mid-to-late 2000s.

It is often reported that increased age (60-65 years old or greater) is a risk factor for CDI. There seems to be some disagreement on this in the literature, and speculation on whether the increased risk is confounded by other acquired comorbidities such as renal failure [38, 39].

## 4.2   Trends in renal failure

The distribution of ESRD has remained consistent since the coding changes in 2005. AKI and CKD have skyrocketed since the coding changes, with nearly

20% of the inpatient population having some form of renal failure, this segment represents a huge portion of inpatients.

Figure 3.6 shows the distribution of ESRD cases by age over time. This does appear to be fairly normal and doesn't appear to be changing drastically over time.

## 4.3   CDI readmission risk factors

In their 2015 meta-analysis, Phatharacharukul, et. al. found that the pooled relative risk of C. *diff*-associated diarrhea in patients with CKD and ESRD were 1.95 (95% CI, 1.81-2.10) and 2.63 (95% CI, 2.04-3.38) respectively [40]. This suggests that CKD and ESRD pose a significant risk to recurrent C. *diff*-associated diarrhea, and thus, potential for readmission.

Our findings support this, showing ESRD as a major predictor for CDI readmissions over the surveyed years 2010-2014. AKI was also consistent across years, though with a smaller effect. CKD stages varied as predictors, and only stages 3, 4, and "unknown" were consistent across years. Perhaps the "unknown" cases err on the more sever side, but we have no methods for making that distinction. CKD stage 5 showed some large effects, but was not consistent across years.

Finally, we saw hospital bedsize and metro/metro teaching categorization as having a very large, highly significant effect in the year 2014 only. This suggests something may have changed in either the healthcare system, or the NRD survey in 2014 that contributed to increased readmissions.

# Chapter 5

# Conclusion and future work

## 5.1 Conclusion

CDI is usually, by nature, a complication, acquired most often in a hospital setting while being treated for other illnesses. The increasing proportion of renal failure patients is bound to contribute to the growing CDI population.

The CDI population in general is trending toward younger people.

Various comorbidities, often associated with age, appear to be strong predictors for CDI readmissions. In particular, ESRD has shown to be a consistently strong predictor of CDI readmissions. In addition, AKI and CKD stages 2 and 3, as well as "unknown" CKD stages were consistent predictors for CDI readmission.

## 5.2 Future work

Some larger questions 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 CDI to spread into a broader population? More work should be done to investigate these trends.

While it is frequently reported that increased age is a risk factor for CDI, is it, in fact, due to age? Is it possibly due to other comorbidities that are, in fact, linked to increased age? It is possible that age is merely correlated with increased risk of CDI, not a cause. On its surface, this might not be practical. Increased age is still a good predictor for CDI risk, but if it confounded by latent factors, it would be more valuable to know what those factors are.

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.

Mortality would also be an outcome variable of interest. If ESRD patients contract CDI, what is their mortality rate? Findings at the National Kidney Foundation showed mortality to be at 3.8% among CDI patients with ESRD, compared to 1.46% for CDI patients without ESRD [41]. These findings should be verified and updated with newer data.

We did not have the data to support treatment studies. Drugs such as *Vanco* and *Dificid* have shown to be effective on CDI, but to what extent do each contribute to reduced risk of readmission? Fecal microbiota transplants have also shown to be extremely effective in patients with recurrent C. *diff*-associated diarrhea. However, none of these treatment codes are available in either the NIS or the NRD. Ideally, these would be factored into a future study to determine the effects of treatments on readmission rates.

# Appendix A

# Regression Models

|  | Dependent variable: | | | | |
|---|---|---|---|---|---|
|  | readmitted (30-day readmission) | | | | |
|  | 2010 | 2011 | 2012 | 2013 | 2014 |
| hosp_hcontrl_govt | 0.051 | −0.135 | −0.078 | −0.032 | 0.056 |
|  | (0.079) | (0.095) | (0.074) | (0.066) | (0.068) |
| hosp_hcontrl_priv_np | 0.084 | −0.041 | −0.021 | −0.003 | 0.060 |
|  | (0.072) | (0.052) | (0.053) | (0.049) | (0.045) |
| hosp_urcat4 | 0.084* | 0.101* | 0.109** | 0.105** | −0.005 |
|  | (0.041) | (0.046) | (0.040) | (0.035) | (0.035) |
| hosp_ur_teach_metro | −0.057 | −0.138 | −0.159 | −0.075 | 0.198* |
|  | (0.094) | (0.111) | (0.091) | (0.088) | (0.087) |
| hosp_ur_teach_metro_teaching | 0.025 | −0.040 | −0.062 | 0.022 | 0.347*** |
|  | (0.103) | (0.119) | (0.097) | (0.089) | (0.084) |
| hosp_bedsize | −0.027 | −0.019 | −0.017 | 0.026 | 0.112*** |
|  | (0.029) | (0.029) | (0.026) | (0.024) | (0.025) |
| female | −0.079* | −0.109** | −0.028 | −0.079* | −0.036 |
|  | (0.038) | (0.034) | (0.034) | (0.032) | (0.031) |
| acute_kidney_failure | 0.193*** | 0.224*** | 0.209*** | 0.147*** | 0.199*** |
|  | (0.046) | (0.042) | (0.042) | (0.038) | (0.036) |
| chronic_kidney_disease2 | −0.033 | 0.064 | 0.511** | 0.128 | 0.262 |
|  | (0.224) | (0.200) | (0.167) | (0.158) | (0.148) |
| chronic_kidney_disease3 | 0.310*** | 0.325*** | 0.231** | 0.262*** | 0.180** |
|  | (0.082) | (0.079) | (0.074) | (0.055) | (0.061) |
| chronic_kidney_disease4 | 0.461*** | 0.259* | 0.473*** | 0.281** | 0.312*** |
|  | (0.122) | (0.101) | (0.101) | (0.087) | (0.085) |
| chronic_kidney_disease5 | 0.696** | 0.407 | 0.574* | 0.656** | 0.490 |
|  | (0.219) | (0.260) | (0.284) | (0.225) | (0.299) |
| chronic_kidney_disease6 | 0.650*** | 0.662*** | 0.715*** | 0.610*** | 0.719*** |
|  | (0.072) | (0.073) | (0.057) | (0.058) | (0.058) |
| chronic_kidney_disease_unk | 0.270*** | 0.178** | 0.224*** | 0.213*** | 0.284*** |
|  | (0.059) | (0.059) | (0.061) | (0.061) | (0.066) |
| renal_failure_unspecified | 0.681 | 0.597 | 0.652 | −0.258 | −2.402 |
|  | (0.482) | (0.429) | (0.577) | (0.535) | (8.869) |
| Constant | −1.649*** | −1.516*** | −1.616*** | −1.818*** | −2.114*** |
|  | (0.131) | (0.119) | (0.101) | (0.097) | (0.099) |
| Observations | 35,103 | 38,412 | 38,847 | 39,737 | 40,115 |

| Note: | $+ p < 0.1$; $* p < 0.05$; $** p < 0.01$; $*** p < 0.001$ |
|---|---|

Table A.1: Logistic regression coefficient estimates for 30-day readmissions, 2010-2014. Each year was fit independently. Standard error reported below the estimate in parentheses.

|                               | *Dependent variable:* |
|                               | readmitted (60-day readmission) |
|                               | 2010 | 2011 | 2012 | 2013 | 2014 |
| ----------------------------- | -------- | -------- | -------- | -------- | -------- |
| hosp_hcontrl_govt             | 0.026    | −0.142*  | −0.092   | −0.067   | 0.031    |
|                               | (0.076)  | (0.070)  | (0.073)  | (0.058)  | (0.062)  |
| hosp_hcontrl_priv_np          | 0.106    | −0.020   | −0.009   | 0.008    | 0.042    |
|                               | (0.066)  | (0.056)  | (0.054)  | (0.044)  | (0.042)  |
| hosp_urcat4                   | 0.090*   | 0.108**  | 0.120**  | 0.121*** | −0.025   |
|                               | (0.041)  | (0.042)  | (0.039)  | (0.033)  | (0.033)  |
| hosp_ur_teach_metro           | −0.027   | −0.137   | −0.104   | −0.042   | 0.317*** |
|                               | (0.096)  | (0.099)  | (0.088)  | (0.086)  | (0.084)  |
| hosp_ur_teach_metro_teaching  | 0.060    | −0.059   | −0.032   | 0.023    | 0.426*** |
|                               | (0.106)  | (0.109)  | (0.094)  | (0.088)  | (0.085)  |
| hosp_bedsize                  | −0.015   | 0.014    | 0.012    | 0.042    | 0.097*** |
|                               | (0.030)  | (0.027)  | (0.027)  | (0.022)  | (0.021)  |
| female                        | −0.033   | −0.114** | −0.003   | −0.064*  | −0.031   |
|                               | (0.035)  | (0.035)  | (0.032)  | (0.030)  | (0.028)  |
| acute_kidney_failure          | 0.300*** | 0.281*** | 0.218*** | 0.174*** | 0.243*** |
|                               | (0.038)  | (0.043)  | (0.041)  | (0.035)  | (0.035)  |
| chronic_kidney_disease2       | 0.024    | 0.084    | 0.496**  | 0.080    | 0.252    |
|                               | (0.204)  | (0.186)  | (0.158)  | (0.145)  | (0.149)  |
| chronic_kidney_disease3       | 0.388*** | 0.287**  | 0.198**  | 0.317*** | 0.184**  |
|                               | (0.077)  | (0.088)  | (0.069)  | (0.058)  | (0.056)  |
| chronic_kidney_disease4       | 0.474*** | 0.302**  | 0.462*** | 0.292*** | 0.366*** |
|                               | (0.114)  | (0.110)  | (0.118)  | (0.081)  | (0.082)  |
| chronic_kidney_disease5       | 0.603**  | 0.884*** | 0.507    | 0.611**  | 0.446    |
|                               | (0.224)  | (0.256)  | (0.270)  | (0.221)  | (0.300)  |
| chronic_kidney_disease6       | 0.637*** | 0.682*** | 0.733*** | 0.671*** | 0.730*** |
|                               | (0.065)  | (0.065)  | (0.061)  | (0.054)  | (0.058)  |
| chronic_kidney_disease_unk    | 0.229*** | 0.233*** | 0.217*** | 0.198*** | 0.271*** |
|                               | (0.061)  | (0.060)  | (0.054)  | (0.058)  | (0.063)  |
| renal_failure_unspecified     | 1.112*   | 0.447    | 0.575    | −1.060   | −1.099   |
|                               | (0.467)  | (0.403)  | (0.525)  | (0.618)  | (0.609)  |
| Constant                      | −1.543***| −1.370***| −1.520***| −1.661***| −1.846***|
|                               | (0.134)  | (0.108)  | (0.105)  | (0.089)  | (0.090)  |
| Observations                  | 31,859   | 34,992   | 35,435   | 36,460   | 36,784   |

*Note:* $^{+}$ p<0.1; $^{*}$ p<0.05; $^{**}$ p<0.01; $^{***}$ p<0.001

Table A.2: Logistic regression coefficient estimates for 60-day readmissions, 2010-2014. Each year was fit independently. Standard error reported below the estimate in parentheses.

|  | Dependent variable: | | | | |
|---|---|---|---|---|---|
|  | readmitted | | | | |
|  | (1) | (2) | (3) | (4) | (5) |
| hosp_hcontrl_govt | 0.108 | −0.073 | −0.108 | −0.041 | −0.001 |
|  | (0.083) | (0.068) | (0.071) | (0.059) | (0.062) |
| hosp_hcontrl_priv_np | 0.150* | 0.038 | −0.028 | 0.023 | 0.013 |
|  | (0.071) | (0.051) | (0.058) | (0.049) | (0.042) |
| hosp_urcat4 | 0.083 | 0.086* | 0.103** | 0.121*** | −0.016 |
|  | (0.042) | (0.037) | (0.039) | (0.033) | (0.032) |
| hosp_ur_teach_metro | −0.0003 | −0.065 | −0.079 | −0.029 | 0.297*** |
|  | (0.100) | (0.096) | (0.089) | (0.087) | (0.085) |
| hosp_ur_teach_metro_teaching | 0.088 | 0.015 | 0.015 | 0.042 | 0.396*** |
|  | (0.109) | (0.101) | (0.095) | (0.089) | (0.086) |
| hosp_bedsize | −0.013 | 0.019 | 0.009 | 0.044* | 0.079*** |
|  | (0.030) | (0.025) | (0.027) | (0.022) | (0.021) |
| female | −0.052 | −0.112** | −0.003 | −0.069* | −0.029 |
|  | (0.034) | (0.037) | (0.035) | (0.031) | (0.029) |
| acute_kidney_failure | 0.306*** | 0.291*** | 0.202*** | 0.198*** | 0.260*** |
|  | (0.043) | (0.041) | (0.044) | (0.036) | (0.036) |
| chronic_kidney_disease2 | 0.055 | 0.049 | 0.452** | −0.069 | 0.277 |
|  | (0.210) | (0.184) | (0.164) | (0.155) | (0.155) |
| chronic_kidney_disease3 | 0.351*** | 0.245** | 0.213** | 0.331*** | 0.193*** |
|  | (0.078) | (0.091) | (0.070) | (0.058) | (0.057) |
| chronic_kidney_disease4 | 0.603*** | 0.323** | 0.519*** | 0.357*** | 0.374*** |
|  | (0.122) | (0.118) | (0.119) | (0.083) | (0.084) |
| chronic_kidney_disease5 | 0.691** | 0.761** | 0.457 | 0.620** | 0.472 |
|  | (0.242) | (0.274) | (0.284) | (0.233) | (0.292) |
| chronic_kidney_disease6 | 0.726*** | 0.641*** | 0.725*** | 0.669*** | 0.757*** |
|  | (0.064) | (0.066) | (0.064) | (0.054) | (0.062) |
| chronic_kidney_disease_unk | 0.237*** | 0.225*** | 0.280*** | 0.196** | 0.267*** |
|  | (0.068) | (0.066) | (0.058) | (0.062) | (0.066) |
| renal_failure_unspecified | 1.074* | 0.309 | 0.571 | −0.398 | −1.343 |
|  | (0.463) | (0.413) | (0.573) | (0.619) | (0.775) |
| Constant | −1.545*** | −1.389*** | −1.427*** | −1.641*** | −1.731*** |
|  | (0.138) | (0.100) | (0.107) | (0.092) | (0.087) |
| Observations | 28,694 | 31,473 | 31,958 | 32,915 | 32,974 |
| *Note:* | | | | $^{+}$ p<0.1; $^{*}$ p<0.05; $^{**}$ p<0.01; $^{***}$ p<0.001 | |

Table A.3: Logistic regression coefficient estimates for 90-day readmissions, 2010-2014. Each year was fit independently. Standard error reported below the estimate in parentheses.
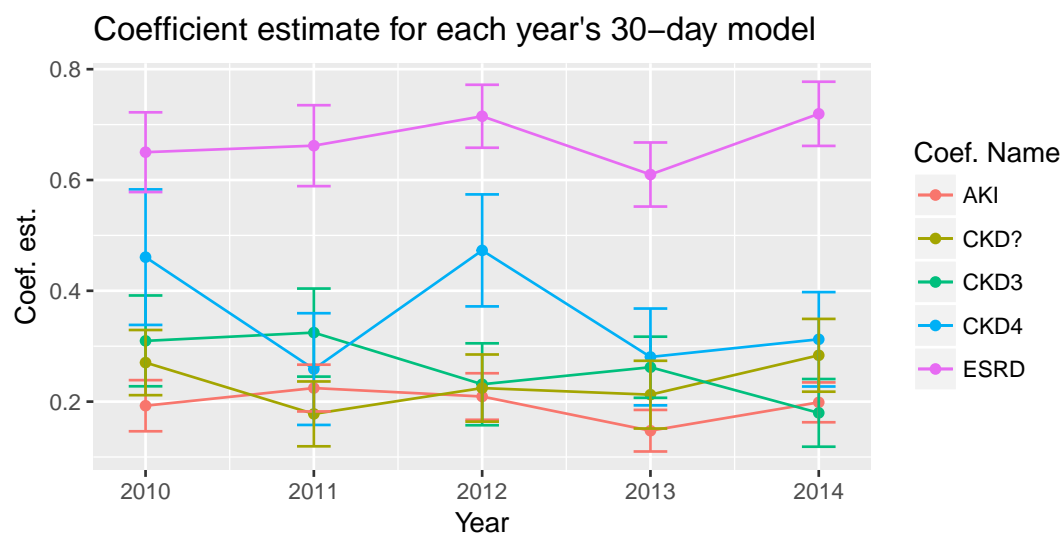
Figure A.1: Coefficient estimates for the 30-day readmission model. Here, we only show coefficients that were statistically significant across all years. ESRD is highly significant and a strong predictor of readmission.
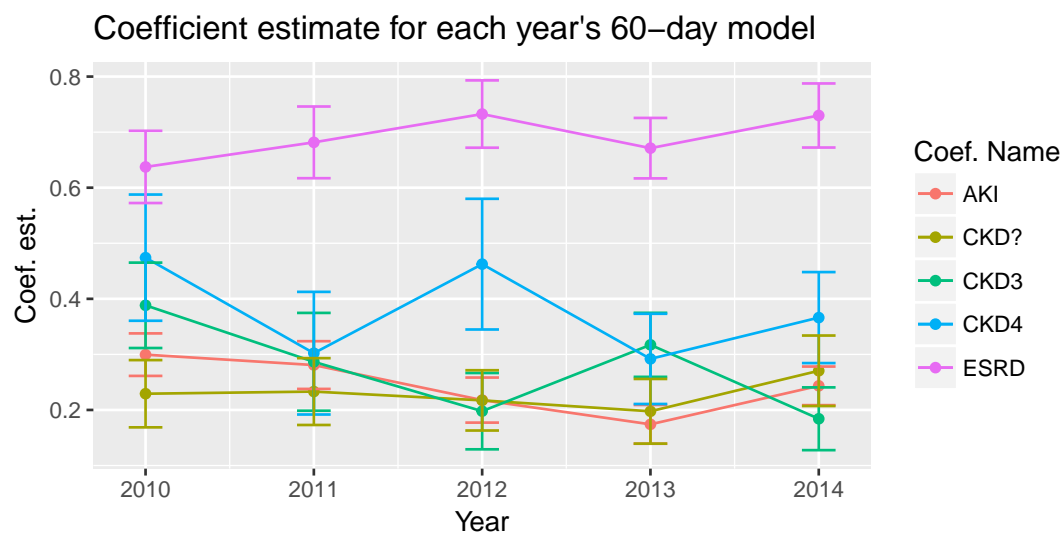


Figure A.2: Coefficient estimates for the 60-day readmission model. Here, we only show coefficients that were statistically significant across all years. ESRD is highly significant and a strong predictor of readmission.
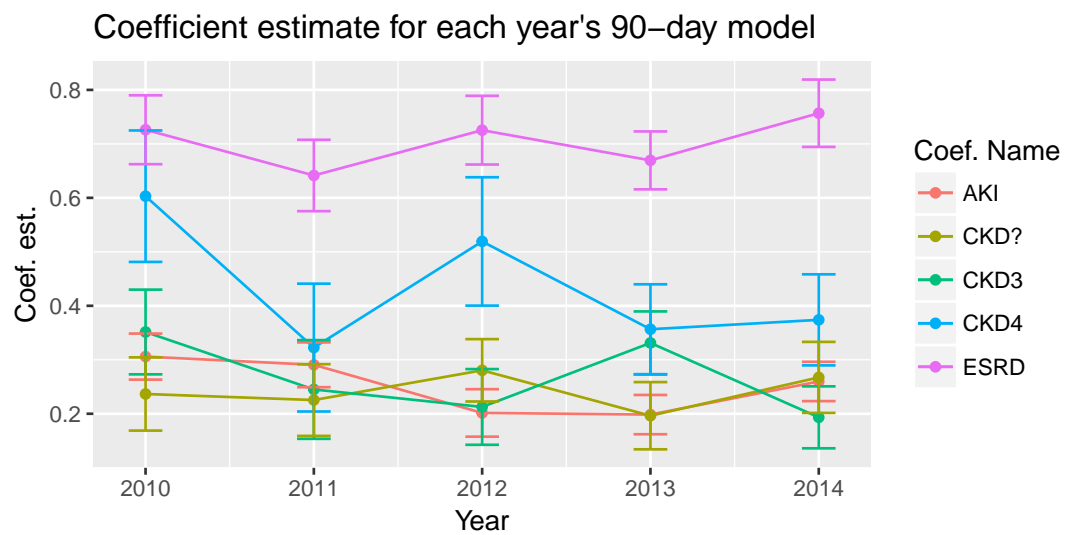
Figure A.3: Coefficient estimates for the 90-day readmission model. Here, we only show coefficients that were statistically significant across all years. ESRD is highly significant and a strong predictor of readmission.

# 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] *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).

[6] *Antibiotic/Antimicrobial Resistance: Biggest Threats.* `https://www.cdc.gov/drugresistance/biggest_threats.html`. 2018. (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] *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).

[9] Andrew S. Levey et al. "A More Accurate Method To Estimate Glomerular Filtration Rate from Serum Creatinine: A New Prediction Equation". In: *Annals of Intern Medicine* 130 (6 1999), pp. 461–470. DOI: `10.7326/0003-4819-130-6-199903160-00002`.

[10] Andrew S. Levey et al. "A New Equation to Estimate Glomerular Filtration Rate". In: *Annals of Internal Medicine* 150 (9), 604612.

[11] *Inpatient or outpatient hospital status affects your costs.* (Visited on 04/14/2018).

[12] 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).

[13] 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).

[14] *Hospital Readmissions Reduction Program Factsheet.* `https://www.aha.org/system/files/2018-01/fs-readmissions.pdf`. 2018. (Visited on 04/14/2018).

[15] *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).

[16] *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).

[17] 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`.

[18] 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).

[19] Steven G. Heeringa, Brady T. West, and Patricia A. Berglund. *Applied Survey Data Analysis, Second Edition.* Taylor and Francis Group, LLC, 2017. Second.

[20] 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).

[21] 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).

[22] *DUA Training - Accessible Version.* `https://www.hcup-us.ahrq.gov/DUA/dua_508/DUA508version.jsp`. (Visited on 04/14/2018).

[23] 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`.

[24] 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.

[25] Thomas Lumley. *Analysis of Complex Survey Samples*. https://cran.r-project.org/web/packages/survey/survey.pdf. Mar. 2018.

[26] Hannes Muehleisen. *In-Process Version of 'MonetDB'*. https://cran.r-project.org/web/packages/MonetDBLite/MonetDBLite.pdf. Jan. 2018.

[27] *Column store features*. https://www.monetdb.org/content/column-store-features.

[28] 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.

[29] *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.

[30] *Introduction To The HCUP Nationwide Readmissions Database (NRD) 2013*. https://www.hcup-us.ahrq.gov/db/nation/nrd/NRD_Introduction_2013.jsp. Nov. 2015. (Visited on 04/14/2018).

[31] Thomas Lumley. *Lonely PSUs*. http://r-survey.r-forge.r-project.org/survey/exmample-lonely.html. (Visited on 04/14/2018).

[32] Sadeq Chowdhury. "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).

[33] 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).

[34] Hadley Wickham and Winston Chang. *Create Elegant Data Visualisations Using the Grammar of Graphics*. https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf. Dec. 2016.

[35] Brian Detweiler. https://github.com/bdetweiler. May 2018.

[36] Brian Detweiler. https://bdetweiler.github.io/projects/projects.html. May 2018.

[37] 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.

[38] 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.

[39] 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.

[40] 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.

[41] Ed Susman. *C. Diff Dangerous in ESRD*. https://www.medpagetoday.com/meetingcoverage/nkf/38255. Apr. 2013. (Visited on 04/14/2018).