Identifying Risk Factors for *Clostridioides difficile* Acquisition from Transmission in Acute-Care Hospitals

by

Deena Ereifej

B.S. Microbiology and Philosophy of Science, University of Pittsburgh, 2022

Submitted to the Graduate Faculty of the

School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Public Health

University of Pittsburgh

2023

UNIVERSITY OF PITTSBURGH

SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Deena Ereifej

It was defended on

December 5, 2023

and approved by

Committee Member: Linda Rose Frank, PhD, MSN, ACRN, FAAN, Professor, Department of Infectious Disease and Microbiology, School of Public Health, School of Medicine, and School of Nursing, University of Pittsburgh

Committee Member: Toan Ha, MD, DrPH, Assistant Professor, Department of Infectious Disease and Microbiology, School of Public Health, University of Pittsburgh

Committee Member: Alexander J. Sundermann, DrPH, MPH, Assistant Professor, School of Medicine, Division of Infectious Disease, Department of Medicine, University of Pittsburgh

Thesis Advisor: Graham Snyder, MD, MS, Associate Professor, School of Medicine, Division of Infectious Disease, Department of Medicine, University of Pittsburgh

Identifying Risk Factors for *Clostridioides difficile* Acquisition from Transmission in Acute-Care Hospitals

Deena Ereifej, MPH

University of Pittsburgh, 2023

Background:

Healthcare-associated infections develop during or soon after receiving healthcare services or being in a healthcare setting. *Clostridioides difficile* is a common and preventable healthcare-associated infection. Whole genome sequencing (WGS) can improve differentiation of *C. difficile* carriage by identifying genetically related isolates acquired from hospital transmission. Risk factors for healthcare-associated *C. difficile* infections (CDIs) have been identified, but no published study has identified epidemiological factors characterizing risk of transmission in healthcare settings. Identifying risk factors for *C. difficile* acquisition from transmission can help guide infection prevention interventions to reduce rates of transmission.

Objectives/Aims:

The goal of this study was to identify risk factors for infection with hospital-acquired *C. difficile* compared to patients with no genetically similar in-hospital patient source.

Method(s) Used/Approach Taken:

Data was collected from the study hospital's electronic health record (EHR) for all patients with healthcare-associated CDIs during the study period. Cases were defined as patients with CDI whose isolate clustered with another genetically similar isolate, excluding index patients. Controls were patients with CDI whose isolates did not cluster or were an index patient. A prediction model was generated for in-hospital *C. difficile* acquisition using EHR data. Elastic net regression (ENR)

was utilized with ten-fold cross validation to select significant risk factors. All pairwise interactions were formulated and tested for association using ENR. A prediction model was determined using the selected risk factors in a multivariate logistic regression analysis.

Results:

Among 809 patients with healthcare-associated *C. difficile*, 114 were excluded from the analysis and the study cohort contained 84 cases and 611 controls. The prediction model identified risk factors (transplant procedure, length of stay, antibiotic receipt and a virulence factor gene) and protective factors (autoimmune disorder, ICU admission and virulence factor genes). The fraction of variance explained by this model for predicting *C. difficile* acquisition was 28.5%. Eight variables predicted $\geq$90% of the model variance.

Summary/Conclusions:

Risk factors for *C. difficile* transmission can help guide infection preventionists in mitigating transmission of *C. difficile* by implementing targeted interventions and protocols. These would include pre-emptive contact precautions when handling high-risk patients, enhanced environmental cleaning or asymptomatic *C. difficile* screening.

Table of Contents

Preface x

Abbreviations xi

1.0 Introduction 1

2.0 Background 3

2.1 Healthcare-Associated *Clostridiodes difficile* 3

2.2 Transmission of *Clostridiodes difficile* in Hospitals 3

2.3 Current *Clostridiodes difficile* Diagnostic Stewardship Methods 4

2.4 Known Risk Factors for Healthcare-Associated *Clostridiodes difficile* Infections ... 6

2.4.1 Patient Characteristics and Clinical History 6

2.4.2 Virulence Factor Genes 7

2.5 Utilizing Whole Genome Sequencing to Detect Outbreaks 10

3.0 Methods and Materials 11

3.1 Study Site Characteristics 11

3.2 Selection of Study Population 11

3.2.1 Patient Exclusion, Case Definitions and Final Cohort 12

3.3 Variables of Interest 13

3.4 *Clostridiodes difficile* Testing 18

3.4.1 PCR Testing 18

3.4.2 Two-Step Testing 19

3.5 EDS-HAT Methods 20

3.5.1 Sample Collection and Isolation 23

3.5.2 Whole Genome Sequencing, Library Preparation and Sequencing.....23

3.5.3 Bioinformatics Analysis.....24

3.6 Statistical Methods 24

3.6.1 Creating the Analysis Data Set 25

3.6.2 Main Effects.....26

3.6.3 Interactions 27

3.6.4 Final Model.....27

4.0 Results..... 28

4.1 Study Population 28

4.2 Main Effects and Interactions 30

4.3 Prediction Model 30

4.4 Risk Factors for *Clostridiodes difficile* Acquisition from Transmission 31

5.0 Discussion..... 33

5.1 Suggested Interventions 33

5.1.1 Suggested Interventions to Prevent Acquisition from Transmission 33

5.1.2 Interventions to Prevent Progression of Carriage to Transmission.....34

5.2 Public Health Significance 35

5.3 Limitations 36

5.4 Future Directions..... 37

6.0 Conclusion 38

Bibliography 39

List of Tables

Table 1: Virulence Factor Gene Descriptions and Risks 8

Table 2: Variables of Interest 15

Table 3: Elixhauser Comorbidities..... 16

Table 4: Expanded List of Variables..... 17

Table 5: Interpreting Results of the Two-Step *Clostridiodes difficile* Test 19

Table 6: R Packages Used in Statistical Analysis..... 25

Table 7: Characteristics of Study Cohort..... 29

Table 8: Prediction Model for *Clostridiodes difficile* Acquisition from Transmission 31

List of Figures

Figure 1: Study Cohort Flowchart 12

Figure 2: *Clostridiodes* difficile Acquisition in Healthcare Settings 13

Figure 3: Flow Chart Describing How to Report Results in the EHR..... 20

Figure 4: EDS-HAT Protocol..... 22

Figure 5: Risk Factors for Acquiring Clostridiodes difficile from Transmission..... 32

Preface

Approval for this study was granted by the Quality Review Committee (Project ID: 4524). This study does not include patient identifiers or release confidential health information. Thank you to the Clinical Analytics team for providing EHR data through the study.

I would like to sincerely thank my thesis committee, Dr. Graham Snyder, Dr. Linda Frank, Dr. Alexander Sundermann and Dr. Toan Ha, for their support throughout this study. Their guidance and wisdom have proven to be invaluable, and this study could not have been successful without them. I want to also extend a special thank you to Dr. Snyder for mentoring me throughout the study design process and providing so much insight to the medical aspect of the world of infection prevention.

Additionally, I would like to thank the MiGEL team for their support and feedback during this process. Specifically, I would like to thank Dr. Lora Pless for her mentorship throughout this study. I am very grateful for the opportunity to have expanded my knowledge of statistical and coding methods under her guidance.

Lastly, I would like to thank the department of Infectious Disease and Microbiology at Pitt Public Health and the department of Infection Control and Hospital Epidemiology for allowing me to pursue my passion for this field.

The completion of this study would not have been possible without the support of these people. I am truly thankful for each person that has helped me through this process.

Abbreviations

C. difficile – *Clostridiodes difficile*

MiGEL – Microbial Genomics Epidemiology Laboratory

CDI – *Clostridiodes difficile* infection

HAI – Healthcare-associated infection

PPE – Personal protective equipment

WGS – Whole genome sequencing

IP – Infection preventionist

CDC – Center for Disease Control and Prevention

NHSN – National Healthcare Safety Network

HA-CDI – Healthcare-associated *C. difficile* infection

VF – Virulence factor

EDS-HAT – The Enhanced Detection System for Healthcare-Associated Transmission

ML – Machine learning

EHR – Electronic Health Record

PCR – Polymerase chain reaction

GDH – Glutamate Dehydrogenase

EIA – Enzyme Immunoassay

CCMA-TAL – Cycloserine Cefoxitin Mannitol Agar with Taurocholate and Lysozyme

dsDNA – Double-stranded DNA

SNP – Single nucleotide polymorphism

ENRR – Elastic net regression

SD – Standard deviation

CI – Confidence interval

SE – Standard error

1.0 Introduction

Clostridiodes difficile (CD), formerly *Clostridium difficile*, is the most common cause of healthcare-associated infections (HAIs) that trigger diarrhea and colitis in the United States [1]. *C. difficile* infections (CDIs) can be caused by acquisition from transmission or triggered by a disruption in the gut microbiome [1]. Within hospitals, resource-intensive measures, such as, environmental cleaning/decontamination and personal protective equipment (PPE), are utilized to prevent transmission while antimicrobial stewardship works to prevent the onset of disease. Diagnostic stewardship is also considered resource-intensive, but rather than preventing the onset of disease, it limits the diagnosis of *C. difficile*-associated colitis when asymptomatic carriage is present.

Whole genome sequencing (WGS) surveillance paired with traditional epidemiological methods has been effective in identifying and confirming transmission in hospitals [2]. Data from WGS has offered insight into the true burden of transmission in hospitals. This data allows infection preventionists (IPs) to identify and investigate transmission events and develop interventions based on transmission routes.

Several risk factors for the onset of healthcare-associated CDIs have been previously identified, such as older age, increased length of stay and broad-spectrum antibiotic use [12], but it is unknown if these risk factors are associated with *C. difficile* transmission in healthcare settings. At the time of this study, no previous publication has identified epidemiological factors associated with risk of *C. difficile* transmission in acute-care hospitals. If risk factors of *C. difficile* transmission can be identified, IPs can develop targeted interventions to mitigate the spread of *C. difficile*. These interventions would include pre-emptive contact precautions, enhanced hand

hygiene compliance/environmental cleaning, and asymptomatic surveillance to identify virulence factors associated with higher risk of transmission. This study aims to investigate the following hypothesis:

I. Among patients with healthcare-associated *C. difficile*, characteristics associated with in-hospital transmission can be identified.

2.0 Background

2.1 Healthcare-Associated *Clostridiodes difficile*

Studies show over 500,000 infections and nearly 29,000 CDI-attributed deaths occur annually within hospitals in the United States [1]. *C. difficile* is a gram-positive, endospore and toxin-producing, obligate anaerobe that primarily infects patients when the natural gut microbiota has been disturbed. Disruption of the gut microbiota can result from antibiotic use, other infections, or comorbidities [1]. *C. difficile* colonizes the gastrointestinal tract of almost 3% of healthy adults [3]. This percentage is higher in select subpopulations such as residents in long-term care facilities.

Clinical symptoms of *C. difficile*-associated colitis include abdominal cramps, fever and an elevated white blood cell count. Additionally, the formation of a pseudomembrane within the colon, causing inflammation, can be observed endoscopically [4]. Severe and sudden *C. difficile* infections can cause the colon to get larger, this is known as toxic megacolon [4].

C. difficile produces two toxins, Toxin A (virulence factor gene TcdA) and Toxin B (virulence factor gene TcdB), that elicit an inflammatory response. This inhibits regulatory cell functions causing loss of the intestinal barrier function [5].

2.2 Transmission of *Clostridiodes difficile* in Hospitals

Transmission primarily occurs through the fecal-oral route as dormant *C. difficile* spores can resist degradation in unhabitable environments and sporulate in ideal conditions [6]. Spores

are taken up by susceptible hosts, germinate within the intestinal tract and adhere to the epithelial layer to infiltrate surrounding mucous layers [7].

The ability of spores to remain on surfaces for an extended period of time has contributed to the virulence of *C. difficile* [7]. Environmental contamination is a common source of spores in a healthcare setting. Spores can live on a variety of high-touch surfaces throughout hospitals. These surfaces include restrooms, bedside tables, doorknobs, bed rails and commonly used equipment [7]. Patients who are colonized with *C. difficile* can shed spores from their stool; however, those with an active CDIs shed much higher levels of spores into the environment [7].

Spores are resistant to many cleaning products, including hand sanitizers, which causes high environmental contamination if proper disinfection measures are not taken [7]. Transmission via healthcare personnel hands occurs because of spore resistance to hand sanitizers [9]. Chlorine-releasing cleaning products are most effective in killing *C. difficile* spores [8]. Hand hygiene with soap and water and proper environmental cleaning are among the most important factors in preventing the spread of *C. difficile* in hospitals [9]. However, antibiotic stewardship remains the most effective method in preventing the onset of disease and subsequent spread of *C. difficile* in hospitals [47].

2.3 Current *Clostridiodes difficile* Diagnostic Stewardship Methods

At the study hospital, steps have been taken to reduce the number of clinically non-indicated CDI tests ordered. These tests result in positive CDIs reported to the CDC's National Healthcare Safety Network (NHSN). Patients who test positive for *C. difficile* on hospital day 3 or later, regardless of the clinical assessment, are reported to the NHSN, thus increasing the number

of inappropriate diagnoses and overall rate of CDIs [10]. These patients are clinically unlikely to have colitis and are more likely to be asymptomatic carriers of *C. difficile*. These instances are inaccurately reported as an HAI to the NHSN. In these scenarios, diagnostic stewardship is needed. This would limit testing for CDIs when a positive test represents asymptomatic carriage rather than infection, thus reducing the reporting of HAIs and clinically non-indicated treatment. Additionally, treating clinically non-indicated CDIs in patients poses a risk of disrupting the patient's gut microbiome, increasing their risk for developing a CDI or development of multidrug resistant organisms through inappropriate antibiotic use [11].

At the study hospital, an intervention was developed to reduce the quantity of clinically non-indicated *C. difficile* testing [10]. An IP was tasked to review *C. difficile* orders prior to specimen processing and processing would not occur without approval from an IP. To obtain approval for a *C. difficile* test to be processed, an IP conducted a chart review of the patient and, if the chart review was inconclusive, reviewed the clinical status of the patient with the bedside nurse to determine if the order met the criteria for *C. difficile* testing [10]. If the order criteria were not met, the IP would consult the ordering physician and provide education on diagnostic stewardship [10]. Orders were then approved if the physician felt strongly about testing for *C. difficile*, or if the physician could not be reached, otherwise orders were cancelled if they did not meet criteria [10].

Throughout the trial period of this intervention, the study hospital saw an 8%-17% decrease in *C. difficile* HAIs and overall use of oral vancomycin decreased as well [10]. This trial period also saw a successful decrease in clinically non-indicated orders [10]. It is important to note that this study period began after the trial period of this intervention. As a result, there may have been clinically non-indicated CDIs included in this study.

2.4 Known Risk Factors for Healthcare-Associated *Clostridiodes difficile* Infections

Identifying risk factors of healthcare-associated CDI (HA-CDI) acquisition were used to guide the selection of variables of interest used in the prediction model to determine risk factors for transmission. For the purpose of this study, risk factors including patient characteristics, clinical history and present virulence factor (VF) genes are considered.

2.4.1 Patient Characteristics and Clinical History

Several well-studied risk factors exist for acquisition of CDI due to carriage in hospitals. Older age coupled with comorbidities are the most well-known risk factors [12]. As people begin to age, their encounters with healthcare systems become more frequent while their ability to fight infections decreases. With older age also comes other health complications that require more intensive care and potential hospitalization. The use of broad-spectrum antibiotics is also a well-documented risk factor of CDI development. β-lactam, fluoroquinolone and carbapenem antibiotic classes have been attributed to disruption of the gut microbiome, enabling *C. difficile* to become pathogenic [12]. Additionally, the length of stay in acute-care hospitals is an important factor to consider when determining risk factors for CDI. Studies have reported that with each additional day of hospitalization there is a 4% increase in risk of developing a CDI [13] and a five-fold increased risk when patients are hospitalized longer than a week [14].

Other high-risk populations include immunocompromised patients, organ transplant recipients, patients with inflammatory bowel disease and chemotherapy patients [12]. Patients with prior CDIs are prone to recurrent infections and are considered high-risk [15]. These risk factors contribute to the disruption of the gut microbiome.

2.4.2 Virulence Factor Genes

Virulence factor (VF) genes have been associated with the onset of CDIs [16]. They can influence several host-pathogen interactions, including elevated production of toxins A and B, the presence of a binary toxin, CDT, and mutations in the negative regulator of toxin expression [16]. Table 1 outlines VF genes and their role in causing disease. In this study. VF genes were identified by a bio-informaticist after whole genome sequencing of bacterial isolates if the DNA sequence matches ≥80% of the nucleotide identify for genes in the VF database (http://www.mgc.ac.cn/VFs/).

Table 1: Virulence Factor Gene Descriptions and Risks

Gene	Description	Risk	Source
CD0873	Surface-exposed lipoprotein	Increased cell wall attachment and colonization factors	[17]
CD2831	Sortase-anchored protein, putative adhesin	Increased cell wall attachment, interactions with matrix proteins (collagen), biofilm formation and immune evasion	[18]
CD3246	Sortase-anchored protein, putative adhesin	Increased cell wall attachment, interactions with matrix proteins (collagen), biofilm formation and immune evasion	[19]
cbpA	Enhancing collagen interaction and extracellular matrix adherence	Increased adhesion	[20]
tcdA	Inactivates Rho GTPases	Disrupts cytoskeleton resulting in disruption of tight junctions and loss of intestinal barrier function	[21]
tcdB	Inactivates Rho GTPases	Disrupts cytoskeleton resulting in disruption of tight junctions and loss of intestinal barrier function	[22]
cwp66	Implicated adhesion and stress tolerance	Release and dissemination of *C. difficile* in the host	[22]
cwp84	Cleavage of adhesions, such as SlpA for the paracrystalline layer assembly	Release and dissemination of *C. difficile* in the host	[22]
fbpA/fbp68	Fibronectin binding protein	Increased cell adherence	[23]
groEL	Heat shock protein	increased cell adherence	[23]
iap	Iota toxin gene	Toxin production	[24]
ibp	Iota toxin gene	Toxin production	[24]
slpA	Presence and low molecular weight subunits	Increased adhesion to gut mucosa	[25]

ToxA	Toxin A gene	Cytotoxic to living cells	[24]
ToxB	Toxin B gene	Cytotoxic to living cells, more so than Toxin A	[24]
zmp1	Secreted protease zinc- metalloprotease	Cleave host proteins, destabilizes fibronectin network	[26]

2.5 Utilizing Whole Genome Sequencing to Detect Outbreaks

Establishing genetic relatedness between organisms can identify potential outbreaks that could be misidentified by traditional epidemiological infection prevention methods. Whole genome sequencing (WGS) has led to improved detection of transmission of healthcare-associated infections [27]. A reactive, multistep process to determine outbreaks is executed in healthcare settings where WGS is not used. IPs investigated suspected instances of transmission using traditional epidemiological methods, primarily examining common transmission routes, which could include procedure rooms, medical devices, healthcare workers and inpatient rooms. Since traditional epidemiological methods are typically reactive to the onset of infections, there is a possibility of missing other patients in an outbreak when the transmission route is not common.

Early studies using WGS have provided insight to the advantage this technology has for identifying genetic relatedness among pathogens. In a study cohort of 957 *C. difficile* isolates over the course of three years, 333 isolates (35%) were genetically related to at least one other isolate in the cohort [48].

The Enhanced Detection System for Healthcare-Associated Transmission (EDS-HAT) uses WGS surveillance and machine learning (ML) of the electronic health record (EHR) to identify previously undetected outbreaks and transmission routes [2]. 524 unique patient isolates of *C. difficile* were collected in a retrospective study using EDS-HAT to determine the burden of transmission in an acute-care hospital. 80 (15.3%) had at least one genetically related isolate within the cohort [2]. Section 3.5 describes the methods used to identify outbreaks using WGS are outlined.

3.0 Methods and Materials

3.1 Study Site Characteristics

This study took place at an adult tertiary care hospital with over 700 total beds and over 100 critical care beds. It is a level 1 regional resource trauma center that specializes in solid-organ transplants, performing over 400 solid-organ transplants annually. The hospital serves the mid-sized city of Pittsburgh, Pennsylvania with over 300,000 people as well as patients from the surrounding regions of western Pennsylvania, Eastern Ohio and West Virginia.

3.2 Selection of Study Population

This study selected patients identified by the EDS-HAT database. Patients whose isolate was collected after a hospital stay of three or more days and/or had a recent healthcare exposure within 30 days of the isolate collection met the EDS-HAT inclusion criteria [2]. Patients who had an HA-CDI from November 1st, 2016 to August 31st, 2019 and November 1st, 2021 to May 31st, 2023 were selected for this study. The two-year time gap is due to the EDS-HAT study team pausing retrospective analysis to begin real-time analysis in November 2021. The initial patient cohort from the EDS-HAT database comprised of 809 patients (Figure 1). The hospital's Clinical Analytics Team provided data on the initial patient cohort, outlining patient demographic data and other variables of interest to provide a better understanding of their healthcare exposure prior to their CDI.

3.2.1 Patient Exclusion, Case Definitions and Final Cohort

Patients with incomplete data for any of the study variables were excluded from this study. Patients with an admission status other than inpatient at the study hospital were also excluded from this study as the goal was to determine risk factors associated with acquisition from transmission for patients only in an inpatient setting. The analysis data set excluded 117 patients (14.5%); 86 patients (10.6%) with incomplete data and 31 patients (3.8%) who were not admitted as inpatients at the time of testing (Figure 1). Additional information regarding creating the analysis data set can be found in section 3.6.1.

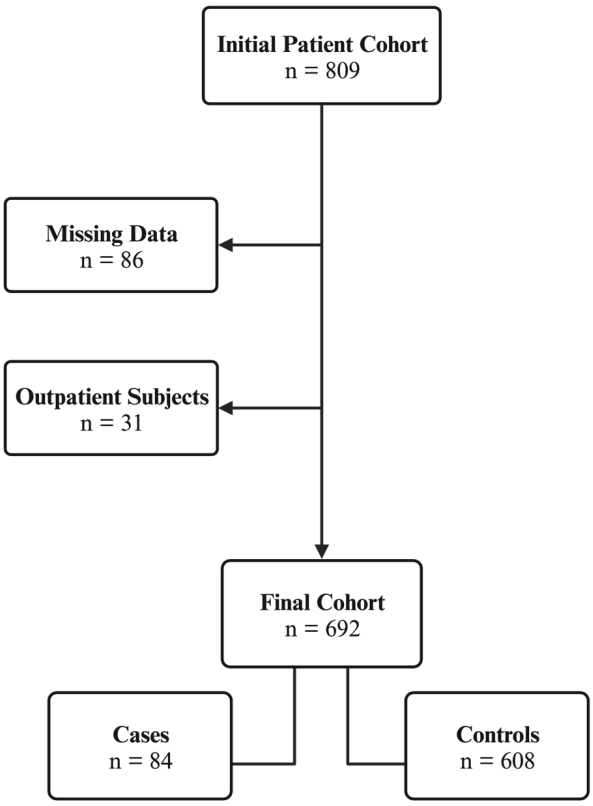

Figure 1: Study Cohort Flowchart

The final cohort consisted of 692 patients (86.1%) (Figure 1). Cases were defined as having at least one other genetically related isolate within the cohort, also known as a cluster, and did not include the index patients within a cluster (*N*=84, 12.1%). The control group included the remaining 608 patients (87.9%) who had CDIs with no genetically related isolates within the cohort. Index patients were included in the control group (Figure 2).

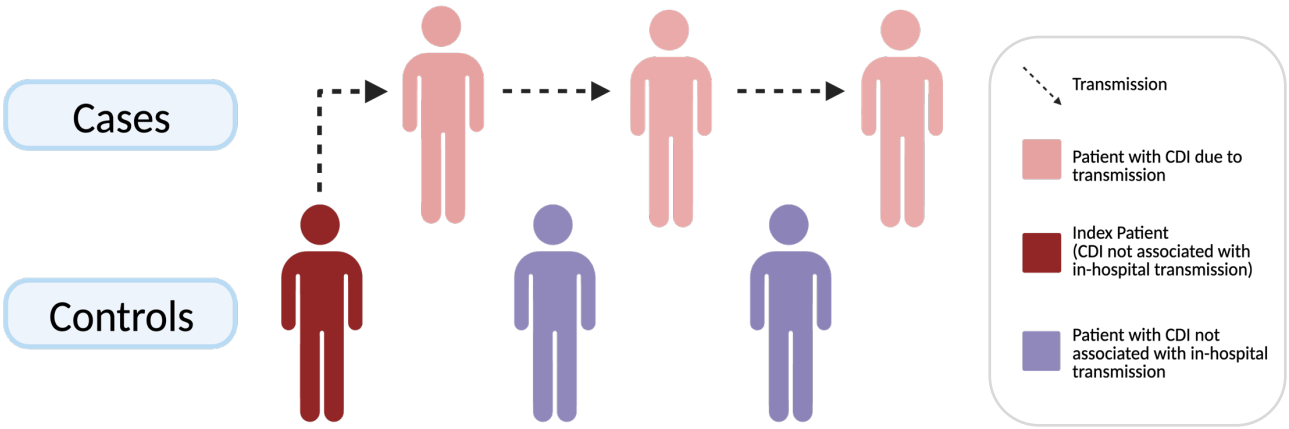

Figure 2: *Clostridiodes* difficile Acquisition in Healthcare Settings

3.3 Variables of Interest

Variables for this study were selected based on known risk factors for *C. difficile* infections. Variables were categorized into four groups; demographic information, extent of care during hospital stay, clinical history and other variables of interest (Table 2). Patient demographics included sex, race and age at isolate collection, were obtained. Variables outlining a patient's extent of care were also obtained: length of stay, number of OR visits, OR procedures (14 unique procedures), antibiotic receipts (three unique antibiotic classes), ICU care and time spent in the

ICU. Variables describing a patient's clinical history include identification of prior CDIs, timing of the earliest and most recent CDI a patient may have had, if vancomycin was used to treat a prior CDI, how previous vancomycin was administered, isolation codes for other significant organisms during the admission period (eight unique isolation codes) and the patient's Summary Elixhauser Comorbidity Score (S-Elixhauser) and comorbidities within the summary (38 unique comorbidities). An expanded list of variables for OR procedures, antibiotic classes, isolation codes and vancomycin admission routes can be found in Table 4. Other variables of interest included virulence factor genes which were coded as the count of virulence factor genes detected in a given patient isolate (Table 1), days from admission to isolate collection and type of CDI test. Variables were coded as binary, categorical or continuous for the purpose of this study (Table 2).

Table 2: Variables of Interest

Variable Group	Binary Variables	Categorical Variables	Continuous Variables
Demographic	Sex	Race	Age at isolate collection
Extent of care during hospital stay	ICU admission, OR visit, OR procedures	NA	Length of stay, days in ICU, number of OR visits
Clinical History	Summary Elixhauser Comorbidities, antibiotic receipt within 14 days of isolate collection, isolation code for clinically significant organisms, prior CDI, vancomycin used to treat prior CDI, route of admission for prior vancomycin use	NA	Summary Elixhauser Comorbidity Score, Number of isolation codes, number of prior CDIs, days to earliest prior CDI, days to most recent prior CDI
Other	Virulence factor genes	NA	Days from admission to isolate collection, total number of virulence factor genes

The S-Elixhauser score was used to capture a patient's individual comorbidities as well as providing a predictor of patient mortality [28]. It has been internally validated in previous studies as a predictor for 30-day, in-hospital and 1-year mortality [28]. The list of the 38 comorbidities are outlined in Table 3.

Table 3: Elixhauser Comorbidities

Neurological/Psychological	Dementia Depression Psychoses Neurologic disorders affecting movement Paralysis Seizures and epilepsy Other neurologic disorders
Vascular	Coagulopathy Deficiency anemia Blood loss anemia Peripheral vascular disease Pulmonary circulation disease Cerebrovascular disease Hypertension, uncomplicated Hypertension, complicated
Endocrine	Hypothyroidism Other thyroid disorders Diabetes, complicated Diabetes, uncomplicated
Cancer	Metastatic cancer Leukemia Lymphoma Solid tumor without metastasis, malignant Solid tumor without metastasis, in situ
Cardiovascular/Pulmonary	Congestive heart failure Valvular disease Chronic pulmonary disease
Renal/Liver	Liver disease, mild Liver disease, moderate to severe Renal failure, moderate Renal failure, severe
Other	Acquired immunodeficiency syndrome/Human immunodeficiency virus Peptic ulcer with bleeding Arthroplasties Weight loss Obesity

	Alcohol misuse Drug misuse

Table 4: Expanded List of Variables

Binary Variables	**OR procedures**	Gastrointestinal Urinary Vascular Cardiac General Transplant Plastic Otolaryngology Anesthesiology Thoracic Cardiology Neurology Orthopedic Trauma
	Antibiotic classes noted within 14 days of isolate collection	Beta-lactam Fluoroquinolone Carbapenem
	Isolation codes for clinically significant organisms	Multi-drug resistant *Actinobacter* species Carbapenem-Resistant *Enterobacteriaceae* Extended Spectrum Beta-Lactamase *Escherichia coli* Multi-drug resistant Gram-negative rods Methicillin-resistant *Staphylococcus aureus* Vancomycin-resistant *Enterococcus* Vancomycin-resistant *Staphylococcus aureus* Emerging pathogen
	Admission route for prior vancomycin use	Oral Intravenous line Jejunostomy Nasogastric Tube Gastrostomy tube
Categorical Variables	**Race**	White* Black Other (Alaskan Native, Native American, Chinese, Other Asian, Indian Unknown)

**Comparison group*

3.4 *Clostridiodes difficile* Testing

It is important to note the diagnostic testing for CDI changed in the study hospital during the study period. Prior to December 2018, a polymerase chain reaction (PCR) test was used to detect a toxin-producing gene present in CD genomes. This testing resulted in patients who were colonized with *C. difficile* testing positive for a CDI but were not experiencing an active infection. Currently the testing protocol for CDI consists of a two-step process to identify the presence of bacteria and the disease-causing toxin produced. The overall purpose of switching to a two-step testing process was to enhance patient care, refine treatment methods, and minimize misdiagnosis of CDIs.

3.4.1 PCR Testing

Prior to the availability of PCR-based testing, toxin tests were favored over culture-based diagnosis due to the rapid testing turnaround. PCR can target toxin-producing genes; however, these genes may be present without substantial toxin expression [29]. In some cases, positive PCR results reflect *C. difficile* colonization, rather than infection. Patients who are primarily colonized with *C. difficile* do not require antibiotic treatment. Treatment of clinically non-indicated *C. difficile* with antibiotics can result in disruption of the gut microbiome and subsequently trigger a clinically-indicated CDI in the patient. As a result, misdiagnosis of CDIs has become a concern for IPs and a new testing method was developed [29].

3.4.2 Two-Step Testing

The two-step protocol provides more information about both the presence of bacteria and of disease-causing toxin production [49]. This test first detects Glutamate Dehydrogenase (GDH) that is present with all CDIs and then the toxin produced; both tests are Enzyme Immunoassay (EIA) type tests. If there is a discrepancy between the two, a PCR test will be performed. Study hospital provides a table for physicians to interpret results of this test (Table 5). Table 5 is provided by the study hospital.

Table 5: Interpreting Results of the Two-Step *Clostridiodes difficile* Test

Test Result			Reported Result	Interpretation
GDH EIA	**Toxin A/B EIA**	**Toxin gene PCR**		
Negative	Negative	[not performed]	Negative	Negative for *C. difficile*
Negative	Positive	[not performed]	Indeterminate	This is a rare result. The test will be performed again to double check, and another specimen may need to be submitted
Positive	Positive	[not performed]	Positive	Toxin-producing *C. difficile* is present.
Positive	Negative	Positive	Positive	Positive for *C. difficile* toxin DNA. This may represent colonization or infection. Clinical correlation required.
Positive	Negative	Negative	Negative	Negative for *C. difficile* Toxin DNA not present.

An algorithm (Figure 3) is also utilized to help guide practitioners on how to report the results of both the two-step test and a PCR test, if necessary. Figure 3 is provided by the study hospital.

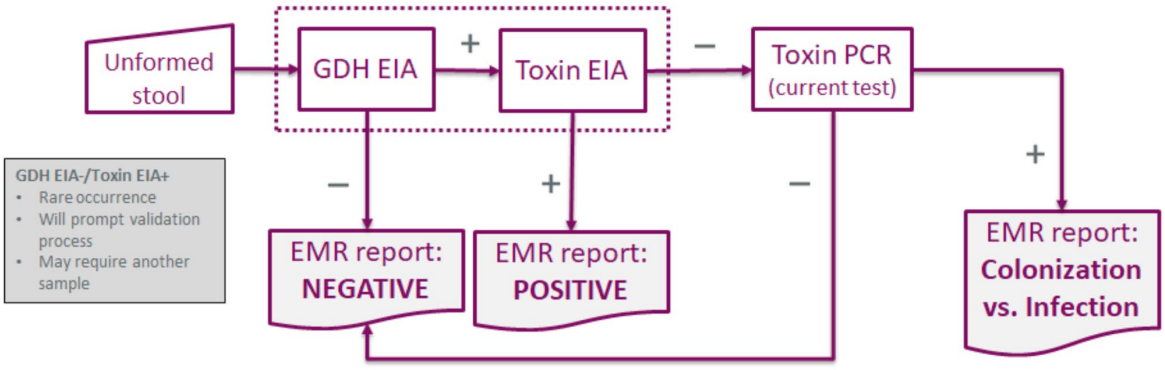

Figure 3: Flow Chart Describing How to Report Results in the EHR

In short, for the EHR to reflect a positive CDI test, the GDH EIA and toxin EIA must both be positive. If the GDH EIA is negative, the EHR will display a negative CDI result. If the GDH EIA is positive, but the toxin EIA is negative a PCR test will be performed; a positive PCR test will result an indication of colonization in the EHR and a negative PCR test will yield in a negative CDI test in the EHR.

3.5 EDS-HAT Methods

EDS-HAT determines patient of interest who are identified based on their exposure to the study hospital's healthcare system in the previously described inclusion criteria for *C. difficile*. EDS-HAT identifies genetically similar organisms that traditional IP methods may not have identified or have misidentified. EDS-HAT has been performing real-time WGS surveillance for the following clinically significant organisms since November 2021: *Acinetobacter* species, *Pseudomonas* species, extended-spectrum B-lactamase-producing *Escherichia coli*, *Klebsiella* species, *Clostridioides difficile*, ESBL *Enterobacter* species, vancomycin-resistant *Enterococcus*,

methicillin-resistant *Staphylococcus aureus*, *Stenotrophomonas* species, *Serratia* species, *Burkholderia* species, *Legionella* species, *Providencia* species, *Proteus* species, and *Citrobacter* species.

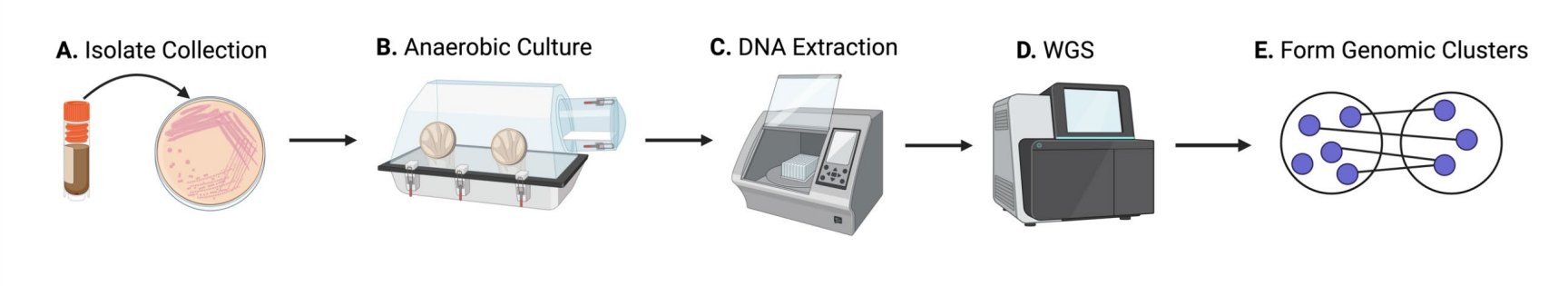

Figure 4: EDS-HAT Protocol

3.5.1 Sample Collection and Isolation

Clinical stool samples were obtained from the study hospital's clinical microbiology laboratory. To isolate *C. difficile* from the clinical sample, stool was cultured on an enriched, selective, and differential media (Figure 4A; Cycloserine Cefoxitin Mannitol Agar with Taurocholate and Lysozyme "CCMA-TAL") and incubated in a Coy Labs Vinyl Anaerobic Chamber for three days (Figure 4B) [30]. A single representative isolate of *C. difficile* was obtained and was confirmed by testing for the production of proline aminopeptidase (PRO disc) [31].

The cells were pelleted for WGS by inoculating half of a 10 µL loop of bacteria from the CCMA plate in microcentrifuge tubes containing 750 µL phosphate buffer saline (PBS). The cells were centrifuged at 6.0 × *g* for 10 minutes and the supernatant was removed. The pellets were stored at -20°C and then resuspended into 500 µL of PBS prior to DNA extraction.

3.5.2 Whole Genome Sequencing, Library Preparation and Sequencing

Genomic DNA was extracted using the MagMAX DNA Multi-Sample Ultra 2.0 extraction kit on the King Fisher Apex (Figure 4C). Briefly, this process uses magnetic beads to isolate and purify nucleic acids. DNA was eluted in 100 µL of elution buffer and was quantified using a Qubit broad range double-stranded DNA (dsDNA) kit. If the sample had a concentration ≥3.5ng/µl, it was sequenced. . For samples that did not meet this threshold, DNA was extracted again.

DNA libraries were prepared on an Eppendorf EpiMotion machine using Illumina DNA Prep (M) Tagmentation kit. Unique 10-mer index adapter sequences were ligated to the DNA fragments. DNA libraries sequenced were amplified and eight samples were pooled together in

equimolar concentrations, followed by DNA quantification using a Qubit high sensitivity dsDNA kit.

DNA libraries were sequenced using an Illumina MiSeq or NextSeq550 platform (v3, 600 cycle or v2.5, 300 cycle kits, respectively; Figure 4D). Libraries were diluted to a final loading concentration of 16 pM (MiSeq) or 1.5-1.6 pM (NextSeq550).

3.5.3 Bioinformatics Analysis

Unicycler (version 0.5.0) was used to assemble WGS reads and Prokka (version 1.14) was used for annotation [32]. Multilocus sequence types were determined using PubMLST typing schemes (mlst version 2.11) [33]. Species was confirmed using Kraken2 (ref). Genomes passed quality control for species match, the assembly was within 20% of the expected length, the number of contigs was ≤350, and the genome had a minimum average depth of 35×. Pairwise core genome single nucleotide polymorphisms (SNPs) were determined using either Snippy (version 4.3.0) [34] or SKA (version 1.0) [35] and SNP distances were calculated. *C. difficile* isolates were defined as a cluster for those containing two or less SNP differences (Figure 4E).

3.6 Statistical Methods

There are several methods available when selecting a prediction model; this study utilized the regularization of generalized linear models. The primary goal was to create a simple, parsimonious model that contained the smallest number of predictors. To accomplish this, I chose elastic net regression (ENR) to account for the correlation structure for some of the potential

prediction variables [45]. ENR penalizes for complex models by shrinking the beta-coefficients using both LASSO and ridge regression penalty factors. By using ENR for variable selection, some variable coefficients shrank moderately while others were set to zero, and thus removed from further consideration in the final model. Additionally, the ENR procedure robustly estimated the coefficients for correlated variables. The ENR procedure utilized a binary outcome the outcome (cases vs controls, where cases were defined as CDI acquired from in-hospital transmission; also see section 3.2.1 for further details on case/control inclusion criteria). The association between patient phenotype and *C. difficile* genomic variables were determined using logistic regression of the outcome on predictors selected by ENR. All statistical analyses were done using R (version 4.3.1) [46]. and the details for the packages that were used are provided in Table 6.

Table 6: R Packages Used in Statistical Analysis

Package Name (Version)	Description	Source
data.table (1.14.8)	Extension of 'data.frame' used toimport data into R	[36]
dplyr (1.1.2)	A grammar of data manipulation	[37]
fmsb (0.7.5)	Used to determine Nagelkerke's pseudo R^2	[38]
glmnet (4.1-8)	Lasso and Elastic-Net regularized generalized linear models	[39]
lmtest (0.9-40)	Testing liner regression models	[40]
Rcpp (1.0.11)	Seamless R and C++ integration	[41]
rsq (2.5)	Used to determine Nagelkerke's partial pseudo R^2 for each variable	[42]
stats (4.3.1)	Used for generalized linear modeling and general statistic measures	[43]

3.6.1 Creating the Analysis Data Set

To prepare the dataset, quality control of the data was performed. Each variable was coded in the following manner: binary, continuous or categorial (Table 2). Variables that could be interpreted as "yes" and "no" were coded as a binary. OR procedures, antibiotic receipt within 14

days of isolate collection, isolation codes, vancomycin admission route and virulence factors were all originally categorical variables that were manually expanded and converted to a binary. For example, each OR procedure became its own variable and patients were assigned 0 or 1 based on the procedures they received, if any. The 38 S-Elixhauser comorbidities were each coded as a binary as well.

Other variables remained as a continuous numerical value or as a categorial variable (Table 2). Variables that were coded as continuous were normalized by Z-score transformation by subtracting the mean and dividing by the standard deviation. Patients with no data for continuous numerical variables were assigned 0. For example, patients who were not admitted to the ICU were assigned a value of 0 for the "days in ICU" variable; whereas patients who were admitted to the ICU varied in their length of stay in the ICU.

The categories for race were consolidated due to sample size. White and Black racial groups remained as their own categories whereas Alaskan Native, Native American, Chinese, Other Asian, Indian and unknown racial groups were collapsed into a group coded as "Other" for the purpose of this analysis. Comparison groups were indicated among categorical variables with the largest sample size within each category. For the race variable, patients who self-identified as White were the comparison group.

3.6.2 Main Effects

The command 'model.matrix' was used to generate the final data frame. This command expands categorical variables with a binary designation. Next, ENR regression was performed using 10-fold cross validation. During cross validation, the data was divided into ten bins that contained an equal number of patients. The patient's assignment into a given bin was randomly

determined. Then, a training and testing procedure was used to determine the penalty factors for the prediction variables. The model is trained on nine of the ten bins and then tested on the remaining bin. This procedure was done until every patient was used for both training and testing. The result of this procedure was a list of eight variables ENR determined as the main effects in the model.

3.6.3 Interactions

In some instances, interactions between main effects have a significant impact on the prediction model. ENR was also used to test whether such interactions were associatedwith the outcome. To do this, all pairwise interactions were determined of the eight main effect ENR-chosen variables. Again, as before, ENR was performed using 10-fold cross validation while forcing the main effects into the model. Significant ENR-chosen interactions chosen were included in the final prediction model.

3.6.4 Final Model

The main effect variables and the interactions selected by ENR were used to generate a prediction model for the outcome variable using a generalized linear model in the form of logistic regression. The fraction of the variance explained by the model and the contribution of each variable to the model were computed using Nagelkerke's pseudo-R-squared [44].

4.0 Results

4.1 Study Population

The study cohort consisted of 692 patients with 608 patients (87.9%) who had CDIs not acquired from in-hospital transmission (controls) and 84 patients (12.1%) with CDIs acquired from in-hospital transmission (cases). The characteristics of the study cohort patients are presented in Table 7. The study cohort patients were mostly white (79.3%) and male (52.0%). The average age of control patients was approximately 60.6 years [SD=15.9] and the average age for case patients was approximately 62.6 years [SD=15.6]. Case patients averaged longer hospital stays (40.3 days, SD=70.2) compared to other control patients (22.4 days, SD=23.6). Although more control patients were admitted to the ICU during their stay (54.9% of controls versus 42.8% of cases), case patients averaged longer stays in the ICU (17.7 days, SD=60.6 for cases versus 7.9 days, SD=16.4 for controls).

Table 7: Characteristics of Study Cohort

Variable	Controls N=608	Cases N=84
Age at isolate collection, mean [SD]	60.6 [15.9]	62.6 [15.6]
Male Sex (%)	312 (51.3)	46 (54.8)
Female Sex (%)	296 (48.7)	38 (45.2)
Race (%)		
White	482 (79.3)	67 (79.8)
Black	80 (13.2)	10 (11.9)
Other (Alaskan Native, Chinese, Other Asian, Native American, Indian, Unknown)	46 (7.6)	7 (8.3)
Hospital Length of Stay (days), mean [SD]	22.4 [23.6]	40.3 [70.2]
ICU, ever admitted (%)	334 (54.9)	36 (42.8)
ICU Length of Stay (days), mean [SD]	7.9 [16.4]	17.7 [60.6]
Elixlhauser Comorbidity Score, mean [SD]	5.1 [2.9]	4.6 [3.1]
Isolation Codes, mean [SD]	0.45 [0.7]	0.19 [1.1]
Antibiotic Receipt ≤14 days prior to isolate collection (%)	41 (6.7)	10 (11.9)
Antibiotic Class Use ≤14 days prior to isolate collection (%)		
β-lactam Antibiotics	28 (4.6)	7 (8.3)
Fluoroquinolone Antibiotics	8 (1.3)	3 (3.6)
Carbapenem Antibiotics	10 (1.6)	0 (0.0)
Prior CDI Diagnosis (%)	88 (14.5)	15 (17.9)
Vancomycin use to treat prior CDI (%)	55 (9.1)	13 (15.5)
Operating Room Visit (%)	345 (42.9)	39 (46.4)
Operating Room Procedure (%)		
GI	73 (12.0)	11 (13.1)
Urinary	13 (92.1)	1 (1.2)
Vascular	25 (4.1)	5 (5.9)
Cardiac	18 (2.9)	1 (1.2)
Transplant	34 (5.6)	10 (11.9)
Plastic	24 (3.9)	2 (2.4)
Otolaryngology	20 (3.3)	3 (3.6)
Anesthesiology	3 (0.5)	1 (1.2)
Thoracic	58 (9.5)	6 (7.1)
Cardiology	4 (0.7)	0 (0.0)
Neurology	37 (6.1)	3 (3.6)
Orthopedic	34 (5.6)	4 (4.8)
Trauma	1 (0.2)	0 (0.0)

4.2 Main Effects and Interactions

The following variables were determined to be main effects in the prediction model for risk factors associated with *C. difficile* acquisition from transmission in a healthcare setting: transplant procedure during hospital stay, antibiotic recipient within 14 days of isolate collection, longer length of stay (days), VF gene *iap*, and ICU admission during hospital stay.

Only one pairwise interaction between the main effects had a significant impact on the prediction model: transplant procedure × increased length of stay.

4.3 Prediction Model

A generalized linear model was used to generate prediction model for *C. difficile* acquisition due to transmission (Table 8). This analysis identified five risk factors and four protective factors. The fraction of variance explained by this model for predicting *C. difficile* acquisition from transmission was 28.5%. Eight variables explain $\geq$90% of the variance in the model. Variables with an odds ratio greater than or equal to one are considered a risk factor, whereas variables with an odds ratio less than or equal to 1 are considered protective factors.

Table 8: Prediction Model for *Clostridiodes difficile* Acquisition from Transmission

Variable	N (%)	Estimate	SE	Z-Score	P-Value	Odds Ratio (95% CI)	Partial Pseudo R^2 (Nagelkerke)	Variance Explained by the Model
Transplant procedure during hospital stay	44 (6%)	1.400	0.447	3.13	0.0017	4.06 (1.69-9.74)	0.027	0.142
Antibiotic recipient within 14 days of isolate collection	51 (7%)	1.270	0.434	2.92	0.0035	3.56 (1.52-8.35)	0.024	0.127
Longer length of stay (days)	692 (100%)	0.553	0.148	3.75	0.0002	1.74 (1.3-2.32)	0.052	0.277
VF gene *iap*	165 (24%)	0.406	0.666	0.61	0.5424	1.5 (0.41-5.53)	0.001	0.006
Transplant procedure × increased length of stay	44 (6%)	-0.290	0.221	-1.32	0.1882	0.75 (0.49-1.15)	0.005	0.025
ICU admission during hospital stay	370 (53%)	-0.870	0.280	-3.11	0.0019	0.42 (0.24-0.72)	0.032	0.169
VF gene CD3246	539 (78%)	-0.937	0.752	-1.25	0.2127	0.39 (0.09-1.71)	0.005	0.027
VF gene *cbpA*	546 (79%)	-0.965	0.471	-2.05	0.0404	0.38 (0.15-0.96)	0.014	0.073
Autoimmune Disorder	57 (8%)	-1.922	0.804	-2.39	0.0168	0.15 (0.03-0.71)	0.029	0.155

SE = standard error, CI = confidence interval

4.4 Risk Factors for *Clostridiodes difficile* Acquisition from Transmission

This study identified five risk variables for *C. difficile* acquisition from transmission in acute-care hospitals (Figure 5). The most significant risk factor was the combination of having a transplant procedure and an increased hospital stay. Patients who meet these criteria were 5.26 times more likely to acquire *C. difficile* from transmission (95% CI, 2.16-12.88; *P*= 0.1882). Additionally, patients who had a transplant procedure, regardless of length of stay, were 4.06 times more likely to acquire *C. difficile* from transmission (95% CI, 1.69-9.74; *P*= 0.0017). Patients with antibiotic use (β-lactam, fluoroquinolone or ccarbapenem antibiotic classes) were 3.56 times more likely to acquire *C. difficile* from transmission (95% CI, 1.52-8.35; *P*=0.0035). For each additional hospital a patient was 1.74 times more likely to acquire *C. difficile* from transmission (95% CI,

1.30-2.23; *P*= 0.0002). Lastly, if the patient *C. difficile* isolate contained the VF gene *iap*, they were 1.50 times more likely to have acquired *C. difficile* from transmission (95% CI, 0.41-5.53; *P*= 0.5424).

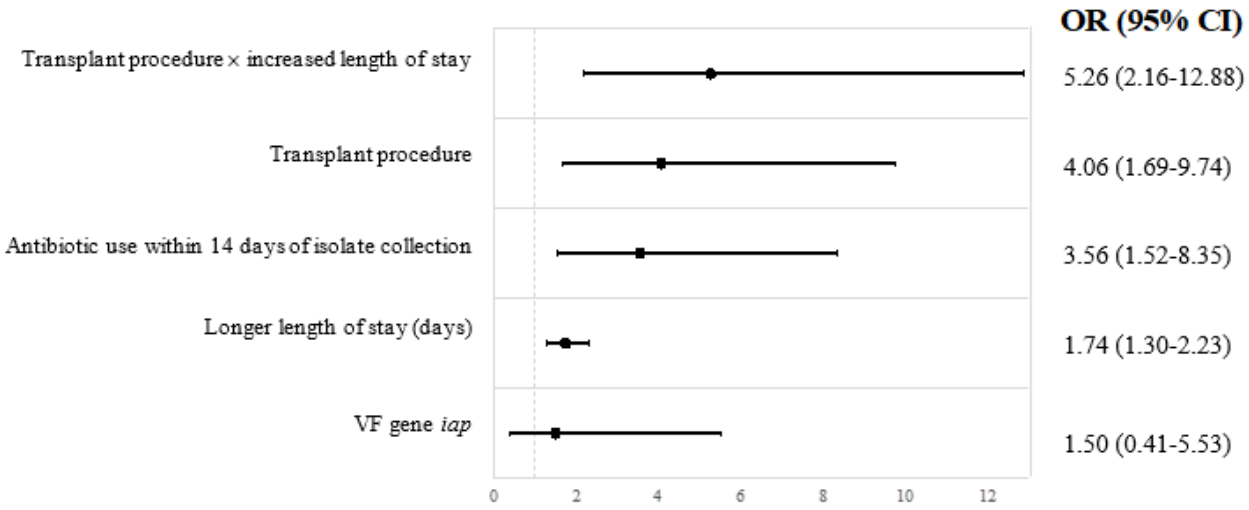

Figure 5: Risk Factors for Acquiring Clostridiodes difficile from Transmission

5.0 Discussion

5.1 Suggested Interventions

Prevention of *C. difficile* transmission is vital for patient safety. This study has identified risk factors present among patients who have acquired CDIs due to in-hospital transmission. Targeted interventions can be implemented that may reduce the rates of *C. difficile* transmission in acute-care hospitals.

5.1.1 Suggested Interventions to Prevent Acquisition from Transmission

The study's findings have highlighted key characteristics of high-risk patients for acquiring *C. difficile* from transmission. Interventions to prevent acquisition from transmission should be explored to mitigate the spread of *C.* difficile in healthcare settings. The first of these interventions includes hospital staff utilizing pre-emptive precautions when handling transplant recipients and patients with longer hospital stays before any sign of infection. These precautions include the use of contact precautions when providing care or transporting patients. Limiting shared rooms and excess healthcare personnel contact would also be included in pre-emptive precautions Additionally, enhanced environmental cleaning in high-risk patient rooms with chlorine-releasing cleaning agents should be implemented as part of the pre-emptive precaution. Hand hygiene is also an important factor in prevention of transmission. Implementing methods to increase the rate of hand hygiene with soap and water should be explored. These interventions are designed to reduce

an at-risk patient's contact with spores that may be present on a healthcare worker's clothes, gloves, or in the environment.

In addition to pre-emptive precautions, increased compliance with the antimicrobial stewardship program should be a priority. Although a provider may need to prescribe antibiotics, all other alternative treatments should be considered before prescribing antibiotics. Any antibiotic being prescribed must by clinically necessary based on the clinical condition of the patient and current guidelines. Prescribing antibiotics as a method for treating an infection with an unknown root cause is not recommended.

5.1.2 Interventions to Prevent Progression of Carriage to Transmission

While this study discussed the risk factors for *C. difficile* acquisition from transmission, interventions to prevent the progression of *C. difficile* carriage to transmission are also necessary. Ensuring proper contact precautions and environmental cleaning protocols are being adhered to when a patient presents with a CDI is vital to reduce *C. difficile* transmission from an index patient. Additionally, adhering to antimicrobial stewardship guidelines can ensure patient gut microbiomes remain intact to prevent *C. difficile* from causing infection. Furthermore, if antibiotics are deemed clinically necessary, implementing primary or secondary prophylaxis against CDIs should be considered to prevent *C. difficile* colonized in a patient's gut microbiome from becoming pathogenic. When colonized *C. difficile* becomes pathogenic, treatment with vancomycin or fidaxomicin are appropriate. Nevertheless, pharmacologic measures to prevent CDI recurrence should also be taken, such as fecal microbiota transplants, gut microbiota restoration or bezlotoxumab.

A unique finding of this study was the identification of the VF gene *iap* as a risk factor among patients who acquired *C. difficile* from transmission. VF genes were found to be conserved from index patients to others in transmission clusters. Thus patients whose isolate had the VF gene *iap* were more likely to have acquired *C. difficile* from transmission. The VF gene *iap* codes for the iota toxins secreted by *C. difficile*. Recent work has shown the effects of iota toxins on adherence of *C. difficile* on host cells and may act as a colonization factor [50]. Although the iota toxin itself may not cause transmission of *C. difficile*, preventing the progression of carriage to transmission is still necessary. Asymptomatic stool surveillance can help IPs identify patient *C. difficile* isolates with the VF gene *iap* and monitor those patients to prevent the progression. This may include limiting antibiotic usage or alternative *C. difficile* treatment if the patient were to develop a CDI. At the time of this study, no other publication has outlined the risk of VF genes associated with acquisition transmission using WGS and EHR data. It is important to note that the VF gene *iap* has an Odds Ratio of 1.50, but the 95% confidence interval contains a lower interval below one. Therefore, further studies with larger sample sizes are needed to narrow the confidence interval.

5.2 Public Health Significance

At the time of this study, no prior publication has identified risk factors for *C. difficile* acquisition from transmission. CDI can cause significant health problems among patients and steps should be taken to reduce the possibility of infection. The use of WGS has provided a tool to distinguish patients who have acquired *C. difficile* due to transmission and those who have not. Identifying risk factors for *C. difficile* acquisition from transmission can guide infection

preventionists in the development of interventions to mitigate the spread of *C. difficile*. Reducing the incidence of CDIs due to transmission can improve patient outcomes and reduce the cost of medical interventions necessary to treat CDIs. Interventions to reduce CDIs due to transmission may be implemented, these include pre-emptive contact precautions with high-risk patients, enhanced environmental cleaning and adherence to antimicrobial stewardship protocols. These interventions may impact how healthcare workers interact with high-risk patients and those who already present with an active CDI. Education on the importance of hand hygiene with soap and water, proper environmental cleaning and adherence to antimicrobial stewardship guidelines are also necessary to mitigate the spread of *C. difficile* in healthcare settings.

5.3 Limitations

Throughout this study, limitations were noted. First, it is unlikely that all outbreaks and outbreak patients were captured due to the inclusion criteria and study period. Although the inclusion criteria are meant to capture most patients with healthcare-associated CDIs, it is unlikely all healthcare-associated CDIs were reported in this study. Patients whose isolate was collected prior to day 3 of their hospital admission and did not have a healthcare exposure within 30 days would not have been captured in this study. Moreover, community-acquired CDIs were not sequenced and may have been the index for a subsequent CDI in a healthcare setting, resulting in misclassification. Additionally, case definitions were dependent on potential sources and without a pre-study recruitment period, cases and controls may have been misclassified.

Furthermore, transmission in this study was assumed based on SNP differences; some isolates did not have an epidemiologically recognized transmission source and genetically related

isolates may not have been the result of transmission. Patients with genetically similar isolates without epidemiological links may have had a common community source outside of a healthcare setting. Community-acquired *C. difficile* was not within the scope of this study.

Lastly, the sample size of this study cohort can impact the statistical significance of risk factors and result in overfitting of the model. The impact of the sample size can be observed in the 95% confidence intervals of the risk factors; a larger sample size would reduce the confidence intervals.

5.4 Future Directions

Prospective studies seeking to determine risk ratios among the identified associations is necessary to validate these findings. Replication to validate this work in future studies can incorporate new variables including non-OR procedures, hand hygiene compliance on units, and drugs such as pantoprazole and other proton pump inhibitors. Additionally, this study should be replicated at other healthcare centers with different patient demographics to identify other variables associated with *C. difficile* acquisition from transmission. Furthermore, expanding the sample size of future replications can provide better insight to other risk factors for *C. difficile* acquisition from transmission. Lastly, developing a prediction score for *C. difficile* acquisition from transmission could identify patients at high risk and guide targeted interventions. If patients with high risk scores and natural colonization with *C. difficile* can be identified, then IPs can implement enhance antimicrobial stewardship interventions to those patients, thus preventing the progression of carriage to infection and transmission.

6.0 Conclusion

The use of WGS can provide invaluable data on the burden of transmission in a healthcare setting Not only does WGS allow for early identification of outbreaks, but it also highlights unsuspected transmission routes. The transmission clusters identified using WGS allowed for categorization of CDI patients. Well-documented risk factors for HA-CDIs paired with WGS data were used to formulate a study and generate a prediction model to identify risk factors for *C. difficile* acquisition from transmission.

This study identified significant risk factors that contributed to *C. difficile* acquisition from transmission in an acute-care hospital. The implications of this study include suggested interventions to mitigate the spread of *C. difficile* in healthcare settings and limit the progression of carriage to active infection. Educating staff on the importance of prevention methods a key factor in reducing the spread of CDIs. This study's findings can help guide the development of targeted interventions and education for staff that may reduce transmission of *C. difficile* in a healthcare setting.

Bibliography

[1] Lessa, F. C., Winston, L. G., McDonald, L. C., & Emerging Infections Program C. difficile Surveillance Team (2015). Burden of Clostridium difficile infection in the United States. The New England journal of medicine, 372(24), 2369–2370. <https://doi.org/10.1056/NEJMc1505190>

[2] Sundermann, A. J., Chen, J., Kumar, P., Ayres, A. M., Cho, S. T., Ezeonwuka, C., Griffith, M. P., Miller, J. K., Mustapha, M. M., Pasculle, A. W., Saul, M. I., Shutt, K. A., Srinivasa, V., Waggle, K., Snyder, D. J., Cooper, V. S., Van Tyne, D., Snyder, G. M., Marsh, J. W., Dubrawski, A., … Harrison, L. H. (2022). Whole-Genome Sequencing Surveillance and Machine Learning of the Electronic Health Record for Enhanced Healthcare Outbreak Detection. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, *75*(3), 476–482. <https://doi.org/10.1093/cid/ciab946>

[3] Vohra, P., & Poxton, I. R. (2011). Comparison of toxin and spore production in clinically relevant strains of Clostridium difficile. *Microbiology (Reading, England)*, *157*(Pt 5), 1343–1353. <https://doi.org/10.1099/mic.0.046243-0>

[4] Bartlett, J. G., & Gerding, D. N. (2008). Clinical recognition and diagnosis of Clostridium difficile infection. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, *46 Suppl 1*, S12–S18. <https://doi.org/10.1086/521863>

[5] Buonomo, E. L., & Petri, W. A., Jr (2016). The microbiota and immune response during Clostridium difficile infection. *Anaerobe*, *41*, 79–84. <https://doi.org/10.1016/j.anaerobe.2016.05.009>

[6] Buddle, J. E., & Fagan, R. P. (2023). Pathogenicity and virulence of *Clostridioides difficile*. *Virulence*, *14*(1), 2150452. <https://doi.org/10.1080/21505594.2022.2150452>

[7] Moore, Luke & Cooley, Nicholas & Gilchrist, Mark. (2012). Clostridium difficile: microbiology & infection. Clinical Pharmacist. 4. 250

[8] Macleod-Glover, N., & Sadowski, C. (2010). Efficacy of cleaning products for C. difficile: environmental strategies to reduce the spread of Clostridium difficile-associated diarrhea in geriatric rehabilitation. *Canadian family physician Medecin de famille canadien*, *56*(5), 417–423.

[9] Czepiel, J., Dróżdż, M., Pituch, H., Kuijper, E. J., Perucki, W., Mielimonka, A., Goldman, S., Wultańska, D., Garlicki, A., & Biesiada, G. (2019). Clostridium difficile infection: review. *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology*, *38*(7), 1211–1221. <https://doi.org/10.1007/s10096-019-03539-6>

[10] Berg, M. L., Ayres, A. M., Weber, D. R., McCullough, M., Crall, V. D., Lewis, C. L., Valek, A. L., Vincent, L. A., Penzelik, J., Sasinoski, C. A., Cheng, A. L., Bradford, C. F., Bell, E. O., Edwards, K. M., Castronova, I. A., Brady, M. B., Slaughter, J., Oleksiuk, L. M., & Snyder, G. M. (2023). Diagnostic stewardship for *Clostridioides difficile* testing in an acute care hospital: A quality improvement intervention. *Antimicrobial stewardship & healthcare epidemiology : ASHE*, *3*(1), e67. https://doi.org/10.1017/ash.2023.141

[11] Boly, F. J., Reske, K. A., & Kwon, J. H. (2020). The Role of Diagnostic Stewardship in *Clostridioides difficile* Testing: Challenges and Opportunities. *Current infectious disease reports*, *22*(3), 7. <https://doi.org/10.1007/s11908-020-0715-4>

[12] Ananthakrishnan A. N. (2011). Clostridium difficile infection: epidemiology, risk factors and management. *Nature reviews. Gastroenterology & hepatology*, *8*(1), 17–26. <https://doi.org/10.1038/nrgastro.2010.190>

[13] Debast, S. B. et al. Successful combat of an outbreak due to Clostridium difficile PCR ribotype 027 and recognition of specific risk factors. Clin. Microbiol. Infect. 15, 427–434 (2009).

[14] Debast, S. B., Vaessen, N., Choudry, A., Wiegers-Ligtvoet, E. A., van den Berg, R. J., & Kuijper, E. J. (2009). Successful combat of an outbreak due to Clostridium difficile PCR ribotype 027 and recognition of specific risk factors. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, *15*(5), 427–434. <https://doi.org/10.1111/j.1469-0691.2009.02713.x>

[15] Song, J. H., & Kim, Y. S. (2019). Recurrent *Clostridium difficile* Infection: Risk Factors, Treatment, and Prevention. *Gut and liver*, *13*(1), 16–24. <https://doi.org/10.5009/gnl18071>

[16] Tijerina-Rodríguez, L., Villarreal-Treviño, L., Morfín-Otero, R., Camacho-Ortíz, A., & Garza-González, E. (2019). Virulence Factors of *Clostridioides* (*Clostridium*) *difficile* Linked to Recurrent Infections. *The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie medicale*, *2019*, 7127850. <https://doi.org/10.1155/2019/7127850>

[17] Kovacs-Simon, A., Leuzzi, R., Kasendra, M., Minton, N., Titball, R. W., & Michell, S. L. (2014). Lipoprotein CD0873 is a novel adhesin of Clostridium difficile. *The Journal of infectious diseases*, *210*(2), 274–284. <https://doi.org/10.1093/infdis/jiu070>

[18] Arato, V., Gasperini, G., Giusti, F., Ferlenghi, I., Scarselli, M., & Leuzzi, R. (2019). Dual role of the colonization factor CD2831 in Clostridium difficile pathogenesis. *Scientific reports*, *9*(1), 5554. <https://doi.org/10.1038/s41598-019-42000-8>

[19] Peltier, J., Shaw, H. A., Couchman, E. C., Dawson, L. F., Yu, L., Choudhary, J. S., Kaever, V., Wren, B. W., & Fairweather, N. F. (2015). Cyclic diGMP regulates production of sortase substrates of Clostridium difficile and their surface exposure through ZmpI protease-mediated cleavage. *The Journal of biological chemistry*, *290*(40), 24453–24469. <https://doi.org/10.1074/jbc.M115.665091>

[20] Tulli, L., Marchi, S., Petracca, R., Shaw, H. A., Fairweather, N. F., Scarselli, M., Soriani, M., & Leuzzi, R. (2013). CbpA: a novel surface exposed adhesin of Clostridium difficile targeting human collagen. *Cellular microbiology*, *15*(10), 1674–1687. <https://doi.org/10.1111/cmi.12139>

[21] Aktories, K., Papatheodorou, P., & Schwan, C. (2018). Binary Clostridium difficile toxin (CDT) - A virulence factor disturbing the cytoskeleton. *Anaerobe*, *53*, 21–29. https://doi.org/10.1016/j.anaerobe.2018.03.001 [22] de la Riva L, Willing SE, Tate EW, Fairweather NF. Roles of cysteine proteases Cwp84 and Cwp13 in biogenesis of the cell wall of Clostridium difficile. J Bacteriol. 2011 Jul;193(13):3276-85. doi: 10.1128/JB.00248-11. Epub 2011 Apr 29. PMID: 21531808; PMCID: PMC3133288.

[23] Dingle, T. C. (2012). Investigating the role of adherence in clostridium difficile pathogenesis (Doctoral thesis, University of Calgary, Calgary, Canada). Retrieved from <https://prism.ucalgary.ca. doi:10.11575/PRISM/4897>

[24] Perelle, S., Gibert, M., Bourlioux, P., Corthier, G., & Popoff, M. R. (1997). Production of a complete binary toxin (actin-specific ADP-ribosyltransferase) by Clostridium difficile CD196. *Infection and immunity*, *65*(4), 1402–1407. <https://doi.org/10.1128/iai.65.4.1402-1407.1997>

[25] Spigaglia P, Barketi-Klai A, Collignon A, Mastrantonio P, Barbanti F, Rupnik M, Janezic S, Kansau I. Surface-layer (S-layer) of human and animal Clostridium difficile strains and their behaviour in adherence to epithelial cells and intestinal colonization. J Med Microbiol. 2013 Sep;62(Pt 9):1386-1393. doi: 10.1099/jmm.0.056556-0. Epub 2013 Mar 21. PMID: 23518658.

[26] Janoir C. (2016). Virulence factors of Clostridium difficile and their role during infection. *Anaerobe*, *37*, 13–24. <https://doi.org/10.1016/j.anaerobe.2015.10.009>

[27] Eyre D. W. (2022). Infection prevention and control insights from a decade of pathogen whole-genome sequencing. *The Journal of hospital infection*, *122*, 180–186. <https://doi.org/10.1016/j.jhin.2022.01.024>

[28] Mehta, H. B., Li, S., An, H., Goodwin, J. S., Alexander, G. C., & Segal, J. B. (2022). Development and Validation of the Summary Elixhauser Comorbidity Score for Use With ICD-10-CM-Coded Data Among Older Adults. *Annals of internal medicine*, *175*(10), 1423–1430. <https://doi.org/10.7326/M21-4204>

[29] Polage, C. R., Gyorke, C. E., Kennedy, M. A., Leslie, J. L., Chin, D. L., Wang, S., Nguyen, H. H., Huang, B., Tang, Y. W., Lee, L. W., Kim, K., Taylor, S., Romano, P. S., Panacek, E. A., Goodell, P. B., Solnick, J. V., & Cohen, S. H. (2015). Overdiagnosis of Clostridium difficile Infection in the Molecular Test Era. *JAMA internal medicine*, *175*(11), 1792–1801. <https://doi.org/10.1001/jamainternmed.2015.4114>

[30] Galdys, A. L., Nelson, J. S., Shutt, K. A., Schlackman, J. L., Pakstis, D. L., Pasculle, A. W., Marsh, J. W., Harrison, L. H., & Curry, S. R. (2014). Prevalence and duration of asymptomatic Clostridium difficile carriage among healthy subjects in Pittsburgh,

Pennsylvania. *Journal of clinical microbiology*, *52*(7), 2406–2409. <https://doi.org/10.1128/JCM.00222-14>

[31] Fedorko, D. P., & Williams, E. C. (1997). Use of cycloserine-cefoxitin-fructose agar and L-proline-aminopeptidase (PRO Discs) in the rapid identification of Clostridium difficile. *Journal of clinical microbiology*, *35*(5), 1258–1259. <https://doi.org/10.1128/jcm.35.5.1258-1259.1997>

[32] Seemann T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*, *30*(14), 2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>

[33] Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. *Genome biology*, *20*(1), 257. <https://doi.org/10.1186/s13059-019-1891-0>

[34] Seemann T. Snippy [Internet]. 2023. Available from: <https://github.com/tseemann/snippy>

[35] Harris SR. SKA: Split Kmer Analysis Toolkit for Bacterial Genomic Epidemiology [Internet]. Genomics; 2018 Oct.

[36] Dowle M, Srinivasan A (2023). data.table: Extension of `data.frame`. R package version 1.14.8, <https://CRAN.R-project.org/package=data.table>.

[37] Wickham H, François R, Henry L, Müller K, Vaughan D (2023). dplyr: A Grammar of Data Manipulation. R package version 1.1.2, <https://CRAN.R-project.org/package=dplyr>

[38] Nakazawa M (2023). fmsb: Functions for Medical Statistics Book with some Demographic Data. R package version 0.7.5, <https://CRAN.R-project.org/package=fmsb>

[39] Friedman J, Tibshirani R, Hastie T (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." Journal of Statistical Software, *33*(1), 1-22. doi:10.18637/jss.v033.i01 <https://doi.org/10.18637/jss.v033.i01>

[40] Achim Zeileis, Torsten Hothorn (2002). Diagnostic Checking in Regression Relationships. R News 2(3), 7-10.

[41] Eddelbuettel D, Francois R, Allaire J, Ushey K, Kou Q, Russell N, Ucar I, Bates D, Chambers J (2023). Rcpp: Seamless R and C++ Integration. R package version 1.0.11, <https://CRAN.R-project.org/package=Rcpp>

[42] Zhang D (2022). rsq: R-Squared and Related Measures. R package version 2.5, <https://CRAN.R-project.org/package=rsq>

[43] R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

[44] Nagelkerke, N. (1991). A note on a General Definition of the Coefficient of Determination. Biometrika, 78 (3), pp. 691-692.

[45] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society Series B: Statistical Methodology, 67(2), 301-320.

[46] R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

[47] Stephenson, B., Lanzas, C., Lenhart, S., Ponce, E., Bintz, J., Dubberke, E. R., & Day, J. (2020). Comparing intervention strategies for reducing Clostridioides difficile transmission in acute healthcare settings: an agent-based modeling study. *BMC infectious diseases*, *20*(1), 799. <https://doi.org/10.1186/s12879-020-05501-w>

[48] Eyre, D. W., Cule, M. L., Wilson, D. J., Griffiths, D., Vaughan, A., O'Connor, L., Ip, C. L. C., Golubchik, T., Batty, E. M., Finney, J. M., Wyllie, D. H., Didelot, X., Piazza, P., Bowden, R., Dingle, K. E., Harding, R. M., Crook, D. W., Wilcox, M. H., Peto, T. E. A., & Walker, A. S. (2013). Diverse sources of C. difficile infection identified on whole-genome sequencing. *The New England journal of medicine*, *369*(13), 1195–1205. <https://doi.org/10.1056/NEJMoa1216064>

[49] Crobach, M. J., Planche, T., Eckert, C., Barbut, F., Terveer, E. M., Dekkers, O. M., Wilcox, M. H., & Kuijper, E. J. (2016). European Society of Clinical Microbiology and Infectious Diseases: update of the diagnostic guidance document for Clostridium difficile infection. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, *22 Suppl 4*, S63–S81. <https://doi.org/10.1016/j.cmi.2016.03.010>

[50] Awad, M. M., Johanesen, P. A., Carter, G. P., Rose, E., & Lyras, D. (2014). Clostridium difficile virulence factors: Insights into an anaerobic spore-forming pathogen. *Gut microbes*, *5*(5), 579–593. <https://doi.org/10.4161/19490976.2014.969632>