Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world-wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Samantha Sefton Date

Evaluating the Relationship between Healthcare Facility Network Characteristics and the Incidence of *Clostridium difficile* Infection (CDI) in Long Term Care Facilities in the Atlanta area, 2016

By

Samantha M. Sefton

Master of Public Health

Department of Epidemiology

_____

Dr. Scott Fridkin, MD

Committee Chair

Evaluating the Relationship between Healthcare Facility Network Characteristics and the Incidence of *Clostridium difficile* Infection (CDI) in Long Term Care Facilities in the Atlanta area, 2016

By

Samantha M. Sefton

B.S., Rhodes College, 2016

Faculty Thesis Advisor: Dr. Scott Fridkin, MD

An abstract of

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health in Epidemiology

2018

Abstract

Evaluating the Relationship between Healthcare Facility Network Characteristics and the Incidence of *Clostridium difficile* Infection (CDI) in Long Term Care Facilities in the Atlanta area, 2016

By Samantha M. Sefton

Background: *Clostridium difficile* infection (CDI) is a common healthcare associated infection (HAI) and is known to cause inflammation of the colon, severe diarrhea and occasionally death. Older age, exposure to healthcare facilities (i.e., environmental contamination), and antibiotic receipt are all risk factors for CDI. Transfer dynamics of patients within healthcare networks incorporate these factors and have also been associated with CDI risk. We evaluated the relationship between facility level characteristics of long term care facilities (LTCF) in the Atlanta area and the incidence of CDI.

Methods: We analyzed incident level CDI data from Georgia Emerging Infections Program (EIP). Incident LTCF-onset (LFTCO) cases are defined as an initial toxin positive test with no previous positive test in the 8 weeks prior. Centers for Medicare and Medicaid (CMS) cost reports were used to gather facility level characteristics, including bed days, which was used to calculate the CDI incidence facility rate. Negative Binomial Regression was used to evaluate LTCF characteristics associated with facility CDI rates (number of CDI/ 10,000 bed days). These characteristics include number of admissions, average length of stay, nursing staff ratio, bed size and connectivity metrics.

Results: There were 64 facilities included in the analysis, reporting a combined 155 incident CDI cases (facility CDI range: 0 -19 cases). A negative binomial model was created including the variables outdegree (no. of facilities receiving patients from the LTCF), number of admissions into the LTCF, and average length of stay. CDI rate was found to increase 69% for each unit decrease in average length of stay, holding number of admissions and outdegree constant.

Conclusions: Average length of stay is the most significant predictor of CDI rate in LTCF. The rate of CDI decreases as patients stay in the facility longer. This is a contrasting finding compared to acute care facilities, indicating that this relationship is unique to LTCF.

Evaluating the Relationship between Healthcare Facility Network Characteristics and the Incidence of *Clostridium difficile* Infection (CDI) in Long Term Care Facilities in the Atlanta area, 2016

By

Samantha M. Sefton

Bachelor of Science

Rhodes College

2016

Faculty Thesis Advisor: Dr. Scott Fridkin, MD

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of

Master of Public Health in Epidemiology

2018

TABLE OF CONTENTS

The primary dataset used in this project was collected by the Georgia Emerging Infections Program (GAEIP). The GAEIP was not involved in the analyses presented in this thesis

LIST OF TABLES

LIST OF FIGURES

BACKGROUND

Clostridium difficile infection (CDI) is a common healthcare associated infection (HAI) that is caused by *C. difficile* bacteria. *Clostridium difficile* is a gram-positive, spore-forming, anaerobic bacterium. The bacterium was first discovered in 1935 when it was isolated from a healthy infant (1). In 1978, *Clostridium difficile* was identified as the primary bacteria for causing psuedomembraneous colitis, a very severe form of inflammation of the colon (2). The bacterium was later determined to be the primary cause for most antibiotic-associated diarrhea (3).

There are many symptoms of CDI. These include, but are not limited to, abdominal pain, diarrhea, increased heart rate, fever, and loss of appetite. Many individuals are also colonized with the *C. difficile* bacteria and spread the bacteria to other individuals without presenting with symptoms. Some reports have suggested around 50% or more of patients that are hospitalized are colonized with *Clostridium difficile* and do not have any symptoms (4-7). Therefore, it is difficult to accurately assess the burden of disease and to accurately trace the sources of CDIs. However, certain risk factors have been identified.

CDI risk factors include older age, exposure to healthcare facilities, environmental contamination in healthcare facilities, and antibiotic receipt (8,9,10). In 2011, the United States alone reported half a million CDIs (11). It was found that markedly higher incidences of CDIs were reported among adults over the age of 65, white individuals, and females. One study found that individuals over the age of 65 were

at a 10 times increased risk for developing CDI compared to those aged 65 or younger (12).

CDIs are prevalent in long term care facilities (LTCFs). It has been estimated that anywhere from 50-75% of residents in LTCFs receive at least one antibiotic a year (13,14,15). Furthermore, 8%-33% of individuals who have been treated with antibiotics while in a LTCF develop CDI (16,17).

Patient transfer networks have recently become an area of interest in regard to HAIs. These networks use metrics such as indegree, outdegree, and betweeness to quantify how connected a specific facility is in regard to other facilities in the surrounding area. A study looking at the influence of transfer network on HAIs found a strong association between such network metrics and CDI (18).

LTCFs admit patients from their homes, but most often admit patients after they are discharged from the hospital. This post-acute care time period (defined as 12 weeks after hospital discharge) is a time period patients are most vulnerable to CDI (19). However, the source of the bacterium may be related to either the hospital, nursing facility, or potentially the home, thus, it is challenging to trace the onset of a CDI. Frequent visitations to multiple healthcare facilities also increase the risk of obtaining a CDI. A study found that CDI infections in LTCF residents occur within a month of admission to the LTCF, which indicates that these patients likely are exposed to the bacterium at the discharging facility (20).

It is necessary to examine and evaluate the incidence of CDIs in different types of LTCFs in order to understand if the burden or risk of CDIs in one of the most vulnerable

populations differs by different types of LTCFs. Towards this end, we sought to evaluate facility level characteristics to predict the CDI incidence rate in LTCFs. We examined the incidence of *Clostridium difficile* infection in long term care facilities in the eight county Atlanta area. By comparing the burden of CDI amongst the long-term care facilities in the Atlanta area, we hoped to identify characteristics of LTCFs to create a predictive model for CDI incidence in LTCFs in the Atlanta area.

METHODS

Analysis was done using SAS statistical Software.

Skilled Nursing Facility Cost Report Data

Centers for Medicare and Medicaid Services (CMS) cost report data was used to determine facility level characteristics. Skilled Nursing Facility (SNF) cost reports from fiscal year 2015 were used for analysis as cost reports from fiscal year 2016 were found to be incomplete. The cost reports were analyzed using a SAS statistical software macro, created by Dr. James Baggs, which allowed for extraction of specific variables of interest from the cost reports. The following variables were considered variables of interest for the facility level analysis: component name, provider CMS Certification Number (CCN) state, city, address, zip code, county, urban or rural, SNF bed size, SNF bed days, SNF inpatient days, SNF discharges, SNF average length of stay, and SNF admissions. The provider CCN is a unique 6-digit facility identifier, which corresponds to the state of the facility and the type of services that have been Medicare certified for that facility.

CDI incident case data

Surveillance for CDI in Georgia is performed by the Georgia's Emerging Infections Program (GEIP) in the eight county Atlanta area, comprised of the following counties: Clayton, Cobb, DeKalb, Douglas, Fulton, Gwinnett, Newton, and Rockdale counties, known as Health District 3 (hd3). 2016 incident CDI case data was retrieved from the GEIP. An incident CDI case is defined as an initial toxin positive test with no previous positive test in the 8 weeks prior. Long Term Care Facility Onset (LTCFO) CDI is defined as a stool collected at a LTCF or Long Term Acute Care Facility (LTACH) or

admitted from a LTCF less than 4 days prior to the stool collection. Due to the large number of incident CDI cases, medical record access and abstraction of case data are limited to only sampled incident CDI cases (i.e., 1 of 3 adult incident cases are randomly selected for full chart review, all pediatric cases are selected).

Access to connectivity values

The GA EIP also had access to summary data from Center for Medicare and Medicaid Services (CMS) on inter-facility transfers of Medicare patients. The GA EIP analyzed the number of Medicare patients admitted or transferred directly and indirectly (within a twelve-month time period) between each pair of facilities using R to calculate connectivity metrics, including weighted indegree, weighted outdegree, indegree, outdegree, and betweeness (see definitions below). Finally, using cluttering algorithms available in igraph, facilities in the network were clustered into groups which had predominately transfers between facilities in the cluster, compared to transfers between facilities outside of a cluster. These were provided by Dr. Scott Fridkin from a parallel study ongoing in the GA EIP.

Identification of Facilities

Facilities of interest were identified for the eight-county metro-Atlanta area. The master list of facilities was created from the CMS 2015 cost reports. The list was narrowed to facilities in the eight county surveillance area (HD3). This facility list included 68 facilities: four facilities were removed from the facilities of interest due to incomplete connectivity values and GA EIP records for these facilities, leaving 64 LTCFs available for analysis.

Cases Aggregated to Facility

An epidemiologic categorization was assigned to each CDI case. These categories included: hospital onset (HO, defined as CDI test date was >3 days into hospital stay), community onset healthcare facility associated (COHCFA, defined as CDI test date ≤3 days into hospital stay, but overnight stay in a healthcare facility in preceding 12 weeks), community associated (CA, defined as CDI test date ≤3 days into hospital stay, but no overnight stay in any healthcare facility in preceding 12 weeks), LTCF onset (LTCFO, defined as CDI test date ≤3 days into hospital stay or in LTCF, and residing in LTCF in the 4 days prior to test date), and unknown. The unknown category are the cases that were not sampled, and therefore cannot be accurately categorized or traced to an origin facility. Only LTCFO cases were the cases used for analysis. Using abstracted data regarding recent facility transfers, each LTCFO case was assigned an origin LTCF based on the LTCF where they were located 4 days prior to their stool sample.

Maximizing Case Counts

Two different methods were used in an effort to maximize the number cases attributed to origin facilities. First, GA EIP CDI incidence data was accessed for the first 6 months of the 2017 surveillance year. However, this data was incomplete and the percent of LTCFO cases were not comparable to the corresponding months of 2016, and believed to be only partially complete by GA EIP staff. Another method used to maximize data was looking at non-sampled cases. Only the laboratory used for CDI test is known for all non-sampled cases; location of onset of the patient is not reliably abstracted but was knows from some non-sampled cases. However, adding non-sampled

cases that had a positive stool sample submitted from a LTCF in the laboratory record did not improve case counts for facilities that previously had zero cases; due to a lack of maximizing LTCF attribution and an unreliable attribution method, this method was not pursued further. In the end, we limited the attribution of origin facility to only those CDI cases that were sampled for full chart review.

CDI rate

To account for GEIP's sampling technique, facilities with zero incident cases were given a case count of 1 and facilities with 1 or more incident CDI cases, had their raw case count multiplied by three to account for the 1:3 sampling strategy. This created a final case count value for each facility in the master facility list. The final case count data was used as the numerator for the CDI rate. The variable encompassing SNF bed days was used from the CMS cost report 2015 data to create a CDI rate. The CDI rate was calculated per 10,000 skilled nursing facility bed days. Two facilities were removed from the analysis due to missing cost report data.

Descriptive Analysis

Descriptive analysis was conducted to identify case-level characteristics. Characteristics include: sex, age, number of hospitalized patients, race, underlying illness, severity of illness, location of stool collection, death, antimicrobial therapy, and previous CDI episode. Facility specific characteristics described included number of SNF admissions, number of CDI cases, CDI rate, number of beds, number of bed days, nursing, connectivity metrics, average length of stay, facility cluster variable, and county. Connectivity metrics include weighted indegree, indegree, weighted outdegree,

outdegree, and betweenness. The incident CDI case count per LTCF and the estimated incidence (accounting for sampling) were illustrated (Figure 1). Univariate analysis was performed comparing facility specific characteristics between facilities that had zero cases (n=17) and facilities that have at least one case (n=47) with corresponding p-values.

Description of Variables

Charlson Comorbidity Index assess the one-year mortality for a patient by creating an index to account for comorbid conditions (21).

Weighted indegree is the weighted sum of patients coming into the index facility. It measures the number of patients the index facility receive from other facilities and it is weighted on the number of different facilities these patients come from.

Indegree is the sum of patients coming into the index facility. It measures the number of patients the index facility receives from other facilities.

Weighted outdegree is the weighted sum of patients leaving the index facility. It measures the number of patients the index facility sends to other facilities and it is weighted on the number of different facilities these transfers represent.

Outdegree is the sum of patients leaving the index facility. It measures the number of patients the index facility sends to other facilities.

Betweeness is a centrality measure that quantifies how often the index facility acts as a connector between two other facilities.

Nursing ratio was calculated by summing Licensed Practical Nurse (LPN) hours for a given facility and Registered Nurse (RN) hours for a given facility and dividing this value

by the total nurse staffing hours per resident per day for that facility. It assess the proportion of clinical providers that have higher levels of nurse training, a proxy for quality of nursing coverage for a given facility.

Bed size is a value provided by the CMS cost reports. This variable was eliminated in the creation of a model since it is highly associated with bed days, which is used in the creation of CDI rate.

Bed Days is a value provided by the CMS cost reports referring to the sum of days for all the patients in the facility for the year.

Average Length of Stay is a continuous variable that is provided in the CMS cost reports. This variable identifies the average length of stay for a LTCF resident in that facility for a given year. This variable is reclassified into an ordinal variable titled "average length of stay-ordinal" in the following categories:

1= less than 8 weeks

2= 8 weeks to less than 12 weeks

3= 12 weeks to less than 6 months

4= 6 months and over

Admissions is a continuous variable that is provided in the CMS cost reports. This variable identifies the average number of LTCF admissions for a given facility. It is reclassified, by quartiles, into an ordinal variable titled "number of admissions-ordinal" in the following categories:

1= less than 224

2= 224 to 354

3= 355 to 547

4= 548 and over

Cluster is a variable that is created based on set tolerances to the degree of connectivity. A group of facilities with more frequent transfers is considered a cluster. Clusters are defined based on their area and direct transfers. This variable is a nominal variable with to define the 9 main clusters identified using igraph clustering algorithms (values 1-9).

Correlation Coefficients

Facility level characteristic variables were assessed to determine their individual correlation to the CDI rate per 10,000 bed days. Scatter plots were created to depict the correlation between each facility characteristic and the CDI rate per 10,000 bed days.

Predictive Modeling

A negative binomial distribution is used to create a model for the aggregate count data. The offset is identified as the log of SNF bed days as each facility had a different number of bed days. The facility level variable bed size is removed from prospective variables as it is associated with the denominator variable bed days in the CDI rate. Eligible facility level variables were analyzed in a simple regression model with CDI rate as the outcome. The significance of each variable as a predictor for CDI rate was assessed. The variable that had the most significant Wald Chi-square value was added to the model first. Forward selection continued in this manner, adding one variable at a time to the model until there were no more significant predictors to add. Backwards selection was used to verify the model. Lastly, the parameter estimates were exponentiated to allow for easier interpretation.

RESULTS

Categorization of CDI cases

In 2016, there were 4,577 reported CDI cases in Atlanta, GA. Of these cases, 37.05% of the cases were randomly selected via the 1:3 sampling technique for further case review. After the sampling, there were 178 LTCFO cases, making up 10.5% of the sampled cases. For 155 of these cases (86.5%), an origin facility was identified and was successfully matched to the CMS cost reports. The remaining 23 facilities (13.5%) could not be attributed to a specific facility (**Table 1**).

Case level analysis of characteristics

There were 154 different individuals representing the 155 CDI cases that could be attributed to a specific facility (**Table 2**). The median age of these SNF residents is 76, with 60% of the cases being female. Of the 155 cases, 47.7% of the cases were white race, while 40.7% were black race. Around 25% of the 155 cases had a previous CDI episode more than 8 weeks before the reported episode (i.e., a prior infectious episode) and 44.5% of the reported cases were hospitalized at the time of the stool collection or within 7 days of their stool collection. Around 10% of patients had a Charlson comorbidity index above zero.

Facility level analysis of characteristics

Among the 64 facilities, the mean SNF CDI rate was 1.8 per 10,000 bed days (range 0.1 - 15.6 per 10,000 bed days) (**Table 3).** The median number bed days was 46,417.5 (range: 5,856-98,820) and the median average length of stay was 129.4 days

(range: 19.8-983.4). Over half of the facilities were located in either Fulton (26.6%) or DeKalb (25.0%). The 155 cases were attributed to 47 unique facilities, leaving 17 (26.6%) facilities without any reported LTCFO CDI case in 2016.

Comparing facilities with and without CDI infection

Forty-seven (73.4%) of the facilities of interest had at least one associated LTCFO CDI case; these facilities differed from the 17 facilities that had zero (**Table 4**). There was a significant difference for the following variables: weighted indegree (p=0.008), indegree (p= 0.002), betweeness (p=0.040), weighted outdegree (p=0.001), outdegree (p=0.001), number of bed days (p=0.049), number of admissions (p=0.007), average length of stay classified as an ordinal variable (p=0.039), and number of admissions classified as an ordinal variable (p=0.017). There was no difference between the groups for the county (p=0.603) or nursing ratio (p=0.157). Location of facility, either by county or by cluster did not differ between the group of facilities that had zero cases and the group of facilities that had at least one case.

Adjustment for sampling

After adjusting for sampling, CDI rates ranged from 0.1 per 10,000 bed days to 15.6 per 10,000 bed days. A facility's relative ranking in CDI burden defined by incidence (taking into account SNF bed-days) generally reflected that defined by raw case counts; exceptions can be identified by comparing the histograms side by side (**Figure 1**). CDI cases for each facility ranged from zero cases to 19 cases (**Figure 1a**); CDI rate per 10,000 bed days for each facility ranged from A to B. The overall trends of

these graphs are comparable; however, there are noticeable variations between the two graphs due to the difference in bed days among the facilities.

Pearson correlation coefficients

Facility level characteristics including weighted indegree, indegree, weighted outdegree, outdegree, betweeness, nursing ratio, bedsize, average length of stay, average length of stay classified as ordinal, number of admissions, and number of admissions classified as ordinal were correlated against the outcome variable, CDI rate. None of the continuous independent variables correlated strongly with CDI incidence, all Pearson Correlation Coeficients were <0.5 (**Table 5**). However, many of these week correlations were statistically significant, including weighted indegree (p=0.001), indegree (p=0.001), outdegree (p<0.001, betweenness (p=0.003), average length of stay categorized ordinally (p< 0.001, number of admissions (p=0.001), and number of admissions categorized ordinally (p=0.021). Scatterplots illustrate these correlations (**Figure 2**).

Negative Binomial model

Each eligible variable was assessed in a single linear regression model to predict the outcome CDI rate (**Table 5).** Average length of stay categorized in an ordinal fashion was the single most significant predictor of CDI rate (p<0.001). After this variable was added to the model, number of admissions categorized in an ordinal manner was the next most significant predictor (p<0.001). Lastly, outdegree was the final variable to included in the model (p=0.007). When backwards selection was conducted on the same eligible variables, the model was identical to the forward selection model. The final model consists of average length of stay as an ordinal variable (p<0.001), number of admissions

as an ordinal variable ($p<0.001$), and outdegree ($p<0.001$) (**Table 6).** The expected increase in the log CDI rate for a one-unit increase in average length of stay was found to be -1.18. The exponentiated estimate values yielded incident rate ratio for the average length of stay of 0.31 (confidence limit: 0.21, 0.44). The incident rate ratio for number of admissions was calculated as 0.40 (confidence limit: 0.27, 0.60). Lastly, an incident rate ratio of 1.09 (confidence limit: 1.02, 1.16) was determined for outdegree.

DISCUSSION

CDI occurred in residents of skilled nursing facilities across the entire Atlanta metropolitan area, not limited to any specific county or group of facilities within the area. CDI occurred in the 73% of these facilities during 2016. This is suspected to be an underestimation of the true CDI counts since 63% of the CDI cases could not be classified, as they were not sampled. Moreover, facilities without any reported cases had a different profile of resident care pathways compared to facilities with at least one case. Facilities without cases tended to admit fewer residents, admit from fewer acute care facilities, keep residents for longer periods of time, and discharge residents to fewer facilities. In contract, the nursing staffing mix was not related to CDI rate; therefore, increasing nursing staff hours would have a null effect on CDI rate and should not be targeted for intervention to decrease CDI rates. CDI cases penetrated every cluster and county, and did not statistically differ between the facilities that had zero cases and the facilities that had at least one case.

Average length of stay proved to be the most significant predictor of CDI rate in our study. A retrospective study in New Jersey found that increasing an individual's length of stay in an acute care hospital by 1 day increased their probability of becoming infected with an HAI by 1.37 percent (22). We found a different relationship in skilled nursing facilities. We found a 69% increase in the rate of CDI for each unit decrease in the average length of stay variable; in the analysis, the unit is moving from high average length of stay to the next quartile for lower average length of stay. Therefore, as length of stay increases, the rate of CDI decreases, while holding the other independent predictors constant. For example, a facility that has an average length of stay less than 8

weeks has a 69% higher rate of CDI compared to a facility that has an average length of stay between 8 and 12 weeks. We believe our analysis is supported by others observations of CDI in skilled nursing facilities. Hunter et. al documented that the vast majority of CDI among SNF residents occur in the first 8 weeks after hospital discharge, suggesting the first few weeks of post-acute care are the highest risk for SNF residents (22). Those, facilities caring for patients with shorter lengths of stay are likely caring for patients at higher risk for CDI, residents that are sicker or of higher acuity.

Other independent facility characteristics were less critical for the predictive model, but of interest. Outdegree assesses the number facilities that residents are admitted to when they require admission to an acute care facility. We found that when controlling for number of admissions and length of stay, that CDI rate increased 1.09 (confidence interval: 1.02, 1.16) times for every unit increase in outdegree. Facilities that send patients to more facilities may do so for a number of reasons. One may be because of infection, including CDI. Other reasons may include lack of resources, lack of staffing, too many patients, or a request to transfer from the patients. These facilities may have higher rates of CDI due to a lack of ability to deal with the demand of the patients and the number of sick patients these facilities have and need to transfer to other facilities. A prospective cohort study analyzing how patient transfer networks contribute to the spread of infection found 90% of all the transfers occur in a span less than 200km and found the correlation between CDI and transfer network structure to be 0.47 (23). They stated that CDI rates were likely higher for facilities that receive a lot of patients because these facilities were receiving some of the sickest patients from many different facilities. Our data suggest that skilled nursing facilities connected to more hospitals tend to have higher

CDI rates, supporting the evolving evidence that connectivity values are an important predictor in assessing the spread of pathogens and infections in a hospital network.

Our final model included an inverse relationship between admissions and CDI incidence, which was opposite of what was observed in univariate analysis. This suggests that when accounting for bed-days, average length of stay and outdegree – the relationship changes for number of admissions. The scatter plot for admissions and CDI incidence suggests the relationship is not linear in nature, and we suspect that when accounting for these other variable, admissions is a proxy measure for some unmeasured variable that is worth pursing further. Therefore, further analysis of this variable is necessary to assess if it should be included as a variable in the model.

Limitations

Due to the sampling technique conducted by the GAEIP on CDI data, an estimation technique was applied to the one third of cases used in the analysis for calculation of CDI incidence. However, this method may not have correctly estimated the actual number of CDI cases that occurred at each facility. Rather than relying on an estimated incidence, another year of data would allow for an accurate calculation of incidence at each facility.

Related to the incidence calculations, there is some ambiguity about the best method for calculating incidence of CDI. Use of the bed-days as the denominator presumes that each day of care in a SNF is of equal risk. However several sources suggest that the period of greatest risk for CDI among SNF residents is in the post-acute care period, or the first 12 weeks of residence at a SNF. SNF predictors of CDI incidence

may differ if the analysis were to be repeated and incidence was defined as the number of CDI cases per 100 admissions, reducing the impact of longer lengths of stay on the incidence. These types of exploratory analysis are needed before firm conclusions can be drawn about the relative importance of the independent predictors on CDI incidence in our network.

Finally, CCNs were used to match facilities from the GAEIP dataset and the CMS cost report dataset; however, these numbers did not always align. For example, the datasets may have had facilities with the same CCN, but the facilities had different names. Although the facilities were still matched by CCN, if these facilities were in fact different than the number of bed days may differ causing an incorrect CDI rate. Additionally, there were facilities that were not in the CMS cost report data and the connectivity data, which were removed from the analysis. However, this number was small, and unlikely to have affected the main conclusions of this study.

Conclusions

CDI incidence within LTCFs in the Greater Atlanta Area can be predicted using the facility characteristic values for average length of stay, number of admissions, and no. of facilities that receive the residents when transferred out. Connectivity within a healthcare network is quickly becoming a widespread topic in regard to hospital acquired infections. Further research is necessary on the LTCF healthcare network to assess the strength of connectivity values in predicting infections.

References

1. Hall I, O'Toole E. Intestinal flora in newborn infants with a description of a new pathogenic anaerobe, *Bacillus difficilis*. *Am J Dis Child.* 1935;49:390.
2. George RH, Symonds JM, Dimock F, et al. Identification of Clostridium difficile as a cause of pseudomembranous colitis. *British Medical Journal*. 1978;1(6114):695.
3. Bartlett JG, Moon N, Chang TW, et al. Role of Clostridium difficile in antibiotic-associated pseudomembranous colitis. *Gastroenterology*. 1978;75(5):778-82.
4. McFarland LV, Mulligan ME, Kwok RY, et al. Nosocomial acquisition of Clostridium difficile infection.*N Engl J Med.* 1989; 320:204-21.
5. Shim JK, Johnson S, Samore MH, et al. Primary symptomless colonisation by Clostridium difficile and decreased risk of subsequent diarrhoea. *Lancet*. 1998;351:633-636.
6. Kyne L, Warny M, Qamar A, et al. Asymptomatic carriage of Clostridium difficile and serum levels of IgG antibody against toxin A. *N Engl J Med.* 2000;342:390-397.
7. Samore MH, DeGirolami PC, Tlucko A, et al. Clostridium difficile colonization and diarrhea at a tertiary care hospital. *Clin Infect Dis.* 1994;18: 181-187.
8. Pépin J, Saheb N, Coulombe MA, et al. Emergence of fluoroquinolones as the predominant risk factor for *Clostridium difficile*-associated diarrhea: a cohort study during an epidemic in Quebec. *Clin Infect Dis*. 2005;41:1254-1260.
9. Lessa FC, Gould CV, McDonald LC. Current status of *Clostridium difficile* infection epidemiology. *Clin Infect Dis*. 2012;55:Suppl 2:S65-S70. (doi: 10.1093/cid/cis319).

10. Simor AE Yake SL Tsimidis K. Infection due to *Clostridium difficile* among elderly residents of a long-term-care facility. *Clin Infect Dis.* 1993; 17:672–8.

11. Lessa FC, Mu Y, Bamberg WM, et al. Burden of Clostridium difficile infection in the United States. *N Engl J Med.* 2015;372(9):825–34 .(doi: 10.1056/NEJMoa1408913).

12. Pépin J, Valiquette L, Cossette B. Mortality attributable to nosocomial *Clostridium difficile*-associated disease during an epidemic caused by a hypervirulent strain in Quebec. *CMAJ*. 2005; 173 (9): 1037-42. (doi: 10.1503/cmaj.050978).

13. Loeb M, Simor AE, Landry L, et al. Antibiotic use in Ontario facilities that provide chronic care. *J Gen Intern Med.* 2001; 16: 376-83. (doi: 10.1046/j.1525-1497.2001.016006376.x).

14. Lee YL, Trupp LD, Lee R, et al. Infection surveillance and antibiotic utlization in a community-based skilled nursing facility. *Aging Clin Exp Res* 1996; 8: 113-22.

15. Mylotte JM. Measuring antibiotic use in a long-term care facility. *Am J Infect Control* 1996: 24: 174-9.(doi: 10.1016/S0196-6553(96)90009-7).

16. Simor AE, Yake SL, Tsimidis K. Infection due to *Clostridium difficile* among elderly residents of a long-term-care facility. *Clin Infect Dis.* 1993; 17:672–8. (doi: 10.1111/j.1532-5415.1990.tb03539.x)

17. Thomas DR, Bennett RG, Laughon BE, et al. Postantibiotic colonization with *Clostridium difficile* in nursing home patients. *J Am Geriatr Soc* 1990; 38: 415-20.

18. Fernandez-Gracia, Onnela JP, Barnett ML, et al. Influence of a patient transfer network of US inpatient facilities on the incidence of nosocomial infections. *Scientific Reports*. 2017; 7: 2930. (doi: 10.1038/s41598-017-02245-7).

19. Hunter JC, Mu Y, Dumyati GK, et al. Burden of Nursing Home- Onset *Clostridium difficile* Infection in the United States: Estimates of Incidence and Patient Outcomes. *Open Forum Infectious Diseases*. 2016; 3(1): 1-8. (doi: 10.1093/ofid/ofv196)

20. Pawar D, Tsay R, Nelson DS, et al. Burden of *Clostridium difficile* infection in long-term care facilities in Monroe County, New York. *Infect Control Hosp Epidemio.l* 2012; 33: 1107-12. (doi: 10.1093/ofid/ofv196)

21. Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987; 40(5): 373-83.

22. Hassan M, Tuckman HP, Patrick RH, et al. Hospital length of stay and probability of acquiring infection. *Int J Pharm Healthc Mark.* 2010; 4 (4): 324-338. (doi: 10.1108/17506121011095182)

23. Fernandez- Gracia J, Onnela JP, Barnett M, et al. 2017. Spread of pathogens in the patient transfer network of US hospitals. In: Lee D., Lin YR., Osgood N., Thomson R. (eds) Social, Cultural, and Behavioral Modeling. SBP-BRiMS 2017. Lecture Notes in Computer Science, vol 10354. Springer, Cham. (doi:10.1007/978-3-319-60240-0_33).

Tables

Table 1: Categorization of CDI cases (n=4577) in Atlanta, GA in 2016.

CDI Categorization	No.	%
Unknown (not-sampled)	2881	63.0
Hospital Onset (HO)	393	8.6
Community Onset Healthcare Facility Associated (COHCFA)	301	6.6
Community Associated (CA)	824	18.0
Long-term Care Facility Onset (LTCFO)	178	3.9
LTCFO attributed to a known facility*	155	3.4

*facility was identified and successfully matched CMS cost reports

Table 2: Characteristics of the skilled nursing facility residents from the CDI cases that can be attributed to a specific facility (n=155) in Atlanta, GA in 2016.

Characteristics	
Age, median yrs (range)	76 (68)
Individuals representing 155 incident cases, count	154
	No. Cases (% total)
Sex	
Male	62(40)
Female	93 (60)
Hospitalized at time of stool collection or within 7 days after stool collection	69 (44.5)
Race	
American Indian/ Alaska Native	1 (0.7)
Asian	0
Black	63 (40.7)
Native Hawaiian/ Other Pacific Islander	0
White	74 (47.7)
Unknown	18 (11.6)
Death	12(7.8)
Underlying illness	
Congestive heart failure	3 (1.9)
Diabetes mellitus	9 (5.8)
Chronic pulmonary disease	4 (2.6)
Charlson comorbidity index	
0	140 (90.3)
1	1 (0.7)
2	4 (2.6)
3	6 (3.9)
4	2 (1.3)
5	1 (0.7)
6	1 (0.7)
Location of Stool Collection	
Hospital Inpatient	54 (34.8)
Long Term Acute Care Hospital	0
Emergency Room	18 (8.4)
Long Term Care/ SNF	88 (56.8)
Outpatient	0
Other	0
Unknown	0
Previous CDI episode (>8 weeks prior to this episode)	39 (25.2)

Table 3: Characteristics of skilled nursing facilities in Atlanta, GA in 2016 (n=64).

	Total across all facilities Sum (N)	**Facility- specific Mean (SD), median (range)**
Number of SNF admissions (continuous)	26,799	418.7 (285.9), 354.5 (10.0-1107.0)
CDI cases	155	2.4 (3.3), 1.5 (0-19.0)
SNF CDI rate per 10,000 Patient-Days	NA	1.8 (2.4), 1.0 (0.1-15.6)
Number of Beds	2684	136.4 (51.8), 129.5 (16.0-268.0)
SNF Bed days	3,097,002	48,390.7 (19,909.1), 46,417.5 (5,856.0-98,820.0)
Nursing Ratio	NA	42.3 (4.6), 42.5 (31.4-54.0)
Connectivity Metrics	NA	
Indegree		15.0 (5.5), 15.5 (3.0-29.0)
Weighted indegree		131.4(129.1), 74.0(3.0-573.0)
Outdegree		9.9 (4.6), 10.0 (0-24.0)
Weighted outdegree		58.6 (38.4), 54.0 (0-179.0)
Betweeness		13.2 (14.0), 8.2 (0-61.2)
Average Length of Stay (continuous)	NA	171.8 (183.6), 129.4 (19.8-983.4)
		No. facilities (% total)
Facility presence of CDI case	NA	
0 reported CDI case		17 (26.6)
≥ 1 reported CDI case		47 (73.4)
Facility breakdown by county	NA	No. facilities (% total)
Clayton		4 (6.3)
Cobb		13 (20.3)
DeKalb		16 (25.0)
Douglas		1 (1.6)
Fulton		17 (26.6)
Gwinnett		9 (14.1)
Newton		2 (3.1)
Rockdale		2 (3.1)

*Nursing Ratio= (Reported LPN Staffing Hours per Resident per Day+ Reported RN Staffing Hours per Resident per day) / (Reported Total Nurse Staffing Hours per Resident per Day. Nursing Ratio had two facilities that did not have data for the variable

Table 4: Comparison of skilled nursing facilities with and without LTCFO CDI incident cases in Atlanta, GA in 2016.

	Total **Mean (SD), median (range)**	**Facilities with 0 Cases (N=17)** **Mean (SD), median (range)**	**Facilities with ≥ 1 case (N=47)** **Mean (SD), median (range)**	**p-value**
Weighted indegree	131.4(129.1), 74.0 (3.0-573.0)	60.7 (59.6), 47.0 (3.0-218.0)	157 (138.0), 103.0 (9.0-573.0)	<0.01
Indegree	15.0 (5.5), 15.5 (3.0-29.0)	11.4 (4.5), 12.0 (3.0-19.0)	16.3 (5.3), 16.0 (8.0-29.0)	<0.01
Betweeness	13.2 (14.0), 8.2(0.0-61.2)	7.2 (11.0), 3.5 (0.0-46.9)	15.3 (14.4), 11.3 (0.6-61.3)	0.04
Weighted outdegree	58.6 (38.4), 54.0 (0.0-179.0)	32.4 (27.3), 26.0 (0.0-104.0)	68.1 (37.6), 60.0 (4.0-179.0)	<0.01
Outdegree	9.9 (4.6), 10.0 (0.0-24.0)	6.8 (3.7), 6.0 (0.0-16.0)	11.0 (4.4), 11.0 (2.0-24.0)	<0.01
Nursing Ratio	42.3 (4.6), 42.5 (31.4-54.0)	40.9 (3.9), 42.2 (33.4-47.2)	42.8 (4.8), 42.6 (31.4-54.0)	0.16
SNF Bed Days	48,390.7 (19,909.1), 46,417.5 (5,856-97,820)	40,228.6 (18,794.0), 36,500.0 (5,856-75,396)	51,342.9 (19,659.0), 50,005.0 (9,882-97,820)	0.05
SNF Admissions	418.7 (285.9), 354.5 (10.0-1,107.0)	259.7 (189.8), 247.0 (10.0-656.0)	476.3 (294.4), 396.0 (38.0-1,107.0)	0.01
SNF Average Length of Stay	171.8 (183.6), 129.4 (19.8-983.4)	243.8 (227.7), 160.7 (52.5-819.6)	145.8 (159.8), 95.8 (19.8-983.4)	0.06
	No. (% of total)	**No. (% of total)**	**No. (% of total)**	
Average Length of Stay **1= < 8 weeks** **2= 8 weeks - < 12 weeks** **3= 12 weeks - < 6 months** **4= 6 months and over**	1= 10 (15.6) 2=13 (20.3) 3= 23 (35.9) 4= 18 (28.1)	1= 1 (5.9) 2= 1 (5.9) 3= 8 (47.1) 4= 7 (41.2)	1= 9 (19.2) 2= 12 (25.5) 3= 15 (31.9) 4= 11 (23.4)	0.03
Number of Admissions **1= less than 224** **2= 224 to 354** **3= 355 to 547** **4= 548 and over**	1= 16 (25.0) 2=16 (25.0) 3=16 (25.0) 4=16 (25.0)	1= 8 (47.1) 2= 5 (29.4) 3= 1 (5.9) 4= 3 (17.7)	1= 8 (17.0) 2= 11 (23.4) 3= 15 (31.9) 4= 13 (27.7)	0.02
County Breakdown	**No. (% of total)**	**No. (% county)**	**No. (% county)**	0.60
Clayton	4 (6.3)	2 (50.0)	2 (50.0)	
Cobb	13 (20.3)	2 (15.4)	11 (84.6)	
DeKalb	16 (25.0)	3 (18.8)	13 (81.3)	
Douglas	1 (1.6)	0 (0.0)	1 (100.0)	
Fulton	17 (26.6)	6 (35.3)	11 (64.7)	
Gwinnett	9 (14.1)	3 (33.3)	6 (66.7)	
Newton	2 (3.1)	1 (50.0)	1 (50.0)	

Rockdale	2 (3.1)	0 (0.0)	0 (100.0)	

Cluster	**No. (% of total)**	**No. (% cluster)**	**No. (% cluster)**	0.13
1	6 (9.4)	2 (33.3)	4 (66.7)	
2	14 (21.9)	2 (14.3)	12 (85.7)	
3	4 (6.6)	1 (25.0)	3 (75.0)	
4	13 (20.3)	2 (15.4)	11 (84.6)	
5	9 (14.1)	3 (33.3)	6 (66.7)	
6	3 (4.7)	1 (33.3)	2 (66.7)	
7	8 (12.5)	2 (25.0)	6 (75.0)	
8	6 (9.4)	3 (50.0)	3 (50.0)	
9	1 (1.6)	1 (100.0)	0 (0.0)	

Table 5: Pearson correlation coefficient for facility level metrics against CDI rate per 10,000 bed days and estimates using facility-specific metric as a single predictor for CDI rate per 10,000 bed days in skilled nursing facilities in Atlanta, GA in 2016.

	Pearson Correlation		SLR model estimates		
Facility-specific metric	coefficient	p-value	Estimate	Wald Chi-Square	p-value
Weighted indegree	0.42	<0.01	<0.01	6.05	0.01
Indegree	0.40	<0.01	0.07	6.50	0.01
Weighted outdegree	0.16	0.22	<0.01	0.01	0.93
Outdegree	0.44	<0.01	0.08	6.12	0.013
Betweeness	0.36	<0.01	0.02	4.81	0.03
Nursing Ratio	0.16	0.23	0.03	0.66	0.42
Bedsize	-0.22	0.08	NA	NA	NA
Average Length of Stay	-0.24	0.05	<-0.01	3.54	0.06
Average Length of Stay- Ordinal	-0.44	<0.01	-0.65	22.26	<0.01
Number of Admissions	0.39	<0.01	<0.01	3.84	0.05
Number of Admissions- Ordinal	0.29	0.02	0.18	1.58	0.21

Table 6: Estimates and incident rate ratios using a negative binomial multivariate model with offset of log bed days predicting the CDI incidence rate for skilled nursing facilities in Atlanta, GA in 2016.

Parameter	**Estimate**	**Standard Error**	**Wald 95% Confidence Limits**	**Wald Chi-Square**	**p-value**	**IRR**	**Confidence limits**
Intercept	-5.70	0.89	-7.44, -3.96	31.36	<0.01	NA	NA
Average Length of Stay*	-1.18	0.19	-1.54, -0.81	40.38	<0.01	0.31	0.21, 0.44
Number of Admissions*	-0.91	0.20	-1.31, -0.52	20.40	<0.01	0.40	0.27, 0.60
Outdegree	0.09	0.03	0.02, 0.15	7.37	<0.01	1.09	1.02, 1.16
Dispersion	0.30	0.14	0.12, 0.76				

*variables categorized in an ordinal fashion

Figures

Figure 1 a: Number of incident LTCFO CDI cases in Atlanta, GA in 2016.

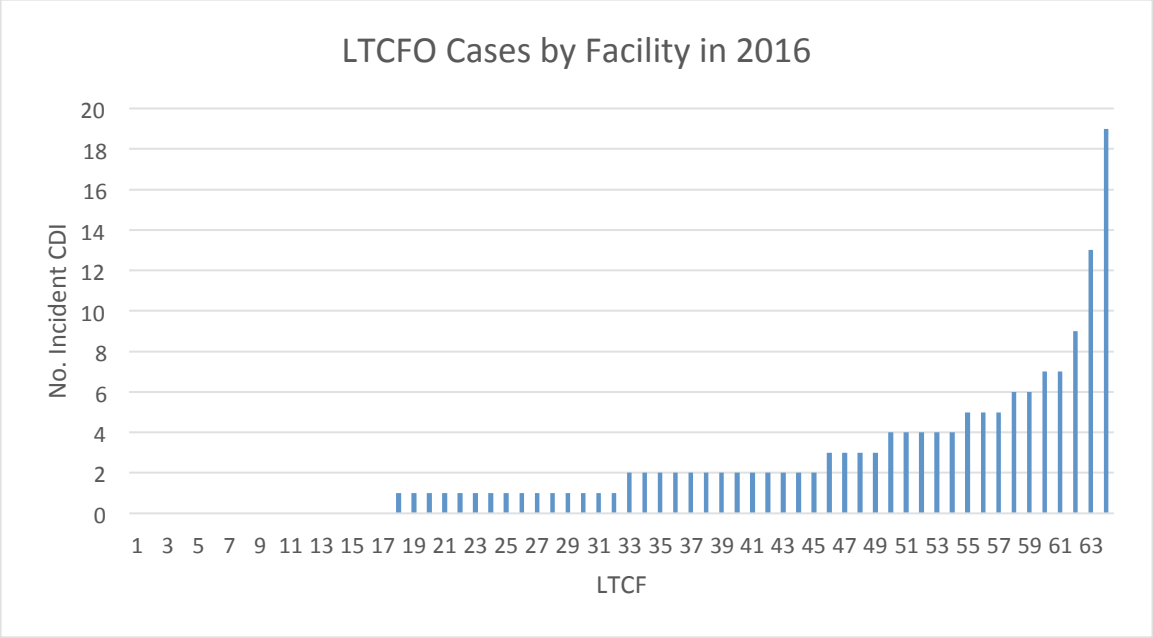


Figure 1 b: CDI rate for LTCFO cases in Atlanta, GA after adjusting for sampling in 2016.

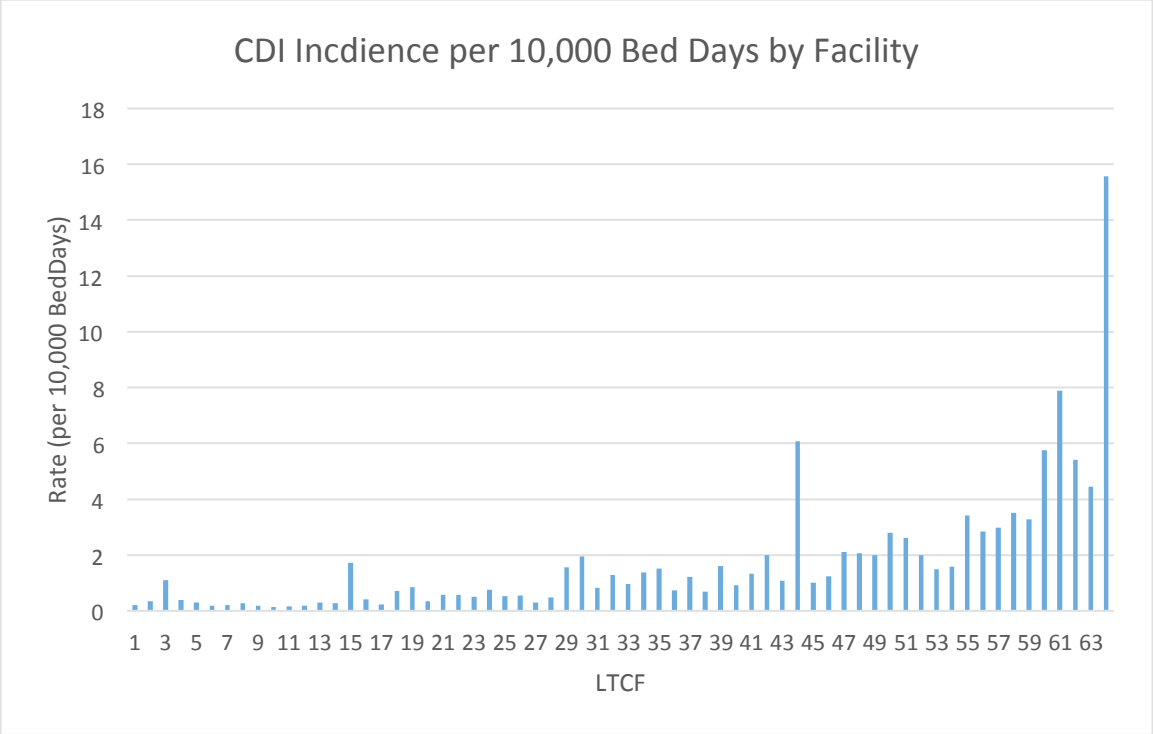


Figure 2 a: Scatter plots depicting the correlations between connectivity metrics and CDI rate per 10,000 bed days in skilled nursing facilities in Atlanta, GA in 2016.

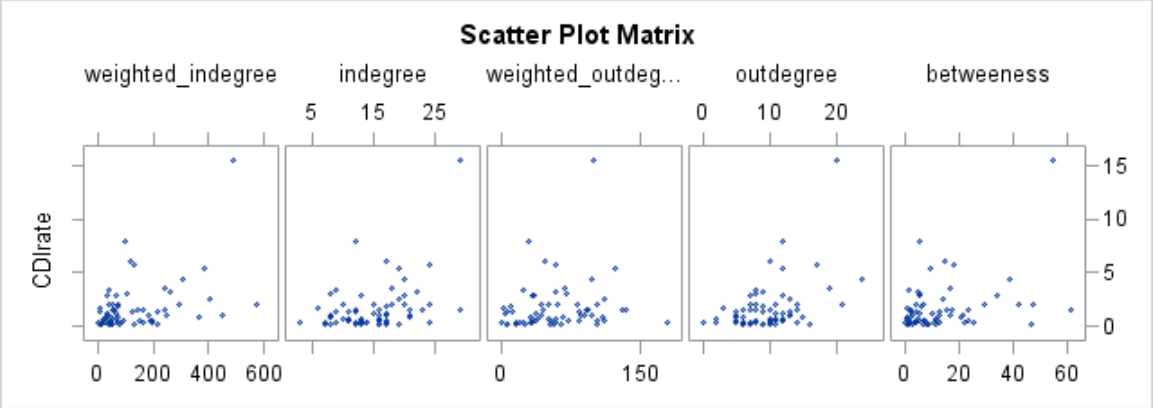
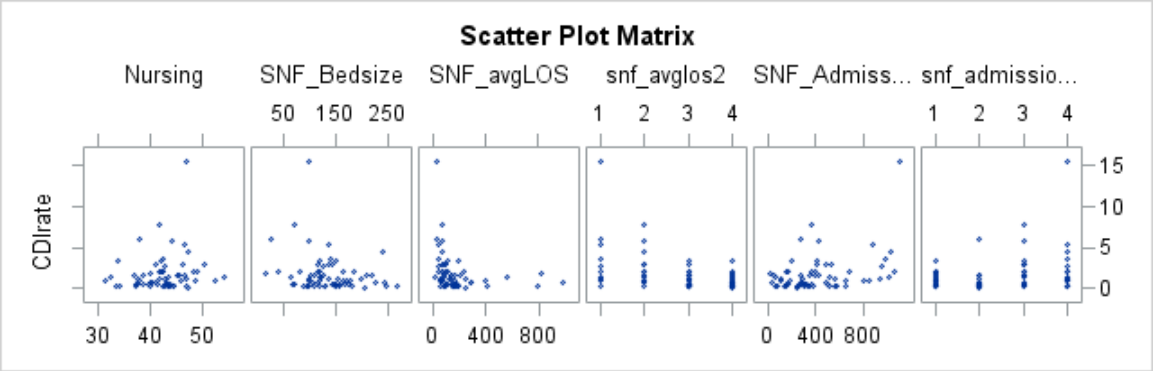


Figure 2 b: Scatter plots depicting the correlation between facility specific variables and CDI rate per 10,000 bed days in skilled nursing facilities in Atlanta, GA in 2016.



Appendices

Appendix A: Acronyms

CA	Community associated
CDI	Clostridium difficile infection
COHCFA	Community onset healthcare facility associated
CCN	Centers for Medicare and Medicaid Services Certification Number
CMS	Centers for Medicare and Medicaid Services
GA	Georgia
GEIP	Georgia Emerging Infections Program
HAI	Hospital acquired infection
HO	Hospital Onset
LPN	Licensed Practical Nurse
LTACH	Long Term Acute Care Hospital
LTCF	Long term care facility
LTCFO	Long Term Care Facility Onset
RN	Registered Nurse
SNF	Skilled Nursing Facility