ORIGINAL ARTICLE

Developing a Clinical Prediction Rule for First Hospital-Onset *Clostridium difficile* Infections: A Retrospective Observational Study

Anne Press, BS;[1] Benson Ku, BS;[1] Lauren McCullagh, MPH;[1] Lisa Rosen, ScM;[2] Safiya Richardson, MD;[1] Thomas McGinn, MD, MPH[1]

BACKGROUND. The healthcare burden of hospital-acquired *Clostridium difficile* infection (CDI) demands attention and calls for a solution. Identifying patients' risk of developing a primary nosocomial CDI is a critical first step in reducing the development of new cases of CDI.

OBJECTIVE. To derive a clinical prediction rule that can predict a patient's risk of acquiring a primary CDI.

DESIGN. Retrospective cohort study.

SETTING. Large tertiary healthcare center.

PATIENTS. Total of 61,482 subjects aged at least 18 admitted over a 1-year period (2013).

INTERVENTION. None.

METHODS. Patient demographic characteristics, evidence of CDI, and other risk factors were retrospectively collected. To derive the CDI clinical prediction rule the patient population was divided into a derivation and validation cohort. A multivariable analysis was performed in the derivation cohort to identify risk factors individually associated with nosocomial CDI and was validated on the validation sample.

RESULTS. Among 61,482 subjects, CDI occurred in 0.46%. CDI outcome was significantly associated with age, admission in the past 60 days, mechanical ventilation, dialysis, history of congestive heart failure, and use of antibiotic medications. The sensitivity and specificity of the score, in the validation set, were 82.0% and 75.7%, respectively. The area under the receiver operating characteristic curve was 0.85.

CONCLUSION. This study successfully derived a clinical prediction rule that will help identify patients at high risk for primary CDI. This tool will allow physicians to systematically recognize those at risk for CDI and will allow for early interventional strategies.

Infect Control Hosp Epidemiol 2016;37:896–900

Clostridium difficile is a spore-forming gram-positive anaerobic bacillus that is the most commonly recognized cause of infectious diarrhea.[1] *Clostridium difficile* infection (CDI) has a significant impact on a patient's clinical course and quality of life. Nosocomial CDI often prolongs hospital stays, with a median increase in length of stay ranging between 2.8 and 16.1 days.[2] According to Rupnik et al,[3] each year approximately 500,000 cases of CDI are acquired in US hospitals and long-term care facilities, which results in 15,000 to 20,000 deaths.

During the past decade, attempts to reduce the rates of CDI have included environmental decontamination, infection control, and education among hospital staff.[4–13] Although prevention of secondary transmissions (ie, the spread of CDI from an infected patient to uninfected patients) is necessary, it alone is not sufficient to reduce its spread.

A clinical prediction rule is a tool that quantifies the effect an individual patient's characteristics have toward the patient's diagnosis, prognosis, or likely response to treatment. These characteristics are based on various components of the history, physical examination, and basic laboratory results. Clinical prediction rules use evidence to guide clinical management by allowing physicians to identify a patient's individual risk of a certain disease on the basis of the patient's personal risk factors. Applying a clinical prediction rule will allow healthcare workers to properly identify patients at high risk for CDI and intervene before these patients become ill. Our study focused on formulating a clinical prediction rule to identify those patients at risk for a primary infection, thus excluding those with a prior history of CDI. By doing so we focused on only those patients whose risk level is uncertain. Those with a previous diagnosis of CDI are clearly a high-risk population regardless of other cofounding factors.

After a thorough literature review, we identified only 5 previous studies that examined the use of clinical risk prediction models specifically created to predict the onset of a

Affiliations: 1. Department of Medicine, Hofstra North Shore-LIJ School of Medicine, Manhasset, New York; 2. Feinstein Institute for Medical Research, Manhasset, New York.

Received January 8, 2016; accepted April 4, 2016; electronically published April 28, 2016

© 2016 by The Society for Healthcare Epidemiology of America. All rights reserved. 0899-823X/2016/3708-0003. DOI: 10.1017/ice.2016.97

https://doi.org/10.1017/ice.2016.97 Published online by Cambridge University Press

primary CDI.[14–18] However, these studies have several limitations for practical use at the patient's bedside. Two studies used length of stay of greater than 7 days as a predictor for CDI onset.[14,15] However, length of stay constantly fluctuates and is unknown at the time of admission. Another study used acute physiology score as a variable,[16] which uses variables that are also not readily available. Garey et al[17] evaluated only a cohort of patients on broad-spectrum antibiotics. Tanner et al[18] applied the Waterlow score to patients to predict their risk of acquiring a primary nosocomial CDI. However, this score is limited by its large number of scored items. Therefore, all of these studies are limited in their practical use. Our current study focused on creating a simple-to-use, adaptable tool that fits well into clinical workflow with a focus on identifying a patient's risk for an initial CDI, upon admission.

METHODS

A retrospective cohort study was performed. We collected all patient data from the electronic health record and administrative data systems from January 1 to December 31, 2013. All patients admitted to the hospital who were above the age of 18 were included in the study. Exclusion criteria included patients with a history of CDI, as defined by an *International Statistical Classification of Disease, Ninth Revision*, billing code, or those with community-acquired CDI on admission (defined as occurring before or within 3 days following admission). After applying the inclusion and exclusion criteria, we identified 61,482 subjects, who were divided into a training set and a validation set. Patient characteristics were defined using electronic and administrative data (*International Statistical Classification of Disease, Ninth Revision*, codes). A positive outcome was defined as the development of CDI while in the hospital, as determined by a positive CDI polymerase chain reaction assay result at least 3 days following hospital admission. Since polymerase chain reaction testing is run only on diarrheal samples, all patients who tested positive had clinical symptoms of CDI. If a patient tested positive for CDI within 3 days of admission the patient was determined to have community-acquired CDI and was excluded from the study. Random sampling was performed by research coordinators during the study to confirm CDI diagnoses.

Descriptive statistics were used to describe demographic and clinical characteristics of the sample. The χ^2 test or Fisher exact test was used to examine the association between these characteristics and hospital-acquired CDI, as appropriate. Logistic regression was used to model hospital-acquired CDI as a function of potential risk factors. Best subsets selection was used as a screening method to identify the best set of predictor variables for the logistic regression model. We performed best subsets selection using SAS, version 9.4 (SAS Institute), by computing the score χ^2 for each model and identified the subset with k variables, where k was the first subset for which the score χ^2 changed from the previous subset by less than 5%.

The β-coefficients for variables significantly associated with in-hospital CDI were used to assign a numerical value to each risk factor and summed to create a total risk score for each patient (1 for $0 < \beta < 1$ and 2 for $1 \leq \beta < 2$). The optimal risk score cutoff for predicting hospital-acquired CDI on the receiver operating characteristic curve was calculated. The optimal point was defined as the risk score closest to the upper left corner on the receiver operating characteristic curve, which represents 100% sensitivity and specificity. The area under the receiver operating characteristic curve (AUC) was also assessed. The model was validated using the validation set, and the misclassification rate, AUC, sensitivity, specificity, and positive and negative predictive values were calculated, along with corresponding 95% binomial confidence intervals. All analyses were performed in SAS, version 9.4. The procedures used were reviewed and approved by the responsible institutional review committee at the authors' institution.

RESULTS

Data for 80,324 admissions were pulled from the electronic health record. After exclusion criteria were applied, 78,945 admissions remained. After accounting for repeated admissions, the total population included in this study was 61,482 unique encounters. The population was then divided into a training set (40,990) and a validation set (20,492).

Demographic Characteristics

There were a total of 40,990 subjects in the training set, with a 0.46% overall incidence of CDI. Most subjects were women (60.7%) and the mean (SD) age was 56.2 (21.1) years. The most common comorbid disease was diabetes mellitus with a prevalence of 19.9%. Antibiotic usage was present among 40.3% of the population (Tables 1 and 2).

C. difficile Infection Clinical Prediction Rule

The following variables were deemed nonsignificant and removed from the model using best subsets selection: race, hypertension, cerebral vascular disease, chronic obstructive pulmonary disease, intensive care unit admission, coronary artery disease, diabetes, obesity, hypothyroidism, asthma, steroid usage, proton-pump inhibitor, chronic kidney disease, Parkinson disease, atrial fibrillation, gender, cardiomyopathy, liver disease, chronic pancreatitis, and diverticulitis.

The factors included in the final tool were age, admission in the past 60 days, mechanical ventilation, dialysis, history of congestive heart failure, and use of antibiotic medications. Age, an admission in the past 60 days, and history of congestive heart failure were assigned 1 point each and antibiotic usage, dialysis, and mechanical ventilation were assigned 2 points each (Table 3).

https://doi.org/10.1017/ice.2016.97 Published online by Cambridge University Press

TABLE 1. Demographic and Clinical Characteristics of Derivation Cohort in Study of *Clostridium difficile* Infection (CDI)

Variable	Value, %			
	Total (n = 40,990)	With CDI (n = 189)	Without CDI (n = 40,801)	*P*
Age				
<65 years	60.7	33.3	60.8	.0001
≥65 years	39.3	66.7	39.2	
Gender				.2080
Female	61.9	57.1	61.9	
Male	37.0	40.7	37.0	
Unknown	1.1	2.1	1.1	
Race				.0086
Black	11.3	15.9	11.3	
Asian	6.4	3.2	6.4	
White	36.6	30.2	36.7	
Other	7.5	4.8	7.5	
Unknown	38.3	46.0	38.2	
Admission in past 60 days	11.6	29.6	11.5	.0001
Intensive care unit admission	0.9	3.2	0.9	.0076
Mechanical ventilation	4.5	33.3	4.3	.0001
Dialysis	2.2	13.2	2.1	.0001
Comorbidities				
Atrial fibrillation	9.1	20.1	9.1	.0001
Hypertension	2.9	4.2	2.9	.2624
Cardiomyopathy	1.2	0.5	1.2	.7323
Coronary artery disease	19.7	28.6	19.7	.0021
Congestive heart failure	7.9	21.7	7.8	.0001
Diabetes	19.9	32.8	19.8	.0001
Hypothyroidism	0.2	0.5	0.2	.2597
Obesity	6.2	6.9	6.2	.6990
Pancreatitis	0.2	1.1	0.2	.0663
Diverticulitis	3.3	6.9	3.3	.0057
Parkinson disease	1.0	0.5	1.0	>.99
Cerebral vascular disease	8.3	13.2	8.3	.0136
Chronic obstructive pulmonary disease	4.4	9.0	4.4	.0018
Asthma	6.6	9.0	6.6	.1852
Chronic kidney disease	5.2	16.4	5.2	.0001
Liver disease	1.1	3.2	1.1	.0168
Antibiotics and other drugs				
Antibiotics	40.3	87.3	40.0	.0001
Proton-pump inhibitors	7.8	20.6	7.7	.0001
Steroids	6.8	15.3	6.8	.0001

TABLE 2. Demographic and Clinical Characteristics of Validation Cohort in Study of *Clostridium difficile* Infection (CDI)

Variable	Value, %			
	Total (n = 20,492)	With CDI (n = 93)	Without CDI (n = 20,399)	*P*
Age				
<65 years	60.7	31.2	60.9	.0001
≥65 years	39.3	68.8	39.2	
Gender				.1807
Female	62.5	53.8	62.5	
Male	36.3	44.1	36.3	
Unknown	1.2	2.2	1.2	
Race				.0091
Black	11.7	11.8	11.7	
Asian	6.4	3.2	6.4	
White	36.1	28.0	36.2	
Other	7.6	2.2	7.6	
Unknown	38.3	54.8	38.2	
Admission in past 60 days	11.8	30.1	11.7	.0001
Intensive care unit admission	0.9	1.1	0.9	.8686
Mechanical ventilation	4.7	26.9	4.6	.0001
Dialysis	2.3	14.0	2.2	.0001
Comorbidities				
Atrial fibrillation	8.9	18.3	8.9	.0015
Hypertension	3.0	4.3	3.0	.3651
Cardiomyopathy	1.3	0.0	1.3	.6348
Coronary artery disease	19.5	22.6	19.5	.4554
Congestive heart failure	8.2	28.0	8.1	.0001
Diabetes mellitus	20.3	24.7	20.3	.2911
Hypothyroidism	0.1	0.0	0.1	>.99
Obesity	6.1	5.4	6.2	.7574
Pancreatitis	0.2	2.2	0.2	.0170
Diverticulitis	3.4	5.4	3.4	.2460
Parkinson disease	0.9	2.2	0.9	.2101
Cerebral vascular disease	8.2	15.1	8.1	.0149
Chronic obstructive pulmonary disease	4.5	14.0	4.5	.0003
Asthma	6.1	5.4	6.1	.7816
Chronic kidney disease	5.4	17.2	5.3	.0001
Liver disease	0.9	1.1	0.9	.5784
Antibiotics and other drugs				
Antibiotics	40.4	81.7	40.2	.0001
Proton-pump inhibitors	7.5	22.6	7.4	.0001
Steroids	7.0	21.5	6.9	.0001

Model Calibration

Using the scoring system, a patient's risk of CDI increased along a continuum. For example, a patient's risk increased from less than 1% with a risk score of 0 to 12% with a risk score of 7 and 39% with a score of 9. Similarly, the positive likelihood ratio increased from 1 with a score of 0 to 216 with a score of 9. The positive likelihood ratios and predicted probability of CDI at each score level are summarized in Table 4.

Model Discrimination

The values derived from Table 3 were applied to the original derivation cohort to calculate a risk score for each subject. The optimal cut-off point on the receiver operating characteristic curve to identify patients at "high risk" for CDI was determined to be a score of 3. Using this cut-off, those with a score of less than 3 were determined to be "low risk" and those with a score of 3 or greater were determined to be "high risk." Thus, subjects with a risk score of 3 or greater were predicted to have hospital-acquired CDI. Using this cut-off, the sensitivity and specificity values were 82.0% and 75.7%, respectively. The positive and negative predictive values were 1.5% and 99.9%, respectively. The AUC was 0.85.

Applying the scoring system with a cut-off of 3 to the validation set, the sensitivity and specificity were 81.7% and 75.1%, respectively. The positive and negative predictive values were 1.5% and 99.9%, respectively. Overall, the misclassification rate was 24.9%. The AUC was calculated as 0.85.

DISCUSSION

This study addresses an important gap in current literature. With the increase in nosocomial CDI nationwide, a more standardized approach to prevention is needed. However, there is currently no commonly used risk prediction tool that can determine a patient's risk of acquiring this deadly infection upon admission to the hospital. Previous studies that have attempted to generate clinical prediction tools for this purpose have proposed prediction rules based on data that are not

https://doi.org/10.1017/ice.2016.97 Published online by Cambridge University Press

TABLE 3. Multivariable Model in Study of *Clostridium difficile* Infection

Parameter	β estimate	Standard error	Wald χ^2	Pr > χ^2	Odds ratio	95% Wald confidence limits		Score
Age (≥65 vs <65)	0.6191	0.1616	14.6745	0.0001	1.857	1.353	2.549	1
Admission in past 60 days (yes vs no)	0.8608	0.1655	27.0405	<.0001	2.365	1.710	3.271	1
Mechanical ventilation (yes vs no)	1.6533	0.1634	102.3653	<.0001	5.224	3.793	7.197	2
Dialysis (yes vs no)	1.2372	0.2306	28.7768	<.0001	3.446	2.193	5.415	2
Congestive heart failure (yes vs no)	0.4837	0.1906	6.4409	0.0112	1.622	1.116	2.357	1
Antibiotic ordered (yes vs no)	1.9781	0.2228	78.8352	<.0001	7.229	4.671	11.187	2

TABLE 4. Positive Likelihood Ratios and Predicted Probabilities at Each Score Level

Score	Positive likelihood ratio	Predicted probability
0	1.00	0.06%
1	1.46	0.14%
2	1.96	0.30%
3	3.37	0.65%
4	5.97	1.38%
5	10.24	2.93%
6	16.88	6.10%
7	24.18	12.28%
8	33.21	23.17%
9	215.88	39.38%

readily available or based on a high number of risk factors. Therefore, they are not practical for everyday use. In order to address this problem, we worked to derive a rule that would predict a patient's risk of acquiring a primary CDI on the basis of a limited amount of clinical risk factors that were easily accessible upon admission. We excluded those patients with a previous CDI owing to their high risk for repeated infections.

Within our patient population, we identified an easy-to-calculate tool based on 6 risk factors. These items include demographic information, history of disease, and prescription medication use, encompassing both reversible and irreversible risk factors. A patient's risk score increased the patient's risk of acquiring a CDI exponentially, as evidenced by the increasingly high probability of infection at each score level; a patient's risk of CDI increased from 0.06% with a risk score of 0 to 39% with a score of 9. Moreover, when a cut-off of 3 was used, the scoring system was both sensitive (82.0%) and specific (75.7%) for identifying patients at high risk for infection. Similarly, the AUC of 0.85 in both the derivation and validation samples revealed excellent discriminatory characteristics.

These results show the tool's ability to alert physicians to those at high risk for CDI and will allow for precautionary measures to take place in these high-risk patients. These measures can include but are not limited to decreasing the usage of proton-pump inhibitors and high-risk antibiotics in association with early detection and treatment of active infections.

Furthermore, the high negative predictive value of this tool (99.9%) will allow physicians to identify those patients with a score below 3 who may not need further CDI testing. Therefore, physicians can more accurately determine which patients are in need of CDI testing and treatment and those who may yield false positive results.

Limitations of this study include its retrospective study design and its reliance on administrative and claims data. Although there was random manual sampling of patient data, not all of the electronic data was validated manually. Furthermore, all patients in the cohort were grouped together regardless of whether they were on a medical or surgical ward. The next step for this project is a longitudinal prospective validation of the model's predictive ability among an external population of patients. Following its validation, interventions could be designed to occur before those patients considered at high risk develop a CDI.

The results of this study speak to the importance of primary prevention in decreasing the incidence of hospital-acquired CDI. Risk factors that we noted as putting patients at high risk for infection are those commonly seen in the hospital setting and are easy to identify. Systematically recognizing the risk in a quantified manner allows for interventions that modify either the patients' individual risk factors or their environment. Examples of possible interventions include decreasing the usages of proton pump inhibitors and high-risk antibiotic treatment in these patients. Further work is needed to validate this rule in a prospective environment.

ACKNOWLEDGMENTS

Financial support. None reported.

Potential conflicts of interest. All authors report no conflicts of interest relevant to this article.

Address correspondence to Anne Press, BS, Hofstra North Shore-LIJ School of Medicine, 300 Community Dr, Manhasset, NY 11030 (anniepress@gmail.com).

REFERENCES

1. Tillotson GS, Tillotson J. *Clostridium difficile*—a moving target. *F1000 Med Rep* 2011;3:6.
2. Mitchell BG, Gardner A. Prolongation of length of stay and *Clostridium difficile* infection: a review of the methods used to examine length of stay due to healthcare associated infections. *Antimicrob Resist Infect Control* 2012;1:14.

https://doi.org/10.1017/ice.2016.97 Published online by Cambridge University Press

3. Rupnik M, Wilcox MH, Gerding DN. *Clostridium difficile* infection: new developments in epidemiology and pathogenesis. *Nat Rev Microbiol* 2009;7:526–536.
4. Cohen SH, Gerding DN, Johnson S, et al. Clinical practice guidelines for *Clostridium difficile* infection in adults: 2010 update by the Society for Healthcare Epidemiology of America (SHEA) and the Infectious Diseases Society of America (IDSA). *Infect Control Hosp Epidemiol* 2010;31:431–455.
5. Pépin J, Valiquette L, Alary ME, et al. *Clostridium difficile*-associated diarrhea in a region of Quebec from 1991 to 2003: a changing pattern of disease severity. *CMAJ* 2004;171:466–472.
6. McDonald LC, Killgore GE, Thompson A, et al. An epidemic, toxin gene-variant strain of *Clostridium difficile*. *N Engl J Med* 2005;353:2433–2441.
7. Loo VG, Poirier L, Miller MA, et al. A predominantly clonal multi-institutional outbreak of *Clostridium difficile*-associated diarrhea with high morbidity and mortality. *N Engl J Med* 2005;353:2442–2449.
8. Freeman J, Bauer MP, Baines SD, et al. The changing epidemiology of *Clostridium difficile* infections. *Clin Microbiol Rev* 2010;23:529–549.
9. Gerding DN, Johnson S, Peterson LR, Mulligan ME, Silva J. *Clostridium difficile*–associated diarrhea and colitis. *Infect Control Hosp Epidemiol* 1995;16:459–477.
10. Johnson S, Gerding DN, Olson MM, et al. Prospective, controlled study of vinyl glove use to interrupt *Clostridium difficile* nosocomial transmission. *Am J Med* 1990;88:137–140.
11. Kaatz GW, Gitlin SD, Schaberg DR, et al. Acquisition of *Clostridium difficile* from the hospital environment. *Am J Epidemiol* 1988;127:1289–1294.
12. Mayfield JL, Leet T, Miller J, Mundy LM. Environmental control to reduce transmission of *Clostridium difficile*. *Clin Infect Dis* 2000;31:995–1000.
13. McMullen KM, Zack J, Coopersmith CM, Kollef M, Dubberke E, Warren DK. Use of hypochlorite solution to decrease rates of *Clostridium difficile*–associated diarrhea. *Infect Control Hosp Epidemiol* 2007;28:205–207.
14. Smith LA, Chan CK, Halm M, Slattery W, Lindquist R, Savik K. Development and validation of a *Clostridium difficile* risk assessment tool. *AACN Adv Crit Care* 2014;25:334–346.
15. Chandra S, Thapa R, Marur S, Jani N. Validation of a clinical prediction scale for hospital-onset *Clostridium difficile* infection. *J Clin Gastroenterol* 2014;48:419–422.
16. Dubberke ER, Yan Y, Reske KA, et al. Development and validation of a *Clostridium difficile* infection risk prediction model. *Infect Control Hosp Epidemiol* 2011;32:360–366.
17. Garey KW, Dao-Tran TK, Jiang ZD, Price MP, Gentry LO, Dupont HL. A clinical risk index for *Clostridium difficile* infection in hospitalised patients receiving broad-spectrum antibiotics. *J Hosp Infect* 2008;70:142–147.
18. Tanner J, Khan D, Anthony D, Paton J. Waterlow score to predict patients at risk of developing *Clostridium difficile*-associated disease. *J Hosp Infect* 2009;71:239–244.

https://doi.org/10.1017/ice.2016.97 Published online by Cambridge University Press