

Prediction of *Clostridium difficile* infection in patients with polymerase chain reaction using a classification and regression tree

Diana Martínez-Ruíz

Fundación Valle del Lili

Luis Gabriel Parra-Lara

Fundación Valle del Lili <https://orcid.org/0000-0002-4877-1117>

Fernando Rosso (✉ fernando.rosso@fvl.org.co)

Fundación Valle del Lili

Research article

Keywords: Clostridium difficile, Polymerase chain reaction, Decision trees, Diagnosis

Posted Date: January 13th, 2020

DOI: <https://doi.org/10.21203/rs.2.20682/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background

Clostridium difficile infection (CDI) is commonly diagnosed with the polymerase chain reaction (PCR), but this test finds a high percentage of false positives, so their use and interpretation in CDI is a challenge in the clinical practice. That is why it is necessary to define an algorithm to optimize the use of PCR that considers clinical characteristics to classify patients with diarrhea as CDI or without CDI.

Objective

To identify a predictive algorithm with the clinical features that best classify patients with CDI vs. without CDI, to help physicians in making decisions to request PCR.

Materials and methods

A case-control study was conducted at Fundación Valle del Lili. The population was inpatients between 2012 and 2016, with 18 or more years, and diarrhea, abdominal pain, or other nonspecific gastrointestinal symptoms who underwent PCR. Cases were defined as patients with positive PCR for *C. difficile* and as controls patients with negative PCR for *C. difficile*. Predictive algorithms to classify patients was constructed using a classification tree, classification and regression tree (CART).

Results

A total of 149 patients were included (48 cases with positive PCR and 99 controls with negative PCR). The CART has a high capacity to classify patients with a negative PCR correctly. It includes variables about the history of antibiotics use, the use of proton-pump inhibitor, the use of ranitidine, and the use of antifungal drugs. The CART showed sensitivity 64.6%, specificity 85.8%, positive predictive value 68.8%, negative predictive value 83.3%. and AUC 79.7%.

Discussion

CART had good specificity and a high negative predictive value; it could be considered as an algorithm to identify conditions that indicate when it is not necessary to perform a PCR test in a patient symptom of CDI.

Introduction

Clostridium difficile infection (CDI) is the most common cause of healthcare-associated diarrhea(1). The risk factors identified for CDI are the use of antibiotics, stomach acid-suppressing medications, primarily proton-pump inhibitors (PPI), advanced age (above 64 years), prior hospitalization, exposure to CDI patients, weak health condition, among others(2).

In the past two decades, the incidence and severity of CDI have increased worldwide. Several reports have shown that *C. difficile* affects 20 to 50% of patients in the hospital environment. Indeed, a recent meta-analysis reported that *C. difficile* accounted for 20% of all antibiotic-associated diarrhea cases among hospitalized patients(3,4). Although there is a lack of data about this issue from Latin America, a Colombian study found that 16.7% of gastrointestinal infections in hospitalized adults were caused by *C. difficile*(5).

CDI is commonly diagnosed with the polymerase chain reaction (PCR), which is specific for genes encoding toxins produced by *C. difficile*(6,7). Although this test is useful in clinical practice, its interpretation and applicability are challenging(7). Approximately 50% of hospitalized patients may be asymptotically colonized by toxigenic *C. difficile*(7). Moreover, a positive PCR result does not distinguish between active toxin production and asymptomatic carriage, which results in a large number of false positives(7,8). All these issues lead to increased costs, unnecessary isolation, and antibiotic overuse(9).

One of the most significant challenges in the management of patients with CDI associated diarrhea is to underpin the advance of diagnostic decision-making algorithms in the clinical practice scenario. Clinical algorithms can increase the pretest likelihood of CDI diagnosis. Furthermore, the use of such models in daily clinical practice may lower the overuse of diagnostic tests, thus reducing costs and improving the correct utilization of specific treatments for CDI. The present study used a classification tree to develop a predictive model for CDI diagnosis using a PCR test among inpatients attended at Cali, Colombia.

Materials And Methods

Data source

We performed a retrospective case-control study from 2012 to 2016 at Fundación Valle del Lili (FVL) in Cali, Colombia. FVL is a level IV university hospital affiliated with Universidad Icesi and serves as a referral facility for the southwest region of Colombia. All data were obtained directly from the clinical records.

Patients and Controls

We included hospitalized patients older than 18 years old with diarrhea, abdominal pain, or other nonspecific gastrointestinal symptoms and that underwent PCR testing for *C. difficile*. Cases were defined as patients with a positive PCR result for *C. difficile*. Controls were defined as patients with negative PCR results.

Laboratory methods

The FVL clinical practice guideline (CPG) for CDI diagnosis include PCR test. A Xpert® *C. difficile* assay (Xpert CD assay; Cepheid, Sunnyvale, CA, USA) is a multiplex real-time PCR assay that uses primers targeted to the cytotoxin gene (*tcdB*), binary toxin genes (*cdtA* and *cdtB*), and a single nucleotide deletion

at position 117 in the *tcdC* gene. As a result, the Xpert CD assay can detect toxigenic *C. difficile* strains and differentiate *C. difficile* presumptive 027/NAP1/BI(10).

Statistical analysis

Predictive algorithms to classify patients with CDI and without CDI were designed using a classification and regression tree (CART). There is a non-parametric technique that determines a set of logical conditions (variables) to perform binary classification of cases in homogeneous subgroups and at the same time, considers the complexity of the relationships between the independent variables. For this study, the CART explains why patients are classified as negative or positive PCR, achieving a low rate of misclassification (11,12).

In each division (node), the tree selects the independent variable that best classifies patients with CDI or without CDI. Then, the algorithm includes variables until achieving the highest possible homogeneity or few observations. The steps for developing the CART were: **1)** Construction of the maximum tree with variables that showed significantly less or equal to 20% in the bivariate analysis. For the bivariate analysis, quantitative data were compared using the Mann-Whitney U test, and qualitative data using the Chi-Squared test or Fisher's exact test; **2)** The maximum tree was built using the Gini index, which was a criterion to determine the best division. The minimum change in the rate of misclassification was a criterion for stopping the tree (adding new variables); **3)** Some ten patients were established as the minimum number of observations per division; **4)** The process of pruning backward, to obtain a parsimonious tree, was done with cross-validation using the same sample, and the cost-complexity mean was zero. The resulting classification tree was evaluated for discriminative capacity through sensitivity, specificity, positive and negative predictive values, overall value, likelihood ratio, and the degree of diagnostic accuracy by calculating the area under the ROC curve. All analyses were performed using software R v.3.16(13).

Results

A total of 149 patients were included in the study after fulfilled criteria, forty-eight cases with positive PCR, and ninety-nine controls with negative PCR. Two cases were excluded for the analysis.

The median age was 55 years (IQR=71-35 years). The majority were women. The most frequent comorbidity was cardiovascular disease. The leading cause of hospitalization was for non-surgical conditions (40,9%, n=61). **Table 1** shows the main clinical characteristics.

Table 2 shows the drugs prescribed to the patients. A total of 62.6% of patients had a history of antibiotics use, mainly meropenem and antifungal drugs. Other frequently used drugs were omeprazole, immunosuppressants, and steroids. Statistically significant differences were found in the use of meropenem, clindamycin, polymyxin B, antifungals, steroids, and chemotherapy. Also, regarding the history of drug use, antibiotics, PPI and H2 antagonists, statistically significant differences were found.

The CART is presented in **Figure 1**. The CART had adequate discrimination capacity to distinguish between positive and negative PCR, only with the history of antibiotic use. However, the CART also included other variables, such as the use of PPI, the use of ranitidine, and the use of antifungal drugs, because these variables help make a better decision. The CART showed the following performance: sensitivity 64.6%, specificity 85.8%, positive predictive value 68.8%, negative predictive value 83.3%, positive likelihood ratio 4.56, negative likelihood ratio 0.41, and AUC 79.7% (**Figure 2**).

The algorithm to identify patients with negative PCR was: if a patient did not have a history of antibiotic use, the probability was 89.1%. On the contrary, if a patient had a history of antibiotic use, the probability was 54.3%. But if a patient had a history of antibiotic use, did not use PPI and ranitidine, the probability of a negative PCR was 78.8%. Finally, if a patient had a history of antibiotic use, and did not use PPI but received ranitidine and did not use antifungal drugs, the probability of a negative PCR was 71,4%.

Discussion

In low and middle-income countries, there is a lack of evidence-based diagnostic algorithms for early CDI diagnosis using a PCR test. In 2011, a study showed that the prevalence of in-hospital *C. difficile* was 10 per 10.000 inpatients in Cali, indicating a high prevalence of this infection and the need to identified better strategies for the control of *C. difficile*(14). Our study shows a straightforward algorithm that includes clinically relevant variables and may help clinicians in the diagnostic process of patients with suspected CDI. The CDI is an important cause of nosocomial infection, so the implementation of diagnostic strategies, such as decision trees, is an important alternative to guide physicians in making decisions.

Our CART showed that if a patient had a history of antibiotic use, the probability of a positive PCR was 45.7%. Previous studies in this matter showed almost all antibiotics can increase vulnerability to CDI, but cephalosporins, fluoroquinolones, clindamycin, and certain penicillins (e.g., amoxicillin/clavulanic acid) increase risk to the greatest extent (8,14,15). Above supports the results of this study regarding consider this exposure in the decision to perform a PCR.

Also, PPI has been associated with increased odds of CDI and recurrence(8). The 2018 meta-analysis by Oshima *et al.*(16) concluded that PPI use was associated with CDI in adult (OR 2.30, 95% CI 1.89-2.80; $p < 0.00001$) and pediatric patients (OR 3.00, 95% CI 1.44-6.23; $p < 0.00001$), and with recurrent CDI (OR 1.73, 95% CI 1.39-2.15; $p = 0.02$). In the CART, PCR performance in patients with a history of use of PPI was 75%. This finding suggests that in a case with a history of antibiotic and PPI use, it could be not necessary to use the PCR test for CDI diagnostic, and the physician will be a big probability to have a positive case of CDI without another diagnostic test.

From our results, we can recommend considering the PCR test when the patients have been exposed to PPI, ranitidine, and antifungal drugs. Instead, if the patient has not a history of antibiotics use, the probability of finding a negative result is higher. This model demonstrated a good capacity to classify healthy patients as healthy (specificity) and a high negative predictive value and thus can be considered

as an algorithm to identify conditions that indicate when it is not necessary to perform a PCR test in a patient with symptoms of CDI.

CDI increases patient healthcare costs due to extended hospitalization, re-hospitalization, laboratory tests, and medications. A systematic review found that CDI to be a significant economic healthcare burden in their respective settings, with an increased length of stay and costs(17). That is why our model emerges as a diagnostic alternative for middle and low-income countries, which allows optimizing the indication of PCR for the diagnosis of CDI and thus reduces the economic burden of the disease in health systems.

The CART of this study has the advantage that it is easy to interpret and implement for health workers in their daily practice because the data that contain it is available by performing a routine clinical review to the inpatients. This methodology compared other statistical methods such as regression, which has a good capacity to discriminate, allows identifying directly from the tree, the interactions between the data and the probability of having a PCR with a positive or negative result. Therefore, this tree can be a tool for the decision to perform a PCR in patients with symptoms of CDI in environments where there are restrictions to perform diagnostic tests as PCR.

Limitations. Data were collected retrospectively, which can lead to an information bias due to the absence of data in some variables. The predictive capacity of the variable “history of antibiotics use” is limited by the time of exposure, because the time from the use of antibiotics to the development of the symptoms was not collected. On the other hand, data about the use of certain antibiotics were collected, but there are unidentified if they were administered before or after the PCR. The results of this study can be only generalized to patients treated in a high complexity referral hospital with similar demographics and clinical characteristics of our research. Despite these limitations, the designed algorithm can be a valuable tool in medium and low-income countries where resources and diagnostic tests are routinely limited.

Conclusions

This CART offers a fast algorithm for deciding to perform a PCR test on patients with symptoms of CDI in a referral health facility in southwestern Colombia. Patients with a low probability of a positive PCR result are those without a history of antibiotics use. Still, if they have been exposed, the use of PPI, ranitidine, and antifungal drugs should be considered, other tests should be performed to rule out CDI.

Abbreviations

AUC: area under curve

CART: classification and regression tree

CDI: *Clostridium difficile* infection

FVL: Fundación Valle del Lili

IQR: interquartile range

IRB: Institutional review board

PCR: polymerase chain reaction

PPI: proton-pump inhibitors

Declarations

Ethics approval and consent to participate

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards. The Institutional Review Board – Comité de Ética en Investigación Biomédica approved the study protocol (Protocol Number: 554).

Consent for publication

Not applicable

Availability of data and materials

Data is property of authors and could be available by request at fernando.rosso@fvl.org.co

Competing interests

The authors declare that they have no competing interests

Funding

No funding. This article received no specific grant from any funding agency in the public, commercial, or non-for-profit sectors.

Authors' contributions

DMM, LGPL, and FR contributed equally to this work. DMM and LGPL designed the research. All authors performed the research, analyzed the data, and wrote the paper. All authors made important intellectual contributions to the manuscript, and all authors approved the final version before submission. FR supervised all the process.

Acknowledgements

The authors would like to thank Jorge Cedano MD, Aida Manuela Galvis MD, Stephania González MD, Carlos Vélez MD, Juan Sebastián Solarte MD, María Paula Hormaza MD for their collaboration in the data collection process. Also, to Centro de Investigaciones Clínicas - Fundación Valle del Lili for their constant support.

References

1. Cohen SH, Gerding DN, Johnson S, Kelly CP, Loo VG, McDonald LC, Pepin J, Wilcox MH. 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 May;31(5):431–55.
2. Chhipi-Shrestha G, Mori J, Hewage K, Sadiq R. *Clostridium difficile* infection incidence prediction in hospitals (CDIIPH): a predictive model based on decision tree and fuzzy techniques. *Stoch Environ Res Risk Assess*. 2017;31(2):417–30.
3. Burnham C-AD, Carroll KC. Diagnosis of *Clostridium difficile* infection: an ongoing conundrum for clinicians and for clinical laboratories. *Clin Microbiol Rev*. 2013 Jul;26(3):604–30.
4. Nasiri MJ, Goudarzi M, Hajikhani B, Ghazi M, Goudarzi H, Pouriran R. *Clostridioides* (*Clostridium*) *difficile* infection in hospitalized patients with antibiotic-associated diarrhea: A systematic review and meta-analysis. *Anaerobe*. 2018 Apr;50:32–7.
5. Zea JW, Salazar CL. Enfermedad asociada a *Clostridium difficile*: prevalencia y diagnóstico por laboratorio. *Infectio*. 2012;16(4):211–22.
6. Tartof SY, Yu KC, Wei R, Tseng HF, Jacobsen SJ, Rieg GK. Incidence of polymerase chain reaction-diagnosed *Clostridium difficile* in a large high-risk cohort, 2011-2012. *Mayo Clin Proc* [Internet]. 2014;89(9):1229–38. Available from: <http://dx.doi.org/10.1016/j.mayocp.2014.04.027>
7. Kufelnicka AM, Kim TJ. Effective utilization of evolving methods for the laboratory diagnosis of *Clostridium difficile* infection. *Clin Infect Dis*. 2011;52(12):1451–7.
8. Mullish BH, Williams HR. *Clostridium difficile* infection and antibiotic-associated diarrhoea. *Clin Med* [Internet]. 2018;18(3):237–41. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/29858434>
9. Mueller CJ, Castagno WJ, Duckworth L, McEver RC, Campbell HJ. Implementation of a Decision Tree to Optimize *Clostridium difficile* Testing. *Am J Infect Control* [Internet]. 2017;45(6):S8. Available from: <http://dx.doi.org/10.1016/j.ajic.2017.04.270>
10. Shin S, Kim M, Kim M, Lim H, Kim H, Lee K, Chong Y. Evaluation of the Xpert *Clostridium difficile* assay for the diagnosis of *Clostridium difficile* infection. *Ann Lab Med* [Internet]. 2012/08/13. 2012 Sep;32(5):355–8. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/22950071>
11. StatSoft. *Electronic Statistics Textbook* [Internet]. StatSoft, editor. Tulsa: StatSoft; 2013. Available from: <http://www.statsoft.com/textbook/>
12. Breiman L. *Classification and Regression Trees*. Boca Raton: Routledge; 2017. 368 p.

13. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna: R Foundation for Statistical Computing; 2019. Available from: www.R-project.org/
14. Oñate-Gutiérrez JM, Villegas MV, Correa A. Prevalencia y factores relacionados con la infección por *Clostridium difficile* en un centro hospitalario de alta complejidad en Cali (Colombia). *Infectio*. 2016;21(1):1–6.
15. Bartlett JG, Gerding DN. Clinical recognition and diagnosis of *Clostridium difficile* infection. *Clin Infect Dis*. 2008 Jan;46(Suppl 1):S12-8.
16. Oshima T, Wu L, Li M, Fukui H, Watari J, Miwa H. Magnitude and direction of the association between *Clostridium difficile* infection and proton pump inhibitors in adults and pediatric patients: a systematic review and meta-analysis. *J Gastroenterol* [Internet]. 2018;53(1):84–94. Available from: <https://doi.org/10.1007/s00535-017-1369-3>
17. Ghantaji SS, Sail K, Lairson DR, Dupont HL. Economic healthcare costs of *Clostridium difficile* infection : a systematic review. *J Hosp Infect* [Internet]. 2010;74(4):309–18. Available from: <http://dx.doi.org/10.1016/j.jhin.2009.10.016>

Tables

Table 1. Clinical characteristics of patients

Data	Total, n=147		PCR+, n =48		PCR-, n=99		p-value
	n	%	n	%	n	%	
Age, median (IQR)	55 (71 - 35)		59.5 (73.5 - 37.5)		55 (71 - 34)		0.44
Sex							
Woman	82	55.8	23	47.9	59	59.6	0.18
Comorbidities							
Cardiovascular disease	60	40.8	20	41.7	40	40.4	0.88
Chronic kidney disease	35	23.8	13	27.1	22	22.2	0.52
Hematologic malignancy	31	21.1	13	27.1	18	18.2	0.21
Diabetes	26	17.7	7	14.6	19	19.2	0.49
Infections	23	15.6	11	22.9	12	12.1	0.09
Non-metastatic tumor	21	14.3	4	8.3	17	17.2	0.15
Liver disease	21	14.3	7	14.6	14	14.1	0.94
Rheumatologic disease	17	11.6	4	8.3	13	13.1	0.39
Solid organ transplant	16	10.9	4	8.3	12	12.1	0.58
Endocrine disorder	15	10.2	4	8.3	11	11.1	0.77
Chronic obstructive pulmonary disease	14	9.5	2	4.2	12	12.1	0.15
Congestive heart failure	11	7.5	3	6.3	8	8.1	1
Bone marrow transplant	11	7.5	6	12.5	5	5.1	0.18
Neurological disorder	9	6.1	3	6.3	6	6.1	1
HIV	4	2.7	1	2.1	3	3	1
Cerebrovascular accident	2	1.4	1	2.1	1	1	0.55
Diverticular disease	2	1.4	1	2.1	1	1	0.55
Peptic ulcer disease	2	1.4	1	2.1	1	1	0.55
Inflammatory bowel disease	2	1.4	0	0	2	2	1
AIDS	1	0.7	0	0	1	1	1
Cause of hospitalization							
Medical non-infectious	81	40.8	26	43.8	55	39.4	0.97
Infectious	44	29.9	15	31.3	29	29.3	
Surgical	22	12.9	7	12.5	15	13.1	
ICU hospitalization	103	70.1	36	75	67	67.7	0.36
Time of previous hospitalization							
Less than a month	80	54.4	22	45.8	58	58.6	0.24
Between 1-12 months	46	31.3	15	31.3	31	31.3	
More than 1 year	14	9.5	7	14.6	7	7.1	
Time between hospitalization and PCR							
≤7 days	59	40.14	17	35.4	42	42.4	0.416
≥7 days	88	59.9	31	64.6	57	57.6	

Table 2. Prescription of drugs in patients

Data	Total, n=147		PCR+, n=99		PCR-, n=48		p-value
	n	%	n	%	n	%	
Antibacterial							
Meropenem	70	47.6	29	60.4	41	41.4	0.03
Piperacillin/Tazobactam	46	31.3	17	35.4	29	29.3	0.45
Ceftriaxone	40	27.2	10	20.8	30	30.3	0.23
Cefepime	31	21.1	14	29.2	17	17.2	0.09
Ciprofloxacin	28	19	8	16.7	20	20.2	0.61
Sulfamethoxazole/Trimethoprim	15	10.2	3	6.3	12	12.1	0.39
Polymyxin B	12	8.2	7	14.6	5	5.1	0.06
Amikacin	11	7.5	3	6.3	8	8.1	1
Clindamycin	9	6.1	0	0	9	9.1	0.03
Ampicillin/Sulbactam	9	6.1	2	4.2	7	7.1	0.72
Cefazolin	9	6.1	4	8.3	5	5.1	0.474
Linezolid	7	4.8	2	4.2	5	5.1	1
Moxifloxacin	5	3.4	0	0	5	5.1	0.17
Oxacillin	4	2.7	1	2.1	3	3	1
Doripenem	4	2.7	2	4.2	2	2	0.6
Doxycycline	4	2.7	2	4.2	2	2	0.6
Tigecycline	3	2	1	2.1	2	2	1
Amoxicillin	3	2	2	4.2	1	1	0.25
Levofloxacin	2	1.4	0	0	2	2	1
Clarithromycin	2	1.4	1	2.1	1	1	0.55
Ertapenem	2	1.4	2	4.2	0	0	0.11
Gentamicin	2	1.4	1	2.1	1	1	0.55
Antifungal	67	45,6	28	58,3	39	39.4	0.02
Antiviral	36	24,5	13	27,1	23	23.2	0.47
Antiparasitic	15	10,2	5	10,4	10	10.1	1
Other drugs							
Omeprazole	113	76.9	33	68.8	80	80.8	0.1
Ranitidine	77	52.4	30	62.5	47	47.5	0.09
Immunosuppressant	76	51.7	21	43.8	55	55.6	0.18
Steroid	69	46.9	17	35.4	52	52.5	0.05
NSAID	56	38.1	22	45.8	34	34.3	0.18
Antidepressant	42	28.6	15	31.3	27	27.3	0.62
Chemotherapy	40	27.2	18	37.5	22	22.2	0.05
Pharmacological history							
Antibiotics	92	62.6	42	87.5	50	50.5	<0.001
Immunosuppressant	48	32.7	15	31.3	33	33.3	0.77
Chemotherapy	36	24.5	13	27.1	23	23.2	0.31
Proton-pump inhibitor	34	23.1	18	37.5	16	16.2	<0.0001
Corticosteroids	33	22.4	11	22.9	22	22.2	0.25
NSAID	10	6.8	1	2.1	9	9.1	0.45
H2 antagonist	10	6.8	8	16.7	2	2	<0.0001
Antidepressant	7	4.8	4	8.3	3	3	0.06
Laxatives	4	2.7	2	4.2	2	2	0.23

Figures

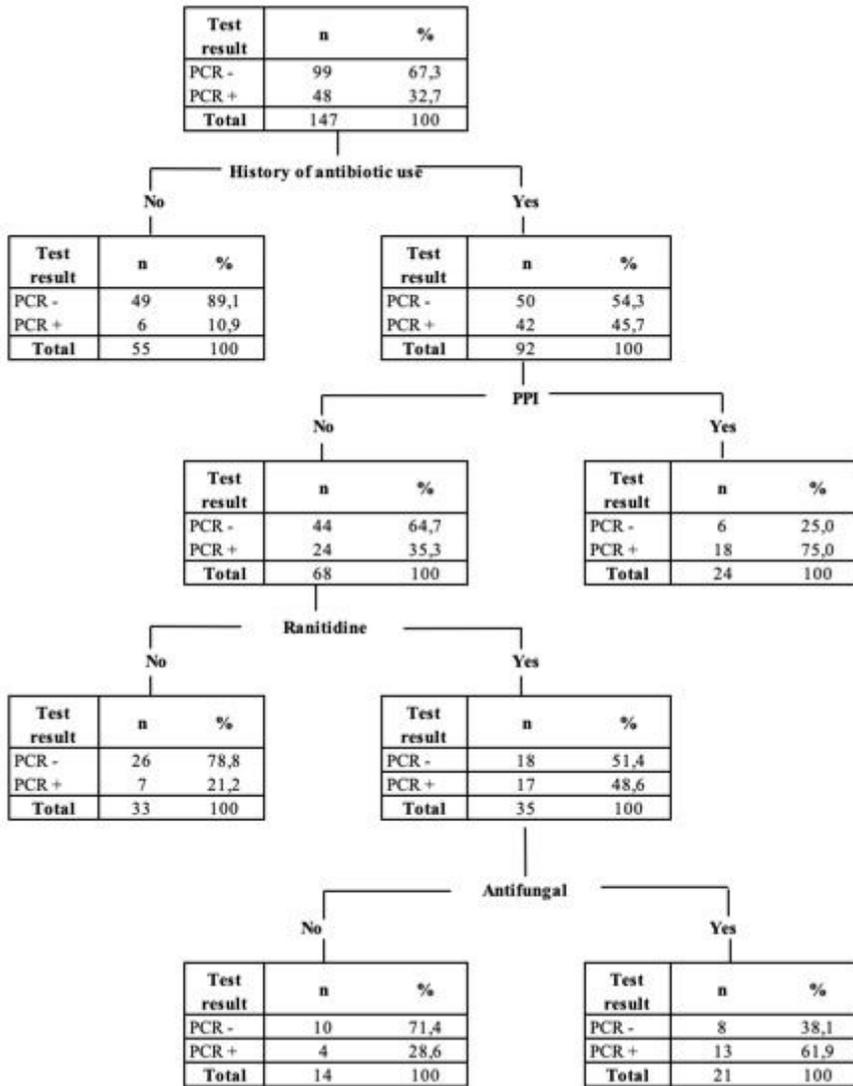


Figure 1

Classification tree for prediction CDI by PCR.

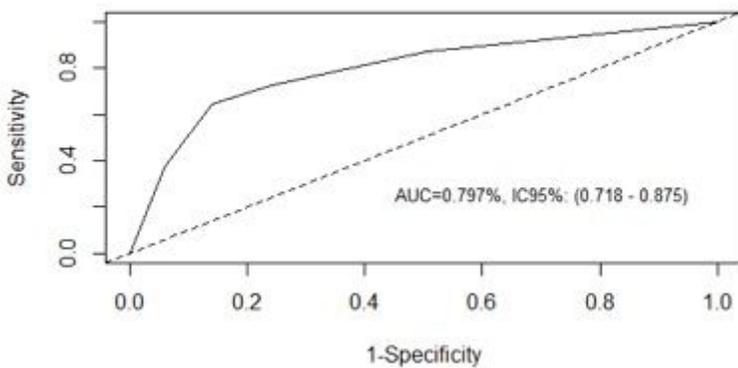


Figure 2

Results from validation of the classification tree.