The American Journal of Surgery 218 (2019) 661–662

Contents lists available at ScienceDirect

The American Journal of Surgery

journal homepage: www.americanjournalofsurgery.com

Letter to the Editor

Prevalence and predictors of *C. difficile* infections in hospitalized patients with major surgical procedures in the USA: Analysis using traditional and machine learning methods

To Editor,

Clostridium difficile infections (CDI) are a major cause of morbidity and mortality in hospitalized patients.[1–3] We sought to examine prevalence and identify predictors of CDI in hospitalized adults aged ≥40 years with any major surgical procedure (MSP). We hypothesize that a mix of patient and hospital factors are associated with occurrence of CDI by using traditional and machine learning (ML) approach.

We conducted a retrospective analysis of the Nationwide Inpatient Sample (NIS)[4] from 2012 to 2014. All hospitalized adults ≥40 years who underwent an MSP were included. The outcome of interest was occurrence of CDI. The independent variables were demographic characteristics and hospital-related variables. A multivariable logistic regression model was used to examine the association between independent and outcome variables. Effects of clustering of outcomes within hospitals were adjusted in the regression model. We then applied an interpretable ML model to the data.[5] As opposed to black-box models, interpretable models generate decisions via transparent decision processes that are easily understandable by humans. We adopted and modified a state-of-the-art interpretable model, Bayesian Rule Set (BRS), which constructs from data a small set of short rules comprised of features that differentiate the positive class (develop CDI) from the negative class (do not develop CDI). A rule is a conjunction of features that are automatically determined by the model to be the most important predictors of CDI. It is predicted that a patient will develop CDI if satisfies any of the rules in a rule set model. The BRS model is learned from a Bayesian framework that consists of a prior that favors a small set of short rules to avoid overfitting and false correlations, and a likelihood that represents how well the rules predict the data. The optimal model with the highest posterior is obtained via a stochastic local search algorithm that guarantees global optimality. Our goal in the machine-learning approach was to find significant associations between features and the outcome (CDI), as well as to build an accurate predictive model. We varied the hyperparameters in BRS to generate four rule set models with different false positive rates (FPR), less than 0.01, less than 0.1, greater than 0.1, and close to 0.5, respectively.

Table 1
Characteristics associated with CDI during hospitalization (traditional multivariable logistic regression model).

Characteristic		Odds Ratio (95% CI)	p-value
Age	Each 1 unit increase	1.01 (1.01–1.01)	<0.01
Sex	Female	0.89 (0.87–0.91)	<0.01
	Male	Reference	
Race	African American	1.15 (1.10–1.19)	<0.01
	Hispanic	0.98 (0.93–1.03)	0.38
	Asian/Pacific Islander	1.20 (1.09–1.31)	<0.01
	Native American	1.10 (0.90–1.34)	0.34
	Other Races	1.02 (0.94–1.11)	0.59
	Caucasian	Reference	
Comorbid burden	Each 1 unit increase	1.46 (1.45–1.47)	<0.01
Insurance status	Medicare	1.23 (1.19–1.28)	<0.01
	Medicaid	1.41 (1.34–1.48)	<0.01
	Uninsured	1.10 (1.01–1.20)	0.03
	Other insurance	0.94 (0.86–1.03)	0.22
	Private insurance	Reference	
Location/teaching status of hospital	Urban teaching hospital	1.39 (1.34–1.44)	<0.01
	Urban non-teaching/rural hospitals	Reference	
Geographic region	Northeast	1.20 (1.14–1.26)	<0.01
	Midwest	0.89 (0.84–0.94)	<0.01
	South	0.88 (0.83–0.92)	<0.01
	West	Reference	
Year of hospitalization	2013	0.95 (0.91–1.00)	0.04
	2014	0.91 (0.87–0.96)	<0.01
	2012	Reference	

https://doi.org/10.1016/j.amjsurg.2018.11.014
0002-9610/© 2018 Elsevier Inc. All rights reserved.

Table 2
Bayesian Rule Set models and their performances (Machine Learning Model).

	Rule set model	FPR	TPR
Rule set 1	- Comorbid ≥ 5, Age ≥ 72, and location = northeast	0.009	0.038
Rule set 2	- Comorbid ≥ 5 and urban teaching hospital	0.085	0.270
Rule set 3	- Comorbid ≥ 5	0.141	0.435
Rule set 4	- Comorbid ≥ 3 - Race = African American, Asian/Pacific Islander, or Native American	0.491	0.813

FPR: False positive rate; TPR: True positive rate.

During the study period, a total of 19,931,347 hospitalized adults underwent an MSP. Half aged 40–64 years and the rest aged ≥65 years. 52.6% were female and 76.3% were Caucasians. The incidence of CDI was 0.7% (0.6% in 40–64 years, 0.9% in ≥65 years). Results from traditional multivariable logistic regression model are summarized in Table 1. Following adjustment of confounders, demographic characteristics including increase in age, male gender, co-morbid burden, African-American and Asian/Pacific Islander race were associated with higher odds for developing CDI. Hospital-related factors including uninsured adults, insurance coverage by Medicare, Medicaid, and those underwent MSP at urban teaching hospitals had higher odds for developing CDI.

The rule sets of the ML model are reported in Table 2. Every rule and every rule set has a p-value <0.01. The first three rule sets consist of one rule and the last rule set consists of two rules in order to cover a larger group of people who either have 3 or more co-morbidities, or whose race is African American, Asian/Pacific Islander, or Native American.

From both the traditional and ML approach, it is apparent that increasing age, increasing co-morbid burden, geographic location, teaching status of hospital, and race are associated with occurrence of CDI. The ML model consistently included co-morbid burden in the model, indicating a large variable importance. Findings from these models can be used to identify those patients that are likely to develop CDI following surgical procedures.

Funding

The study was funded by the Brodie Craniofacial Endowment.

Conflict of interest

None for all authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.amjsurg.2018.11.014.

References

1. Eckmann C, Wasserman M, Latif F, Roberts G, Beriot-Mathiot A. Increased hospital length of stay attributable to Clostridium difficile infection in patients with four co-morbidities: an analysis of hospital episode statistics in four European countries. *Eur J Health Econ*. 2013;14(5):835–846.
2. To KB, Napolitano LM. Clostridium difficile infection: update on diagnosis, epidemiology, and treatment strategies. *Surg Infect*. 2014;15:490–502.
3. Goudarzi M, Seyedjavadi SS, Goudarzi H, Mehdizadeh Aghdam E, Nazeri S. Clostridium difficile infection: epidemiology, pathogenesis, risk factors, and therapeutic options. *Scientifica*. 2014;2014:916826.
4. HCUP NIS Database Documentation. *Healthcare Cost and Utilization Project (HCUP)*. Rockville, MD: Agency for Healthcare Research and Quality; September 2018. www.hcup-us.ahrq.gov/db/nation/nis/nisdbdocumentation.jsp.
5. Wang T, Allareddy V, Rampa S, Allareddy V. *Interpretable Patient Mortality Prediction with Multi-value Rule Sets*. London, United Kingdom: SIGKDD'18; Aug 2018. https://doi.org/10.475/123_4.

Veerajalandhar Allareddy
Chief of Pediatric Cardiac Critical Care, Duke University, Durham, NC, USA

Tong Wang
Department of Management Sciences - Tippie College of Business, The University of Iowa, Iowa City, IA, USA

Sankeerth Rampa
Health Care Administration Program, School of Business – Rhode Island College, Providence, RI, USA

Jennifer Caplin
Department of Orthodontics, University of Illinois at Chicago College of Dentistry, Chicago, IL, USA

Romesh Nalliah
Department of Cariology, Restorative Sciences and Endodontics, School of Dentistry – University of Michigan, Ann Arbor, MI, USA

Aditya Badheka*
Division of Pediatric Critical Care, Stead Family Children's Hospital, University of Iowa, Iowa City, IA, USA

Veerasathpurush Allareddy
Department Head and Brodie Craniofacial Endowed Chair, Department of Orthodontics – University of Illinois at Chicago College of Dentistry, Chicago, IL, USA

* Corresponding author. Division of Pediatric Critical Care, Department of Pediatrics, University of Iowa Stead Family Children's Hospital, 200, Hawkins Dr, Iowa City, IA, 52242, USA.
E-mail address: aditya-badheka@uiowa.edu (A. Badheka).

11 November 2018