Journal of Critical Care 45 (2018) 95–104

Contents lists available at ScienceDirect

Journal of Critical Care

journal homepage: www.jccjournal.org

Healthcare-associated ventriculitis and meningitis in a neuro-ICU: Incidence and risk factors selected by machine learning approach

Ivan Savin [a], Ksenia Ershova [b,c,*], Nataliya Kurdyumova [a], Olga Ershova [a], Oleg Khomenko [b], Gleb Danilov [a], Michael Shifrin [a], Vladimir Zelman [b,c]

[a] Burdenko Neurosurgery Institute, 16 4th Tverskaya-Yamskaya Street, Moscow 125047, Russia
[b] Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Building 3, Moscow 143026, Russia
[c] Department of Anesthesiology, Keck School of Medicine, University of Southern California, 1975 Zonal Ave, Los Angeles, CA 90033, USA

ARTICLE INFO

Available online xxxx

Keywords:
Meningitis
Bacterial
Risk factors
Cross infection
Machine learning
Intensive care unit
Infection control

ABSTRACT

Purpose: To define the incidence of healthcare-associated ventriculitis and meningitis (HAVM) in the neuro-ICU and to identify HAVM risk factors using tree-based machine learning (ML) algorithms.
Methods: An observational cohort study was conducted in Russia from 2010 to 2017, and included high-risk neuro-ICU patients. We utilized relative risk analysis, regressions, and ML to identify factors associated with HAVM development.
Results: 2286 patients of all ages were included, 216 of them had HAVM. The cumulative incidence of HAVM was 9.45% [95% CI 8.25–10.65]. The incidence of EVD-associated HAVM was 17.2 per 1000 EVD-days or 4.3% [95% CI 3.47–5.13] per 100 patients. Combining all three methods, we selected four important factors contributing to HAVM development: EVD, craniotomy, superficial surgical site infections after neurosurgery, and CSF leakage. The ML models performed better than regressions.
Conclusion: We first reported HAVM incidence in a neuro-ICU in Russia. We showed that tree-based ML is an effective approach to study risk factors because it enables the identification of nonlinear interaction across factors. We suggest that the number of found risk factors and the duration of their presence in patients should be reduced to prevent HAVM.

1. Introduction

Healthcare-associated ventriculitis and meningitis (HAVM) may take place in association with invasive neurosurgical procedures (post-neurosurgical meningitis), penetrating head trauma (post-traumatic meningitis), or miscellaneous causes on occasion [1]. HAVM significantly impairs patient outcomes, enhancing morbidity and mortality [2]. The development of post-neurosurgical meningitis can increase mortality rate approximately 3 times (13.7% vs. 4.7%) compared to non-meningitis neurosurgical patients [3]. Moreover, HAVM increases the cost of care. In 2014, Schweizer et al. [4] showed in a large study (>50,000 analyzed operations) a 1.93-fold increase ($23,755 more per case) of attributable health care expenditures to patients with post-neurosurgical meningitis compared to those without infections after neurosurgery. For effective HAVM prevention it is necessary to have reliable data regarding HAVM incidence in different patient cohorts, and learn associated risk factors.

We assigned three primary objectives to this study: (1) to determine the incidence of HAVM in the high-risk population, i.e. patients who stayed in the neuro-ICU for >48 h, (2) to compare HAVM incidence in patients who were exposed to different risk factors during their stay in the ICU and assess relative risk (RR) for each of the factors, and (3) to identify and range HAVM risk factors using regression and machine learning (ML) approaches. We hypothesized that during patients' stay in the ICU a few independent factors would emerge over time, increasing the probability of HAVM development.

The first objective includes the study of HAVM incidence that is usually analyzed depending on risk factors or diagnosis. In the literature, the cumulative incidence of post-neurosurgical meningitis varies

Abbreviations: CCI, Charlson Comorbidity Index; CDC, the Centers for Disease Control; CI, confidence interval; CSF, cerebrospinal fluid; CSFL-NE, CSF leak from nose and/or ears; CSFL-SS, CSF leak from surgical site (also, if appropriate, from ventricular drain pin-channel or from nose in case of transnasal neurosurgery); EETS, endoscopic endonasal transsphenoidal surgery; EVD, external ventricular drain; HAVM, healthcare-associated ventriculitis and meningitis; HAI, healthcare-associated infection; ICPM, intracranial pressure monitoring; ICU, intensive care unit; INSD, implantation of neurosurgical devices; ML, machine learning; NSI, Burdenko Neurosurgery Institute; OR, odds ratio; PCA, principal component analysis; RF, Random Forest classifier; ROC-AUC, area under the receiver-operating characteristic curve; RR, relative risk; SSSI, superficial surgical site infection; XGBoost, Extreme Gradient Boosting classifier; Q1;Q3, first and third quartiles.

* Corresponding author at: Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Building 3, Moscow 143026, Russia.
E-mail address: ksenia.ershova@skolkovotech.ru (K. Ershova).

https://doi.org/10.1016/j.jcrc.2018.01.022
0883-9441/

dramatically. While one study demonstrated 0.3% HAVM incidence in 1587 neurosurgical cases [5], another study found 8.6% incidence in 755 pediatric patients after neurosurgery [6]. The incidence of post-traumatic meningitis also varies and depends on trauma features. One study recently reported this rate as 3.1% for all types of brain trauma [7]. Nationwide statistics for HAVM cases in Russia are not publicly available. According to a 2016 state report, 24771 cases of healthcare-associated infections (HAIs) were registered throughout Russia (including 5623 cases of surgical site infections) without distinguishing which were HAVM [8]. The goal of our study is to fill this gap.

Studying of HAVM incidence is particularly relevant for high-risk ICU patients because the admission to the ICU independently increases the risk of HAIs, according to a 2002 report of the U.S. Centers for Disease Control (CDC) [9]. Approximately 25% of all nosocomial infections in the U.S. occurred among adults and children in ICUs [9] despite the fact that ICU beds account for just 8.5% of all hospital beds [10].

The second and third objectives include HAVM risk factors analysis. To date, several different factors have been suggested as possibly increasing the incidence of HAVM. While some of them are well-established, other factors are less certain, and many detected associations are controversial and are not well-supported by data. The craniotomy has been considered to be the main risk factor of HAVM since 1977 [11]. Additionally, invasive devices, e.g. external ventricular drains (EVD), shunts, external lumbar catheters have been shown to increase the rate of HAVM in multiple studies [1,12-14]. The role of other risk factors, including bedside ICPM, reoperation, the duration of neurosurgery, tracheal intubation, central line, and infectious complications of other localizations remains controversial [1,3,5,6].

Typically, researchers use statistical regression models to select disease risk factors [3,6,12,15,16]. However, it has been argued that regression models are not an optimal approach for such a complex problem as HAIs, where nonlinearity can be broadly presented [17]. In addition, linear models have many disadvantages, including sensitivity to data noise and multicollinearity, that can yield misleading conclusions [18]. Thus, the methods used to assess risk factors need to be improved in order to increase reliability and accuracy of the results and prevent HAVM as a final goal.

If we generalize the task of risk factor identification, we come to the well-known data science problem of feature importance ranking [19], a problem that is effectively solved by using ML [19]. We selected the decision tree-based ML algorithms for the study purpose because they are highly effective in feature selection and in dealing with nonlinearity. Specifically, we applied Random Forest (RF) classifier and Extreme Gradient Boosting classifier (XGBoost) to our data set. XGBoost is one of the most successful ML techniques, because it is computationally efficient, scalable, and prevents over-fitting. For instance, feature ranking was successfully performed by using XGBoost in e-commerce, facilitating a reduction in the number of features four times without performance quality loss [20]; the general task was very similar to ours. In medical research, XGBoost is getting more popular for solving binary classifications combined with feature selection [18,22,23]. For example, this approach identified atypical language fMRI patterns in patients with epilepsy and accurately (ROC-AUC = 0.91) distinguished between people with and without disease [21]. The ML algorithms have several advantages over regression models. Particularly important advantages offered by ML include robustness to highly correlated features and noise and the ability to retrieve nonlinear interactions across features and deal with imbalanced data. Moreover, no normalization is needed and fine-tuning parameters can reduce the impact of class imbalance in a training set without rebalancing data.

For the above mentioned reasons, the usage of ML algorithms for the selection of disease-associated risk factors is likely to grow in the future. To the best of our knowledge, no studies using tree-based ML to identify HAVM risk factors have been performed. Herein we proposed XGBoost-based ML algorithm for HAVM risk factors learning in comparison with regression models and RR analysis.

2. Materials and methods

2.1. Study setting and design

This study is a prospective observational single-center cohort study performed in the neuro-ICU at Burdenko Neurosurgery Institute (NSI) in Moscow, Russia. The NSI ICU has 38 single-bed rooms for patients with neurosurgical diseases and cares for approximately 3000 patients per year. In 2010, the program of infection prevention and control was implemented in the ICU. The study analyzed the data collected within this program. We compared two groups of patients, with and without HAVM. Both groups were selected from the high-risk patients' population (see next section).

2.2. Patients and diagnoses

The study lasted for 80 months, from October 1st, 2010 to June 30th, 2017. Only patients who stayed in the ICU for >48 h were eligible. We considered these patients to be a high-risk population and accordingly limited the study to this group only. The exclusion criteria included infections presenting on admission (according to the CDC/NHSN definition [24]) and the duration of ICU stay longer than 1000 days. All qualified patients regardless their age, conditions, disease, etc., were enrolled and participated in the study until discharge or death. Participants were enrolled starting their third to sixth day in the ICU.

HAVM was defined clinically based on the presence of at least three out of eight criteria: (1) CSF glucose level below 2.2 mmol/l or below 50% of plasma glucose in hyperglycemic patients, (2) CSF neutrophils count above 50/μl, (3) CSF protein above 220 mg/dl, (4) CSF lactate above 4.0 mmol/l, (5) positive CSF culture, (6) visualization of bacteria in CSF by Gram staining, (7) SIRS syndrome, (8) negative neurological dynamics. Infection was defined as healthcare-associated if it met the CDC criteria [24]. The case of HAVM was considered to be a factor-related if the patient had the factor (e.g. EVD, ICPM, etc.) for >48 h prior to the development of HAVM, otherwise it was deemed factor-unrelated. At the end of study, we revised HAVM cases for compliance with diagnostic criteria and confirmed them retrospectively.

Due to the open nature of the study, patients were enrolled and then left the study at different points in time. The therapeutic approach and the ICU team remained constant throughout the study. In the follow-up period (after the patient's discharge from the ICU and until the discharge from the hospital) the information regarding the total length of stay and the outcome was collected.

2.3. Data collection and preprocessing

We prospectively collected 54 parameters for each patient, including demographics, exposure to risk factors, infections, etc. (Table 1 Supplementary). The Charlson Comorbidity Index (CCI) value [25] on admission was used to assess the severity of pre-existing conditions. The data were anonymized and stored electronically as a part of NSI's health record system.

For data preprocessing we first inspected data for missing or out-of-range values. We found some occasional missing values and filled them in after retrieving the information from the health record system. If there was no information available in the system, the group mean was substituted for the missing value. Then we expanded the number of variables by generating 175 new clinically relevant aggregation features, and composed a new analytical dataset (available at https://doi.org/10.5281/zenodo.1021503).

2.4. Statistical analysis

Statistical analysis was performed in Python 3.6 using StatsModels [26], SciPy Scientific Tools [27], and scikit-learn [28]. Qualitative variables for dichotomous events are reported as number of events of one