Psychiatry Research 268 (2018) 217–222

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

A data science approach to predicting patient aggressive events in a psychiatric hospital

Robert Suchting[a,*], Charles E. Green[a,c], Stephen M. Glazier[b], Scott D. Lane[a,b]

[a] Department of Psychiatry and Behavioral Sciences, UTHealth McGovern Medical School, 1941 East Road, Behavioral and Biomedical Sciences Building 1316, Houston, TX, United States
[b] UTHealth Harris County Psychiatric Center, Houston, TX, United States
[c] Center for Clinical Research and Evidence-Based Medicine, UTHealth McGovern Medical School, Houston, TX, United States

ARTICLE INFO

Keywords:
Aggression
Machine learning
Retrospective study
Cost optimization
EHR

ABSTRACT

Recent advances in data science were used capitalize on the extensive quantity of data available in electronic health records to predict patient aggressive events. This retrospective study utilized electronic health records (N = 29,841) collected between January 2010 and December 2015 at Harris County Psychiatric Center, a 274-bed safety net community psychiatric facility. The primary outcome of interest was the presence (1.4%) versus absence (98.6%) of an aggressive event toward staff or patients. The best-performing algorithm, penalized generalized linear modeling, achieved an area under the curve = 0.7801. The strongest predictors of patient aggressive events included homelessness (b = 0.52), having been convicted of assault (b = 0.31), and having witnessed abuse (b = −0.28). The algorithm was also used to generate a cost-optimized probability threshold (6%) for an aggressive event, theoretically affording individualized hospital-staff coverage on the 2.8% of in-patients at highest risk for aggression, based on available hospital operating costs. The present research demonstrated the utility of a data science approach to better understand a high-priority event in psychiatric in-patient settings.

1. Introduction

Patient aggression in mental health care settings presents an ongoing challenge to healthcare organizations and practitioners. Aggression can result in physical and psychological trauma to other patients, staff, and visitors. The sequelae of aggressive incidents can include a variety of problems including utilization costs, increased staffing needs (e.g., isolation, individual observation), loss of staff time due to injury, and staff morale and turnover (Al-Sagarat et al., 2016; Lanctôt and Guay, 2014). In forensic psychiatry settings, up to 70% of staff report being assaulted by patients (Kelly et al., 2014). The potential suffering and cost of aggressive incidents render them priority concerns for practitioners and administrators. However, relatively low base rates complicate the prediction that is requisite for prevention of aggressive events (Bader and Evans, 2015; Raja and Azzoni, 2005). This necessitates that investigators mine large data sets in which any candidate data analytic techniques must countenance large numbers of observations and predictors. While traditional data analytic techniques have notable limitations addressing these large data sets (Alwee et al., 2013; Hvistendahl, 2016; Kuhn and Johnson, 2013), within the past decade, growth in computational power and machine learning techniques have advanced to provide data scientists with additional data analytic tools to address such problems (Hastie et al., 2009; Witten et al., 2011).

The present study explored patient aggression in the psychiatric inpatient hospital setting by employing a battery of data science techniques. Data science is a broad term for methodology that utilizes machine-learning algorithms for prediction typically with large, complex datasets involving an extensive number of variables. These tools have been applied to a wide variety of “big data” problems across a diverse range of phenomena. Examples include automated webpage ranking, probability of loan default, detecting oil slicks from illegal dumping using satellite imagery, electrical supply load forecasting (Witten et al., 2011), and competitive awards to develop algorithms to improve online movie recommendations (Feuerverger et al., 2012). In the neurobehavioral sciences, such techniques are often employed to integrate and reduce dimensionality in large, complex brain imaging and genetic datasets in order to predict mental health status/diagnosis and disease progression/treatment outcomes using classification- or regression-based data science techniques (Dmitrzak-Weglarz et al.,

* Corresponding author.
E-mail address: robert.suchting@uth.tmc.edu (R. Suchting).

https://doi.org/10.1016/j.psychres.2018.07.004
Received 13 February 2018; Received in revised form 3 July 2018; Accepted 4 July 2018
Available online 21 July 2018

2013; Hajek et al., 2015; Huys et al., 2016; Patel et al., 2016; Pettersson-Yeo et al., 2013; Veronese et al., 2013).

To our knowledge, there is limited previous work examining patient aggression in mental health facilities utilizing machine learning algorithms for prediction based on available electronic health record (EHR) data. The present research sought to derive a predictive model of aggressive incidents in mental health hospital using a large (N = 29,841) data sample. By design, "black-box" machine learning tools leverage intense exploration of available data and inductive discovery of predictor variables. While hypothesis-driven methods are the *sine qua non* for advancing scientific disciplines, the present search space of 328 predictor variables rendered the use of machine learning algorithms both attractive and necessary. These techniques simultaneously engage in variable selection and optimization of prediction (with appropriate cross-validation). Moreover, the degree to which resulting predictive models provide indices of variable importance enables the results to function as hypothesis generating. The present goal was to develop a model that produced an estimate of the probability of an aggressive incident for any patient admitted for the first time to the facility, based only on data available from the initial intake and assessment process. Development of a successful model would serve as a first step toward the use of efficient probabilistic estimates to enable risk-stratification and potentially optimized allocation of hospital resources for those patients at highest risk for aggression.

2. Methods

2.1. Participants & procedure

The present analyses utilized intake / assessment data from 29,841 patients admitted to UTHealth Harris County Psychiatric Center (HCPC) between January 2010 and December 2015. HCPC is a 274-bed community safety-net hospital serving Harris and surrounding counties in the greater Houston area. The hospital, based in a large urban setting with an average of >9000 admissions annually, representative of public inpatient psychiatric hospitals in many large U.S. metropolitan areas (geographic differences in diversity not withstanding). Patient admission to HCPC occurred across a range of varying conditions (i.e., psychiatric symptom presentation, voluntary / involuntary status, time of day, referral sources). Data were not differentially included or excluded based on any specific diagnostic or admission criteria. Patients provided information as part of standardized hospital admitting and assessment procedures, summarized below. Aggregation of data for the present analysis followed a structured investigation of available data, wherein a team of subject matter experts in psychiatry and data science evaluated the hospital's existing EHR databases to maximize potentially useful patient information. This process produced a dataset of 328 predictors (313 categorical, 15 continuous) and one outcome measure (positive/negative for an aggressive incident). An aggressive event is coded into the hospital medical record following any episode of uncontrolled verbal or physical aggression that required intervention by and assistance from additional hospital staff to manage the event. In cases of verbal aggression, such intervention would indicate that physical aggression was deemed imminent by staff. Data were missing in approximately 10% of observations, with the majority of missingness occurring in the categorical predictors. Assuming that missingness is itself a potentially important factor, missingness in each categorical predictor was addressed through the creation of a new categorical level to indicate "missing." Remaining missing data (less than 1%) in the continuous predictors were handled natively by each machine learning algorithm (e.g., imputation, data partitioning).

2.2. Outcome variable – Aggression

Aggressive events were obtained from the patient EHR. All episodes of aggression are mandatorily recorded by hospital staff and coded in the EMR as patient-on-patient or patient-on-staff; both types were included in the data analyses. All aggressive events included codes for the type of action taken in response to the aggressive incident (e.g., "Transferred to:", "Family notified", "Plan of care revised", "Education/ Training", etc.). Type of action taken in response to the incident was not included in the data analyses.

2.3. Predictor variables

Upon admission, all patients were given a comprehensive assessment and initial psychiatric examination by the hospital admissions and medical staff. This examination included a full demographic profile, patient vitals (i.e., height, weight, blood pressure), and a comprehensive psychosocial assessment, including histories of early development, education, military service, vocation/work, medical status, psychiatric status, drug/substance use and treatment, nicotine/tobacco use and counseling, abuse (victim or perpetrated physical/verbal/emotional/ sexual abuse), legal status, marital status, religious beliefs, financial status, and living situation. A nursing assessment collected information regarding sleep habits, pain status, patient behavior during interview, a risk assessment, and evaluation of patient mood (via the Affective Disorders Rating Scale (ADRS; Swann et al., 2004). The initial psychiatric evaluation/mental status exam assessed general appearance (i.e., hygiene), musculoskeletal system, speech pattern, thought processes and content, perception, depression, affect, insight, judgment, skin integrity, head trauma, suicidal/homicidal/assault ideation, deterioration in function, chemical dependency, hallucinations, and delusions.

From these initial assessments, the final set of 328 predictor variables was derived. Names, descriptions, and (where applicable) standardized instruments used to obtain the 328 predictor variables are provided in online Supplement 1 and presented in an order that closely matches the description above.

2.4. Data analytic strategy

The present study used the H2O software v. 3.18.0.11 (Aiello et al., 2016b) as scripted in the R statistical computing environment (R Core Team, 2018). H2O provides a powerful Java-based platform for use with large sample sizes. The software includes implementations of several machine learning algorithms; the present study utilized four of these: penalized generalized linear modeling (GLM), random forest (RF), gradient boosting machine (GBM), and deep neural networks (DNN). A detailed account of the data analytic strategy (including a description of each algorithm) is provided in online Supplement 2. Extensive descriptions of these algorithms may be found in H2O's documentation (Aiello et al., 2016a). A summary of the overall data science approach applied in the present study is presented in Fig. 1.

The data science workflow proceeded by splitting the full data set into a training set and a test set, and within the training set, further partitions were made to tune model hyperparameters (algorithm "tuning knobs;" see Supplement 2 for an account of hyperparameter search space). These hyperparameters determine how the algorithm parses data. For example, the penalized GLM has a tuning knob related to how strong the penalization needs to be. The training set was used to tune each of the four algorithms, and each tuned algorithm was then evaluated using the test set. Imbalance in the outcome variable (aggressive incidents were rare) was addressed using options related to class rebalancing native to each algorithm (e.g., over-sampling, under-sampling, or a combination). Knowledge discovery was driven by examining variable importance metrics of the four tuned algorithms. Model performance during tuning and testing was determined by the highest area under the receiver operating characteristic curve (AUC). The best-performing algorithm was considered optimal for the present analysis and was given primary focus for evaluation.