Journal of Applied Microbiology

Journal of Applied Microbiology ISSN 1364-5072

ORIGINAL ARTICLE

Identifying predictors of antimicrobial exposure in hospitalized patients using a machine learning approach

A.S. Chowdhury[1], E.T. Lofgren[2,3], R.W. Moehring[4] and S.L. Broschat[1,2,5]

1 School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA, USA
2 Paul G. Allen School for Global Animal Health, Washington State University, Pullman, WA, USA
3 Department of Mathematics and Statistics, Washington State University, Pullman, WA, USA
4 Department of Medicine, Duke University School of Medicine, Durham, NC, USA
5 Department of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA, USA

Keywords
antimicrobial resistance, antimicrobial stewardship, antimicrobial utilization, cubist regression, data imputation, machine learning, patient features, standardized antimicrobial administration ratio, support vector regression.

Correspondence
Abu Sayed Chowdhury, School of Electrical Engineering and Computer Science, Washington State University, PO Box 642752, Pullman, WA 99164-2752, USA.
E-mail: abu.chowdhury@wsu.edu

2019/1491: received 22 August 2019, revised 17 October 2019 and accepted 22 October 2019

doi:10.1111/jam.14499

Abstract

Aims: Analysis and tracking of antimicrobial utilization (AU) are crucial in antimicrobial stewardship efforts which are used to find effective interventions for controlling antimicrobial resistance. In antimicrobial stewardship, standard risk adjustment models are needed for benchmarking appropriate AU and for fair inter-facility comparison. In this study we identify patient- and facility-level predictors of antimicrobial usage in hospitalized patients using a machine learning approach, which can be used to inform a risk adjustment model to facilitate assessment of AU. To our knowledge, this is the first time machine learning has been applied for this purpose.

Methods and Results: Patient admission records were retrieved from the Duke Antimicrobial Stewardship Outreach Network which include clinical data for 27 community hospitals in the southeastern United States. Candidate features (predictors) were then generated from these records. The number of features was reduced using a statistical approach, and missing values of the reduced feature set were imputed using bootstrapping and expectation-maximization algorithm. Finally, support vector regression (SVR) and cubist regression (CB) models were applied to find root-mean-square error values which were used to evaluate the selected feature set. The performance of the SVR and CB models was found to be better than that of linear null and negative binomial null models, thereby demonstrating the effectiveness of our selected features.

Conclusions: Relevant patient- and facility-level predictors of antimicrobial usage in days of therapy were obtained and evaluated. The potential predictor set can be used in risk adjustment strategies for benchmarking antimicrobial use.

Significance and Impact of the Study: One reason for the rapid emergence of antimicrobial resistance is inappropriate use of antibiotics in hospitalized patients. Identifying predictors of antimicrobial exposure using a machine learning technique can improve the use of AU, enhance patient health outcomes, and reduce the infection spread caused by antimicrobial-resistant organisms.

Introduction

Millions of people contract infectious diseases every year, and several thousand of them die due to antimicrobial resistance (Centers for Disease Control and Prevention (US) 2013). Hence, there is an urgency for hospital and clinical antimicrobial stewardship programs (ASPs) to analyze and track antimicrobial utilization (AU) which will enable identification of interventions and targets for resource allocation (MacDougall and Polk 2005; Doron

and Davidson 2011; Standiford *et al.* 2012; Yu *et al.* 2014). AU estimation benchmarked to an external comparator can assist ASPs to identify opportunities for improving AU in their facilities. However, before benchmarking can be achieved we need to determine the most suitable strategy for diverse hospitals and clinics to compare AU.

Use of national estimates of AU through the National Healthcare Safety Network (NHSN) (Centers for Disease Control and Prevention 2016) AU option provides facilities an external comparator to help determine where to focus further investigation to find target areas for ASP intervention and to provide more meaningful comparisons for stewards attempting to engage clinical staff. This external comparator is called the standardized antimicrobial administration ratio (SAAR) (Pollack and Srinivasan 2014; Centers for Disease Control and Prevention 2016). The SAAR provides an observed-to-expected ratio and allows a facility or unit to compare to the risk-adjusted national baseline. Current SAAR risk adjustment models (van Santen *et al.* 2018) include a limited number of facility- and unit-level predictors such as bed size and location type (ICU or ward; medical, medical surgical or surgical). The literature on benchmarking AU describes a few patient- and hospital-level predictors derived from academic medical centres participating in the University Health Consortium (now Vizient, Inc.) (Pakyz *et al.* 2009; Polk *et al.* 2011; Ibrahim and Polk 2014), large administrative databases (MacDougall and Polk 2008), standalone padiatric hospitals (Gerber *et al.* 2013) and a single health system (Stenehjem *et al.* 2016). However, further investigation is needed to identify potential predictors in hospitalized patients. Examples of existing patient-level predictors include diagnosis codes (International Classification of Diseases, 10th Revision (ICD-10)) and diagnosis-related groups (APR-DRG) for infectious diseases which are primarily used for hospital billing purposes. However, these predictors might be applied to correctly identify patients who have clinical evidence of acute infection requiring antimicrobial treatment (Ibrahim and Polk 2014; Fridkin 2016).

To determine which predictors need to be collected in a national system, it is necessary to build a robust prediction model for improvement of the facility-to-NHSN comparisons and to increase the ability of the SAAR to identify stewardship opportunities for an individual facility. In this work, we present a machine learning approach to identify patient- and facility-level predictors of antibiotic exposure in hospitalized patients that can be used in risk adjustment strategies for hospital benchmarking. To our knowledge, this is the first time machine learning has been applied for such an objective. We generated features (predictors) from electronic health records and then assessed these features to identify the optimal ones for prediction. Data imputation was performed for missing values in the reduced feature set. Finally, we divided the dataset into training and test sets to evaluate the performance of support vector regression (SVR) (Flake and Lawrence 2002; Lin 2001; Smola and Schölkopf 2004; Chowdhury *et al.* 2015; Chowdhury *et al.* 2019b) and cubist regression (CB) (Quinlan 2005; Kuhn *et al.* 2012) models.

Materials and methods

Data collection

We retrieved patient admission records from the Duke Antimicrobial Stewardship Outreach Network (DASON) (Duke University School of Medicine 2019) which includes 27 hospitals in the southeastern United States. DASON contains detailed electronic medical administration records (eMAR) for antimicrobials, patient movement data (bed flow), demographics and billing data (for example, DRG and diagnosis codes). We gathered 382 943 admission records for adult patients ≥18 years old who had hospital stays between October 1, 2015, and September 30, 2017, and used this dataset to train and test our machine learning models. A summary of key demographic and clinical variables for the DASON analytical dataset are given in Table S1. We considered six adult group AU outcomes (as defined by the NHSN) for all-antibacterials, narrow spectrum beta-lactam, *Clostridium difficile* (CDI), as well as agents used for community-onset, hospital-onset and resistant Gram-positive infections, respectively, generating six separate datasets for these SAAR groups. Note that the latter five SAAR groups are subsets of the all-antibacterials SAAR group. SAAR agents are listed in Table S2.

Feature extraction

We considered the features available in the DASON database and also derived new features from the database. Table 1 lists all the features with their dimensions. Elixhauser Score, one of the features we used, measures the presence or absence of a large number of common comorbidities which is a proxy for the complexity of each patient case. The Diagnosis-related Group (DRG)-specific weight (DRG Weight) was obtained directly from the DASON database; it provides a measure of the average resources used for the treatment of medicare patients for the corresponding DRG. This numeric score is a proxy for the intensity of a particular patient's condition. We derived 116 indicator variables (0 or 1) for 116 comorbidity groups. The resulting feature vector corresponds to

Table 1 List of features for analysis of the DASON dataset

Feature	Feature dimension
Elixhauser score	1D
DRG weight	1D
Comorbidity group code	116D
CCS single level category	283D
Total LOS	1D
Number of bedflow	1D
Season	1D
Days from reference date	1D
LOS_nhsnunit	42D
DRG	1D
Number of diagnoses	1D
Gender	1D
Race	1D
Calculated age	1D
Reported height	1D
Reported weight	1D
Death indicator	1D

the medical conditions of a given patient—for example, if they have chronic pulmonary disease, anaemia due to blood loss, etc. We also generated 283 indicator variables (0 or 1) based on the 283 categories of all single-level diagnosis codes (ICD-10-CM codes) generated by the Clinical Classification Software (Agency for Healthcare Research and Quality 2019). These features give specific diagnoses for each patient. The binary features, 0 and 1, indicate absence or presence, respectively.

We computed Total Length of Stay (Total LOS) defined as the number of days a patient spends in any inpatient location for each admission to the hospital. The Number of Bedflow feature is the number of times a patient appears in the 'Bedflow' table in the DASON database. This number reflects whether or not a patient is frequently transferred within a hospital, potentially indicating a rapidly changing clinical trajectory or an increased risk of infection due to exposure to a larger number of fellow patients. The Season feature gives the month a patient was admitted, and Days from Reference Date is defined as the number of days counted from October 1, 2015, and accounts for potential temporal trends within the data. LOS_nhsnunit gives the length of stay for each patient in 42 potential NHSN unit types. The DRG feature is used for classification of inpatient stay and is required for payment. This feature is also used to identify the primary reason for a patient's admission.

The Number of Diagnoses feature for each patient admission acts as a proxy for the complexity of the patient's conditions, with a greater number of diagnoses assumed to indicate a more complicated case. Patient demographics such as gender, race, age, height and weight are taken directly from the DASON database. Gender can be male, female, other and unknown. Types of race can be White/Caucasian, Black or African American, Asian, American Indian/Alaskan Native, Native Hawaiian/Other Pacific Islander, Hispanic or Latino, Other and Unknown. The Death Indicator feature is an indicator variable (0 or 1) indicating whether a patient died during their admission. '0' and '1' stand for alive and died status, respectively. Finally, our outcome variable is the Days of Therapy (DOT) which is based on administration of an antimicrobial agent. One DOT is equivalent to one or more doses of the same agent administered on the same calendar day. A single calendar day can count as more than one DOT if more than one agent is given on that same day.

In summary, we generated 455 features from the DASON database listed in Table S3. Of these, 51 are numeric and the remaining 404 are categorical.

Feature selection and data imputation

A greater number of features does not necessarily imply more accurate results in machine learning. In fact, redundant and irrelevant features may actually degrade the performance of a predictive model and compromise accuracy (Chowdhury *et al.* 2019b). To remove redundant features and to find the most relevant features, we applied feature reduction techniques for mixed type data. We started with 354 579 complete adult records with no missing values and eliminated features as follows.

First we deleted features having the same value for all records regardless of whether an individual was or was not treated with an antimicrobial because such features cannot contribute to the prediction of the outcome variable, in this case the DOT. Then, to remove redundant features, we computed Pearson's correlation coefficient between features, that is, the linear correlation between two feature vectors u and v are estimated. We calculated Pearson's correlation coefficient $\rho_{u,v}$ as described in our previous work (Chowdhury *et al.* 2019a). The two features are considered highly correlated when the value of $|\rho_{u,v}|$ is large, and in this situation we can use either of the features for regression tasks. For correlation values between two features greater than 0·95, we eliminated one of the features.

Next, we used the chi-squared test to remove redundant categorical features. This test is applied to determine if there is a significant correlation between two categorical variables based on their distributions. The chi-squared statistic is computed using Eqn (1).

$$\chi^2 = \sum_{k=1}^{n} \frac{(O_k - E_k)^2}{E_k} \quad (1)$$

Here O_k indicates the observed frequency count for each category, E_k stands for the expected frequency count for each category and n is the total number of categories.

For two features with *P*-values less than 0·001, we kept one of the two features and ignored the other.

Finally, we performed linear regression on all remaining features *vs* DOT (our numeric output variable) for further feature reduction. In this step, features having *P*-values less than 0·05 were removed from further consideration. Table 2 shows the number of features eliminated at each step. The same number of features were eliminated in all steps except for the last, and no numeric features were removed. The reason for elimination of the same number of features in the first three steps is because the narrow spectrum beta-lactam, CDI, community-onset, hospital-onset and resistant Gram-positive SAAR groups are subsets of the all-antibacterials SAAR group. The final list of features and common features identified in all six adult SAAR groups are listed in Tables S4–S10.

After feature reduction using the complete records, we identified missing values for several features in our selected feature set as shown in Table 3. We used Amelia II (Honaker *et al.* 2011), which employs a bootstrap expectation-maximization (EM) algorithm (Dempster *et al.* 1977; Efron and Tibshirani 1994), to impute missing values in our dataset. Amelia is a probabilistic approach that gives unbiased estimates of missing values. It bootstraps existing data samples and estimates means and variances via the EM algorithm which are used in turn to impute missing values using regression.

SVR model

To measure the effectiveness of our final feature set, we used SVR (Lin 2001; Flake and Lawrence 2002; Smola and Schölkopf 2004; Chowdhury *et al.* 2015; Chowdhury *et al.* 2019b), a machine learning model based on the support vector machine (SVM). It differs from the SVM in that the dependent variable of SVR is numeric. SVR is a nonparametric technique that depends on the kernel function used rather than on the distributions of the underlying dependent and independent variables. We used L2-regularized SVR (dual) with a linear kernel to measure the accuracy of our final feature sets. We tuned the SVR model to find the best cost C and loss L parameters (Koshiba and Abe 2003; Tang 2013) by considering various parameter combinations.

Table 2 Feature elimination statistics

	Number of eliminated features				
Adult SAAR group	Step 1	Step 2	Step 3	Step 4	Remaining features
All-antibacterials	39	0	337	22	57
Beta-lactam	39	0	337	31	48
CDI	39	0	337	27	52
Community-onset	39	0	337	31	48
Hospital-onset	39	0	337	28	51
Resistant Gram-positive	39	0	337	26	53

Table 3 Missing values in the feature set

Feature	Number of missing values
DRG weight	2440
Reported height	24 544
Reported weight	13 083

CB model

In addition to the SVR model, we used a rule-based CB model (Quinlan 2005; Kuhn *et al.* 2012) which combines Quinlan's M5 model tree (Quinlan 1992) with nearest-neighbour correction and boosting (Quinlan 1993; Quinlan 2014). In our CB model, a tree is initially created and every path of the tree is collapsed into a rule. The model fits a linear regression model for a rule depending on the data subset represented by the rule. The set of rules is pruned or merged, and linear regression models found in the terminal leaves of the constructed tree are used to make a prediction. This prediction is improved using predictions from earlier nodes of the tree (which occurs recursively up the tree). Although CB regression is consistent with Quinlan's M5 model tree, it generalizes the model by adding boosting (Bühlmann and Hothorn 2007; Sagi and Rokach 2018) and instance-based corrections. We tuned our CB model empirically to find the best committees C' and neighbours N parameters using various parameter combinations.

Linear model and negative binomial generalized linear model

In a linear model (LM), a response variable is represented as a linear function of all predictor variables; the negative binomial generalized linear model (NB-GLM) is a generalization of a LM. Its response variable is the count of occurrences of an event whose distribution follows a negative binomial distribution (Madsen and Thyregod 2010). A null model uses the null hypothesis for all regression parameters equal to zero except for the intercept—that is, the mean of the outcome is the only useful predictor.

R scripts and packages

We wrote scripts in R (https://cran.r-project.org/mirrors.html) to implement feature selection, data imputation and classification. We utilized the R *stats* (ver. 3.5.1)

package to perform Pearson's correlation and chi-squared test measurements and to implement the null LM model, the *MASS* (ver. 7.3-50) R package to implement the null NB-GLM model, the *Amelia* (ver. 1.7.5) R package to perform data imputation, the *caret* (ver. 6.0-80) and *LiblineaR* (ver. 2.10-8) R packages to implement our SVR model, and the *Cubist* (ver. 0.2.2) and *caret* (ver. 6.0-80) R packages to implement our CB model.

Performance measurement

We measured the performance of our machine learning models with our reduced feature set by computing the root-mean-square error (RMSE) calculated using Eqn (2).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{j=1}^{N}(y_j - \hat{y}_j)^2}{N}} \quad (2)$$

where y_j is the actual DOT value, $\hat{y}_j$ is our predicted DOT value, and N is the number of test samples. Lower RMSE values indicate better DOT prediction.

Data availability statement

The data that support the findings of this work are available on request from the corresponding author. The data are not publicly available due to privacy issues.

Code availability statement

R scripts written to implement our method are available from the corresponding author upon reasonable request.

Results

Training performance analysis

We divided each of the six SAAR group datasets into training and test sets with 75% of the data used to train the SVR and CB models and the remaining 25% used to test them. To train our models, we applied k-fold repeated cross validation with $k = 5$ and eight repetitions. We set tuneLength = 5, the total number of parameter combinations to be evaluated. We selected the best SVR and CB models based on RMSE values and these were used with our test set. The parameter values for the best models (that is, cost C and loss L values for SVR and committees C' and neighbours N parameters for CB) with respective RMSE values are given in Table 4. Note that for the SVR, the L_1 loss function performs better than L_2 for all six datasets. This is because L_1 is more robust to outliers in the data than L_2 which squares the sum of slack variables in the regularization term of the objective function used to optimize the SVR. A slack variable is a term added to change an inequality to an equality in an equation, and the regularization term quantifies the degree of importance of misclassification. Outliers are more likely when calculating DOT, which is heavily skewed toward zero but with the presence of some extreme values. The C' and N values for the CB model are 20 and 9, respectively, for the majority of the datasets. The training performance of the CB model is much better than that of the SVR for all six datasets because of its use of boosting. Boosting reduces bias and variance and strengthens machine learning models. As an illustration, the training RMSE for the CB model for all-antibacterials is 1·66 less than its SVR counterpart.

Table 4 Parameter values used to obtain the best SVR and CB models

	SVR			CB		
Adult SAAR Group	C	L	RMSE	C'	N	RMSE
All-antibacterials	4	L_1	7·17	20	9	5·51
Beta-lactam	0·25	L_1	1·62	20	9	1·50
CDI	2	L_1	3·20	20	9	2·43
Community-onset	2	L_1	3·01	10	9	2·12
Hospital-onset	0·25	L_1	3·08	20	9	2·32
Resistant Gram-positive	2	L_1	2·28	20	9	1·97

Feature importance analysis

Figures 1 and 2 show the feature scores for the five most important features acquired from training our two models. The higher the feature score, the more influence it has during prediction. Total LOS is found to be the most important feature for the SVR model for all six SAAR groups. However, for the CB model, DRG Weight is the most important feature for four of the SAAR groups while Total LOS is the most important feature for only two of them.

Test performance analysis

Using our test dataset, we evaluated our best SVR and CB models. We compared the RMSE values for the SVR and CB models with those found using a null LM and a null NB-GLM. The RMSE values of the null models are included to provide a point of comparison using a very simple mean-based prediction model for evaluation of the machine learning model's performance in the appropriate context. As shown in Table 5, both the SVR and CB models show better predictive accuracy than the null LM and null NB-GLM models for all SAAR groups. The

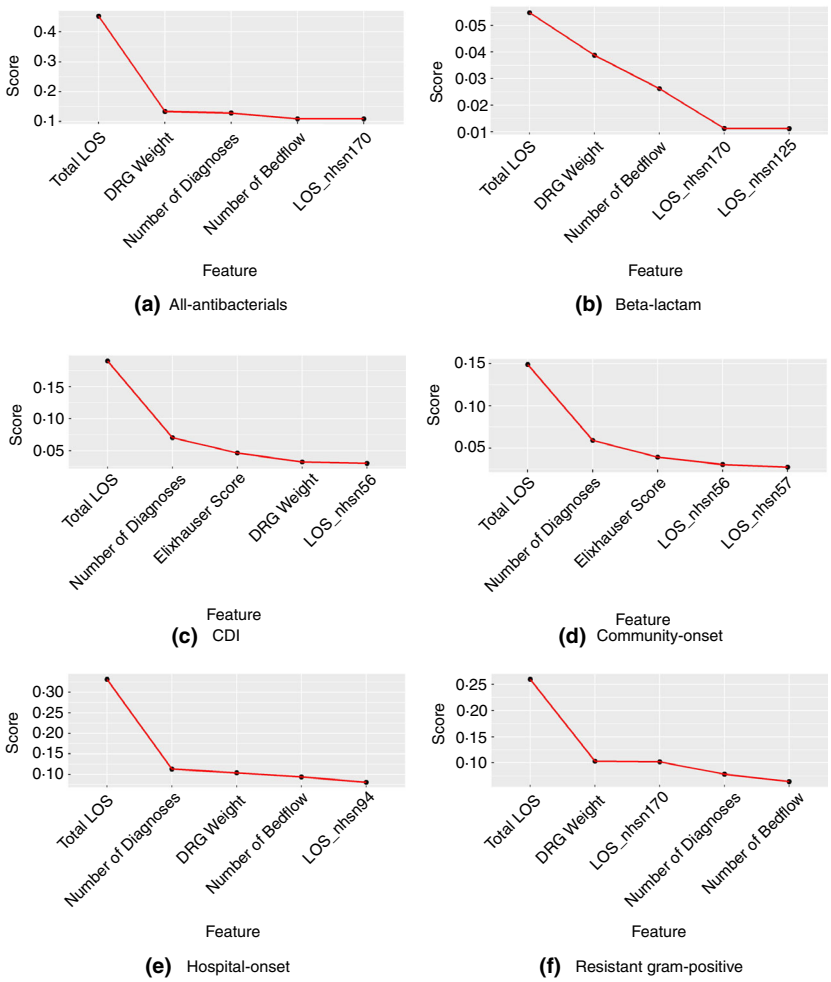

Figure 1 SVR scores for the top five features for six adult SAAR datasets. Total length of stay (LOS) is the most significant feature for all six datasets.

test performance of the CB model surpasses that of the SVR model for all but one of the SAAR groups for which the same RMSE values are obtained.

Comparison of the training and test results presented in Tables 4 and 5 shows the RMSE values for the SVR and CB models to be in agreement, which is especially true for the CB model. This suggests that our machine learning models were not overfit to the training data, that is, their accuracy is not affected by noisy data and, thus, they should work well for arbitrary datasets. Overall, our machine learning models resulted in better performance than the null statistical models, especially our CB model.

Discussion

Identifying predictors for AU from patient admission records is important for benchmarking AU in SAAR risk adjustment methods. Unlike traditional approaches, machine learning techniques provide a means of efficiently measuring the importance of a large number of candidate predictors and are especially useful in situations for which the primary objective of the model is prediction. Thus, they can be used to determine the specific predictive factors needed for collection in a nationwide system when the benefits of improved risk adjustment are weighed against the burden of data collection and standardization. In this paper we demonstrated that a machine learning approach can be used to identify and derive potential predictors from hospital datasets. We utilized both patient and facility information from 27 different hospitals procured over a 2-year period to extract features and then applied statistical methods to find the optimal feature set. We used Amelia II to impute missing values, and SVR and CB models were trained and tested with all our data. Smaller RMSE values obtained using the SVR and CB models relative to those obtained from two null models established the predictive power of our selected feature sets. The results obtained using our CB model were particularly impressive; it outperformed our SVR model for all cases. We used our approach to

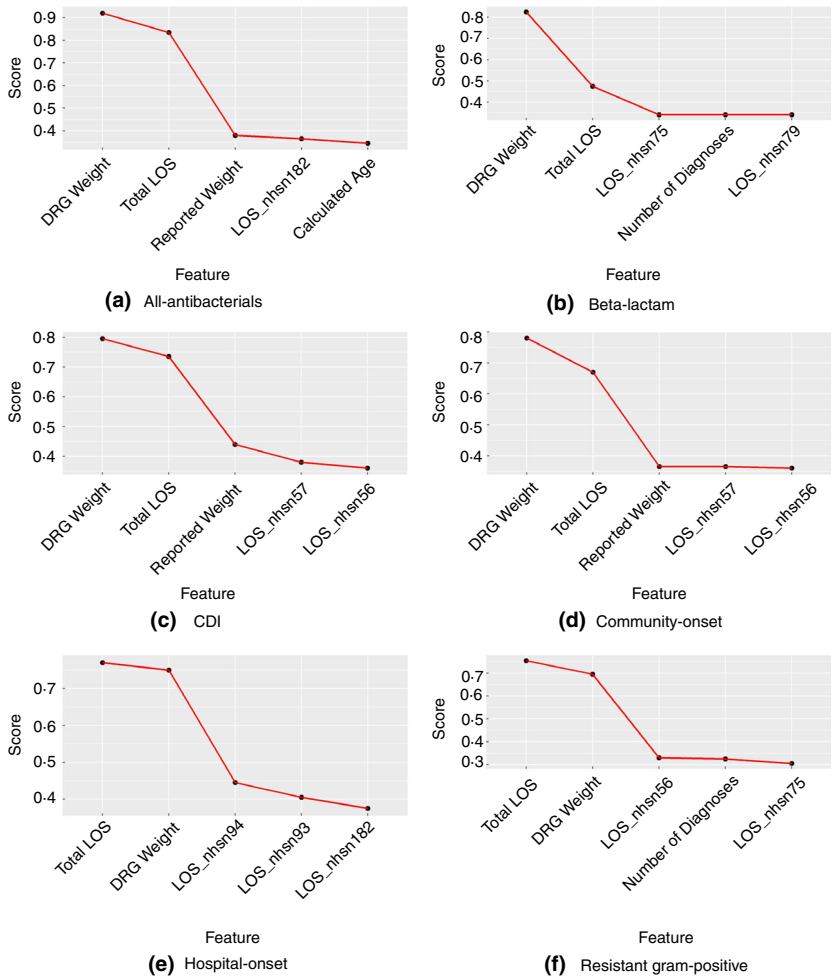

Figure 2 CB scores for the top five features for six adult SAAR datasets. DRG Weight and Total length of stay (LOS) are the most significant features for the six datasets.

determine that the DRG weight and Total LOS are the top ranked predictors for all six different adult SAAR groups. Other features such as the Elixhauser score, number of bedflow, length of stay in medical/surgical ward/critical care and number of diagnoses are also important predictors for all six different adult SAAR groups. The findings of our method support the goal of improving antimicrobial stewardship data for action by providing relevant and robust patient- and facility-level predictors of antimicrobial exposure to be employed in SAAR methodology from a diverse set of hospitals. This knowledge can help in the determination of how to expand the NHSN AU infrastructure to improve and refine the SAAR. In future work, we will consider a more robust feature selection approach for heterogeneous data to confirm these results.

Table 5 Test performance of the best SVR and CB models. Values given are RMSE. In general, the CB model results are better than those of the SVR model

Adult SAAR group	Null LM	Null NB-GLM	SVR	CB
All-antibacterials	8·16	8·60	5·86	5·17
Beta-lactam	1·50	1·90	1·48	1·48
CDI	2·73	2·90	2·47	2·42
Community-onset	2·28	2·48	2·24	2·09
Hospital-onset	3·22	3·37	2·62	2·45
Resistant Gram-positive	2·39	2·61	2·32	1·98

Ongoing stewardship efforts within DASON and elsewhere present the opportunity to validate our predictive models with fresh data in a rapidly changing field and to explore the utility of our models not only in risk adjustment, but potentially in informing patient care. Hospitals in the United States will benefit from the development of SAAR methodology which will allow them to further refine and focus their antimicrobial stewardship activities for maximum effect in optimizing antimicrobial use. By using the identified predictors, ASPs will be able to take action to meet their ultimate goal of decreasing inappropriate AU with subsequent reduction or containment of antimicrobial resistance and ultimately to avert negative patient outcomes.

Conflict of Interest

We declare that no conflict of interest exists.

Acknowledgements

This work was supported by the U.S. Center for Disease Control and Prevention's (CDC's) investments to combat antibiotic resistance under award number 200-2018-96423.

References

Agency for Healthcare Research and Quality (2019) Clinical classifications software (CCS) for ICD-10-CM/PCS. https://www.hcup-us.ahrq.gov/toolssoftware/ccs10/ccs10.jsp. Accessed January 6, 2019.

Bühlmann, P. and Hothorn, T. (2007) Boosting algorithms: regularization, prediction and model fitting. *Stat Sci* **22**, 477–505.

Centers for Disease Control and Prevention (US) (2013) Antibiotic resistance threats in the United States, 2013. US Department of Health and Human Services: Centers for Disease Control and Prevention.

Centers for Disease Control and Prevention (2016) National Healthcare Safety Network: antimicrobial use and resistance (AUR) options. Accessed March, 31.

Chowdhury, A.S., Alam, M.M. and Zhang, Y. (2015) A biomarker ensemble ranking framework for prioritizing depression candidate genes. In 2015 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB) (pp. 1–6) IEEE.

Chowdhury, A.S., Khaledian, E. and Broschat, S.L. (2019a) Capreomycin resistance prediction in two species of mycobacterium using a stacked ensemble method. *J Appl Microbiol* **127**, 1656–1664.

Chowdhury, A.S., Call, D.R. and Broschat, S.L. (2019b) Antimicrobial resistance prediction for Gram-negative bacteria via game theory-based feature evaluation. *Sci Rep* **9**, 1–9.

Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J Roy Stat Soc: Ser B* **39**, 1–22.

Doron, S. and Davidson, L.E. (2011). Antimicrobial stewardship. In *Mayo Clinic Proceedings* ed. Nath, K.A. (vol. 86, no. 11, pp. 1113–1123). Amsterdam, The Netherlands: Elsevier.

Duke University School of Medicine (2019) Duke antimicrobial stewardship outreach network (DASON). https://dason.medicine.duke.edu/. Accessed January 6, 2019.

Efron, B. and Tibshirani, R.J. (1994) *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press.

Flake, G.W. and Lawrence, S. (2002). Efficient SVM regression training with SMO. *Machine learning* **46**, 271–290.

Fridkin, S. (2016) Advancing Metrics for Benchmarking in Stewardship Research: Is perfection our enemy? Presentation at SHEA Antimicrobial Stewardship Research Workshop, San Diego, CA. November 2016. Centers for Disease Control and Prevention.

Gerber, J.S., Kronman, M.P., Ross, R.K., Hersh, A.L., Newland, J.G., Metjian, T.A. and Zaoutis, T.E. (2013) Identifying targets for antimicrobial stewardship in children's hospitals. *Infect Control Hosp Epidemiol* **34**, 1252–1258.

Honaker, J., King, G. and Blackwell, M. (2011) Amelia II: A program for missing data. *J Stat Softw* **45**, 1–47.

Ibrahim, O.M. and Polk, R.E. (2014) Antimicrobial use metrics and benchmarking to improve stewardship outcomes: methodology, opportunities, and challenges. *Infect Dis Clin* **28**, 195–214.

Koshiba, Y. and Abe, S. (2003) Comparison of L1 and L2 support vector machines. In Proceedings of the International Joint Conference on Neural Networks, 2003 (vol. 3, pp. 2054–2059). IEEE.

Kuhn, M., Weston, S., Keefer, C. and Coulter, N. (2012) Cubist models for regression. R package Vignette R package version18.

Lin, C.J. (2001) On the convergence of the decomposition method for support vector machines. *IEEE Trans Neural Networks* **12**, 1288–1298.

MacDougall, C. and Polk, R.E. (2005) Antimicrobial stewardship programs in health care systems. *Clin Microbiol Rev* **18**, 638–656.

MacDougall, C. and Polk, R.E. (2008) Variability in rates of use of antibacterials among 130 US hospitals and risk-adjustment models for interhospital comparison. *Infect Control Hosp Epidemiol* **29**, 203–211.

Madsen, H. and Thyregod, P. (2010) *Introduction to General and Generalized Linear Models*. Boca Raton, FL: CRC Press.

Pakyz, A.L., Gurgle, H.E., Ibrahim, O.M., Oinonen, M.J. and Polk, R.E. (2009) Trends in antibacterial use in hospitalized pediatric patients in United States academic health centers. *Infect Control Hosp Epidemiol* **30**, 600–603.

Polk, R.E., Hohmann, S.F., Medvedev, S. and Ibrahim, O. (2011) Benchmarking risk-adjusted adult antibacterial drug use in 70 US academic medical center hospitals. *Clin Infect Dis* **53**, 1100–1110.

Pollack, L.A. and Srinivasan, A. (2014) Core elements of hospital antibiotic stewardship programs from the Centers for Disease Control and Prevention. *Clin Infect Dis* **59** (suppl. 3), S97–S100.

Quinlan, J.R. (1992) Learning with continuous classes. In *5th Australian Joint Conference on Artificial Intelligence* (vol. **92**, pp. 343–348). Singapore: World Scientific.

Quinlan, J.R. (1993) Combining instance-based and model-based learning. Proceedings of the Tenth International Conference on Machine Learning (pp. 236–243). Burlington, MA: Morgan Kaufmann.

Quinlan, J.R. (2005) An overview of cubist. RuleQuest Research. https://www.rulequest.com/cubist-info.html

Quinlan, J.R. (2014) *C4. 5: Programs for Machine Learning*. Amsterdam, the Netherlands: Elsevier.

Sagi, O. and Rokach, L. (2018) Ensemble learning: a survey. *Wiley Interdiscip Rev* **8**, e1249.

van Santen, K.L., Edwards, J.R., Webb, A.K., Pollack, L.A., O'Leary, E., Neuhauser, M.M., Srinivasan, A. and Pollock, D.A. (2018) The standardized antimicrobial administration ratio: a new metric for measuring and comparing antibiotic use. *Clin Infect Dis* **67**, 179–185.

Smola, A.J. and Schölkopf, B. (2004) A tutorial on support vector regression. *Stat Comput* **14**, 199–222.

Standiford, H.C., Chan, S., Tripoli, M., Weekes, E. and Forrest, G.N. (2012) Antimicrobial stewardship at a large tertiary care academic medical center: cost analysis before, during, and after a 7-year program. *Infect Control Hosp Epidemiol* **33**, 338–345.

Stenehjem, E., Hersh, A.L., Sheng, X., Jones, P., Buckel, W.R., Lloyd, J.F., Howe, S., Evans, R.S. *et al.* (2016) Antibiotic use in small community hospitals. *Clin Infect Dis* **63**, 1273–1280.

Tang, Y. (2013) Deep learning using linear support vector machines. *arXiv* preprint arXiv:1306.0239.

Yu, K., Rho, J., Morcos, M., Nomura, J., Kaplan, D., Sakamoto, K., Bui, D., Yoo, S. *et al.* (2014) Evaluation of dedicated infectious diseases pharmacists on antimicrobial stewardship teams. *Am J Health Syst Pharm* **71**, 1019–1028.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Summary of the adult DASON dataset and key variables used in the machine learning analysis.

Table S2. List of agents.

Table S3. List of all 455 features.

Table S4. Selected feature set for all-antibacterial agents.

Table S5. Selected feature set for narrow spectrum beta-lactam agents.

Table S6. Selected feature set for CDI agents.

Table S7. Selected feature set for community-onset agents.

Table S8. Selected feature set for hospital-onset agents.

Table S9. Selected feature set for resistant Gram-positive agents.

Table S10. List of potential features common in all six SAAR groups.