


Original Article

Machine learning for the prediction of antimicrobial stewardship intervention in hospitalized patients receiving broad-spectrum agents

Rachel J. Bystritsky MD¹ , Alex Beltran BS², Albert T. Young BA³, Andrew Wong BA³, Xiao Hu PhD² and Sarah B. Doernberg MD, MAS¹

¹Department of Medicine, Infectious Diseases, University of California–San Francisco, San Francisco, California, ²Department of Bioengineering, University of California–San Francisco, San Francisco, California and ³School of Medicine, University of California–San Francisco, San Francisco, California

Abstract

Objective: A significant proportion of inpatient antimicrobial prescriptions are inappropriate. Post-prescription review with feedback has been shown to be an effective means of reducing inappropriate antimicrobial use. However, implementation is resource intensive. Our aim was to evaluate the performance of traditional statistical models and machine-learning models designed to predict which patients receiving broad-spectrum antibiotics require a stewardship intervention.

Methods: We performed a single-center retrospective cohort study of inpatients who received an antimicrobial tracked by the antimicrobial stewardship program. Data were extracted from the electronic medical record and were used to develop logistic regression and boosted-tree models to predict whether antibiotic therapy required stewardship intervention on any given day as compared to the criterion standard of note left by the antimicrobial stewardship team in the patient's chart. We measured the performance of these models using area under the receiver operating characteristic curves (AUROC), and we evaluated it using a hold-out validation cohort.

Results: Both the logistic regression and boosted-tree models demonstrated fair discriminatory power with AUROCs of 0.73 (95% confidence interval [CI], 0.69–0.77) and 0.75 (95% CI, 0.72–0.79), respectively ($P = .07$). Both models demonstrated good calibration. The number of patients that would need to be reviewed to identify 1 patient who required stewardship intervention was high for both models (41.7–45.5 for models tuned to a sensitivity of 85%).

Conclusions: Complex models can be developed to predict which patients require a stewardship intervention. However, further work is required to develop models with adequate discriminatory power to be applicable to real-world antimicrobial stewardship practice.

(Received 25 October 2019; accepted 2 March 2020; electronically published 18 June 2020)

Antimicrobial resistance is a growing problem in the care of hospitalized patients, and it is driven by the overuse of antimicrobials. Prior studies have shown that ~30% of antibiotics prescribed in the inpatient setting are inappropriate—either unsuitable or unnecessary.^{1,2} Antimicrobial stewardship programs aim to improve the appropriate use of antimicrobial agents by promoting the selection of optimal antibiotic regimens.³ Postprescription review with feedback (PPRF), or real-time review of antibiotic prescriptions with feedback to prescribers, is a cornerstone of many antimicrobial stewardship programs. PPRF strategies have been shown to reduce inappropriate antibiotic use, resulting in decreased antibiotic resistance and improved clinical outcomes.^{3–7} A major drawback of PPRF is that it is labor intensive and requires a significant time commitment on the part of experienced antimicrobial stewardship pharmacists and/or physicians.⁸

Much of the inefficiency is due to the fact that many prescriptions require review to identify targets appropriate for feedback. Thus, new approaches are needed to improve the efficiency of daily PPRF. One approach to improving efficiency and effectiveness of PPRF is through computerized clinical decision support systems.⁹ However, these systems generally require programming priorities for patients to be reviewed (eg, those on dual anaerobic coverage or those with positive blood cultures). More complex factors or interactions of factors may more accurately predict which patients are receiving inappropriate or unnecessary antibiotics.

Machine learning, the ability of a computer to learn without being explicitly programmed,¹⁰ is a complex set of approaches that are being applied in medicine for a wide variety of purposes. Machine-learning methods are increasingly being applied to infectious diseases, including to identify drug-resistance genes in multidrug-resistant tuberculosis¹¹ and to predict recurrent *Clostridium difficile* infection.¹² The goal of this study was to apply traditional statistical modeling and machine-learning methods utilizing information contained within the electronic health record

Author for correspondence: Rachel J. Bystritsky, E-mail: Rachel.Bystritsky@ucsf.edu

Cite this article: Bystritsky RJ, et al. (2020). Machine learning for the prediction of antimicrobial stewardship intervention in hospitalized patients receiving broad-spectrum agents. *Infection Control & Hospital Epidemiology*, 41: 1022–1027, <https://doi.org/10.1017/ice.2020.213>

(EHR) to determine which patients on antibiotics required stewardship intervention. Our ultimate goal was to develop a model that could be employed by antimicrobial stewardship programs (ASPs) to improve efficiency and effectiveness of PPRF.

Methods

This project was approved by the UCSF Institutional Review Board for the Protection of Human Subjects.

Study sample

This study cohort included adult patients hospitalized between December 1, 2015, and August 1, 2017, at the University of California San Francisco Medical Center (UCSF), a 600-bed academic tertiary-care medical center. The eligible population included adults ≥ 18 years of age who received at least 1 antimicrobial from a list of those routinely tracked by the ASP (or “tracked antimicrobial,” listed in Appendix 1 online). Patients could be included multiple times if there were multiple admissions or courses of antimicrobials during the study period. The unit of analysis was patient days. Days on which patients were being seen by the infectious diseases consultation team were excluded (these patients are not reviewed by the ASP). Only weekdays were included in the data sets used for model development because our ASP does not perform reviews on weekends. New antibiotic courses were defined as courses that were started at least 72 hours after the most recent antibiotic dose administration. Multiple antibiotics given concurrently were defined as a single course.

PPRF description and study outcome

At UCSF, PPRF occurs every weekday and includes review of inpatients receiving any of the broad-spectrum antimicrobials being tracked. The ASP team marks patients on a flowsheet as their records are reviewed, and they make notes to other ASP providers and set a time for the next review. Not all charts are reviewed every day. A chart may not have been reviewed on a given day either because it had been reviewed on a prior day and follow-up had been set for a future date or because there was insufficient time available for the ASP to review all charts on that day. When the ASP suggests a change in antimicrobial management, a note is left in the chart in addition to direct feedback. Our primary outcome was whether the ASP left a note on any given patient day, a proxy measure for whether the ASP deemed antimicrobial management to require a stewardship intervention. Antimicrobial management could be deemed to require intervention by the ASP team if the antibiotic choice was felt to be inappropriate (ie, unnecessarily broad spectrum, drug-pathogen mismatch, redundant therapy) or unnecessary.

Data extraction and processing

Our hospital uses an electronic health record (EHR) system consisting of a combination of commercially available software (Epic, Epic Systems, Verona, WI) and locally developed databases for infection control purposes, which are derived from the EHR. We extracted data from both the Epic-based relational database management system (Clarity) and our infection control database.

For each day a patient was administered at least 1 tracked antimicrobial, we extracted >200 potential predictors from the EHR. We split the data into 2 main categories: time-invariant and time-varying variables. Time-invariant variables included patient demographics (eg, gender, age), admitting service, and

statistics on prior admission (eg, number of admissions in the prior 90 days). Time-varying variables extracted for each day included antimicrobial administration, length of stay, procedures (ie, whether an incision and drainage had been performed), and laboratory and vital sign data. Daily laboratory and vital sign data were extracted corresponding to a time range of 9:00 AM the day prior to 9:00 AM the day of interest to replicate the data used by the stewardship team during their review.

The raw data set included highly granular data including multiple vital sign readings per day, which were simplified to include only the maximal or minimal value per day, depending on the variable (ie, maximum temperature, maximum creatinine, minimum absolute neutrophil count). We mapped all categorical data to binary features (eg, if a patient received a particular antimicrobial on a given day, then the binary feature associated with that medication was set to “1,” or “0” if they did not). For many of the continuous variables, we dichotomized the variables into binary features based on well-established reference ranges (eg, fever if maximum temperature was $>38^{\circ}\text{C}$, see Appendix 2 online for all criteria). We extracted data for each day a patient was prescribed a tracked antimicrobial.

Statistical methods

Univariate analyses were performed using χ^2 tests for categorical data and t tests for continuous variables. Data were split randomly into derivation (80%) and validation (20%) data sets by patient (based on medical record number). Values of variables varying over time (eg, vital signs, most lab values, and antibiotics administered) were analyzed on the day of interest. Fixed variables that were considered relevant to an entire course of antibiotics (eg, initial positive urinalysis or cultures, admission to intensive care unit, admitting service, or demographics) were reported by antibiotic course.

For the logistic regression model, we included a final set of 26 predictors based on a priori assumptions and statistical performance ($P < .05$ in univariate analysis). We used a stepwise logistic model and calculated a C statistic (area under the receiver operating characteristic curve (AUROC)) for comparison to the machine-learning models. Significance levels for removal and addition to the model were 0.15 and 0.10, respectively.

For the machine-learning model, all ordinal and categorical predictors were binarized via one-hot encoding i.e. a new binary column was added to represent each possible predictor/value pair. However, feature selection was handled internally by an approximate algorithm based on feature distributions. Each model consisted of an ensemble of shallow decision trees. They were trained using 10-fold cross validation on the derivation set across 160 different parameter combinations. All parameters were hard-coded except the number of trees, which was determined by an early stopping rule. Trees were added to each model until AUROC for predictions on the hold-out sets did not increase for 50 training rounds. The final model (ie, parameter set) was identified by the best average AUROC across all hold-out sets.

A Brier score,¹³ the mean squared prediction error, was calculated for each model. The Brier score measures differences between observed events and predicted probabilities. Scores vary from 0 to 1 with a score of 0 indicating perfect predictive performance.¹⁴ Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated via standard methods for the corner point of each model as well as prespecified thresholds of sensitivity. The number needed to

Table 1. Characteristics of the Study Population

| By Course | Total Antibiotic Courses (n=18,275) | ASP Recommendation Made (n=684) | No Recommendation (n=17,591) | P Value |
|--------------------------------------|----------------------------------------|------------------------------------|---------------------------------|---------|
| Female, no. (%) | 7,922 (43.4) | 288 (42.1) | 7,634 (43.4) | .45 |
| Age, mean y (SD) | 56.8 (± 0.1) | 57.5 (± 0.6) | 56.8 (± 0.1) | .27 |
| Service, no. (%) | | | | |
| Medical | 11,587 (63.4) | 446 (65.2) | 11,141 (63.3) | |
| Surgical | 4,693 (25.7) | 157 (23.0) | 4,536 (25.8) | .23 |
| Solid-organ transplant | 1,995 (10.9) | 81 (11.8) | 1,914 (10.9) | |
| Positive blood culture, no. (%) | 1,373 (7.5) | 130 (19.0) | 1,243 (7.0) | <.05 |
| Positive urinalysis, no. (%) | 1,861 (10.2) | 117 (17.1) | 1,744 (9.9) | <.05 |
| ICU, no. (%) | 4,504 (24.7) | 285 (41.7) | 4,219 (24.0) | <.05 |
| Prior ID consultation | 2,491 (13.6) | 45 (6.7) | 2,446 (13.9) | <.05 |
| By Day of Tracked Antibiotic Therapy | Total Antibiotic Days (n=62,095) | Recommendation (n=684) | No Recommendation (n=61,411) | P Value |
| WBC count, mean (SD) | 10.1 (± 0.1) | 10.2 (± 0.1) | 10.1 (± 0.1) | .87 |
| Maximum temperature, mean | 37.3 (± 0.1) | 37.4 (± 0.1) | 37.3 (± 0.1) | .07 |
| Daily no. of antibiotics, mean (SD) | 1.5 (± 0.1) | 1.9 (± 0.1) | 1.5 (± 0.1) | <.05 |
| Antibiotic days, no. (%) | | | | |
| Meropenem | 5,611 (9.0) | 152 (22.2) | 5,459 (8.9) | <.05 |
| Piperacillin-Tazobactam | 18,334 (29.5) | 303 (44.3) | 18,031 (29.4) | <.05 |
| Cefepime | 7,184 (11.6) | 76 (11.1) | 7,108 (11.6) | .75 |
| Vancomycin | 27,106 (43.7) | 414 (60.5) | 26,692 (43.5) | <.05 |

Note. ASP, antimicrobial stewardship program; SD, standard deviation; ID, infectious diseases; WBC, white blood cell; ICU, intensive care unit.

Table 2. Predictors Associated with a Recommendation Being Made by the Stewardship Team

| Predictor | Unadjusted OR (95% CI) | Adjusted OR (95% CI) |
|------------------------------------------|---------------------------|-------------------------|
| Abnormal creatinine | 1.70 (1.39–2.06) | 1.21 (0.99–1.47) |
| Culture + for resistant GNR ^a | 2.67 (1.52–4.67) | 2.37 (1.34–4.20) |
| Positive blood culture | 2.89 (2.12–3.95) | 2.50 (1.82–3.44) |
| Fever (>38.0°C) | 0.67 (0.54–0.82) | 0.61 (0.49–0.76) |
| Prior ID consult | 0.42 (0.30–0.57) | 0.39 (0.28–0.53) |
| Tachycardia ^b | 1.61 (1.35–1.92) | 1.31 (1.09–1.58) |
| Meropenem | 2.90 (2.37–3.55) | 3.54 (2.81–4.45) |
| Piperacillin-Tazobactam | 2.00 (1.70–2.36) | 2.04 (1.70–2.45) |
| Vancomycin | 2.16 (1.82–2.56) | 1.88 (1.57–2.25) |
| Ampicillin | 1.93 (0.99–3.77) | 1.73 (0.87–3.41) |
| Clindamycin | 2.94 (1.74–4.96) | 2.16 (1.27–3.68) |
| Metronidazole | 1.54 (1.05–2.25) | 1.64 (1.12–2.42) |

Note. OR, odds ratio; CI, confidence interval; GNR, gram-negative rod; ID, infectious diseases.

^aResistant GNR was defined as ESBL or Amp-C-producing gram-negative rod.

^bTachycardia was defined as a heart rate >110 beats per minute.

review (NNR) is the number of patient days needed to review to identify 1 patient requiring stewardship intervention and was calculated as 1/PPV. The logistic regression and boosted-tree models were compared using the McNemar test¹⁵ using the threshold determined by the corner point of the area's receiver operating curve on the test set.

We used STATA version 15.1 software (StataCorp, College Station, TX) for statistical analyses. For the machine-learning analysis, XGBoost 0.72 was used with Python version 3.6.5 software (Python Software Foundation, Wilmington, DE). Sensitivity, specificity, and positive and negative predictive values were determined using Python, with output from STATA for the logistic regression model.

Results

Between December 15, 2015, and August 1, 2017, 9,651 adult hospitalized patients without infectious diseases consultations received at least one “tracked” antimicrobial for a total of 18,275 antimicrobial courses (mean duration, 3.4 days) over 62,095 patient days. The ASP left recommendations during 684 antimicrobial courses (3.7%). Each patient received 1.9 courses of antibiotics on average during the study period. The most frequently administered antimicrobials in our cohort were vancomycin (43.7%) and piperacillin-tazobactam (20.5%) by patient days. ASP recommendations were made, on average, on day 4.9 of therapy (SD ± 4.3). Baseline demographics of age and gender were similar between patients who did and did not receive an ASP note (Table 1). By univariate analysis, variables that were associated with whether an intervention was made included positive cultures, positive urinalysis, admission to the intensive care unit, lack of prior infectious diseases consultation, international normalized ratio (INR), and type of antibiotic (Table 2).

A list of 25 predictors (see Appendix 1 and Supplementary Materials online) that had been selected a priori and identified as significant in the unadjusted analysis were included in a stepwise

Table 3. Performance Characteristics of Logistic Regression and Boosted-Tree Models^a

| Model | AUC | Brier Score | Sensitivity | Specificity | PPV | NPV | NNR |
|---------------------|------|-------------|-------------|-------------|------|------|------|
| Logistic regression | 0.73 | 0.15 | 0.57 | 0.71 | 0.03 | 0.99 | 33.3 |
| | | | 0.85 | 0.45 | 0.02 | 0.99 | 41.7 |
| | | | 0.90 | 0.40 | 0.02 | 0.99 | 43.5 |
| | | | 0.95 | 0.33 | 0.02 | 0.99 | 45.5 |
| Boosted trees | 0.75 | 0.19 | 0.68 | 0.72 | 0.04 | 0.99 | 27.8 |
| | | | 0.85 | 0.42 | 0.02 | 0.99 | 45.5 |
| | | | 0.90 | 0.34 | 0.02 | 0.99 | 47.6 |
| | | | 0.95 | 0.27 | 0.02 | 0.99 | 50.0 |

Note. AUC, area under the curve; PPV, positive predictive value; NPV, negative predictive value; NNR, number needed to review.

^aThe first line for each model shows the performance characteristics at the corner point with performance characteristics at prespecified sensitivities below.

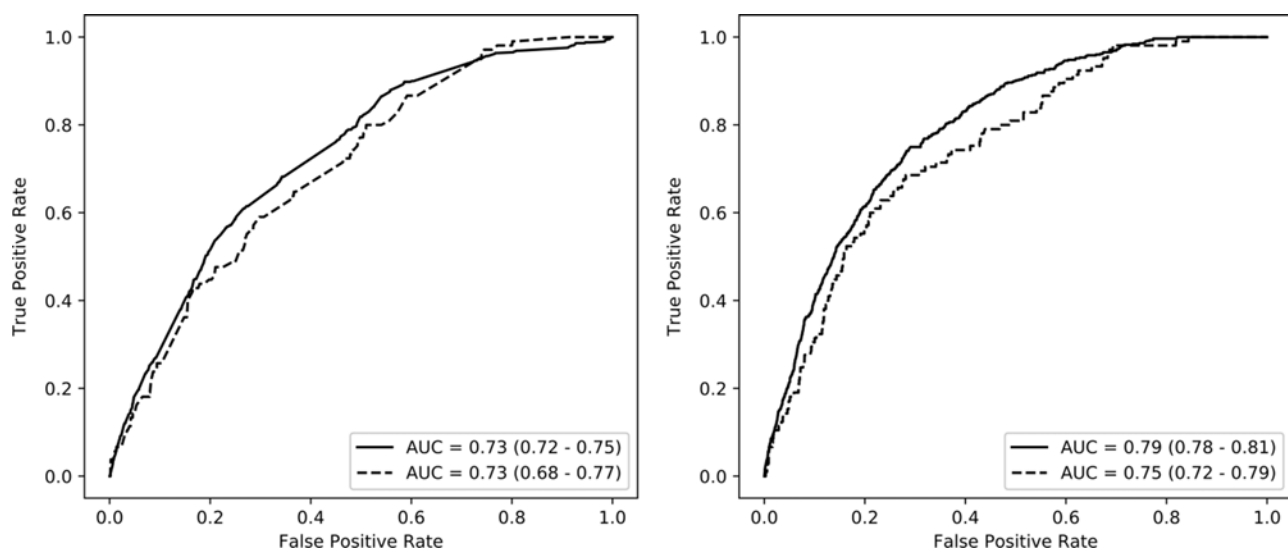


Fig. 1. Area under the receiver operating characteristic curve (AUC) for logistic regression (dotted line) and boosted trees (solid line) for the derivation (left) and validation sets (right).

logistic regression model using the development data set. Using the prespecified criteria for addition or removal of predictor variables, 12 variables remained in the final model (Table 2).

All features extracted from the EHR were used to construct the boosted-trees model using an unbiased approach. The final boosted-tree model contained 232 trees, with a depth of trees of 1.

Both models were then applied to the validation data set, and C-statistics (AUROCs) of 0.73 (95% CI, 0.69–0.77) and 0.75 (95% CI, 0.72–0.79) were obtained for the logistic regression and boosted-tree models ($P = .07$) (Fig. 1A and 1B). Decision thresholds were selected at the corner points for each model as well as at prespecified sensitivities of 0.85, 0.90, and 0.95, which resulted in diagnostic performance shown in Table 3. We tuned the models to higher sensitivity thresholds under the assumption that a stewardship program would want to capture a comprehensive list of patients on inappropriate antibiotics. The number needed to review ranged from 33.3 to 45.5 for the logistical model and from 27.8 to 50 for the boosted-trees model. For comparison, based on our data, the number needed to review using our current process would be 99 based on recommendations being made on 1.1% of cases.

The types of variables that were included in the logistic regression model compared to the boosted-tree model are shown in Table 4 (for full set of variables used in both models see Appendix 3 in the Supplementary Materials online).

Discussion

In this cohort, both the machine-learning (boosted trees) and logistic regression models demonstrated modest performance for predicting which patients required stewardship intervention on a given day. Both methods exhibited similar performance, with AUROCs favoring the boosted-tree model, although this difference was not statistically significant. Both models had high NPVs, and therefore performed well in regard to identifying cases that do not need further manual review, though the paucity of outcomes limits the interpretation of the NPVs. However, both models showed low PPV and high NNR to identify patients requiring stewardship intervention on a given day. The high NNR may diminish the utility of making the PPRF process more efficient, although this may be mitigated in part by the fact that a larger number of patients can be screened with the use of an automated algorithm.

Table 4. Predictors Used in Each Model

| Predictor | Logistic | Boosted Trees |
|--------------------------------------------------------------------------|----------|---------------|
| Arterial blood gas pH \leq 7.2 | X | |
| Antibiotic | X | X |
| Creatinine | X | X |
| Currently in intensive care unit | X | |
| Fever (>38°C) | X | |
| High C-reactive protein | X | |
| Systolic blood pressure \leq 90 mmHg | X | |
| International normalized ratio | X | X |
| Length of stay | X | X |
| White blood cell count >10 | X | |
| No. of concurrent antibiotics | X | X |
| Positive blood culture | X | X |
| Positive urinalysis ^a | X | X |
| Currently on pressors | X | |
| Prior infectious diseases consult | X | |
| Resistant GNR isolated ^b | X | |
| Heart rate >110 beats per minute | X | |
| Arterial blood gas pH | | X |
| Arterial blood pressure | | X |
| Alanine aminotransferase level | | X |
| Blood glucose | | X |
| Currently admitted to intensive care unit | | X |
| Day of antibiotic course | | X |
| Days since last <i>C. difficile</i> infection | | X |
| Days since last infectious diseases note | | X |
| Days since last pulmonary note | | X |
| Department (location in hospital) | | X |
| Heart rate | | X |
| Platelet count | | X |
| Noninvasive systolic blood pressure | | X |
| Miscellaneous culture ^c positive for Amp-C-producing organism | | X |
| Sputum positive for gram-positive rods | | X |
| Sputum positive for nontuberculous mycobacteria | | X |
| Service | | X |
| Maximum temperature | | X |
| Minimum temperature | | X |
| White blood cell count | | X |

Note. GNR, gram-negative rod; ESBL, extended-spectrum beta-lactamase.

^aPositive urinalysis defined as >10 white blood cells per high powered field.

^bResistant GNR was defined as ESBL or Amp-C producing gram-negative rod.

^cMiscellaneous cultures is defined as sterile or nonsterile site cultures other than blood, urine, respiratory or cerebrospinal fluid culture.

This study had several limitations that may have compromised the performance of the predictive models. First, the number of outcomes was sparse. Optimal tuning of machine-learning models requires a large number of outcomes for training and validation.

Although our data set included >60,000 patient days, this is a relatively small data set by machine-learning standards. Additionally, not all patient days in the data set were reviewed by the ASP team, even during the business week, due to limited resources, but they were still included in the model. These omissions may have compromised the ability of the models to discriminate patients on requiring stewardship intervention because some cases may not have been intervened upon simply because resources were insufficient to review them. Because our stewardship team schedules dates for follow-up review of patients, a significant subset of patients not reviewed on a given weekday were not reviewed intentionally because the antibiotic course was deemed appropriate for that day a priori. For example, a patient on appropriate definitive therapy would be scheduled for follow-up review on the day therapy is scheduled to be completed to ensure appropriate duration, but the days prior to this would not have been reviewed further. Thus, including in the model the days not reviewed provided additional information about appropriateness of antibiotics despite the important drawback that some of the days lacking review were not due to presumed appropriateness but instead were due to limited ASP resources for review that day. Whether an intervention was made also depends on additional factors such as perceived probability of acceptance of recommendations by the prescribing team and the availability of data on which a decision to intervene can be based. Pragmatically speaking, these considerations may be important when choosing whether a stewardship intervention should be made but may not be fully recognized by the model. We must be careful when interpreting the output of the model to take into consideration the circumstances under which it was trained. During the study period, most recommendations were made by a small number of individuals staffing the ASP team (ie, 1–2), which further limits the generalizability of the model.

One limitation of machine learning in general is the “black box” nature of the models. The models provide probabilistic outputs but offer no explanation of how these predictions are made.¹⁶ Explainable models are being developed that incorporate hypothesis-driven experiments with the limitation that they require large data sets (tens of thousands to millions of subjects).¹⁷

The advantage of machine learning over other methods includes the ability to handle a large number of variables and samples. In this study, we used a preprocessed data set, which is labor intensive and limits implementation feasibility. We also simplified the data set to include only the maximum or minimum values per day for each variable, which allowed for logistic regression modeling but may have eliminated additional data that would have been predictive in the machine-learning model. In addition, having a well-characterized data set of patients on antibiotics and interventions made will be critical and may necessitate comprehensive point-prevalence style review of patients on antibiotics to achieve. This study used a basic boosted-trees machine-learning model. Other models exist, such as learning to rank models,¹⁶ which model the relative ordering of inputs according to their true labels or scores and may be better suited to the practice of antimicrobial stewardship by allowing prioritization of patients by likelihood of antimicrobial inappropriateness. Future directions include building a model that can use highly granular raw data output from the EHR that would allow for availability of larger sets of data for training as well as real-time updating of the model after implementation, which could then be integrated into the EHR to allow for identification of high-yield targets for manual review.

PPRF has the potential to improve antimicrobial use and to improve outcomes, but optimal implementation remains hampered by the labor intensiveness of manual review. This study serves as an example of an approach to leveraging statistical and machine-learning models to identify patients who may require stewardship intervention and can be targeted for manual review. Our models suggest that machine learning may be able to outperform statistical models when dealing with complex data sets using large numbers of interacting variables. Further research is needed to develop advanced models to predict need for stewardship intervention in hospitalized patients and optimize the efficiency of post-prescription review with feedback.

Acknowledgments. None.

Financial support. This study was supported by an NIH Research Training Grant (grant no. T32 2T32AI007641) to R.B.

Conflicts of interest. All authors report no conflicts of interest relevant to this article.

Supplementary material. To view supplementary material for this article, please visit <https://doi.org/10.1017/ice.2020.213>

References

- Hecker MT, Aron DC, Patel NP, Lehmann MK, Donskey CJ. Unnecessary use of antimicrobials in hospitalized patients: current patterns of misuse with an emphasis on the antianaerobic spectrum of activity. *Arch Intern Med* 2003;163:972–978.
- Cosgrove SE, Seo SK, Bolon MK, *et al*; CDC Prevention Epicenter Program. Evaluation of postprescription review and feedback as a method of promoting rational antimicrobial use: a multicenter intervention. *Infect Control Hosp Epidemiol* 2012;33:374–380.
- Barlam TF, Cosgrove SE, Abbo LM, *et al*. Implementing an antibiotic stewardship program: guidelines by the Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America. *Clin Infect Dis* 2016;62(10):e51–e77.
- Honda H, Murakami S, Tagashira Y, *et al*. Efficacy of a postprescription review of broad-spectrum antimicrobial agents with feedback: a 4-year experience of antimicrobial stewardship at a tertiary care center. *Open Forum Infect Dis* 2018;5(12):ofy314.
- Tamma PD, Avdic E, Keenan JF, *et al*. What is the more effective antibiotic stewardship intervention: preprescription authorization or postprescription review with feedback? *Clin Infect Dis* 2017;64:537–543.
- Lesprit P, *et al*. Postprescription review improves in-hospital antibiotic use: a multicenter randomized controlled trial. *Clin Microbiol Infect* 2015;21(2):180.e1–e7. Epub 2014 Oct 14.
- Davey P, *et al*. Interventions to improve antibiotic prescribing practices for hospital inpatients. *Cochrane Database Syst Rev* 2017;2:CD003543.
- Doernberg SB, Abbo LM, Burdette SD, *et al*. Essential resources and strategies for antibiotic stewardship programs in the acute care setting. *Clin Infect Dis* 2018;67:1168–1174.
- Catho G, De Kraker M, Waldispul Suter B, *et al*. Study protocol for a multicentre, cluster randomized, superiority trial evaluating the impact of computerized decision support, audit and feedback on antibiotic use: the COMPuterized Antibiotic Stewardship Study (COMPASS). *BMJ Open* 2018;8(6):e022666.
- Deo RC. Machine Learning in Medicine. *Circulation* 2015;132:1920–1930.
- Huang H, Ding N, Yang T, *et al*. Cross-sectional whole-genome sequencing and epidemiological study of multidrug-resistant *Mycobacterium tuberculosis* in China. *Clin Infect Dis* 2018; ciy883.
- Escobar GJ, Baker JM, Kipnis P, *et al*. Prediction of recurrent *Clostridium difficile* infection using comprehensive electronic medical records in an integrated healthcare delivery system. *Infect Control Hosp Epidemiol* 2017;38:1196–1203.
- Brier GW. Verification of forecasts expressed in terms of probability. *Mon Weather Rev* 1950;78:1e3.
- Cohen ME, Ko CY, Bilimoria KY, *et al*. Optimizing ACS NSQIP modeling for evaluation of surgical quality and risk: patient risk adjustment, procedure mix adjustment, shrinkage adjustment, and surgical focus. *J Am Coll Surg* 2013;217:336–346.
- Deitterich TG. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput* 1998;10:1895–1923.
- Watson DS, Krutzinna J, Griffiths CE, McInnes IB, Barnes MR, Floridi L. Clinical applications of machine learning algorithms: beyond the black box. *BMJ* 2019;364:I886.
- Vu MT, Adali T, Ba D, *et al*. A shared vision for machine learning in neuroscience. *J Neurosci* 2018;38:1601–1607.
- Deng J, Yuan Q, Mamitsuka H, Zhu S. Drug E-Rank: Predicting Drug-Target Interactions by Learning to Rank. *Methods Mol Biol* 2018;1807:195–202.