

Original Article

A machine learning model to predict optimal antibiotic use in hospital medicine patients

Nathan Radakovich^{1,2} , Priya Prasad¹, Daniel Escobar², Ripal Jariwala² , Allison Bond^{1,2}, Sandra Oreper¹, Junghwan Lee³, Xiao Hu⁴  and Sarah Doernberg² 

¹Department of Medicine, University of California, San Francisco, CA, USA, ²Division of Infectious Diseases, Department of Medicine, University of California, San Francisco, CA, USA, ³Georgia Institute of Technology, Atlanta, GA, USA and ⁴Emory University, Center for Data Science, Atlanta, GA, USA

Abstract

Objective: Unnecessary and suboptimal antibiotic use causes adverse outcomes at both the level of individuals and health systems. Prospective audit and feedback, a core aspect of antibiotic stewardship program (ASP) efforts, reduces such use, but is inefficient in the absence of a pre-screening process. To address this, we used a machine learning approach to stratify antibiotic orders based on the likelihood of benefiting from ASP review, and to identify the factors most influential in the model's predictions.

Design: Machine learning model developed using expert-labeled point-prevalence data.

Setting: Single-center adult academic hospital.

Participants: Hospitalized patients 18 years or older on internal medicine services between May 2021 and August 2022.

Methods: Infectious disease experts assessed antibiotic orders for necessity and optimal use to create labels which were used as ground truth to train a machine learning model that uses automatically queried data from the electronic health record including vital signs, laboratory values, microbiological data, and medication administration records.

Results: Our model achieved an area under the receiver operating characteristic curve of 0.89 for antibiotic necessity and 0.80 for optimal use. Model predictions were driven largely by markers of clinical instability, inflammation, and infection. Simpler clinical indices of infection had no predictive ability.

Conclusions: We describe a model that predicts antibiotic necessity and optimal use using routinely available data that can be automatically aggregated from medical records. This makes it a promising option for early identification and intervention on orders most likely to benefit from ASP review.

(Received 4 May 2025; accepted 7 August 2025)

Introduction

As many as 50% of hospitalized patients receive antibiotics during the course of their hospitalization.¹ Despite antibiotics' life-saving role in treating bacterial infections, approximately one fifth of hospitalized patients who receive them experience adverse effects, including hypersensitivity reactions, direct toxicity, and *Clostridioides difficile* infection.^{2,3} On a systems level, broad-spectrum antibiotic use drives increased rates of multi-drug resistance, contributing to worldwide morbidity and mortality.⁴ In hospital settings, where an estimated 30% of antibiotic use is unnecessary, some of this harm may be avoidable.⁵ Antibiotic stewardship programs (ASPs) seek to balance antibiotics' utility with the harms stemming from unnecessary or suboptimal use

through restriction of certain antibiotics, prospective audit and feedback (PAF), and co-implementation with infection control measures.⁶ ASP interventions have been shown to decrease rates of both multidrug-resistant organisms and incidence of *C. difficile* infection, and have an important role to play in reducing the adverse consequences of suboptimal antibiotic use.⁷

Prospective audit and feedback (PAF), a backbone of ASP efforts, optimizes antibiotic use via expert review of medical records for appropriateness and via proactive outreach to prescribing teams.⁸ This includes identifying cases where antibiotics are not indicated, or where modifying them might reduce toxicity or improve efficacy. PAF reduces the use of broad-spectrum antibiotic use and might reduce rates of antibiotic resistance without negatively affecting outcomes.^{2,9,10} Despite PAF's benefits, its labor-intensive nature limits higher-volume implementation.¹¹ An automated method to stratify orders based on their need for review might refocus ASP audits toward orders most likely to benefit from interventions.

Corresponding author: Sarah Doernberg; Email: sarah.doernberg@ucsf.edu

Cite this article: Radakovich N, Prasad P, Escobar D, *et al.* A machine learning model to predict optimal antibiotic use in hospital medicine patients. *Antimicrob Steward Healthc Epidemiol* 2025. doi: 10.1017/ash.2025.10142

© The Author(s), 2025. Published by Cambridge University Press on behalf of The Society for Healthcare Epidemiology of America. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

To this end, we developed a machine learning model to flag suboptimal or non-indicated antibiotic orders for ASP audit and feedback. The model uses data that can be automatically queried from the electronic health record (EHR), and makes clinician-interpretable predictions about indications and optimal antibiotic use.

Methods

Study population

Data collection and analysis for this study were approved by the University of California, San Francisco (UCSF) internal review board. Our study cohort included adult patients receiving at least one antibiotic who were treated on the hospital medicine service between May 2021 and August 2022 at UCSF Parnassus Medical Center, an adult tertiary academic hospital. Separate medical services (intensive care, cardiology and malignant hematology) were excluded, as were surgical services. Three infectious diseases experts (two faculty physicians and one infectious diseases specialty pharmacist) completed quarterly periodic point prevalence surveys of all patients on a given medicine team receiving antibiotics on the day of the survey. A single clinician assessed orders for (1) whether antibiotics were indicated and (2) whether the antibiotic selected was optimal. To ensure uniformity, all reviews were performed using the National Antimicrobial Prescribing Survey (NAPS) tool, a standardized, clinically validated questionnaire.¹² National Antimicrobial Prescribing Survey (NAPS) assesses antimicrobial agent indication and optimal use including dose, route of administration, clinical syndrome, patient allergies/comorbidities, and microbiological data, if applicable (a complete version of the tool is in the supplemental appendix).

Reviewers at our institution assessed orders based on patient notes, vital signs, laboratory values, imaging, medicine administration record (MAR), and microbiological data. Data were collected and processed at the level of individual antibiotic orders. Two concurrent orders for the same patient were considered separately—due to the intermittent point prevalence design of the reviews, repeat orders for the same patient were assessed at intervals of 3 months or greater, and were therefore considered independently of one another. Patients receiving infectious disease consultation were excluded from review.

Data extraction, labeling, and descriptive statistics

Data were extracted from the medical record via a reporting database, Clarity, attached to our institution's EHR (Epic Systems Corporation, Verona, Wisconsin, United States). To match the model's intended use case (ie, automated screening of existing/new antibiotic orders), data were censored after 11:59 PM on the day prior to review. Subsequent data processing and machine learning model development were done using the Python programming language (Python 3.9.2, Python Software Foundation, Wilmington, Delaware, United States); libraries used are detailed in the supplement. Student's *T*-test or the Kruskal–Wallis test were used to compare continuous variables, the chi-squared test or Pearson's exact test for categorical ones.

Model generation

We selected the initial pool of training features based on their relevance to infection. Candidate features included: markers of hemodynamic and respiratory status (eg, vital signs, pulse

oximetry); inflammatory markers (white blood cell count (WBC), C-reactive protein (CRP), eg.); and markers of end organ function (creatinine, bilirubin, eg.). A complete medication administration record (MAR), including antibiotics, immunosuppressants, and microbiological data were also included. Medication administration record (MAR) data contained date and time of administration for antibiotics, allowing duration of treatment to be determined for each. Microbiological data included species, presence of extended-spectrum beta lactamase producing enterobacterales (ESBL; designated by presence of ceftriaxone resistance), identification of methicillin-resistant *Staphylococcus aureus* isolates (MRSA; designated by either phenotypic testing or rapid molecular diagnostics), and for negative cultures, the quantity of time with no growth detected.

We selected a machine learning approach to develop our model based on machine learning models' ability to capture hierarchical, nonlinear relationships without explicit *a priori* instruction about how to relate variables to one another.¹³ The model chosen, XGBoost, makes predictions by iteratively creating decision trees to generate and refine predictions.¹⁴ The model used a 10-fold cross-validation scheme with non-repeating partitions of 90% train and 10% test data to prevent overfitting. The performance metrics reported herein are aggregated from the 10 different partitions of test data, which were not used to directly train models.

Area under the receiver operating characteristic curve (AUROC) was used to adjudicate predictive ability—AUROC measures the ability to distinguish between different types of objects, with an AUROC of 1.0 representing a perfect predictor and an AUROC of 0.5 representing a predictor that performs no better than random chance (an AUROC 0.5–0.6 is generally considered poor, 0.6–0.7 weak, 0.7–0.8 fair, 0.8–0.9 good, and 0.9–1.0 excellent, although these are not formal definitions and performance needs to be considered in context). SHAP (SHapley Additive exPlanations), an approach to machine learning model interpretation that uses game theory to identify variables' relative influence on model predictions, was used both to iteratively select variables during model development and to interpret the final model's predictions.

After model generation and evaluation we assessed the performance of the systemic inflammatory response syndrome (SIRS) criteria, a commonly used diagnostic tool for infections in hospitalized patients, to query whether our model's performance exceeded that of simpler clinical decision tools; the quick sepsis-related organ failure assessment score was not used due to the quantity of missing data for Glasgow Coma Scale measurements.^{15–17}

Study design, model generation and data reporting were executed in alignment with the TRIPOD guidelines for clinical prediction models.¹⁸ The supplemental appendix describes additional details about variable selection, feature engineering and model development.

Results

425 antibiotic orders were reviewed and annotated between May 2021 and August 2022. 176 (41%) of patients were female, the average age was 64 years (interquartile range (IQR): 53–74), and the median duration of hospitalization prior to review was 11 days (IQR: 7–15). Of orders reviewed, 317 (75%) and 236 (56%) were assessed as indicated and optimal, respectively. Demographic, vital sign, laboratory and antibiotic data are summarized in Table 1. No statistically significant differences were seen between antibiotics

Table 1. Cohort characteristics

	Optimal		Indicated	
	Yes	No	Yes	No
Female	96 (22)	80 (18)	136 (32)	37 (8)
BSHO	100 (23)	71 (16)	135 (32)	35 (8)
BSCA	159 (37)	141 (33)	218 (52)	74 (17)
Gram positive	172 (40)	142 (33)	239 (57)	65 (15)
MRSA	172 (40)	142 (33)	239 (57)	65 (15)
ESBL	93 (21)	59 (13)	124 (29)	26 (6)
<i>Pseudomonas</i>	117 (27)	96 (22)	160 (38)	50 (12)
Positive blood culture	58 (13)	34 (8)	75 (18)	11 (2)
Positive urine culture	74 (17)	72 (16)	112 (26)	30 (7)
Age	65 (53–74)	64 (54–75)	64 (55–75)	63 (52–73)
Days of antibiotics	1 (1–5)	1 (1–4)	1 (1–4)	1 (0–3)
ASI at review	9 (5–13)	10 (6–14)	9 (5–13)	9 (5–13)
Median SBP, 24h	121 (110–135)	119 (106–132)	121 (107–133)	120 (108–137)
Median HR, 24h	84 (71–96)	89 (77–99)	84 (73–98)	90 (80–99)
Median RR, 24h	18 (16–18)	18 (16–18)	18 (16–18)	18 (16–18)
Median temperature, 24h	98 (97–98)	98 (97–98)	98 (97–98)	98 (97–98)
Median WBC, 24h	8 (6–13)	9 (7–14)	8 (6–13)	9 (6–15)
Median platelet count, 24h	221 (143–338)	240 (156–324)	228 (142–338)	231 (158–290)
Median lactate, 24h	1 (1–2)	1 (1–1)	1 (1–1)	1 (1–2)

Parentheses show column percentages for categorical variables and interquartile ranges for continuous ones. BSHO: broad-spectrum hospital onset, HSCA: broad-spectrum community acquired, MRSA: methicillin-resistant *Staphylococcus aureus*, ESBL: extended-spectrum beta lactamase, ASI: antibiotic spectrum index, SBP: systolic blood pressure, HR: heart rate (beats per minute), RR: respiratory rate (breaths per minute), WBC (1 000 cells per microliter).

received, vital sign values, laboratory values, or demographic characteristics between patients who met the “indicated” versus “optimal” criteria (unadjusted $P > 0.05$ for all comparisons).

Stepwise feature selection identified optimum performance at 18 features for both the indicated and optimal criteria, yielding respective AUROCs (95% confidence interval) of 0.89 (0.87–0.91) and 0.80 (0.77–0.83) (Figure 1). To explore which types of clinical data were most informative for making predictions, we also trained separate models using: vital signs only; laboratory values only; and microbiology/antibiotic administration data only (these were paired to assess how well the model could learn to identify concordance between culture data and spectra of coverage).

For the final models (ie, those using all available data that had undergone feature selection to optimize performance), we also assessed performance on different subsets of antibiotics as defined by the National Healthcare Standardized Antimicrobial Administration Ratios antimicrobial categories: broad-spectrum hospital onset; broad-spectrum community onset; MRSA; ESBL and *Pseudomonas aeruginosa*.¹⁹ In brief, model performance was highest using all available data, and lower when trained individually on vitals, laboratory data, and microbiological/MAR data, in descending order of performance. This held true for both predicting whether antimicrobials were indicated and for whether the optimal antimicrobial was being used. Model performance across spectra of coverage appeared similar. For ease of readability, model performance is detailed in Table 2. Systemic inflammatory response syndrome (SIRS) criteria distributions are shown in Figures S3a and S3b.

Table 2. Model performance

	Indicated		Optimal	
	AUROC	95% CI	AUROC	95% CI
Training data				
All data	0.89	(0.87–0.91)	0.8	(0.77–0.83)
Vital signs	0.87	(0.70–0.98)	0.77	(0.58–0.91)
Laboratory data	0.76	(0.57–0.93)	0.76	(0.540–0.90)
Microbiology/antibiotics	0.74	(0.45–0.98)	0.66	(0.48–0.82)
SIRS criteria	0.42	(0.37–0.49)	0.42	(0.36–0.47)
Spectrum of coverage	AUROC	95% CI	AUROC	95% CI
BSHO	0.86	(0.78–0.92)	0.77	(0.70–0.86)
BSCA	0.87	(0.80–0.90)	0.81	(0.75–0.85)
MRSA	0.87	(0.82–0.92)	0.77	(0.71–0.84)
ESBL	0.89	(0.78–0.97)	0.73	(0.64–0.82)
PsA	0.82	(0.73–0.88)	0.78	(0.70–0.83)

Area under the receiver operating characteristic curve (AUROC) with 95% confidence intervals (CI) are listed, with performance stratified by the profile of data used to train the model (all available data vs subsets) as well as model performance by spectrum of coverage (broad-spectrum hospital onset (BSHO), broad-spectrum community onset (BSCO), methicillin-resistant *Staphylococcus aureus* (MRSA), extended-spectrum beta-lactamase producing organisms (ESBL) and *Pseudomonas aeruginosa* (PsA). The systemic inflammatory response (SIRS) criteria are listed as a comparator.

The top 5 features identified by SHAP for each outcome were, in descending order: diastolic blood pressure range over the most recent 8 hours, median diastolic blood pressure at –48 hours,

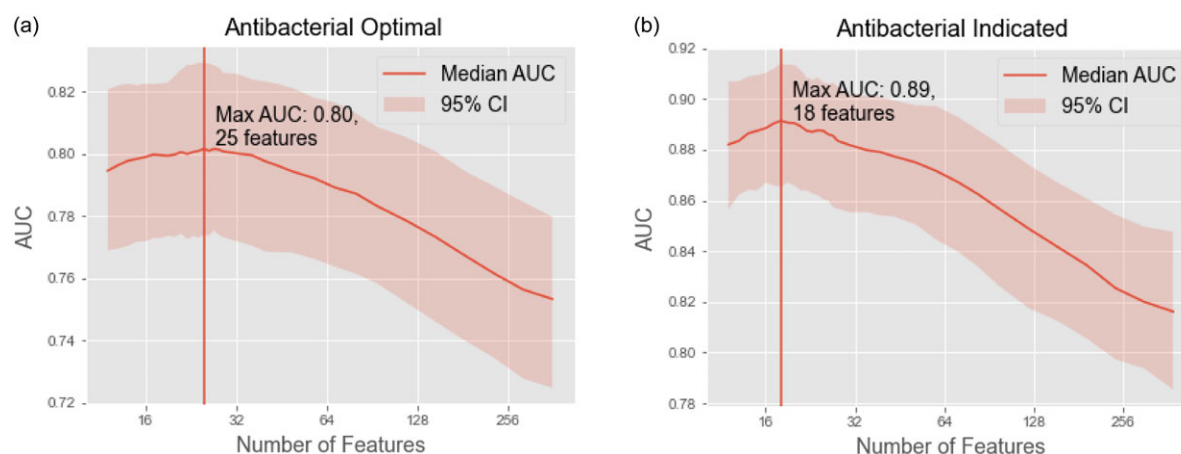


Figure 1. Feature selection and optimization. 1a and 1b show AUROC trend as non-informative features are iteratively removed; trend lines represent a rolling average of AUROC, with 95% confidence interval (CI) depicted by shading.

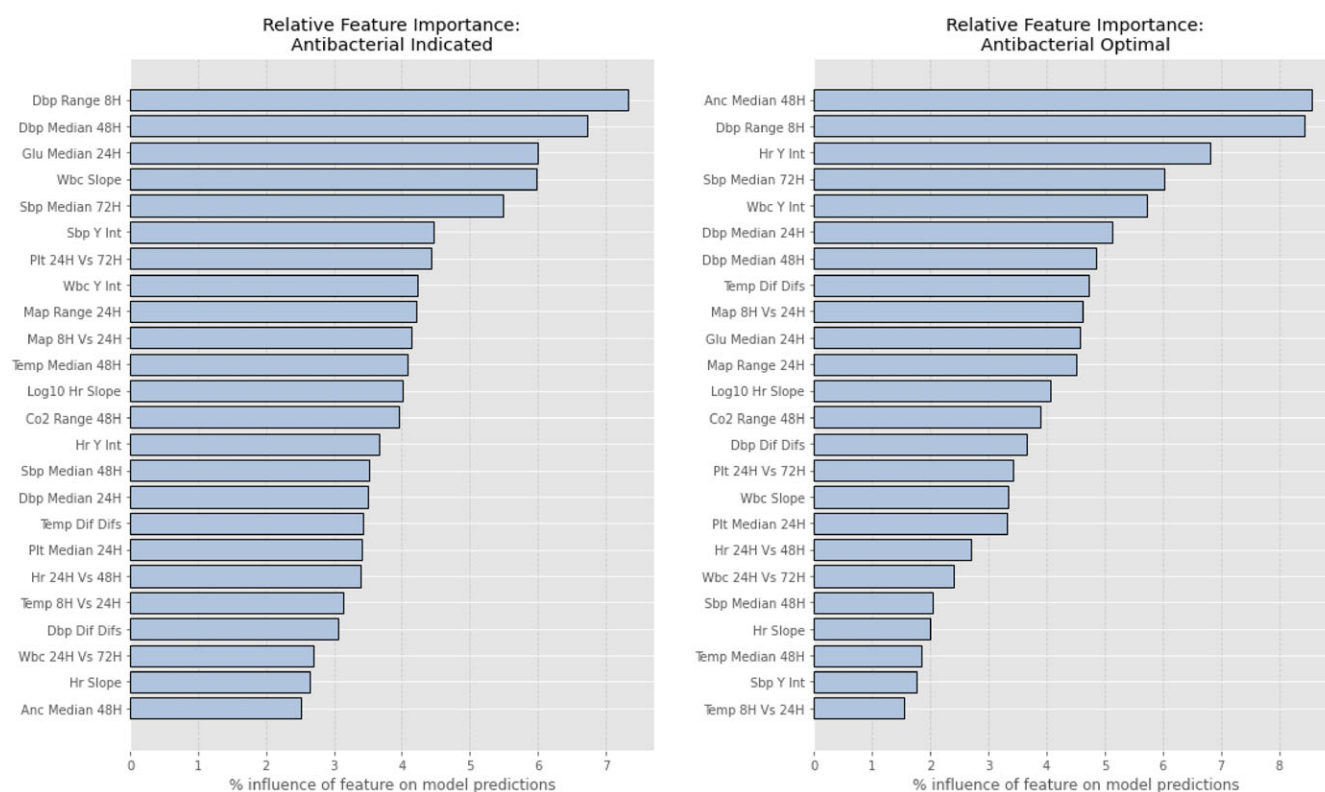


Figure 2. global feature performance. The x axis depicts the relative contribution of a feature as determined by SHAP values, normalized to a total of 100%.

median glucose at –24 hours, estimated WBC slope, and median systolic blood pressure at –72 hours for the “indicated” outcome; and median absolute neutrophil count at –48 hours, diastolic blood pressure range over the most recent 8 hours, estimated heart rate intercept, median systolic blood pressure at –72 hours, and estimated WBC intercept for the “optimal” outcome (Figure 2).

To identify clinical correlates for those predictions, we examined the relationship between feature values (ie, clinical values) and SHAP values (ie, the values used to summarize the

relative importance of those features). Illustrative results from the “optimal” model are shown in Figure 3. Figures 3a and 3b depict diastolic blood pressure range in the 8 hours prior to data collection and the rate of change in temperature. Both examples show a positive relationship between markers of clinical instability and a positive prediction (ie, a order predicted as optimal). Figure 3c shows correspondence between both elevated and lower platelet levels and positive predictions. Figure 3d shows correspondence between a rise in white blood cell count in the

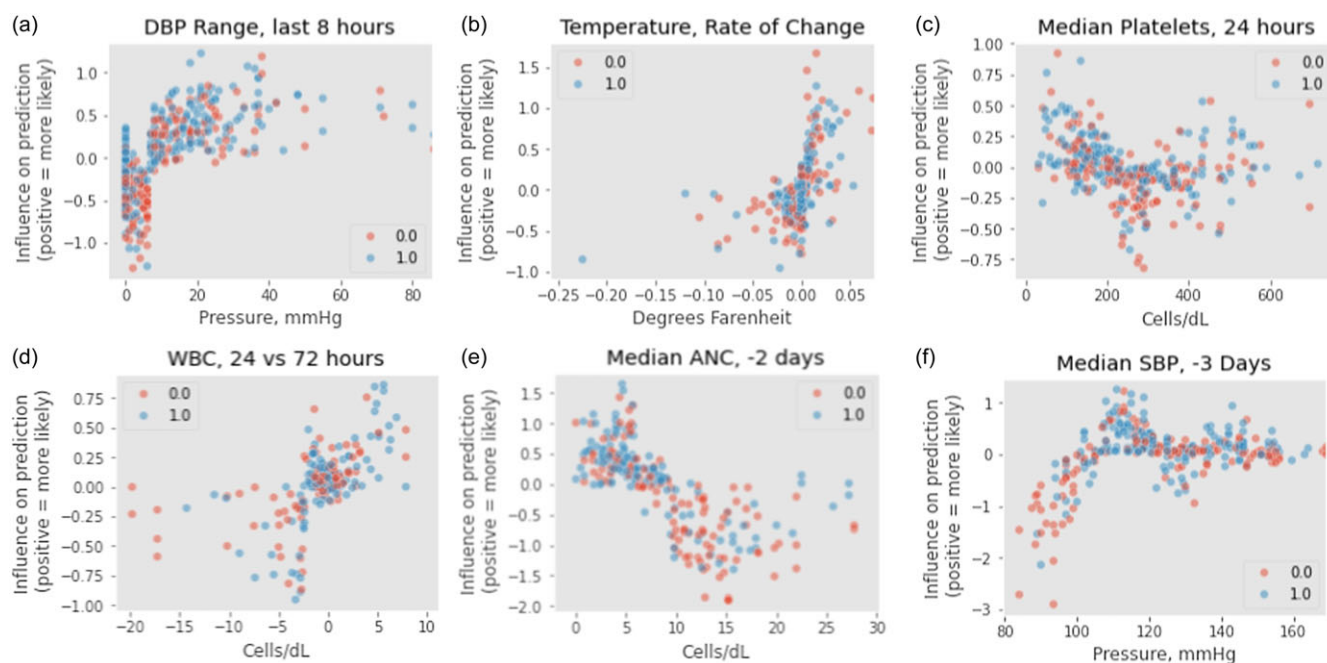


Figure 3. illustrative feature-prediction relationships for optimal use. Each dot represents a single prediction by the model (red dots represent orders in need of review based on ASP staff review; blue dots those which do not). The y axis represents a given value's impact on model predictions as determined by SHAP. Negative values predict suboptimal/non-indicated orders in need for review, and vice versa.

72 hours preceding review and positive predictions. Figures 3e and 3f, interestingly, show that elevated absolute neutrophil count at -48 hours and decreased systolic blood pressure at -72 hours correspond to negative predictions. A complete set of feature-variable graphs for each model is provided in the supplemental materials.

Discussion

We describe a machine learning model that accurately stratifies inpatient antibiotic orders according to predicted likelihood of being not indicated or suboptimal and thus needing review by an ASP team. The model employs ubiquitously available data that can be automatically extracted from the EHR without the need for manual extraction, and can be interrogated to make interpretations explainable. Its predictions demonstrate accuracy over several spectra of antibiotics, although its performance appears slightly higher in community-acquired compared to hospital-acquired infections, potentially reflecting differences in complexity between infections arising in the community versus those in patients with significant healthcare exposure.

The most influential clinical features in our model include markers of hemodynamic instability, inflammation, and end-organ dysfunction. Relatively little model performance comes from antibiotic orders themselves. This emphasis on markers of infection mirrors clinicians' approach to empiric antibiotic use; empiric treatment is life-saving in newly septic patients, and is often guided by clinical status rather than microbiological data. Interestingly, markers of instability or inflammation did not yield positive predictions more than 48 hours in advance of review. In these cases abnormal values (leukocytosis, low systolic blood pressure) actually correlate with predictions that antibiotic use might not be indicated/optimal. One potential explanation is that abnormal values in this setting can arise from preexisting non-

infectious causes, such as reactive leukocytosis in malignancy or stable hypotension in a patient with cirrhosis. Alternatively, abnormal values 2–3 days prior to review may function as a proxy for prior antibiotic use, and the need to revisit prior empiric treatment.

We have previously described a machine learning approach to screen for antibiotic orders in need of ASP review, describing an approach that similarly found markers of clinical instability or systemic inflammation as key predictors.²⁰ In contrast to this prior work, which used data from a single time point, our model makes use of time series data, which lends context to data that would otherwise be interpreted in isolation. This likely accounts for the increased discriminative ability of the model described in this paper.

Microbiological and MAR data contributed little to model predictions in our study. Models trained on such data alone did retain predictive ability but with poorer performance compared to vital sign or laboratory data when each was used in isolation. In contrast, work recently published by Tran-The et al, who describe a model for recommending antibiotic discontinuation, oral switch, or de-escalation, places greater emphasis on specific classes of antibiotics as well as susceptibility data.²¹ Beyond slightly different use cases for the models, this may be in part due to differences in the time between antibiotic initiation and ASP review. 60% of orders in our data set were less than 48 hours old with a median of 1.2 days, whereas data described in the authors' paper had a mean antibiotic duration of 13.2 days in the discontinuation arm, 8.2 in their early de-escalation, and 17.3 in their late de-escalation (ibid). Microbiological data are unlikely to yield actionable information in fewer than 48 hours, a period in which decisions about antibiotics are often made empirically and influenced by markers of clinical instability or active infection. Sample size likely also influenced which features Goodman et al (2022), in a similar paper modeling antibiotic appropriateness, were able to make use of.²²

Ultimately, the goal of a model such as this one would be to use an automated pipeline to triage antimicrobial orders in order for ASP staff to make better use of finite resources by prioritizing orders for review. Surveys of antibiotic use in United States hospitals suggest more than half of patients admitted will receive antibiotics, with a median of 776 days of therapy per 1,000 patient-days.²³ Given the high proportion of patients receiving antibiotics, manually reviewing all orders is impractical, and our institution's current approach is to select limited subsets of the hospital at a time in tandem with periodic point prevalence studies. While local implementation might vary in its specifics, the core data used by a model such as this one would be able to be updated on a daily basis, providing a regular, automated means for directing ASP staff's attention to high-impact areas. This might extend ASP reach to a much broader swath of hospitalized patients.

Several limitations pertain to this model. While NAPS is a standardized and validated tool, inter-rater variability is higher at our institution for infectious syndromes with clear diagnostic and treatment guidelines compared to those where more weight is placed on clinical gestalt.^{24,25} In cases such as documented bacteremia where diagnostic criteria are straightforward, data labeling—and therefore model predictions—are likely more reliable than in less clear-cut instances like undifferentiated sepsis. Sample size, use of hospital medicine patients for training, and the single-center nature of the model, both of which reflect the time-intensive nature of ASP review, limit the model's generalizability and complexity. Many potentially useful features, such as considering antibiotics or microbes individually rather than by spectrum of coverage, would also be more feasible with a larger sample size. While we chose to focus on clinical variables relevant to the need for antibiotics and the spectrum of coverage required, several other factors contribute to “optimal” use, including considerations around patient comorbidities, toxicities, and social factors that are not accounted for in the model. External validation, both in academic and community settings, would strengthen the case for widespread use of a model such as this one.

From an implementation standpoint, while the model was optimized to identify high-yield orders for ASP review, unstructured data such as clinical notes detailing the prehospitalization use of suppressive/prophylactic antibiotics might lead to divergent assessments. Structured data such as vital signs are key factors in identifying and diagnosing infection. However, clinical notes, radiology reports, and pathology reports lend additional context—nuances gleaned from these sources cannot be accounted for by our model. Large language models, a type of neural network that can be appropriated for text classification tasks, could be used to abstract some of this information if given a large enough corpus of training data and represent a promising development in future efforts to more effectively implement stewardship interventions, although outstanding questions remain regarding privacy and regulatory structures.^{26,27}

In summary, we describe an approach toward automatically reviewing new inpatient antibiotic orders using machine learning. In contrast to labor-intensive manual review, this model uses data that can be abstracted from the medical record without the need for clinician involvement. The model's higher throughput might be used to optimize ASP teams' limited staffing resources; antibiotics predicted to be non-indicated or suboptimal could be prioritized, freeing up ASP resources for other tasks. Bearing our model's strengths and limitations in mind, it is best framed as a tool for helping ASP teams triage the highest-yield orders for PAF, and defer review of those least likely to need intervention. Future

extensions of this work include incorporating additional data sources, validation in broader practice settings, and investigating similar models' impact when used to provide direct feedback to providers.

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/ash.2025.10142>.

Acknowledgements. No parties aside from the named authors had a role in designing or executing the research described, or in drafting and submitting the manuscript.

Financial support. None.

Competing interests. The authors report no relevant funding sources or conflicts of interest related to this work.

Research transparency and reproducibility. An anonymized dataset can be made available on request.

References

- Magill SS, O'Leary E, Ray SM, et al. Antimicrobial use in US hospitals: comparison of results from emerging infections program prevalence surveys, 2015 and 2011. *Clin Infect Dis Off Publ Infect Dis Soc Am* 2021;72:1784–1792. doi:10.1093/cid/ciaa373.
- Tamma PD, Avdic E, Li DX, Dzintars K, Cosgrove SE. Association of adverse events with antibiotic use in hospitalized patients. *JAMA Intern Med* 2017;177:1308–1315. doi:10.1001/jamainternmed.2017.1938.
- Hagiya H, Kokado R, Ueda A, et al. Association of adverse drug events with broad-spectrum antibiotic use in hospitalized patients: a single-center study. *Intern Med* 2019;58:2621–2625. doi:10.2169/internalmedicine.2603-18.
- Murray CJL, Ikuta KS, Sharara F, et al. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *The Lancet* 2022;399:629–655. doi:10.1016/S0140-6736(21)02724-0.
- 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. doi:10.1001/archinte.163.8.972.
- Davey P, Marwick CA, Scott CL, et al. Interventions to improve antibiotic prescribing practices for hospital inpatients. *Cochrane Database Syst Rev* 2017;2017:CD003543. doi:10.1002/14651858.CD003543.pub4.
- Baur D, Gladstone BP, Burkert F, et al. Effect of antibiotic stewardship on the incidence of infection and colonisation with antibiotic-resistant bacteria and clostridium difficile infection: a systematic review and meta-analysis. *Lancet Infect Dis* 2017;17:990–1001. doi:10.1016/S1473-3099(17)30325-0.
- 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 Off Publ Infect Dis Soc Am* 2016;62:e51–77. doi:10.1093/cid/ciw118.
- Chen JZ, Hoang HL, Yaskina M, et al. Efficacy and safety of antimicrobial stewardship prospective audit and feedback in patients hospitalised with COVID-19 (COVASP): a pragmatic, cluster-randomised, non-inferiority trial. *Lancet Infect Dis* 2023;23:673–682. doi:10.1016/S1473-3099(22)00832-5.
- Ohashi K, Matsuoka T, Shinoda Y, et al. Evaluation of long-term pharmacist-led prospective audit and feedback in antimicrobial stewardship: an 8-year study. *Am J Infect Control* 2024;52(6):670–677. doi:10.1016/j.ajic.2023.12.013.
- Pulcini C, Morel CM, Tacconelli E, et al. Human resources estimates and funding for antibiotic stewardship teams are urgently needed. *Clin Microbiol Infect* 2017;23:785–787. doi:10.1016/j.cmi.2017.07.013.
- James R, Nakamachi Y, Morris A, et al. The feasibility and generalizability of assessing the appropriateness of antimicrobial prescribing in hospitals: a review of the Australian national antimicrobial prescribing survey. *JAC-Antimicrob Resist* 2022;4(1). doi:10.1093/jacamr/dlac012.
- Machine learning: Trends, perspectives, and prospects | Science. <https://www.science.org/doi/10.1126/science.aaa8415>. Accessed February 8, 2024.

14. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. KDD '16. Association for Computing Machinery. 2016:785–794. doi:[10.1145/2939672.2939785](https://doi.org/10.1145/2939672.2939785).
15. Churpek MM, Snyder A, Han X, *et al*. Quick sepsis-related organ failure assessment, systemic inflammatory response syndrome, and early warning scores for detecting clinical deterioration in infected patients outside the intensive care unit. *Am J Respir Crit Care Med* 2017;195:906–911. doi:[10.1164/rccm.201604-0854OC](https://doi.org/10.1164/rccm.201604-0854OC).
16. Prasad PA, Fang MC, Martinez SP, Liu KD, Kangelaris KN. Identifying the sickest during triage: using point-of-care severity scores to predict prognosis in emergency department patients with suspected sepsis. *J Hosp Med* 2021;16:453–461. doi:[10.12788/jhm.3642](https://doi.org/10.12788/jhm.3642).
17. Seok H, Jeon JH, Park DW. Antimicrobial therapy and antimicrobial stewardship in sepsis. *Infect Chemother* 2020;52:19–30. doi:[10.3947/ic.2020.52.1.19](https://doi.org/10.3947/ic.2020.52.1.19).
18. Collins GS, Moons KGM, Dhiman P, *et al*. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 2024;385:e078378. doi:[10.1136/bmj-2023-078378](https://doi.org/10.1136/bmj-2023-078378).
19. O'Leary EN, Edwards JR, Srinivasan A, *et al*. National healthcare safety network standardized antimicrobial administration ratios (SAARs): a progress report and risk modeling update using 2017 Data. *Clin Infect Dis Off Publ Infect Dis Soc Am* 2020;71:e702–e709. doi:[10.1093/cid/ciaa326](https://doi.org/10.1093/cid/ciaa326).
20. Bystritsky RJ, Beltran A, Young AT, Wong A, Hu X, Doernberg SB. Machine learning for the prediction of antimicrobial stewardship intervention in hospitalized patients receiving broad-spectrum agents. *Infect Control Hosp Epidemiol* 2020;41:1022–1027. doi:[10.1017/ice.2020.213](https://doi.org/10.1017/ice.2020.213).
21. Tran-The T, Heo E, Lim S, *et al*. Development of machine learning algorithms for scaling-up antibiotic stewardship. *Int J Med Inf* 2024;181:105300. doi:[10.1016/j.ijmedinf.2023.105300](https://doi.org/10.1016/j.ijmedinf.2023.105300).
22. Goodman KE, Heil EL, Claeys KC, Banoub M, Bork JT. Real-world antimicrobial stewardship experience in a large academic medical center: using statistical and machine learning approaches to identify intervention “Hotspots” in an antibiotic audit and feedback program. *Open Forum Infect Dis* 2022;9:ofac289. doi:[10.1093/ofid/ofac289](https://doi.org/10.1093/ofid/ofac289).
23. Polk RE, Fox C, Mahoney A, Letcavage J, MacDougall C. Measurement of adult antibacterial drug use in 130 US hospitals: comparison of defined daily dose and days of therapy. *Clin Infect Dis* 2007;44:664–670. doi:[10.1086/511640](https://doi.org/10.1086/511640).
24. Bystritsky R, Gruenberg K, Abdoler E, Hiltz-Horeczko A, Doernberg SB. Inter-rater reliability assessment of antibiotic prescription quality by infectious diseases physicians, fellows, and pharmacists. *Antimicrob Steward Healthc Epidemiol ASHE* 2023;3:e227. doi:[10.1017/ash.2023.509](https://doi.org/10.1017/ash.2023.509).
25. Khanina A, Douglas AP, Yeoh DK, *et al*. Validation of the Antifungal National Antimicrobial Prescribing Survey (AF-NAPS) quality assessment tool. *J Antimicrob Chemother* 2023;78:1367–1377. doi:[10.1093/jac/dkad085](https://doi.org/10.1093/jac/dkad085).
26. Bhayana R. Chatbots and large language models in radiology: a practical primer for clinical and research applications. *Radiology* 2024;310:e232756. doi:[10.1148/radiol.232756](https://doi.org/10.1148/radiol.232756).
27. Yang J, Liu C, Deng W, *et al*. Enhancing phenotype recognition in clinical notes using large language models: phenoBCBERT and phenoGPT. *Patterns* 2024;5:100887. doi:[10.1016/j.patter.2023.100887](https://doi.org/10.1016/j.patter.2023.100887).