A pragmatic machine learning model to predict carbapenem resistance

Ryan J McGuire
Washington University School of Medicine in St. Louis

Sean C Yu
Washington University School of Medicine in St. Louis

Philip R O Payne
Washington University School of Medicine in St. Louis

Albert M Lai
Washington University School of Medicine in St. Louis

M Cristina Vazquez-Guillamet
Washington University School of Medicine in St. Louis

See next page for additional authors

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs

Recommended Citation
McGuire, Ryan J; Yu, Sean C; Payne, Philip R O; Lai, Albert M; Vazquez-Guillamet, M Cristina; Kollef, Marin H; and Michelson, Andrew P, "A pragmatic machine learning model to predict carbapenem resistance." Antimicrobial agents and chemotherapy. 65, 7. e0006321 (2021).
https://digitalcommons.wustl.edu/open_access_pubs/10781

This Open Access Publication is brought to you for free and open access by Digital Commons@Becker. It has been accepted for inclusion in Open Access Publications by an authorized administrator of Digital Commons@Becker. For more information, please contact vanam@wustl.edu.
A Pragmatic Machine Learning Model To Predict Carbapenem Resistance

Ryan J. McGuire,a Sean C. Yu,b Philip R. O. Payne,b Albert M. Lai,b M. Cristina Vazquez-Guillamet,c,d Marin H. Kollef,c Andrew P. Michelsonb,c

aDepartment of Internal Medicine, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA
bInstitute for Informatics, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA
cDivision of Pulmonary and Critical Care Medicine, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA
dDivision of Infectious Disease, Washington University School of Medicine in St. Louis, St. Louis, Missouri, USA

ABSTRACT Infection caused by carbapenem-resistant (CR) organisms is a rising problem in the United States. While the risk factors for antibiotic resistance are well known, there remains a large need for the early identification of antibiotic-resistant infections. Using machine learning (ML), we sought to develop a prediction model for carbapenem resistance. All patients ≥18 years of age admitted to a tertiary-care academic medical center between 1 January 2012 and 10 October 2017 with ≥1 bacterial culture were eligible for inclusion. All demographic, medication, vital sign, procedure, laboratory, and culture/sensitivity data were extracted from the electronic health record. Organisms were considered CR if a single isolate was reported as intermediate or resistant. Patients with CR and non-CR organisms were temporally matched to maintain the positive/negative case ratio. Extreme gradient boosting was used for model development. In total, 68,472 patients met inclusion criteria, with 1,088 patients identified as having CR organisms. Sixty-seven features were used for predictive modeling. The most important features were number of prior antibiotic days, recent central venous catheter placement, and inpatient surgery. After model training, the area under the receiver operating characteristic curve was 0.846. The sensitivity of the model was 30%, with a positive predictive value (PPV) of 30% and a negative predictive value of 99%. Using readily available clinical data, we were able to create a ML model capable of predicting CR infections at the time of culture collection with a high PPV.

KEYWORDS antibiotic resistance, carbapenem, electronic health record, machine learning, predictive modeling

With 2.8 million antibiotic-resistant infections occurring each year in the United States, costing 2.3 billion dollars annually and leading to over 35,000 deaths, antibiotic resistance remains one of the biggest challenges facing health care today (1, 2). Despite knowing that inappropriately broad antibiotic coverage promotes antibiotic resistance (3–5), current guidelines for undifferentiated infections recommend the rapid initiation of empirical antibiotics, often unnecessarily broad in spectrum (6, 7). A subset of infections that is of particular concern is that caused by carbapenem-resistant species, as these isolates are associated with increased morbidity and mortality (8–10).

Despite carbapenem-resistant infection rates remaining stable over the last 8 years, carbapenem resistance (CR) remains a substantial threat (1). Additionally, the rise of extended-spectrum-beta-lactamase-producing organisms (1, 11) has the potential to lead to more frequent carbapenem use (12), which is a major risk factor for the development of further resistance (13–15).

Even with a plethora of well-known risk factors for antibiotic resistance (3–5, 9, 10, 13, 16–19), the ability to quickly and accurately identify which patients will go on to...
develop CR infections remains lacking in today’s clinical practice. Because CR is associated with prolonged hospitalization, increased mortality, and increased health care cost (8–10, 18, 19), the ability to accurately predict CR at the time of culture collection has many potential benefits, including better patient outcomes through appropriate empirical antibiotic selection (20, 21), improved antibiotic stewardship, and decreased health care utilization.

The use of machine learning (ML) in health care has seen a significant growth in the last decade, particularly in the fields of diagnosis, prognosis, and drug development (22, 23). ML is a branch of artificial intelligence that brings together statistical learning and computer science to facilitate knowledge discovery. These advanced computational methods are commonly used to analyze large volumes of information and derive highly accurate predictive models that function even in the presence of missing data.

As technology and ML algorithms have progressed, it is becoming increasingly more feasible for these advanced computational methods to be integrated into the electronic health record (EHR), though a standardized framework is lacking for how this integration could be accomplished, especially with respect to antibiotic resistance and stewardship. With this in mind, we sought to develop a novel ML algorithm to predict patients at risk for CR infections at the time of culture collection.

(This work was presented in part at the CHEST Annual Meeting, 18 to 21 October 2020 [24].)

RESULTS

In total, 68,472 patients met eligibility criteria, with 1,088 patients with carbapenem-resistant infections identified. After subsampling, 58,752 patients with non-carbapenem-resistant (NCR) infections were used for model training and testing. Baseline demographics and their differences are shown in Table 1. Table 2 shows a representative sample of cohort differences by feature; a breakdown of all 67 features can be found in Table S2 in the supplemental material. Patients with carbapenem-resistant infections were more likely to be male (58% versus 49%; \( P < 0.0001 \)), to have had a previous admission within the last

### TABLE 1 Baseline characteristics

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Value for patients with infection type(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Nonresistant</td>
</tr>
<tr>
<td>Total no. of patients</td>
<td>58,752</td>
</tr>
<tr>
<td>No. (%) of females</td>
<td>29,898 (51)</td>
</tr>
<tr>
<td>No. (%) of race</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>38,856 (66)</td>
</tr>
<tr>
<td>Black</td>
<td>16,211 (28)</td>
</tr>
<tr>
<td>Other</td>
<td>3,250 (6)</td>
</tr>
<tr>
<td>Asian</td>
<td>435 (1)</td>
</tr>
<tr>
<td>Age (yrs)</td>
<td>59 (46–70)</td>
</tr>
<tr>
<td>BMI (kg/m(^2))</td>
<td>28 (23–33)</td>
</tr>
<tr>
<td>Wt (kg)</td>
<td>80 (66–97)</td>
</tr>
<tr>
<td>No. (%) receiving service at time of culture</td>
<td></td>
</tr>
<tr>
<td>Medicine</td>
<td>31,726 (54)</td>
</tr>
<tr>
<td>Surgery</td>
<td>17,038 (29)</td>
</tr>
<tr>
<td>Emergency</td>
<td>2,937 (5)</td>
</tr>
<tr>
<td>Charlson score(^c)</td>
<td>4 (3–6)</td>
</tr>
<tr>
<td>Elixhauser score(^c)</td>
<td>11 (3–20)</td>
</tr>
</tbody>
</table>

\(^a\)All continuous variables are expressed as median (interquartile range) and were compared using Student’s \( t \) test unless marked. Categorical variables are expressed as number (percent) and were compared using the chi-square test.

\(^b\)\( P < 0.0001 \).

\(^c\)Compared using the Mann-Whitney test.
180 days (50% versus 26%; \( P < 0.0001 \)), to have received systemic antibiotics in the last 180 days (45% versus 18%; \( P < 0.0001 \)), to have had a central venous catheter (CVC) placed within 24 h of culture collection (65% versus 32%; \( P < 0.0001 \)), and to have had inpatient surgery (45% versus 29%; \( P < 0.0001 \)). The most common specimen sources for the resistance cohort are shown in Table 3. The most common carbapenem-resistant species belonged to the genera *Pseudomonas* (58%), *Enterobacter* (10%), and *Acinetobacter* (8%) (Table S3). Information regarding missing data can be found in Table S4.

Figures 1 and 2 show the area under the receiver operator curve (AUROC) and area under the precision recall curve (AUPRC) for the 20% held-out testing data compared to the mean values of their respective training curves. The cross validation training curves are available in the supplemental material (Fig. S1 and S2). The mean AUROC and mean AUPRC for the training data were 0.843 and 0.218, respectively. The AUROC and AUPRC for the test data were 0.846 and 0.224, respectively.

Figure 3 shows various performance metrics compared to increasing threshold cutoffs. Using a probability of 0.16, the specificity of the model was 99%, sensitivity was 30%, positive predictive value (PPV) was 30%, and negative predictive value (NPV) was 99%. The model calibration curve is available in the supplemental material (Fig. S3).

The most important features for model prediction were CVC placement within 24 h of culture, total number of antibiotic days, the presence of a vancomycin-resistant *Enterococcus* swab during admission, mechanical ventilation within 24 h of culture, and length of surgery during admission. These features remained relatively stable during cross validation training. Further details about feature importance can be found in the supplemental material (Fig. S4 and S5).

### Table 2: Cohort differences by feature

<table>
<thead>
<tr>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value for patients with infection typea</td>
</tr>
<tr>
<td>Nonresistant (( n = 58,752 ))</td>
</tr>
<tr>
<td>No. of prior admissions since start of data</td>
</tr>
<tr>
<td>No. of days since last admission</td>
</tr>
<tr>
<td>No. (%) admitted within 180 days,</td>
</tr>
<tr>
<td>No. (%) with ICU stay within 180 days</td>
</tr>
<tr>
<td>No. (%) with surgery during admission</td>
</tr>
<tr>
<td>No. (%) with surgery within 180 days</td>
</tr>
<tr>
<td>Total antibiotic days since start of data</td>
</tr>
<tr>
<td>No. (%) receiving antibiotics within 180 days</td>
</tr>
<tr>
<td>No. (%) receiving carbapenem within 180 days</td>
</tr>
<tr>
<td>No. (%) intubated within 24 h of culture</td>
</tr>
<tr>
<td>No. (%) receiving CVC within 24 h of culture</td>
</tr>
<tr>
<td>No. (%) receiving CVC within 180 days</td>
</tr>
<tr>
<td>No. (%) receiving vasopressorb within 24 h of culture</td>
</tr>
</tbody>
</table>

*Abbreviations: CVC, central venous catheter; ICU, intensive care unit.*

*aAll continuous variables are expressed as median (interquartile range) and were compared using the Mann-Whitney test. Categorical variables are expressed as number (percent) and were compared using the chi-square test.*

*bAll \( P \) values were <0.0001.*

*cIncludes vasopressin, phenylephrine, norepinephrine, epinephrine, and dopamine.*

### Table 3: Most common carbapenem-resistant specimen data

<table>
<thead>
<tr>
<th>Specimen source</th>
<th>No. (%) in resistance cohort (( n = 1,088 ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantitative respiratory culturesa</td>
<td>395 (36)</td>
</tr>
<tr>
<td>Urine</td>
<td>272 (25)</td>
</tr>
<tr>
<td>Wound</td>
<td>120 (11)</td>
</tr>
<tr>
<td>Blood</td>
<td>103 (9)</td>
</tr>
<tr>
<td>Tissue</td>
<td>46 (4)</td>
</tr>
</tbody>
</table>

*aIncludes bronchoalveolar lavage, sputum, tracheal aspirate, and bronchial washings.*
DISCUSSION

There are multiple well-described risk factors for antibiotic resistance (3–5, 9, 10, 13, 16–19) but few documented methods that allow the accurate, real-time prediction of CR. In this study, we developed and validated an ML-based algorithm that uses readily available clinical data to predict the growth of a carbapenem-resistant species at the time of culture collection. In this cohort, the model achieved a sensitivity of 30%, a PPV of 30%, and an NPV of 99%.

In this population, CR was found in 1.6% of all isolates. The primary risk factors for CR were similar to those previously identified for antibiotic resistance (5, 9, 14, 16, 17) including male sex, mechanical ventilation, CVC placement, recent hospital admission, recent intensive care unit (ICU) stay, recent antibiotic administration, and prolonged antibiotic exposure. These findings suggest that CR is largely driven by an increased exposure to the health care system as a whole versus prolonged antibiotic exposure alone.

The most common sources of CR were the respiratory (36%) and urinary (25%) tracts (Table 3). This distribution of CR is similar to that reported in other publications (5). The most common carbapenem-resistant organisms identified were *Pseudomonas* species (58%) and *Enterobacter* species (10%), which can be viewed in the supplementary material (Table S3). Interestingly, *Pseudomonas* species exhibited a much higher incidence of resistant infections than that reported in other models (5, 25). This difference likely reflects a geographical and/or institutional variance in prevalence.

The most promising implementation of this model would be real-time integration into the EHR to be run at the time of all culture collections. While the PPV and sensitivity

![Testing receiver operating curve compared to mean training curve. Abbreviations: ROC, receiver operating curve; AUROC, area under the receiver operating characteristic curve.](https://journals.asm.org/journal/aac)
were relatively low, our model was still able to achieve an AUPRC of 0.224 during testing, despite an excessively large class imbalance. A PPV of 30% represents a significant improvement over other available metrics, especially considering the time of prediction and breadth of culture sources included. In the same vein, a CR incidence of 1.6% means that our model will rarely trigger positive, limiting the absolute number of patients who may be overtreated, but may substantially improve outcomes when the high morbidity and mortality associated with CR and inadequate antibiotic coverage are taken into account. Additionally, the NPV of 99% has a strong potential for de-escalating therapy or preventing the overuse of more novel and often last-resort anticarbapenemase agents.

While different institutions will have different technological capabilities, all features used for the training of this model are generally ubiquitous across most EHR systems, and we see this paper as representing a framework for this type of analysis. As with all machine learning models, our model is limited by all data being from a single academic center, and it is difficult to know how it would perform at another facility. However, now that a framework has been established, retraining the model on additional data is relatively straightforward and allows our model to be periodically updated and tuned to better reflect ever-changing microbiological patterns.

Looking at model performance in more detail, our training data produced a mean AUROC of 0.843 and mean AUPRC of 0.218. Both curves were relatively stable across all four cross validation runs (Fig. S1 and S2). After optimization, the model produced an AUROC of 0.846 and AUPRC of 0.224 from the 20% held-out testing data. Both test AUROC and test AUPRC correlated well with their corresponding training curves, which implies a well-fitted model and greater generalizability. Additionally, the calibration

**FIG 2** Testing precision recall curve compared to mean training curve. Abbreviations: PRC, precision recall curve; AUPRC, area under the precision recall curve.
curve (see the supplemental material) demonstrated good prediction correlation and reliability, with an $R^2$ value of 0.845 ($P = 0.001$), again signifying better model translatability to clinical practice.

Not surprisingly, many of the features that had the strongest influence on model prediction are also some of the best-documented risk factors for antibiotic resistance. These findings were to be expected given that all features were selected a priori and most of these features are proxies for increased health care exposure (5). More importantly, the features that were the strongest predictors of resistance remained relatively consistent across all four cross validation runs (Fig. S5). Feature consistency conveys model stability and is reinforced by the narrow range of training AUROC and AUPRC curves.

As evidenced in Fig. 3, choosing a threshold of 0.16 resulted in a sensitivity of 30% and PPV of 30%. A threshold of 0.16 was used for reporting, as this allowed us to maximize sensitivity and PPV without diminishing the NPV. However, the strength of our model versus a regression-based model is that the threshold for a positive prediction can be easily adjusted to maximize the most suited performance variable depending on application. The results reported are similar to previously published regression-based models (5, 26, 27), which are even further limited by their use of specific pathogen-based characteristics that are not known until well after culture collection; additionally, these models are not able to incorporate missing data.

Our model represents a large improvement over other models, as we included both positive and negative cultures as well as all bacterial pathogens and a broader range of infectious sources. This unfiltered inclusion of all culture results potentially allows our model to be executed in real time and return a result at time of culture collection.
(a more clinically significant time of prediction) versus waiting on the collected culture to become positive up to the standard 48 h later. We also see our model working in synergy with newer PCR-based and mass spectrometry-based technologies that allow more rapid identification of bacterial species (28, 29). Most of these technologies still require several hours after species identification to identify resistance genes, and our model could fill this time gap to help guide providers if an identified species is at risk of being carbapenem resistant. However, further prospective studies will need to be done to substantiate this hypothesis.

Unfortunately, XGBoost models are mathematically complex and cannot be easily decrypted. Despite the model’s concealed nature, it is promising that the most important variables identified for model prediction are similar to those known to be strongly associated with antibiotic resistance (Fig. S4), suggesting a level of clinical validity.

The strengths of our study included a relatively large cohort of over 68,000 patients and an observation period spanning 5 years. Additionally, we were able to closely follow the TRIPOD guidelines (30), including the use of cross-validation during training and using held-out data for validation. In contrast to using regression-based models, our use of gradient-boosted decision trees allowed us to include missing data and minimized the need for imputation, improving model generalizability and avoiding potential bias (31). Furthermore, all features were chosen a priori, and no $P$ value pruning was used for feature selection (31). Finally, the use of gradient-boosted decision trees allowed us to modify the predictive threshold and optimize the most clinically meaningful performance metrics.

Our study is not without limitations. First, we have already highlighted how a single center can limit generalizability based on differences in technological capabilities and infection prevalence. Second, the model was trained and tested on a single data set without external validation, further limiting the model’s generalizability. Third, we had access only to inpatient data and potentially missed relevant outpatient factors that could be used to improve model performance. Additionally, the constraints of our data set required us to use discharge data for comorbidity and surgery-based data rather than being able to screen these features at the time of culture procurement, though this limitation can be negated when running in real time. Last, we chose to define cut-offs for certain continuous features such as lab values, procedures, and medication administration, which potentially introduced bias (31).

**Conclusion.** Using readily available data, we have demonstrated that CR can be accurately predicted at the time of culture collection. With all of our predictive features being easily obtained from the EHR, our model can potentially be run in real time and improve antibiotic stewardship and patient outcomes. Further prospective research is being undertaken to investigate these claims and to assess the feasibility of model use in everyday clinical practice.

**MATERIALS AND METHODS**

**Study population.** A retrospective cohort analysis was performed on all patients 18 years of age and older admitted to Barnes-Jewish Hospital/Washington University in St. Louis, a large tertiary-care academic medical center, between 1 January 2012 and 10 October 2017, with one or more bacterial or mycology cultures. Patients with a diagnosis of cystic fibrosis were excluded due to differences in treatment strategies for antibiotic resistance. This study was conducted according to the Washington University School of Medicine Institutional Review Board Protocol (IRB no. 201804121).

**Data source.** Data were extracted directly from the EHR. Collected elements included demographic, medication, vital sign, laboratory, billing code, procedure, culture, and sensitivity data.

**Outcomes.** An isolate with CR was defined as any organism with reported in vitro resistance or intermediate sensitivity to one or more carbapenems. *Pseudomonas* species resistant to ertapenem alone were not included, given their inherent resistance.

**Cohort selection.** All patients with a documented isolate resistant to a carbapenem were analyzed at the time of their first resistant culture; data from subsequent admissions following a patient’s sentinel resistant admission could not be used. Any carbapenem-resistant culture (whether Gram negative or Gram positive) was considered for use in the positive cohort, as long as it was the first carbapenem-resistant culture for a given patient. We chose to look for first carbapenem resistance because we felt that providers inherently know that patients with prior carbapenem resistance are at greater risk for growing a similar isolate again.

All other patients who had at least one culture collected at any admission during the study window were used for selecting the negative cohort. Because our model will eventually be used in real time, no
specific microbiology result was required for patients in the negative cohort. This lack of filtering was used because a real-time model will encounter all possible culture outcomes, including Gram-positive species with or without CR, Gram-negative species with or without CR, fungal species, and no growth. Mycology culture data were also considered for both cohorts, as bacterial species are capable of growing on fungal media.

To account for temporal changes in prevalence and antibiotic use, patients with carbapenem-resistant pathogens (positive cohort) were matched to patients with non-carbapenem-resistant pathogens (negative cohort) at random, in a ratio that preserved the prevalence of resistance (1:54). Admission dates were used to match each positive-cohort patient to 54 corresponding negative-cohort patients who were admitted within 6 months of each other. While patients in the negative cohort may have had multiple admissions that met inclusion criteria, patients could be analyzed only once, and data were screened at the time of their matched admission.

**Feature selection and data processing.** The list of features used can be found in Table S1. Features were selected preprediction based on known risk factors for multidrug resistance and clinical availability. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidelines (30), as well as recent reporting guidelines set by Leisman et al. (31), were followed for method development and data reporting.

Time-dependent features, such as intubation or absolute neutrophil count within 24 h of culture, could use either pre or postculture data, with preculture data being preferentially selected if available. Besides body mass index (BMI) and weight, in which prior data were carried forward if missing, all other missing data were left as null. Medications, catheters, and procedures were presumed to have been completed or given if ordered. Only systemic antibiotics were used for calculating antibiotic-based features, as it was felt that topical or nonabsorbable oral antibiotics, such as oral vancomycin, were less likely to promote CR. To more accurately reflect what the model would encounter in real time, all features were screened at time of culture collection except for comorbidity-based features, surgery-based features, and features specifically discussed above due to the constraints of our data set. Comorbidity data were calculated using ICD (International Classification of Disease) codes at time of discharge, as there was no feasible way in our data set to determine what diagnoses were present prior to admission.

Data processing and analyses were performed with Python version 3.7.8 within Jupyter Lab (32). Additional libraries used included Pandas (33), Numpy (34), Scikit (35), Shapley Additive Explanations (36), and matplotlib (37).

**Model training and testing.** Extreme gradient boosting (XGBoost) classification was used to develop the prediction model. XGBoost is a form of gradient-boosted tree algorithms in which decision trees are sequentially built, with each iteration learning from past errors (38, 39). The final trained model uses input data from unknown patients to calculate a probability of CR. The data were randomly split into training and test data sets, such that 80% of patients were assigned to the training data set and 20% to the test data set. The model was derived using the training cohort, Bayesian hyperparameter tuning, and 4-fold cross validation. Stratification was used to ensure that relatively equal percentages of resistant patients were used for both model training and testing.

**Statistical analysis.** Between-group differences were calculated using Student’s t test, the Mann-Whitney test, or the chi-squared test of independence depending on the type and distribution of data. Continuous variables are reported as medians. The performance of the final, hyperparameter-optimized model was evaluated using the 20% held-out test data set. Performance was assessed using the area under the receiver operating curve (AUROC) and area under the precision recall curve (AUPRC). After model training, a probability cutoff was chosen that maximized performance. A P value of <0.001 was considered statistically significant.

**Data availability.** The ability to publish data to a public data repository is restricted due to the confidential nature of human subject data.

---

**SUPPLEMENTAL MATERIAL**

Supplemental material is available online only.

**SUPPLEMENTAL FILE 1, PDF file, 1.3 MB.**

**ACKNOWLEDGMENTS**

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. We have no conflicts of interest to declare.

**REFERENCES**


4. Fortin É, Platt RW, Fontela PS, Buckeridge DL, Quach C. 2015. Predicting antimicrobial resistance prevalence and incidence from indicators of antimicrobial use: what is the most accurate indicator for surveillance in...


https://doi.org/10.1038/s42256-019-0138-9.


