



RESEARCH ARTICLE

Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia [version 1; referees: 2 approved]

Mathupanee Oonsivilai ¹, Yin Mo^{1,2}, Nantasit Luangsanatip¹, Yoel Lubell ¹, Thyl Miliya³, Pisey Tan³, Lorn Loek³, Paul Turner ^{3,4}, Ben S. Cooper^{1,4}

¹Mahidol-Oxford Tropical Medicine Research Unit, Faculty of Tropical Medicine, Mahidol University, Bangkok, Thailand

²Division of Infectious Disease, University Medicine Cluster, National University Hospital, Singapore, Singapore

³Cambodia-Oxford Medical Research Unit, Angkor Hospital for Children, Siem Reap, Cambodia

⁴Center for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK

v1 First published: 10 Oct 2018, 3:131 (<https://doi.org/10.12688/wellcomeopenres.14847.1>)

Latest published: 10 Oct 2018, 3:131 (<https://doi.org/10.12688/wellcomeopenres.14847.1>)

Abstract

Background: Early and appropriate empiric antibiotic treatment of patients suspected of having sepsis is associated with reduced mortality. The increasing prevalence of antimicrobial resistance reduces the efficacy of empiric therapy guidelines derived from population data. This problem is particularly severe for children in developing country settings. We hypothesized that by applying machine learning approaches to readily collect patient data, it would be possible to obtain individualized predictions for targeted empiric antibiotic choices.

Methods and Findings: We analysed blood culture data collected from a 100-bed children's hospital in North-West Cambodia between February 2013 and January 2016. Clinical, demographic and living condition information was captured with 35 independent variables. Using these variables, we used a suite of machine learning algorithms to predict Gram stains and whether bacterial pathogens could be treated with common empiric antibiotic regimens: i) ampicillin and gentamicin; ii) ceftriaxone; iii) none of the above. 243 patients with bloodstream infections were available for analysis. We found that the random forest method had the best predictive performance overall as assessed by the area under the receiver operating characteristic curve (AUC). The random forest method gave an AUC of 0.80 (95%CI 0.66-0.94) for predicting susceptibility to ceftriaxone, 0.74 (0.59-0.89) for susceptibility to ampicillin and gentamicin, 0.85 (0.70-1.00) for susceptibility to neither, and 0.71 (0.57-0.86) for Gram stain result. Most important variables for predicting susceptibility were time from admission to blood culture, patient age, hospital versus community-acquired infection, and age-adjusted weight score.

Conclusions: Applying machine learning algorithms to patient data that are readily available even in resource-limited hospital settings can provide highly informative predictions on antibiotic susceptibilities to guide appropriate empiric antibiotic therapy. When used as a decision support tool, such approaches have the potential to improve targeting of empiric therapy, patient outcomes and reduce the burden of antimicrobial resistance.

Open Peer Review

Referee Status:

| | Invited Referees | |
|---------------------------------------|------------------|------------|
| | 1 | 2 |
| version 1 published 10 Oct 2018 | report | report |

1 **Quentin Leclerc**, London School of Hygiene & Tropical Medicine, UK

Gwen Knight , London School of Hygiene & Tropical Medicine, UK

2 **Randall C Wetzel**, Children's Hospital Los Angeles, USA

Melissa Aczon, Children's Hospital Los Angeles, USA

Discuss this article

Comments (0)

Keywords

resistance, antimicrobial, antibiotic, machine learning, prediction, neonate



This article is included in the [Mahidol Oxford Tropical Medicine Research Unit \(MORU\)](#) gateway.

Corresponding author: Paul Turner (pault@tropmedres.ac)

Author roles: **Oonsivilai M:** Formal Analysis, Methodology, Software, Visualization, Writing – Original Draft Preparation; **Mo Y:** Writing – Original Draft Preparation, Writing – Review & Editing; **Luangasanatip N:** Formal Analysis, Methodology; **Lubell Y:** Supervision; **Miliya T:** Investigation, Resources; **Tan P:** Investigation, Resources; **Loeuk L:** Investigation, Resources; **Turner P:** Conceptualization, Data Curation, Funding Acquisition, Supervision, Writing – Review & Editing; **Cooper BS:** Conceptualization, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This work was part of the Wellcome Trust-Mahidol University-Oxford Tropical Medicine Research Programme [106698/Z/14/Z]. BSC is supported by the UK Medical Research Council and Department for International Development [MR/ K006924/1]. MY is supported by the Singapore National Medical Research Council Research Fellowship and National University Hospital [grant number NMRC/Fellowship/0051/2017].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2018 Oonsivilai M *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution Licence](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Oonsivilai M, Mo Y, Luangasanatip N *et al.* **Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia [version 1; referees: 2 approved]** Wellcome Open Research 2018, 3:131 (<https://doi.org/10.12688/wellcomeopenres.14847.1>)

First published: 10 Oct 2018, 3:131 (<https://doi.org/10.12688/wellcomeopenres.14847.1>)

Introduction

There is consistent evidence that early and appropriate treatment of sepsis can reduce mortality¹. Since the definitive identification of a bacterial pathogen and its antibiotic susceptibility typically take three to four days using conventional culture methods, empiric antibiotic therapy (i.e. therapy that starts before the causative organism and its antibiotic susceptibility is known) is recommended. Choice of empirical antibiotic aims to balance two objectives: first, to cast a wide spectrum of coverage effective against the most likely causative organisms; second, to minimize the selection of resistance to reserve antibiotics for the wider population². Balancing the consequences associated with these two concerns - immediate patient outcomes and long-term resistance patterns impacting on future patients - represents a major challenge.

Empiric antibiotic choice for invasive bacterial infections in hospitalized children in low-to-middle income countries (LMICs) constitutes a particularly stark example of this problem owing to the high attributable mortality³, and the high prevalence of antimicrobial resistance, particularly in neonates⁴.

Current World Health Organization (WHO) guidelines for suspected sepsis or serious bacterial infection in newborns recommend empirical usage of gentamicin and ampicillin as the first line therapy, and change to third-generation cephalosporins if there is a lack of improvement in 24–72 hours^{5,6}. However, a systematic review of community-acquired neonatal sepsis in developing countries in 2012 found that of the causative pathogens in older infants (1–12 months), only 63% and 64% showed *in vitro* susceptibility to ampicillin and gentamicin, and third-generation cephalosporins, respectively⁶. For neonates, susceptibilities were even lower, with only 57% and 56% of pathogens susceptible to ampicillin and gentamicin and third-generation cephalosporins, respectively.

The potential harms of widespread antimicrobial resistance in children were illustrated in a recent study performed between 2007 and 2016 in a Cambodian children's hospital, which found those infected with third-generation cephalosporin-resistant bacteria were less likely than others to receive appropriate antimicrobial therapy (57% vs. 94%), and when appropriate therapy was administered, it was initiated later⁷. While anticipated clinical efficacy is the primary deciding factor in empirical antibiotic choices⁸, there are other important considerations as well. These include side effect profile⁹, cost, ease of administration and risks of promoting resistance emergence in hospital settings².

The adoption of antimicrobial stewardship programmes in hospitals is widely advocated internationally. This is true both in LMICs and high income countries^{10,11}. Locally-adapted hospital antibiotic policies are important components of such programmes, and typically contain recommendations for empiric antibiotic use. In most cases, these recommendations are derived from expert opinion and informal (non-quantitative) syntheses of available evidence¹². In some cases simple decision support systems based on logistic regression models and scoring

systems have been developed to help identify patients at high risk of being infected with multidrug-resistant pathogens. These approaches have primarily been developed in high- and upper middle-income countries^{13–18}. The use of predictive modelling as part of clinical decision support systems for antimicrobial management remains rudimentary, with only one example identified in a recent systematic review¹⁹. It has, however, been demonstrated in a randomized trial (in Israel, Germany and Italy) that a computerized decision support system making use of an underlying causal probabilistic network model can lead to more appropriate empiric antibiotic prescribing²⁰.

We hypothesized that applying modern machine learning approaches to readily collected patient data can surpass the performance of those based on logistic regression or simple decision trees, and derive patient-specific predictions for antibiotic susceptibility. Improved predictions directing empirical antibiotic therapy may contribute to better patient outcomes while avoiding the overuse of in-appropriate antibiotics that select for resistance.

In this study, we propose a locally adapted decision support system for a Cambodian children's hospital by applying an array of machine learning algorithms to patient-level data. We evaluated the ability of the algorithms to predict whether the causative organisms are susceptible to: i) ampicillin and gentamicin; ii) ceftriaxone; iii) none of the above. We specifically focus on the value of using the predictive models to identify patients at high risk of being infected with organisms resistant to ceftriaxone, a third-generation cephalosporin, the most commonly prescribed empirical antibiotic in practice at our study site.

Methods

Data collection

Retrospective data were collected from the Angkor Hospital for Children, a non-governmental hospital in Siem Reap, Northwestern Cambodia with approximately 100 beds, and its Satellite Clinic situated 30km away, with 20 inpatient beds. The hospital provides free surgical and general medical care to children less than 16 years of age and is equipped with an intensive care unit (ICU). Admitted neonates and children come from both urban and rural settings, with about two thirds residing in Siem Reap province. Over 90% of inpatients come from the community, and the rest are transferred from another hospital. None of the children are born in the hospital as there is no obstetric service.

Blood cultures are routinely taken from febrile inpatients (axillary temperature > 37.5°C) in accordance with clinical algorithms. Processing of these cultures including *in vitro* antibiotic susceptibility testing has been described elsewhere²¹. Children with at least one positive blood culture between February 2013 and January 2016 were included in the present study. Bloodstream infections with organisms that are likely skin contaminants such as coagulase-negative *Staphylococci*, Gram-positive bacilli, and mixed growths of environmental Gram-negative bacilli were excluded. We collected routine

clinical and living conditions data, including household size, presence of domestic animals, and factors relating to water and sanitation.

The study was approved by the Angkor Hospital for Children Institutional Review Board (AHC-IRB, 290) and the Oxford Tropical Research Ethics Committee (OxTREC, 08-12). Written consent for the use of the patient data was obtained from the guardians of the children.

Data analysis

We evaluated a suite of machine learning algorithms based on their ability to predict the invasive pathogens' Gram stain and *in vitro* susceptibility to antibiotics using available information prior to receiving culture results. Specifically, we considered susceptibility to: i) ampicillin and gentamicin; ii) ceftriaxone; iii) none of the above. In the event that more than one organism was grown from the same blood culture, they were categorized as susceptible to the specified antibiotics only if all organisms were susceptible to at least one.

To predict the above antibiotic susceptibilities, we selected 35 independent variables (predictors) from patient records by coding quality and relevance. Dichotomous predictors where all but ten or fewer patients had the same value were excluded. Missing data for binary predictors were treated as negative.

Weight for age standard deviations (z-score), a measure of malnutrition, was calculated using the lambda, mu, and sigma (LMS) method²² based on growth charts from the Centers for Disease Control. An earlier version of this article is available on BioRxiv as a preprint <https://doi.org/10.1101/367037>.

Training the algorithms

We first performed a logistic regression with backwards step-wise AIC model selection²³. Additional machine learning algorithms were then explored, including decision trees constructed via recursive partitioning²⁴, random forests²⁵, boosted decision trees using adaptive boosting²⁶, linear support vector machines (SVM)²⁷, polynomial SVMs, radial SVMs²⁸ and *k*-nearest neighbours²⁹. All analysis was done in R version 3.5.1³⁰ using the following packages: MASS³¹ (stepwise logistic regression); rpart³² (decision trees); ranger³³ (random forest); fastAdaboost³⁴ (boosted decision trees); kkn³⁵ (*k*-nearest neighbors); kernlab³⁶ (polynomial and radial SVM); and LiblineaR³⁷ (linear SVM and regularized logistic regression).

Machine learning models were five-fold cross-validated. Data were randomly partitioned into five parts, with one part randomly held out for error estimation. An average of three repeats was taken to calculate the error for parameter fitting. Parameters were fitted for highest Kappa based on a grid search³⁸.

The data set was split 80/20 for training and testing purposes. For categorical variables we ensured that each category is represented by at least one record in the training set. To assess performance in predicting antibiotic susceptibility patterns,

each model was refitted to 1,000 random selections of training and testing data sets. Performance was compared based on area under the receiver operating characteristic curve (AuROC) from the test set. We select the best method overall, then consider its probability calibration and the most important predictors. Variable importance in random forests was calculated using the method described in Janitza *et al.*³⁹.

Identifying the optimum cut-off

The ROC curve describes the diagnostic ability of a binary classifier system, and plots the true positive rate (or sensitivity, i.e. the chance of correctly identifying a non-susceptible infection) against the false positive rate (or 1-specificity, i.e. the chance of incorrectly concluding an infection is non-susceptible). From this it would be possible to derive an optimal cutoff to maximize the overall test *accuracy* (i.e. the chance the test gives a true positive or true negative results). However, choosing the cut-off in this way would fail to account for the different health and economic costs of the two types of misclassification error (predicting resistance to an antibiotic when an organism is susceptible, and predicting susceptibility when an organism is resistant). A more rational approach is to choose the test cutoff to maximize overall utility, taking into account the different numbers of expected false positives and false negatives associated with different cutoffs and the different health-economic impacts of these two misclassification errors. These include costs of antibiotic prescriptions, excess length of stay, mortality as a result of inappropriate empiric antibiotic prescriptions and, most challengingly, future impact of resistance selection resulting from different antibiotic prescribing decisions. Because the cost of future resistance is difficult to quantify, we adopt an alternative approach by considering willingness to pay (WTP) for avoiding unnecessary use of carbapenems (where such use is considered unnecessary if the organism is susceptible to a first line antibiotic). With this economic framework, and using conventional recommendations for WTP per quality adjusted life year (QALY) gained⁴⁰, health impact and monetary costs can be combined on the same scale and represented as net monetary value (monetary loss + QALY loss × WTP). In this way, we can assign different net monetary values to each of the four possible test outcomes (true positive, true negative, false positive, false negative). The optimal cutoff for utility will be a value of the specificity that minimizes this net monetary loss. We provide illustrative examples of these calculations (see [S3 Appendix](#) for further details) and provide a user-friendly web application to enable optimal cutoffs to be determined under different assumptions, available at <http://moru.shinyapps.io/ahc-ml-amr-cost/>.

Results

[Figure 1](#) shows the selection of cases used for model training and testing. Of 245 cases, two cases were excluded; one due to missing target outcome data, and the other due to a biologically impossible value.

Based on the AuROC derived from the test data set, the random forest method is the most frequently ranked first ([Figure 2A and 2C](#)), and was consistently superior to decision trees, boosted

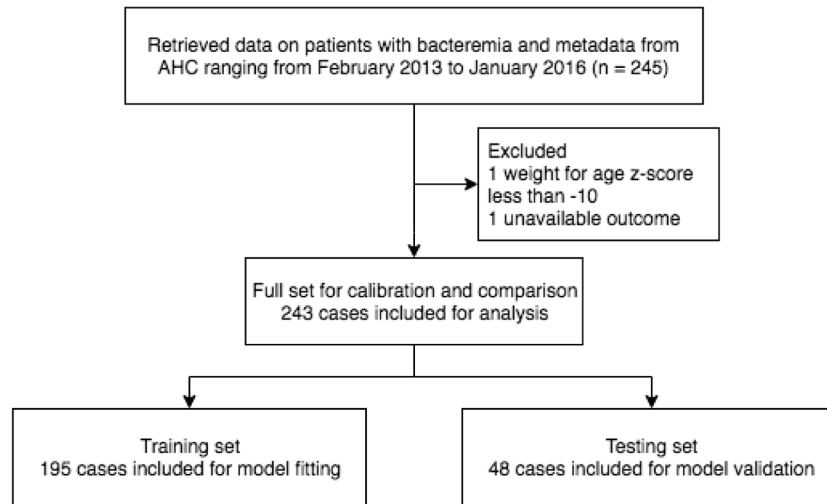


Figure 1. Selection of records.

decision trees, k -nearest neighbours, and the widely-used step-wise logistic regression. The performance of SVM approaches is generally good, but varies with the kernel the models were based on and which outcomes were being considered. For example, an SVM with polynomial kernel has similar performance to the random forest approach when predicting resistance to ceftriaxone (Figure 2B), but performed poorly in predicting lack of susceptibility to all three of the antibiotics (Figure 2C).

Ranking, although a good indicator of relative performance, does not necessarily indicate prediction ability itself. A comparison between multiple low AuROCs could still give a top ranked winner despite having low AuROC values. Figure 3 shows the Receiver Operating Characteristic (ROC) curves for predicting lack of susceptibility to ceftriaxone for all methods. This figure highlights the disconnect between predictive performance on the training data set (blue dotted line) and that on the test set (black dashed lines), highlighting the importance of separating the training and testing data. ROC curves show tradeoffs between specificity and sensitivity. If ROC curves for different methods were plotted on the same plot, it is possible for ROC curves for different methods to cross, indicating that optimal methods may vary depending on the cutoff used, and that the methods with the highest AUC may not always be the best for a given application. Importantly, the random forest test set ROC curve did not cross with other test set ROC curves. S1 Appendix shows ROC curves for remaining outcomes.

To be effective in supporting decisions, it is useful to not only rank well (predict correctly), but also to be well-calibrated (i.e. the estimated probabilities that pathogens lack susceptibility to an antibiotic should be similar to observed frequencies). Calibration refers to coherence between these estimated probabilities and the observed frequencies. To illustrate this, a calibration plot for the prediction of resistance to ceftriaxone with the random forests algorithm is shown in Figure 4. This

shows that even though the random forests method gives high accuracy (i.e. has a high AuROC), in this particular case it tends to be overconfident in its prediction probability. This overconfidence in prediction could not be improved even after adjustment with isotonic regression or Platt scaling⁴¹.

Figure 5 illustrates the influence of each independent variable on the random forest model in predicting antibiotic susceptibilities^{39,42}.

This shows that the most important predictor for resistance to ceftriaxone is patient age (leaving this out would decrease the model accuracy 100% of the time). Patient age is closely followed by days from hospital admission to blood sample, age-adjusted weight score, and the classification of the infection as hospital- or community-acquired (omitting this variable would decrease model accuracy 75% of the time). Other variables had much smaller effects.

The most important predictors in the random forest model for the other three outcomes were broadly similar. Interestingly, the classification of infection as hospital- or community-acquired had less importance for predicting resistance to ampicillin and gentamicin compared to ceftriaxone, but household size was found to much more important.

Figure 6 illustrates how, used as part of a decision support system, the choice of test threshold to inform antibiotic prescribing decisions would impact on the number of patients treated empirically with appropriate antibiotics. Taking a test threshold of 0.21 for the predicted probability that ceftriaxone would not be an effective treatment (so above this value, patients would be recommended to receive a second-line antibiotic, typically a carbapenem, instead of ceftriaxone), 15 out of 15 (100%) patients in the test data set who have ceftriaxone-resistant infections would be correctly identified (true positives). This threshold choice would also lead to 14 of the 33 (42%) patients with ceftriaxone-susceptible infections unnecessarily

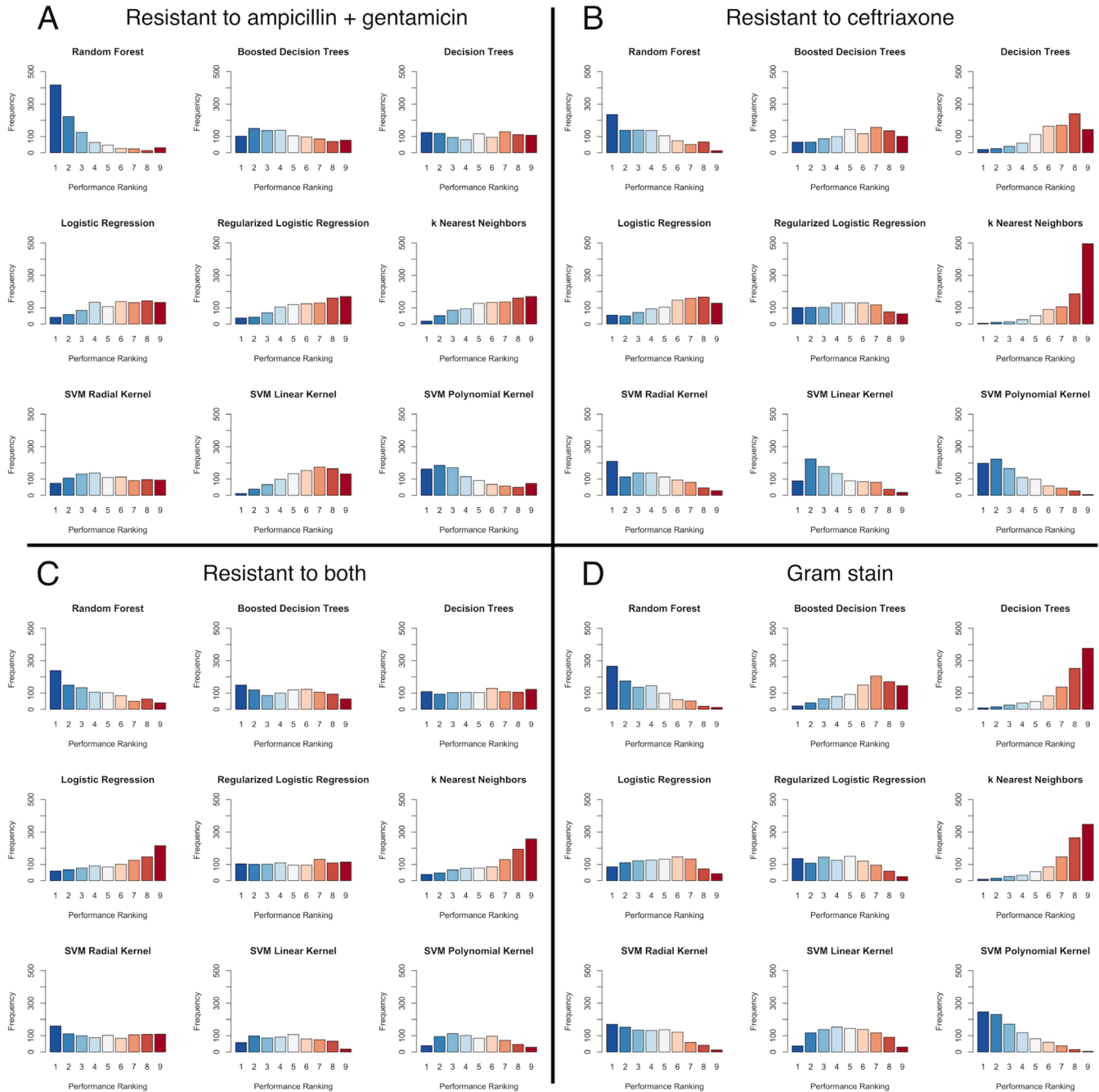


Figure 2. Comparison of performance rankings. Histograms of performance rankings obtained with 1000 random splits of the data into training (80%) and testing (20%) sets for the eight machine learning algorithms for predicting four outcomes (**A**) Resistance to ampicillin and gentamicin (**B**) Resistance to ceftriaxone (**C**) Resistance to ampicillin and gentamicin, and ceftriaxone (**D**) Gram stain. A ranking of 1 (blue) is best, 9 (red) is worst, based on the area under the receiver operating characteristic curve (AuROC) with the test data.

receiving the second-line antibiotic (over-treatment). Adjusting the threshold corresponds to moving the red line in [Figure 6A–B](#) up and down, changing the numbers of patients over- and under-treated. The choice of this threshold has an impact on patient outcomes and costs; their combined impact can be represented as the net utility loss (expressed as a net monetary value) due to infection ([Figure 6D](#)). A rational

approach would be to choose the threshold to minimize this utility loss. However, quantifying utility loss due to future selection for resistance when using antibiotics is challenging⁴³, so an alternative approach is to choose a prediction threshold based on clinical judgment, and work backwards to determine how this implicitly values the utility loss due to over-treatment. In this example, we find a threshold of 0.21 implies that we

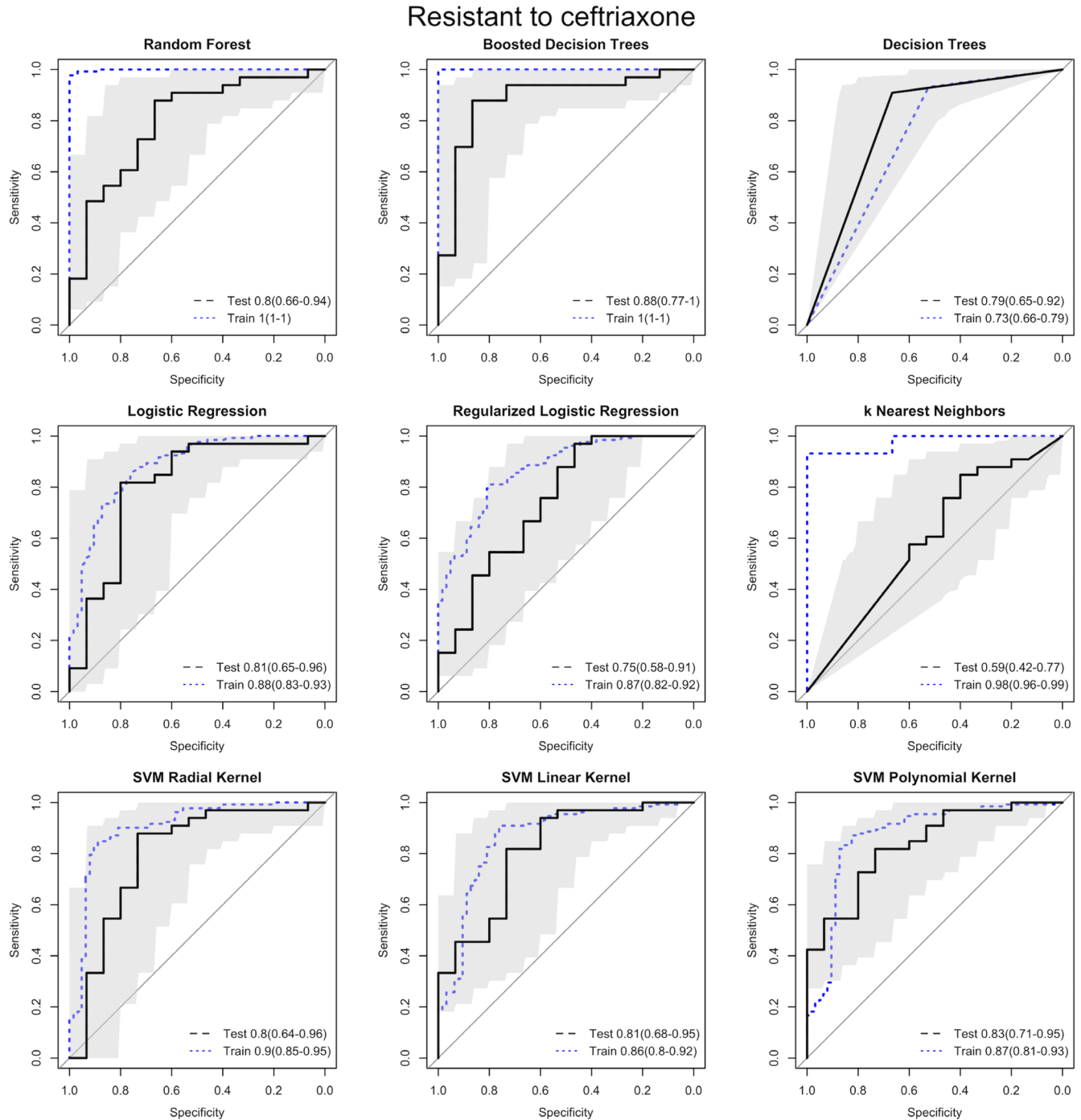


Figure 3. Receiver Operating Characteristic (ROC) curves for predicting resistance to ceftriaxone. Training set (blue dotted line), testing set, i.e. predictive performance (black solid line with 95% confidence intervals shown by shading). The diagonal line represents the line of no-discrimination, or the expected performance of a random guess.

would be willing to pay \$US 200 to avoid one unnecessary course of a carbapenem. Details of the calculations can be found in the supplementary text ([S3 Appendix](#)).

Discussion

Our results show that modern machine learning algorithms can outperform widely-used logistic regression models and provide

predictions about antibiotic susceptibility that could potentially be used to improve empirical antibiotic prescribing. We found that the random forest approach performed particularly well, especially for predicting ceftriaxone resistance, the most widely used empiric antibiotic for our study patients. To our knowledge this is the first time such machine learning algorithms have been applied to this problem in a hospital setting.

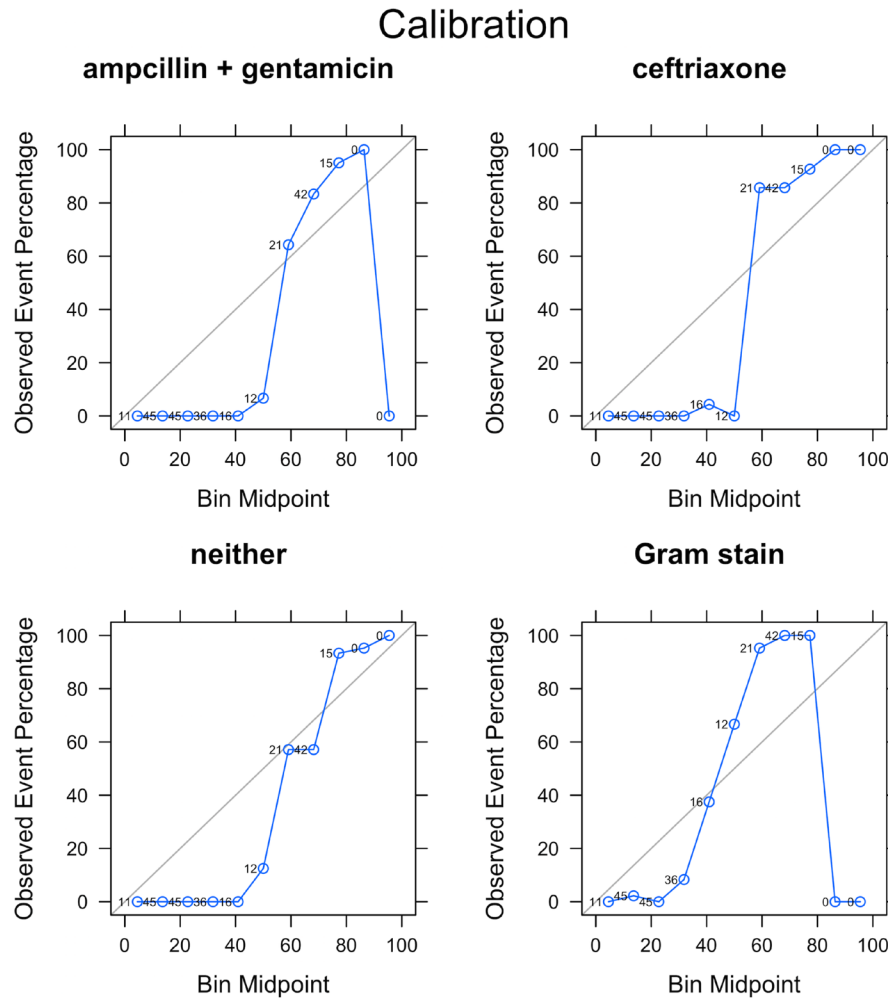


Figure 4. Calibration plots for random forest algorithm. This compares predicted probabilities (grouped into 10 equal-sized bins on x-axis) to observed event frequencies in real data (y-axis) using the entire data set. Points close to the grey diagonal line indicate that the predicted probability is close to the observed frequency. Numbers above points indicate the number of records contributing to each point.

The most important variables for predicting antibiotic susceptibility were found to be time from admission to blood culture, patient age, age-adjusted weight score, and hospital versus community-acquired infection. These are objective and routinely collected variables available in most clinical settings. All other variables included in the models are also easily collected at minimal cost through short questionnaires. The computations underlying the predictions can readily be performed in a few seconds on a low-cost computer, or remotely via any device connected to the Internet. This makes the approach highly suitable for other LMIC settings, which typically face the highest disease burden and the most urgent problems with antimicrobial resistance⁴⁴. These machine-learning models, which are often assumed to depend on large datasets more commonly available in high-income settings⁴⁵, may be of considerable value even in resource-limited and relatively data-poor settings.

Wider implications

Used as part of a decision support system, the best machine learning approaches should, in theory, make it possible to

substantially increase the proportion of patients who receive effective empiric antibiotics, while minimizing the risks of increased resistance selection that would be associated with a blanket change in the default choice of empiric antibiotics for all patients. Clearly, further work is needed to evaluate such deployment in practice⁴⁶.

Rapid microbiological diagnostic tests offer an alternative pathway for improving the precision of early antibiotic prescribing. Affordable and accurate tests are not currently available, but this situation may change in the coming years. While machine learning approaches proposed here could be considered a stopgap, we think it is more likely that the two approaches will be complementary. Results from future rapid diagnostic tests could be used as inputs in machine learning algorithms along with other patient variables, and would be expected to lead to more reliable predictions than those from the rapid tests alone.

Utility

A common dilemma in designing diagnostic systems is to identify the optimal cutoff point for sensitivity and specificity on

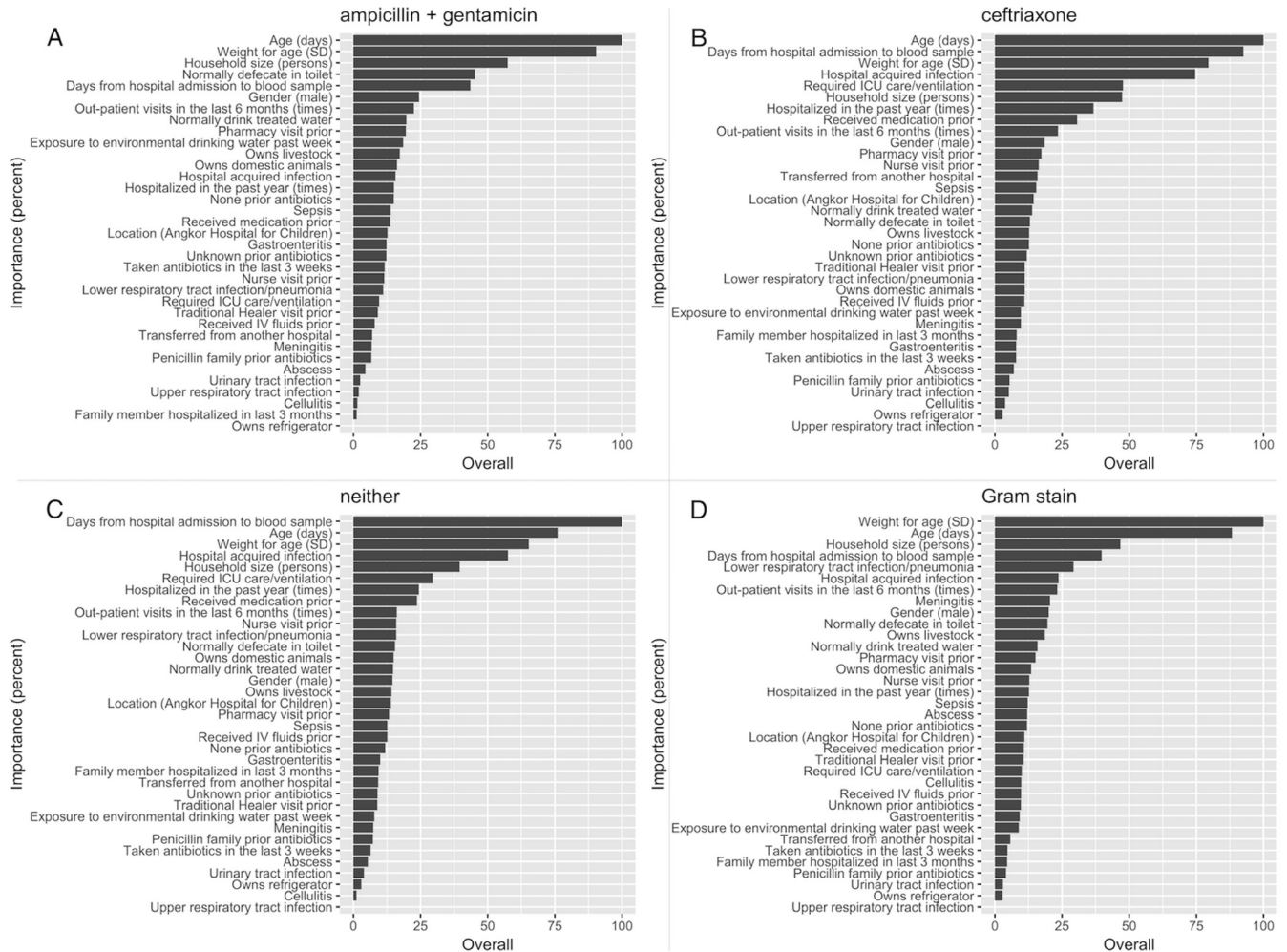


Figure 5. Importance of predictors in random forest models. Results show the relative importance of variables for predicting resistance to ampicillin + gentamicin (**A**) resistance to ceftriaxone (**B**) resistance to all three antibiotics (**C**) Gram stain (**D**).

the ROC curve. Increasing the sensitivity threshold for detecting antibiotic resistance will capture more cases of resistance, but will inadvertently lead to more false positives, resulting in increased prescriptions of unnecessary broad-spectrum antibiotics and selection for resistance. Conversely, while setting the threshold at higher specificity will reduce false positives, the model will miss more patients with resistant bacterial infections, leading to delayed prescription of appropriate antibiotics. A natural approach would be to choose the cutoff to maximise utility (which includes health outcomes and opportunity costs associated with economic costs). While quantifying the direct health care cost components is relatively straightforward, the costs of resistance are far more challenging to calculate. Shrestha *et al.* estimated the costs of resistance per antibiotic consumed, assigning a cost of \$US 0.8 and \$US 1.5 per standard unit of carbapenem in Thai and US settings, respectively⁴³. However, these estimates did not take into account the potentially grave potential consequence of losing a 'last-line' antibiotic to resistance. Better quantification of the cost of resistance is an important area of future research⁴⁷.

Strengths and limitations

We systematically evaluated a number of machine learning algorithms to determine the algorithms with the best predictive performance. Most currently available clinical scoring systems rely on logistic regression models, probably for historical reasons. No method is universally better than another method^{48,49}, however different algorithms have strengths and weaknesses and our results suggest that by focusing on a single learning algorithm, much of the previous literature may have missed an important opportunity.

It is possible that a more extensive exploration of logistic regression models would have yielded better results (for example by including interaction terms and variable transformations). However, such complexities are rarely considered in practice and would impose a substantially greater burden on the analyst than the simple "cookbook" approaches considered in this study.

A second important strength of our work is that algorithm training and evaluation were performed on different data sets. Though

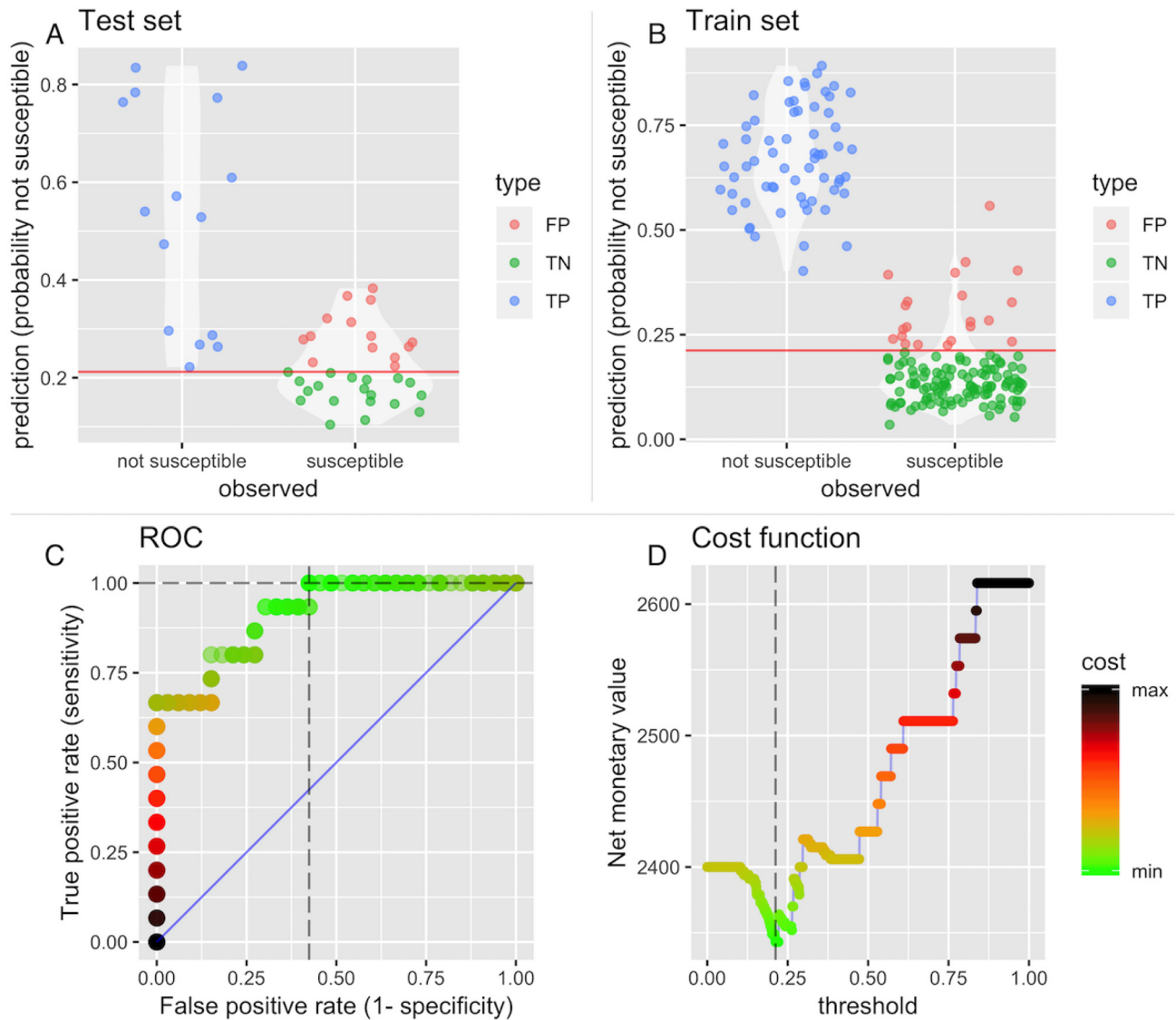


Figure 6. Effects of test cutoff on decision outcomes and utility. Panels **A** and **B**: Impact of test threshold (horizontal red line in panels **A** and **B**) on classification of resistance to ceftriaxone into false negatives (FN), false positives (FP), true negatives (TN) and true positives (TP) in test (panel **A**) and training (panel **B**) data. Panel **C**: The ROC curve is shaded according to utility loss at different cutoffs, where horizontal dashed lines correspond to the threshold selected by minimizing the cost function (**D**), i.e. maximising utility. Higher utilities, i.e. lower costs (expressed as a net monetary value) are shaded in green. Interactive version available at <http://moru.shinyapps.io/ahc-ml-amr-cost/>

there are some notable exceptions^{13,15,18}, this separation has not always been done in previous attempts to predict antibiotic susceptibility. As is clearly shown in Figure 3, if this separation is not done true predictive power is likely to be substantially lower than reported.

Thirdly, our analysis uses single-hospital data for formulating and evaluating models. If we had used a large dataset aggregated from multiple settings in the hope of increasing generalisability, the algorithm performance would have likely suffered. Scoring systems developed in one setting have been found to have substantially worse performance in different settings^{14,50}. By applying many different models to the same data set, our approach focuses on generalizing predictions toward new events within the same setting⁵¹.

There are several limitations to our study. The trade-off with a setting-specific predictive system is the likely poor predictive value when applied in another setting^{14,50}. Wider deployment of such approaches would require models to be tailored to local data. The model may also become less relevant as time passes. Identifying the most appropriate temporal and spatial selection windows for training data is an important area for future research.

Understanding the algorithms

One potential obstacle to the wider adoption of machine learning algorithms is that, to many, they are a black box. An intuitive way to understand them is to consider a geometric interpretation. Suppose we have a dataset with two predictors, height and weight and one binary outcome, diseased or healthy. We then plot a graph with weights on the x-axis and height on the

Table 1. Distribution of variables for logistic regression for susceptibility to ceftriaxone.

| Characteristics | Treatable n = 127 (No/Yes) | Resistant n = 68 (No/Yes) | OR (univariate) | 95% CI | P-value |
|---|----------------------------------|---------------------------------|--------------------|------------|---------|
| Age (days) | 703; 1063* | 1616; 1613* | 1.00 | 1.00-1.00 | <0.001 |
| Complication during admission | | | | | |
| Required ICU care/ventilation | 32/31 | 23/109 | 0.20 | 0.10-0.40 | <0.001 |
| Transfer from another hospital | 18/45 | 18/114 | 0.47 | 0.22-1.02 | 0.057 |
| Admission differential diagnosis | | | | | |
| Sepsis | 37/26 | 82/50 | 1.19 | 0.64-2.21 | 0.581 |
| Meningitis | 4/59 | 23/109 | 2.63 | 0.86-8.06 | 0.090 |
| Lower respiratory tract infection/pneumonia | 17/46 | 33/99 | 0.83 | 0.42-1.65 | 0.596 |
| Upper respiratory tract infection | 3/60 | 6/126 | 1.29 | 0.33-5.04 | 0.714 |
| Gastroenteritis | 9/54 | 18/114 | 0.95 | 0.40-2.25 | 0.902 |
| Cellulitis | 4/59 | 11/121 | 0.95 | 0.28-3.29 | 0.937 |
| Abscess | 2/61 | 10/122 | 1.08 | 0.32-3.65 | 0.902 |
| Urinary tract infection | 2/61 | 12/120 | 7.36 | 0.95-57.25 | 0.057 |
| Weight for age (SD) | -2.2; 1.7* | -2.1; 1.7* | 1.00 | 0.84-1.20 | 0.968 |
| Hospitalised in the last year (times) | 0; 0-3‡ | 0; 0-3‡ | 0.50 | 0.30-0.84 | 0.009 |
| Out-patient visits in the last 6 months (times) | 0; 0-3‡ | 0; 0-3‡ | 1.12 | 0.71-1.79 | 0.620 |
| Treatment prior to current admission | | | | | |
| Pharmacy | 8/55 | 43/89 | 3.80 | 1.67-8.66 | 0.001 |
| Nurse | 22/41 | 64/68 | 1.64 | 0.88-3.03 | 0.117 |
| Traditional Healer (Khru Khmer) | 8/55 | 15/117 | 0.77 | 0.32-1.87 | 0.562 |
| Received IV fluids | 11/52 | 31/101 | 1.23 | 0.59-2.55 | 0.576 |
| Received medication | 34/29 | 105/27 | 4.81 | 2.44-9.48 | <0.001 |
| Household size | 6; 3-10‡ | 6; 3-10‡ | 1.06 | 0.93-1.20 | 0.403 |
| Owns domestic animals | 49/14 | 92/40 | 0.56 | 0.27-1.17 | 0.122 |
| Owns livestock | 44/19 | 89/43 | 0.79 | 0.42-1.49 | 0.463 |
| Normally defecate in a toilet | 33/30 | 62/70 | 0.83 | 0.45-1.51 | 0.537 |
| Owns refrigerator | 4/59 | 5/127 | 1.46 | 0.38-5.60 | 0.578 |
| Taken antibiotics in the last 3 weeks | 4/59 | 22/110 | 3.78 | 1.08-13.20 | 0.037 |
| Family member hospitalized in last 3 months | 5/58 | 9/123 | 0.85 | 0.27-2.65 | 0.777 |
| Exposure to environmental drinking water in past week | 7/56 | 24/108 | 2.15 | 0.88-5.24 | 0.091 |
| Normally drink treated water | 26/37 | 61/71 | 1.43 | 0.78-2.63 | 0.249 |
| Hospital acquired infection | 34/29 | 11/121 | 0.07 | 0.03-0.16 | <0.001 |
| Days from hospital admission to blood sample | 0; 0-104‡ | 0; 0-104‡ | 0.87 | 0.81-0.94 | <0.001 |
| Gender (Male) | 33/30 | 79/53 | 1.27 | 0.70-2.33 | 0.430 |
| Location (Angkor Hospital for Children) | 54/9 | 94/38 | 0.28 | 0.12-0.66 | 0.004 |
| Taken antibiotics prior to admission | | | | | |
| None (antibiotics) | 39/24 | 62/70 | 0.51 | 0.28-0.95 | 0.033 |
| Penicillin Family | 4/59 | 17/115 | 2.32 | 0.84-6.45 | 0.106 |
| Unknown | 17/46 | 53/79 | 1.95 | 0.99-3.84 | 0.053 |

*Mean; SD for normal distributions, ‡Mode; Range for exponential distributions,

SD, standard deviation; CI, confidence interval; OR, odds ratio; Inf, infinity

Odds ratio from multivariate logistic regression analysis prior to step-wise backward elimination

See [S2 Appendix](#) for other outcomes

y-axis. We can imagine each data point inhabiting a point in this 2-dimensional graph plane, *feature space*. Each point would have a label of the class we are trying to predict (i.e. diseased/healthy). A classification problem can be likened to a search to find a line (or lines) which best separates the data points with different labels on its feature space. In this example, this refers to a line which splits between the diseased and healthy on the height-weight graph. For two independent variables this plane is 2 dimensional. For n predictors this would require n -dimensions. For $n > 3$ this is harder to visualize, but the geometric interpretation still holds.

A geometric visualization allows us to appreciate the varying performances of each method by considering how each method arrives at the conclusion as to which line (or combination of lines) is best. A decision tree can be considered a combination of decisions, each represented by a line in our feature plane (i.e. *is weight > 50 kg?* can be considered a line at 50 on the weight axis). A combination of simple lines allows for more complex decision boundaries. However, because of their ability to create complex boundaries, they tend to over-fit. Random forests are designed to correct for the over-fitting by decision trees by building a consensus of a multitude of decision trees, and averaging these trees by giving the majority vote after polling all component decision trees based on classification.

Conclusions

Decision support systems, informed by readily available setting-specific data, have the potential to lead to evidence-based hospital antibiotic policies which could improve appropriate prescribing of empiric antibiotics. This would be expected to lead to better patient outcomes and minimize the risk of

antibiotic resistance emergence. While guidelines for developing a hospital antibiotic policy advocate conducting literature reviews and basing recommendations on local cumulative surveillance antibiograms¹², we have shown that machine learning algorithms informed by relatively small amounts of patient-level data can be used to derive patient specific predictions for empirical antibiotic therapy. Such a prediction system can be developed cheaply, using easily-collected data, and is well-suited to LMIC settings.

Data availability

Zenodo: Manuscript dataset - Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia, <http://doi.org/10.5281/zenodo.1256967>⁵².

Grant information

This work was part of the Wellcome Trust-Mahidol University-Oxford Tropical Medicine Research Programme [106698/Z/14/Z]. BSC is supported by the UK Medical Research Council and Department for International Development [MR/K006924/1]. MY is supported by the Singapore National Medical Research Council Research Fellowship and National University Hospital [grant number NMRC/Fellowship/0051/2017].

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We thank the staff and patients at Angkor Hospital for Children for their invaluable contribution and support that made this work possible.

Supplementary material

S1 Appendix. ROC comparisons for predicting for treatability by remaining outcomes: ampicillin + gentamicin, neither, and Gram stain. Blue dotted lines: training set, black dashed lines: testing test (actual performance).

[Click here to access the data.](#)

S2 Appendix. Distribution of variables for logistic regression for treatability by remaining outcomes: ampicillin + gentamicin, neither, and Gram stain.

[Click here to access the data.](#)

S3 Appendix. Economic model to identify the optimum cut off.

[Click here to access the data.](#)

References

1. Liu VX, Fielding-Singh V, Greene JD, *et al.*: **The Timing of Early Antibiotics and Hospital Mortality in Sepsis.** *Am J Respir Crit Care Med.* 2017; **196**(7): 856–863. [PubMed Abstract](#) | [Publisher Full Text](#)
2. de Man P, Verhoeven B, Verbrugh HA, *et al.*: **An antibiotic policy to prevent emergence of resistant bacilli.** *Lancet.* 2000; **355**(9208): 973–978. [PubMed Abstract](#) | [Publisher Full Text](#)
3. Liu L, Oza S, Hogan D, *et al.*: **Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals.** *Lancet.* 2016; **388**(10063): 3027–3035. [PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
4. Lubell Y, Ashley EA, Turner C, *et al.*: **Susceptibility of community-acquired pathogens to antibiotics in Africa and Asia in neonates—an alarmingly short**

- review. *Trop Med Int Health*. 2011; **16**(2): 145–151.
[PubMed Abstract](#) | [Publisher Full Text](#)
5. Organization WH: **Pocket book of hospital care for children: guidelines for the management of common childhood illnesses**. *World Health Organization*. 2013.
[Reference Source](#)
 6. Downie L, Armiento R, Subhi RW, *et al.*: **Community-acquired neonatal and infant sepsis in developing countries: efficacy of WHO's currently recommended antibiotics--systematic review and meta-analysis**. *Arch Dis Child*. 2013; **98**(2): 146–54.
[PubMed Abstract](#) | [Publisher Full Text](#)
 7. Fox-Lewis A, Takata J, Miliya T, *et al.*: **Antimicrobial Resistance in Invasive Bacterial Infections in Hospitalized Children, Cambodia, 2007-2016**. *Emerg Infect Dis*. 2018; **24**(5): 841–851.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 8. Mitimila EI, Cooke RW: **Antibiotic regimens for suspected early neonatal sepsis**. *Cochrane Database Syst Rev*. 2004; (4): CD004495.
[PubMed Abstract](#) | [Publisher Full Text](#)
 9. Fuchs A, Zimmermann L, Bickle Graz M, *et al.*: **Gentamicin Exposure and Sensorineural Hearing Loss in Preterm Infants**. *PLoS One*. 2016; **11**(7): e0158806.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 10. Johannsson B, Beekmann SE, Srinivasan A, *et al.*: **Improving antimicrobial stewardship: the evolution of programmatic strategies and barriers**. *Infect Control Hosp Epidemiol*. 2011; **32**(4): 367–374.
[PubMed Abstract](#) | [Publisher Full Text](#)
 11. Hersh AL, Beekmann SE, Polgreen PM, *et al.*: **Antimicrobial stewardship programs in pediatrics**. *Infect Control Hosp Epidemiol*. 2009; **30**(12): 1211–1217.
[PubMed Abstract](#) | [Publisher Full Text](#)
 12. World Health Organization: **Step-by-step approach for development and implementation of hospital antibiotic policy and standard treatment guidelines**. World Health Organization Regional office for South-East Asia. 2011.
[Reference Source](#)
 13. Tumbarello M, Trecarichi EM, Bassetti M, *et al.*: **Identifying patients harboring extended-spectrum-beta-lactamase-producing Enterobacteriaceae on hospital admission: derivation and validation of a scoring system**. *Antimicrob Agents Chemother*. 2011; **55**(7): 3485–3490.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 14. Pan A, Lee A, Cooper B, *et al.*: **Risk factors for previously unknown methicillin-resistant Staphylococcus aureus carriage on admission to 13 surgical wards in Europe**. *J Hosp Infect*. 2013; **83**(2): 107–113.
[PubMed Abstract](#) | [Publisher Full Text](#)
 15. Lee AS, Pan A, Harbarth S, *et al.*: **Variable performance of models for predicting methicillin-resistant Staphylococcus aureus carriage in European surgical wards**. *BMC Infect Dis*. 2015; **15**(1): 105.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 16. Kengkla K, Charoensuk N, Chaichana M, *et al.*: **Clinical risk scoring system for predicting extended-spectrum β -lactamase-producing Escherichia coli infection in hospitalized patients**. *J Hosp Infect*. 2016; **93**(1): 49–56.
[PubMed Abstract](#) | [Publisher Full Text](#)
 17. Johnson SW, Anderson DJ, May DB, *et al.*: **Utility of a clinical risk factor scoring model in predicting infection with extended-spectrum β -lactamase-producing enterobacteriaceae on hospital admission**. *Infect Control Hosp Epidemiol*. 2013; **34**(4): 385–392.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 18. Tumbarello M, Trecarichi EM, Tumietto F, *et al.*: **Predictive models for identification of hospitalized patients harboring KPC-producing Klebsiella pneumoniae**. *Antimicrob Agents Chemother*. 2014; **58**(6): 3514–20.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 19. Rawson TM, Moore LSP, Hernandez B, *et al.*: **A systematic review of clinical decision support systems for antimicrobial management: are we failing to investigate these interventions appropriately?** *Clin Microbiol Infect*. 2017; **23**(8): 524–532.
[PubMed Abstract](#) | [Publisher Full Text](#)
 20. Paul M, Andreassen S, Tacconelli E, *et al.*: **Improving empirical antibiotic treatment using TREAT, a computerized decision support system: cluster randomized trial**. *J Antimicrob Chemother*. 2006; **58**(6): 1238–1245.
[PubMed Abstract](#) | [Publisher Full Text](#)
 21. Stoesser N, Moore CE, Pocock JM, *et al.*: **Pediatric bloodstream infections in Cambodia, 2007 to 2011**. *Pediatr Infect Dis J*. 2013; **32**(7): e272–e276.
[PubMed Abstract](#) | [Publisher Full Text](#)
 22. Cole TJ: **The LMS method for constructing normalized growth standards**. *Eur J Clin Nutr*. 1990; **44**(1): 45–60.
[PubMed Abstract](#)
 23. Venables WN, Ripley BD: **Modern Applied Statistics with S**. New York: Springer, fourth ed., 2002.
[Reference Source](#)
 24. Breiman L, Friedman J, Stone C, *et al.*: **Classification and Regression Trees**. *The Wadsworth and Brooks-Cole statistics-probability series*, Taylor & Francis, 1984.
[Reference Source](#)
 25. Breiman L: **Random forests**. *Machine Learning*. 2001; **45**(5): 32.
 26. Freund H, Schapire RE: **Experiments with a new boosting algorithm**. In *lml*. Bari, Italy, 1996; **96**: 148–156.
[Reference Source](#)
 27. Hearst MA, Dumais ST, Osuna E, *et al.*: **Support vector machines**. *IEEE Intelligent Systems and their applications*. 1998; **13**(4): 18–28.
[Publisher Full Text](#)
 28. Scholkopf B, Smola AJ: **Learning with kernels: support vector machines, regularization, optimization, and beyond**. MIT press, 2001.
[Reference Source](#)
 29. Mitchell TM: **Machine learning**. Burr Ridge, IL: McGraw Hill, 1997; **45**(37): 870–877.
[Reference Source](#)
 30. R Core Team: **R: A Language and Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2016.
[Reference Source](#)
 31. Venables WN, Ripley BD: **Modern Applied Statistics with S**. New York: Springer, fourth ed., 2002.
[Reference Source](#)
 32. Therneau T, Atkinson B, Ripley B: **rpart: Recursive Partitioning and Regression Trees**. R package version 4.1-11. 2017.
 33. Wright MN, Ziegler A: **ranger: A fast implementation of random forests for high dimensional data in C++ and R**. *J Stat Softw*. 2017; **77**(1): 1–17.
[Publisher Full Text](#)
 34. Chatterjee S: **fastAdaboost: a Fast Implementation of Adaboost**. R package version 1.0.0. 2016.
[Reference Source](#)
 35. Schliep K, Hechenbichler K: **kkn: Weighted k-Nearest Neighbors**. R package version 1.3.1. 2016.
[Reference Source](#)
 36. Karatzoglou A, Smola A, Hornik K, *et al.*: **kernel – an S4 package for kernel methods in R**. *J Stat Softw*. 2004; **11**(9): 1–20.
[Publisher Full Text](#)
 37. Helleputte T: **LiblineAR: Linear Predictive Models Based on the LIBLINEAR C/C++ Library**. R package version 2.10-8. 2017.
[Reference Source](#)
 38. McHugh ML: **Interrater reliability: the kappa statistic**. *Biochem Med (Zagreb)*. 2012; **22**(3): 276–282.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 39. Janitza S, Celik E, Boulesteix AL: **A computationally fast variable importance test for random forests for high-dimensional data**. *Advances in Data Analysis and Classification*. 2016; 1–31.
[Publisher Full Text](#)
 40. WH Organization: **Macroeconomics and health: investing in health for economic development: report of the commission on macroeconomics and health**. In *Macroeconomics and health: investing in health for economic development: report of the commission on macroeconomics and health*. 2001.
[Reference Source](#)
 41. Niculescu-Mizil A, Caruana R: **Predicting good probabilities with supervised learning**. In *Proceedings of the 22Nd International Conference on Machine Learning*. ICML '05, (New York, NY USA), ACM, 2005; 625–632.
[Publisher Full Text](#)
 42. Altmann A, Toloşi L, Sander O, *et al.*: **Permutation importance: a corrected feature importance measure**. *Bioinformatics*. 2010; **26**(10): 1340–1347.
[PubMed Abstract](#) | [Publisher Full Text](#)
 43. Shrestha P, Cooper BS, Coast J, *et al.*: **Enumerating the economic cost of antimicrobial resistance per antibiotic consumed to inform the evaluation of interventions affecting their use**. *Antimicrob Resist Infect Control*. 2018; **7**(1): 98.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 44. Okeke IN, Laxminarayan R, Bhutta ZA, *et al.*: **Antimicrobial resistance in developing countries. Part I: recent trends and current status**. *Lancet Infect Dis*. 2005; **5**(8): 481–493.
[PubMed Abstract](#) | [Publisher Full Text](#)
 45. Salathé M: **Digital epidemiology: what is it, and where is it going?** *Life Sci Soc Policy*. 2018; **14**(1): 2018.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 46. Rawson TM, Moore LSP, Hernandez B, *et al.*: **A systematic review of clinical decision support systems for antimicrobial management: are we failing to investigate these interventions appropriately?** *Clin Microbiol Infect*. 2017; **23**(8): 524–532.
[PubMed Abstract](#) | [Publisher Full Text](#)
 47. Leibovici L, Paul M, Andreassen S: **Balancing the benefits and costs of antibiotic drugs: the TREAT model**. *Clin Microbiol Infect*. 2010; **16**(12): 1736–1739.
[PubMed Abstract](#) | [Publisher Full Text](#)
 48. Wolpert DH, Macready WG: **No free lunch theorems for optimization**. *IEEE transactions on evolutionary computation*. 1997; **1**(1): 67–82.
[Publisher Full Text](#)
 49. Caruana R, Niculescu-Mizil A: **An empirical comparison of supervised learning algorithms**. In *Proceedings of the 23rd international conference on Machine learning*. ACM 2006; 161–168.
[Publisher Full Text](#)
 50. Slekovec C, Bertrand X, Leroy J, *et al.*: **Identifying patients harboring extended-spectrum- β -lactamase-producing Enterobacteriaceae on hospital admission is not that simple**. *Antimicrob Agents Chemother*. 2012; **56**(4): 2218–2219.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
 51. Shaikhina T, Lowe D, Daga S, *et al.*: **Machine learning for predictive modelling based on small data in biomedical engineering**. *IFAC-PapersOnLine*. 2015; **48**(20): 469–474.
[Publisher Full Text](#)
 52. Oonsivilai M, Yin M, Luangsanatip N, *et al.*: **Manuscript dataset - Using machine learning to guide targeted and locally-tailored empiric antibiotic prescribing in a children's hospital in Cambodia [Data set]**. *Zenodo*. 2018.
<http://www.doi.org/10.5281/zenodo.1256967>

Open Peer Review

Current Referee Status:  

Version 1

Referee Report 29 January 2019

<https://doi.org/10.21956/wellcomeopenres.16176.r34568>



Randall C Wetzel , Melissa Aczon

VPICU, Children's Hospital Los Angeles, Los Angeles, CA, USA

This is a well written and potentially important paper analyzing the results of a Machine Learning approach to predicting antibiotic resistance in a children's hospital in Cambodia with multiple models including Random Forests. This study compared performance of several off-the-shelf machine learning methods in classifying bacterial infections. There were 243 cases, 35 independent variables and 4 clinical (binary) outcomes: Gram stain; resistance to ampicillin & gentamicin; resistance to ceftriaxone; and resistance to both. The results are pretty much exactly what you would expect (with random forests and GBMs outperforming other methods). Overall, the paper is well-written and offers some interesting cost and utility discussions. The methods section reads like a text book comparison of several ML algorithms.

We applaud this useful work which is well done and well described.

Methods:

1. Recommendation for authors: for each of the 4 tasks, please provide the class-split (resistant or non-resistant, etc) of your data.
2. It is nice to see the full cost analysis, instead of only sensitivity-specificity.

Results:

Figure 4: Presumably, this calibration plot corresponds to a single iteration (out of the 1000) of the random forest (RF), and that much like the ROCs in Figure 3, there'd be variance to them. Please state this explicitly in the text.

In figure 4 how was the calibration plot generated, was this a GiVitl or was a Hosmer Lemehsow methodology used. Check labels and spelling.

Figure 5: Similar to above comment – that this corresponds to a single iteration of the RF?

Figure 6: It's really nice to see this analysis. However, similar comment again as before: Is this from the RF model, and if so, a single iteration?

In general the figures should be similar with labels and formatting to aid the readers understanding. They take a lot of time to digest.

Were models with fewer, key observations considered. In figure sic the importance varies by antibiotic. Where the predictor importance logical – any comments, for example owning livestock and refrigeration

are clearly relevant but were these data prospectively gathered.

Utility: It is really good to see this section; it shows the authors are thinking about the problem holistically.

Congratulations on this nice comparison of ML techniques and a useful approach to suggesting appropriate antibiotic therapy.

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Referee Expertise: Machine learning approaches to using ICU data to predict outcomes and therapies.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 29 October 2018

<https://doi.org/10.21956/wellcomeopenres.16176.r34056>



Quentin Leclerc , Gwen Knight 

Centre for the Mathematical Modelling of Infectious Diseases, Department of Infectious Disease Epidemiology, Faculty of Epidemiology and Population Health, London School of Hygiene & Tropical Medicine, London, UK

This study attempts to evaluate the potential of machine-learning techniques to guide empiric antibiotic prescription in hospitalized children in a resource-limited hospital setting. This is an admirable aim as the impact on patient outcomes and resistance could be great.

The authors gathered data on the antibiotic resistance profile of bacteria infecting patients, as well as a range of information on the patients themselves which only required the use of a simple questionnaire to

be obtained. Different machine learning algorithms were trained and tested on these data, and their predictive capabilities were compared. The study successfully shows the potential of these techniques to identify patients with resistant bacteria, but also to avoid prescribing second-line antibiotics when unnecessary. The conclusions are encouraging regarding the applicability of this currently uncommon approach in a wider setting.

The manuscript is well written, with clarity in explanations. The authors clearly put a lot of effort into making their methods and results as comprehensive as possible, even for an audience without any previous knowledge of machine learning principles. Although the choice of figures is good, to ensure the reader's comprehension we believe some minor corrections should be made to these, and a few methodological points should be further clarified (see below). These do not alter the main analysis but are only for clarity in further understanding.

Minor points

Abstract

- Throughout: "AUC" is used three times as an abbreviation instead of "AuROC" which is defined as the correct abbreviation in the text, suggest replacing "AUC" with "AuROC"

Data

- In terms of the data collected, 35 independent variables seems like a lot – is this unusual for such a hospital to have such data? How generalisable is it?

Methods

- Could you please expand on why Gram staining was predicted? The reasoning for attempting to predict antimicrobial resistance is clear, but no justification or discussion is given for the importance of predicting Gram stain.
- In "Methods – Data Analysis", you mention that "missing data for binary predictors were treated as negative". Could you please justify this choice, and comment on the potential impact that this has on the results? (i.e. might be overestimating the absence of the predictor, which would lead to incorrectly judging its value as a predictor)
- In "Methods – Training the algorithm": Are "error estimation" and "testing purposes" the same thing? As in were the data split five ways to give the 80/20 split? the second paragraph feels like it shouldn't be there, aren't you actually explaining the k fold validation you did in the third one?

Results

- What is a "biologically impossible value"?
- It was unclear what random forest test set ROC "did not cross with other test set ROC curves" means and whether true or not. As in, by eye, it seems very close to crossing other curves – can you plot this in the supplementary perhaps?
- Figure 4: We found this figure difficult. Mainly, in terms of what the bins mean on the x axis. Did you calculate for each strain what the probability of resistance was and then compare to the actual probability of resistance? The bins then grouped strains with the same probability?
- Figure 4: the points appear to be misaligned with the x-axis?
- Figure 4: spelling mistake in "ampicillin"
- Figure 4&5: Could the titles match Figure 3 (i.e. "Resistant to ampicillin + gentamicin", "Resistant to both" ...)? especially since "neither" might induce the reader in thinking this is looking at fully susceptible bacteria.
- Figure 5: Should the y axis label not be "predictors"?

- Figure 5: To aid understanding, could you add the interpretation you give in the paragraph describing Figure 5's results i.e. that 100% means reduction in accuracy 100% of the time as I was unsure of the units for the x axis. Could you also expand on the methodology used to assess the relative importance of the predictors?
- Correction: "but household size was found to BE much more important"
- Figure 6: could you make the y axis the same for Figure A & B to make it clearer that the same cutoff is being used? Also, why is the test set on the left and the train set on the right? Wasn't the train set used to determine the threshold and then tested on the test set? So it makes more sense for them to be switched around?
- Figure 6: what does the vertical dashed line correspond to? It looks like the point that minimises the cost function?
- Figure 6: you include "false negatives (FN)" in the caption, but this doesn't actually appear in the figure, perhaps consider redoing it with different values to actually show false negatives? (In reality you would want to avoid having these, but here it would be beneficial for the reader's comprehension)
- Table 1 is not referred to in the text.
- Table 1: "Inf, infinity" in the caption, but this doesn't appear anywhere in the table
- Online tool – could you also allow the threshold to be varied in this?
- Online tool – what does the variation in the top two plots show? i.e. when you input a new cost value?
- Using the online tool, our understanding of the last result is that the cost function is minimised at a threshold of 0.21 when the WTP to avoid carbapenem use is linked to \$200? However, if this WTP decreases the minimum cost also decreases. i.e. the minimum point when, for example WTP is zero, is \$2290 but at \$200, more like \$2350. Could you provide further explanation of this?

Discussion

- Wider implications: it is unclear to us how machine learning would be needed with rapid diagnostic tests if the latter tell you bug and resistance. Could you expand more on what you think machine learning would add?
- Correction: "potentially grave potential consequence" (remove one)

S3 appendix

- Incorrectly labelled S4 on the pdf.
- *Table S2*: The "WTP for avoiding unnecessary imipenem use" is confusing terminology. At first read, it seems that this should multiply both the TN and the FN. Could you instead somehow link it to the more intuitive first explanation here that imipenem use is linked to a "cost" for future resistance?
- *Table S3*: could you make it clearer that this "WTP" is for QALY gain and different to the WTP for avoiding unnecessary imipenem use?

Is the work clearly and accurately presented and does it cite the current literature?

Yes

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

We have read this submission. We believe that we have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
