

Towards personalized guidelines: using machine-learning algorithms to guide antimicrobial selection

Ed Moran^{1*}, Esther Robinson², Christopher Green^{3,4}, Matt Keeling⁵ and Benjamin Collyer⁵

¹Southmead Hospital, North Bristol NHS Trust, Bristol BS10 5NB, UK; ²Birmingham Public Health Laboratory, Public Health England, Birmingham Heartlands Hospital, Bordesley Green East, Birmingham B9 5SS, UK; ³Birmingham Heartlands Hospital, University Hospitals Birmingham NHS Foundation Trust, Bordesley Green East, Birmingham B9 5SS, UK; ⁴Institute of Microbiology and Infection, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK; ⁵Zeeman Institute, Mathematics Institute and School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

*Corresponding author. E-mail: ed.moran@nhs.net

Received 30 September 2019; returned 29 October 2019; revised 1 May 2020; accepted 1 May 2020

Background: Electronic decision support systems could reduce the use of inappropriate or ineffective empirical antibiotics. We assessed the accuracy of an open-source machine-learning algorithm trained in predicting antibiotic resistance for three Gram-negative bacterial species isolated from patients' blood and urine within 48 h of hospital admission.

Methods: This retrospective, observational study used routine clinical information collected between January 2010 and October 2016 in Birmingham, UK. Patients from whose blood or urine cultures *Escherichia coli*, *Klebsiella pneumoniae* or *Pseudomonas aeruginosa* was isolated were identified. Their demographic, microbiology and prescribing data were used to train an open-source machine-learning algorithm—XGBoost—in predicting resistance to co-amoxiclav and piperacillin/tazobactam. Multivariate analysis was performed to identify predictors of resistance and create a point-scoring tool. The performance of both methods was compared with that of the original prescribers.

Results: There were 15695 admissions. The AUC of the receiver operating characteristic curve for the point-scoring tools ranged from 0.61 to 0.67, and performed no better than medical staff in the selection of appropriate antibiotics. The machine-learning system performed statistically but marginally better (AUC 0.70) and could have reduced the use of unnecessary broad-spectrum antibiotics by as much as 40% among those given co-amoxiclav, piperacillin/tazobactam or carbapenems. A validation study is required.

Conclusions: Machine-learning algorithms have the potential to help clinicians predict antimicrobial resistance in patients found to have a Gram-negative infection of blood or urine. Prospective studies are required to assess performance in an unselected patient cohort, understand the acceptability of such systems to clinicians and patients, and assess the impact on patient outcome.

Introduction

When treating patients for infection, the physician must balance the survival benefit that may result from the prompt initiation of effective antibiotic therapy against the risk of adverse side effects, complications, potential for the development of resistance and the increased costs that may follow the use of unnecessary broad-spectrum agents.¹ The modern health system has access to enormous amounts of information about each patient. Recent studies have demonstrated the ability of open-source machine-learning algorithms to use such data in the prediction of antibiotic resistance in different clinical settings.^{2,3}

In this proof-of-concept study, we assessed the accuracy of the XGBoost machine-learning algorithm trained in predicting antibiotic resistance for *Escherichia coli*, *Klebsiella pneumoniae* and *Pseudomonas aeruginosa* isolated from blood and urine cultures obtained from patients within the first 48 h of admission.⁴ In addition, we identified those factors most strongly associated with antimicrobial resistance (AMR) and developed a simple point-scoring (SPS) tool such as might be used by a clinician when assessing a patient for the risk of resistant Gram-negative infection. We compared the performance of this scoring tool and the XGBoost system with the antibiotic choices made by medical staff in the same patients.

Methods

This was a retrospective study using information collected as part of routine care from three hospitals between January 2010 and October 2016. Representing 1500 beds, they served an ethnically diverse population of 1.2 million people in Birmingham and Solihull, UK. The organization's microbiology laboratory served the majority of local primary care providers. All patients from whose blood or urine cultures *E. coli*, *K. pneumoniae* or *P. aeruginosa* was isolated within 48 h of admission were identified. Hospital demographic, microbiology and electronic-prescribing databases were then searched for information concerning these patients, including admission and discharge dates, diagnostic coding, primary care provider, antibiotics administered during prior admissions dating from January 2010 and all recorded instances of the three organisms of interest including antibiotic susceptibilities. The primary care provider's antibiotic-prescribing behaviour, represented by the total number of prescribed antibiotic items per 1000 registered patients (2016–17), was obtained from the PHE website. Datasets were pseudo-anonymized with a unique study number.

Ethics approvals and study registrations

Ethics approval was provided by the NHS Health Research Authority (reference 17/WM/0406). All clinical data were anonymized before sharing with academic partners.

Training of the gradient-boosted decision tree (GBDT) algorithm

A random allocation of 80% of the data ('training data') was used to train an XGBoost GBDT in the prediction of resistance of the microbiological isolates to each of co-amoxiclav and piperacillin/tazobactam. The remaining 20% was used to assess the GBDT performance ('test data'). This equated to 3064 testing and 12255 training cases for co-amoxiclav, and 2166 testing and 8663 training cases for piperacillin/tazobactam. Further details regarding GBDT training are provided in the [supplementary data](#) (Table S1 and Figure S1, available as [Supplementary data](#) at JAC Online).

Development of an SPS tool for the prediction of resistance

Multivariate analysis was performed on the dataset to identify significant predictors of resistance to co-amoxiclav and piperacillin/tazobactam (Table S2). A naive Bayes classification methodology was then used to generate scores for each variable. The five variables that in combination produced the most accurate predictions [as measured by the AUC for the receiver operating characteristic (ROC) curve] were identified, and an SPS tool was created (Table 1).

Assessing performance

To allow comparison of the two systems with medical staff, patients in the dataset who had received co-amoxiclav, piperacillin/tazobactam or a carbapenem as initial therapy in the first 72 h were identified. Prescribed agents from this period were then compared with the antibiotic susceptibility results. A prescribing decision was classified as 'correct' if the organism was susceptible to the agent chosen and was the narrowest effective agent of the three (co-amoxiclav, piperacillin/tazobactam or a carbapenem), with co-amoxiclav considered narrowest and carbapenems broadest.⁵ An antibiotic decision was classified as 'under' if the organism was resistant to the prescribed agent and 'over' if the organism was susceptible to the prescribed agent but a narrower-spectrum agent would have been suitable. At the time of data collection carbapenems were the institutional agent of choice for those with penicillin allergy and so all patients with recorded penicillin allergy were excluded from this analysis.

Table 1. The point-scoring tools

	Yes	No
A. Prediction of co-amoxiclav resistance (incorporating electronic-prescribing data)		
previous co-amoxiclav prescription	+3	-1
comorbidity: other inflammatory condition of skin; chronic ulcer of skin; other skin disorders	+1	-1
previous co-amoxiclav susceptibility	-2	0
comorbidity: septicaemia, shock	+3	0
previous co-amoxiclav resistance	+7	-1
B. Prediction of piperacillin/tazobactam resistance (incorporating electronic-prescribing data)		
sample type: blood	-2	0
previous co-amoxiclav susceptibility	-1	0
previous co-amoxiclav resistance	+4	-1
comorbidity: diabetes mellitus without complication	+2	-1
previous piperacillin/tazobactam resistance	+7	-1
C. Prediction of co-amoxiclav resistance (excluding electronic-prescribing data)		
comorbidity: mental retardation, senility and organic mental disorders	+2	0
consultant specialty: general medicine	+1	-1
previous co-amoxiclav susceptibility	-2	0
comorbidity: septicaemia, shock	+3	0
previous co-amoxiclav resistance	+7	-1
D. Prediction of piperacillin/tazobactam resistance (excluding electronic-prescribing data)		
comorbidity: COPD and bronchiectasis	+2	0
comorbidity: skin and subcutaneous tissue infections	+3	0
previous co-amoxiclav resistance	+5	-1
number of comorbidities >5	+1	-1
previous piperacillin/tazobactam resistance	+8	-1

The presence of a factor adds the value indicated in the 'Yes' column to the patient's score, the absence the addition of the value in the 'No' column. If the resulting total reaches a pre-determined threshold, the patient is considered at risk of resistance to that antibiotic. In this study, the threshold was set at a point producing a level of 'under' or 'over'-prescribing similar to that of medical staff.

The number of patients falling into each of the three categories as a result of medical staff prescribing in the first 72 h was compared with the predictions of the SPS and the GBDT applied to each individual patient in the test dataset.

Results

The total dataset represented 15695 patient admissions and antibiotic initiations and 9352 individual patients. The median age was 69 years. There were 15580 individual isolates of the three organisms of interest and the average number of isolates per patient over the time period studied was 1.55. Of the 15208 *E. coli* isolates, 26.3% were resistant to co-amoxiclav and 11.8% were resistant to piperacillin/tazobactam. This compared with 9.3% and 8.8%, respectively, of the 194 *K. pneumoniae* isolates. Of the 178 *P. aeruginosa* isolates, 9.4% were resistant to piperacillin/tazobactam and 12.0% were resistant to meropenem. Of the 15364 co-

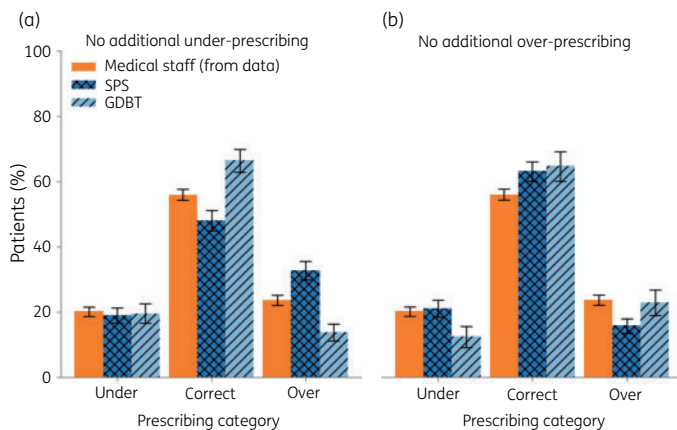


Figure 1. Proportion of patients prescribed antibiotic agents to which their isolate was resistant ('under-prescribed'), was appropriate ('correct') or given unnecessary broad-spectrum agents ('over-prescribed') by medical staff compared with the predictions of the SPS and GBDT algorithms. Tool performance was adjusted such that there was no additional (a) under-prescribing or (b) over-prescribing compared with medical staff. Error bars are 95% Wald CIs. This figure appears in colour in the online version of *JAC* and black and white in the print version of *JAC*.

amoxiclav and 10829 piperacillin/tazobactam resistance tests used in training, 2638 and 2698, respectively, were blood cultures and the remainder urine. The most important factor for predicting co-amoxiclav and piperacillin/tazobactam resistance was the presence of a similar previous culture result. Other important positive predictors for resistance included a diagnosis of diabetes or cancer (for piperacillin/tazobactam resistance) or a diagnosis of sepsis (for co-amoxiclav resistance).

Performance of the tools

The AUC of the ROC for the SPS tool ranged from 0.61 to 0.67 (Figure S2), with the GBDT performing marginally better with an AUC of 0.70 for both piperacillin/tazobactam and co-amoxiclav ($P < 0.001$). GBDT performance was best for urine cultures, with an AUC of 0.71 for co-amoxiclav and 0.70 for piperacillin/tazobactam compared with 0.66 and 0.67, respectively, for blood cultures ($P < 0.001$).

There were 3160 admissions where patients were initially prescribed co-amoxiclav, piperacillin/tazobactam or a carbapenem. Medical staff under-prescribed agents later shown to be ineffective on 20.3% (95% CI 19.1%–21.5%) of occasions, chose correctly on 56.0% (95% CI 54.5%–57.5%) and over-prescribed unnecessary broad-spectrum agents on 23.7% (95% CI 22.4%–25.0%). Figure 1(a) illustrates the level of over-prescribing (false positives) from the tools when adjusted to produce a similar level of under-prescribing (false negatives, 20.3%) to that of medical staff. The SPS tool increased over-prescribing compared with medical staff. In contrast the GBDT would have decreased over-prescribing from 23.7% to 13.4%, a 41.4% reduction ($P < 0.001$), recommending appropriate antibiotics for 66.5% of admissions while maintaining under-prescribing at 20.3%. This would have reduced carbapenem prescribing by 64.6% at the expense of a 16.0% and 7.1% increased use for piperacillin/tazobactam and co-amoxiclav, respectively (Figure S3). For GBDT carbapenem and piperacillin/tazobactam use, 17.2% and 37.2%, respectively, would be considered

appropriate ('correct') compared with 12.2% and 15.1%, respectively, for medical staff prescribing. Adjusting the tools to produce a similar level of over-prescribing (false positives) to that of medical staff (23.7%)—as might be the case if the aim was primarily reducing the number of patients given insufficiently broad empirical treatment on admission—resulted in under-prescribing rates of 12.4% for the GBDT and 19.1% for the SPS tool (Figure 1b).

Discussion

The combination of longitudinal clinical data and machine-learning systems has tremendous potential for improving the antibiotic management of those with infection. Open-source methods can be readily applied to locally held data creating algorithms suitable for each institution and population.^{6–8} Our proof-of-concept study demonstrates that a trained open-source machine-learning algorithm can predict antibiotic resistance among patients admitted to hospital and found to have a Gram-negative infection in blood or urine. Accuracy was greater than that of medical staff and a simpler five-factor risk assessment tool failed to achieve this improvement.

The hospital organization in which this study was conducted has observed increasing use of broad-spectrum antibiotic agents despite static rates of resistance to co-amoxiclav (E. Moran, E. Robinson, C. Green, M. Keeling and B. Collyer, unpublished data). There is a clear role for a point-of-care decision support system that would direct prescribers to a real-time individualized treatment recommendation based upon provisional diagnosis and patient risk factors.

This study has a number of limitations. Training the algorithm necessitated selecting patients with infection known to be caused by one of three Gram-negative organisms. Whilst evidence from those with septic shock suggests that culture-negative and culture-positive patients have similar outcomes, performance is likely to vary in unselected cohorts.⁹ We assessed the performance of each system against 'real-life' prescribing decisions made by medical staff for the 20% of the cohort given co-amoxiclav, piperacillin/tazobactam or a carbapenem. The remaining 80% of patients were managed with different agents or combination treatment. Ultimately performance would need to be assessed prospectively with well-defined endpoints among patients with unconfirmed infection receiving a larger number of empirical choices.¹⁰ It is also important to understand the impact on patients of under- and over-prescribing decisions. Host and disease factors may be the primary determinants of morbidity and mortality in some cohorts and it should not be assumed that initial empirical under-prescribing inevitably leads to an avoidable adverse outcome.¹¹

In conclusion, this study indicates that the use of machine learning could lead to a reduction in the inappropriate use of broad-spectrum antibiotic agents within the hospital setting. Improving antibiotic stewardship in this way may lead to reduced costs, complications and improved clinical outcomes.

Acknowledgements

The authors wish to acknowledge the useful comments provided by Joht Chandan, Noel McCarthy and Massimo Cavallaro.

Funding

This work was supported by the Engineering and Physical Sciences Research Council's (EPSRC) 'Bridging the Gaps' initiative and the University of Warwick INTEGRATE AMR initiative (grant reference EP/M027503/1) and by Health Data Research UK, which is funded by the UK Medical Research Council, EPSRC, Economic and Social Research Council, Department of Health and Social Care (England), Chief Scientist Office of the Scottish Government Health and Social Care Directorates, Health and Social Care Research and Development Division (Welsh Government), Public Health Agency (Northern Ireland), British Heart Foundation and the Wellcome Trust.

Transparency declarations

None to declare.

Author contributions

E.M., E.R. and M.K. initially conceived the project. C.G. advised on project development and data collection. B.C. developed the project and performed all data analysis and algorithm development.

Supplementary data

[Supplementary Information](#), Tables [S1](#) and [S2](#) and Figures [S1](#) to [S3](#) are available as [Supplementary data](#) at [JAC Online](#).

References

1 Vazquez-Guillamet MC, Vazquez R, Micek ST et al. Predicting resistance to piperacillin-tazobactam, cefepime and meropenem in septic patients

with bloodstream infection due to Gram-negative bacteria. *Clin Infect Dis* 2017; **65**: 1607–14.

2 Chow AL, Lye DC, Arah OA. Mortality benefits of antibiotic computerised decision support system: modifying effects of age. *Sci Rep* 2015; **5**: 17346.

3 Curtis CE, Al Bahar F, Marriott JF. The effectiveness of computerised decision support on antibiotic use in hospitals: a systematic review. *PLoS One* 2017; **12**: e0183062.

4 Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. 2018. <http://dmml.cs.washington.edu/xgboost.html>.

5 Weiss E, Zahar JR, Lesprit P et al. Elaboration of a consensual definition of de-escalation allowing a ranking of β -lactams. *Clin Microbiol Infect* 2015; **21**: 649.e1–10.

6 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. *Wellcome Open Res* 2018; **3**: 131.

7 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**: 1238–45.

8 Yelin I, Snitser O, Novich G et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. *Nat Med* 2019; **25**: 1143–52.

9 Kethireddy S, Bilgili B, Sees A et al. Culture-negative septic shock compared with culture-positive septic shock: a retrospective cohort study. *Crit Care Med* 2018; **46**: 506–12.

10 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**: 524–32.

11 Fitzpatrick JM, Biswas JS, Edgeworth JD et al. Gram-negative bacteraemia; a multi-centre prospective evaluation of empiric antibiotic therapy and outcome in English acute hospitals. *Clin Microbiol Infect* 2016; **22**: 244–51.