Development of a Machine-learning Model to Predict Antibiotic Resistance using Urine Culture and Antibiotic Susceptibility Data

Chungsoo Kim[1], Sandy Jeong Rhie[2, *], Rae Woong Park[1,3, *]

[1]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Republic of Korea;
[2]College of Pharmacy, Ewha Womans University, Suwon, Republic of Korea;
[3]Department of Biomedical Informatics, Ajou University School of Medicine, Suwon, Republic of Korea

[*]Co-corresponding authors

Background

Increasing antimicrobial resistance and the emergence of superbug are problems globally. Inappropriate empiric antibiotic use is one of the reasons cause antibiotic resistance.[1] It has been recommended that the local antibiogram be considered.[2, 3] In addition, patients' medical history and previous drug exposure should be beneficial to be considered. However, it has been a challenge to prescribe empiric antibiotics, as it is difficult to identify the causative organism with sensitivity and other factors beforehand.

In this study, we aimed to develop a prediction model to estimate the risk of antibiotics resistance using urine culture tests. We hypothesized that it would predict antibiotic resistance regardless of the causative infection strain in hospitalized patients suspected urinary tract infections while waiting for urine culture results.

Methods

We used a single EHR database of Ajou University School of Medicine (AUSOM) formatted an observational medical outcome partnership – common data model (OMOP–CDM) between 1994 and 2020.

The study population included adult patients who had at least one of the results from urine culture tests and antibiotic susceptibility tests (from ampicillin, ceftriaxone, ciprofloxacin, gentamicin, levofloxacin, nitrofurantoin, tetracycline, trimethoprim/sulfamethoxazole) on admission at Ajou University Medical Center. Outcomes were binarized as a resistant (including intermediate susceptibility) or susceptible for each antimicrobial agent regardless bacteria strain.

Candidate predictors were records of diagnosis, prescription, visit, laboratory (especially urine culture data and antibiotic susceptibility data), procedure, and urine antibiogram data. The urine antibiogram was calculated as resistant (R) or intermediate (I) susceptibility per total number of tests using susceptibility data between one year before admission date. The target bacteria for the antibiogram included *Escherichia coli (E. coli), Pseudomonas aeruginosa* and *Klebsiella pneumoniae*. We selected final covariates with the lasso shrinkage method and clinical guidelines. We split data to 75:25 for training and test and performed 3-fold cross-validation. Lasso logistic regression (LLR), extreme gradient boosting machine (XGB), Random Forest (RF) were used as model algorithms. Hyperparameters were selected with a grid search in each algorithm.

The Performance of a prediction model was evaluated as an aspect of discrimination and calibration. The area under the curve of receiver operator characteristics curve (AUROC) and precision-recall curve (AUPRC) were estimated for model discrimination. Calibration was measured with a calibration slope and calibration plot for comparing mean predicted and observed risks.

Model development and statistical analyses were conducted using R 4.0.2 and open-source HADES

packages of Observational Health Data Sciences and Informatics (OHDSI). All source codes are available in the GitHub repository: https://github.com/ABMI/AbxBetterChoice.

Results

The number of total cases and resistant cases in each susceptibility test were shown in Table 1. The antibiotics with the highest number of susceptibility tests was tetracycline (n = 20,219). However, resistant cases were the most common in the ampicillin (68.5%). The mean overall resistant rate was 37.5%.

Table 1 Model performance of antibiotic resistance prediction models with three different machine-learning algorithms

	Antibiotics susceptibility tests							
	AMP	**CRO**	**CIP**	**GEN**	**LVX**	**NIT**	**TET**	**TMP /SMX**
Total cases (n)	14,455	3,132	16,650	3,132	2,028	5,918	20,219	14,419
Resistant cases (n, %)	9,896 (68.5)	542 (17.3)	8,309 (49.9)	542 (17.3)	1,096 (54.0)	1,464 (24.7)	5,432 (26.9)	6,019 (41.7)
AUROC								
LLR	0.66 (0.64-0.68)	0.83 (0.79-0.87)	0.74 (0.72-0.75)	0.84 (0.80-0.88)	0.77 (0.73-0.81)	0.71 (0.68-0.74)	0.72 (0.70-0.74)	0.63 (0.61-0.64)
RF	0.62 (0.60-0.64)	0.83 (0.79-0.87)	0.71 (0.69-0.72)	0.81 (0.77-0.86)	0.80 (0.76-0.84)	0.71 (0.68-0.74)	0.72 (0.70-0.73)	0.62 (0.60 – 0.63)
XGB	0.65 (0.63-0.67)	0.80 (0.76-0.84)	0.72 (0.70-0.73)	0.83 (0.80-0.87)	0.77 (0.73-0.81)	0.73 (0.70-0.76)	0.74 (0.73-0.76)	0.61 (0.59-0.63)
AUPRC								
LLR	0.81	0.59	0.73	0.61	0.81	0.49	0.47	0.56
RF	0.78	0.60	0.70	0.58	0.85	0.48	0.46	0.55
XGB	0.81	0.55	0.72	0.58	0.82	0.50	0.50	0.55
Calibration slope								
LLR	1.06	1.01	1.04	1.07	0.82	0.97	1.00	1.05
RF	0.97	1.07	1.06	0.98	1.38	1.04	0.98	0.98
XGB	1.03	0.97	0.96	1.01	1.95	1.01	1.10	1.54

AMP: ampicillin, CRO: ceftriaxone, CIP: ciprofloxacin; GEN: gentamicin; LVX: levofloxacin; NIT: nitrofurantoin; TET: tetracycline; TMP/SMX: trimethoprim-sulfamethoxazole; AUROC: area under the receiver operating characteristics curve; AUPRC: area under the precision and recall curve; LLR: lasso logistic regression; RF: random forest; XGB: extreme gradient boosting machine

Total 44 covariates were selected for model development. They were ten types as following: demographic factors, urinary tract related conditions (e.g., prostatitis), infectious diseases (e.g., urinary tract infection), other chronic diseases (e.g., hypertension), antibiotic use, corticosteroid use, prior exposure to multidrug-resistant organisms (e.g., vancomycin-resistance enterococci), invasive procedures (e.g., continuous intravenous injection), prior visits, and urine antibiogram (e.g., resistance rate of E. coli for ciprofloxacin). The change in the resistance rate for each susceptibility test of E. coli is shown in Figure 1. The E. coli showed the highest mean resistance rate against ampicillin (74.9%). The

resistance rate of E. coli to ciprofloxacin gradually increased by mean of 1% per year (coefficient = 0.003 %/day in generalized least squared model). The result of gentamicin and nitrofurantoin tests were censored due to test panel change.

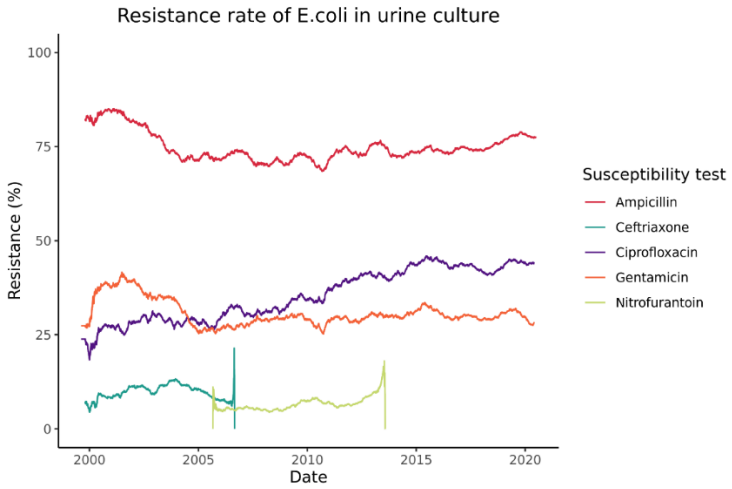

Figure 1 Resistance rate of E. coli in antibiotic susceptibility test from urine sample

The discriminative and calibration performance for predicting a resistance of specific antibiotic of models were presented in Table 1. The LLR algorithm showed the highest AUROC in the ampicillin, ceftriaxone, ciprofloxacin, gentamicin, trimethoprim/sulfamethoxazole models (0.66, 0.83, 0.74, 0.84, 0.63, respectively). The RF presented the highest AUROC for the levofloxacin model (0.80). The AUROC of XGB was highest in the nitrofurantoin and tetracycline model (0.73, 0.74, respectively). The AUROC was the highest in the gentamicin model among LLR and XGB models with different antibiotics (0.84, 0.83, respectively) except RF models (ciprofloxacin model, AUROC; 0.83). The AUPRC was highest in four LLR models (ampicillin, 0.81; ciprofloxacin, 0.73; gentamicin, 0.61; trimethoprim/sulfamethoxazole, 0.56). Calibration performance, shown as calibration slope and calibration plot, was acceptable in all prediction models (Table 1, Figure 2).

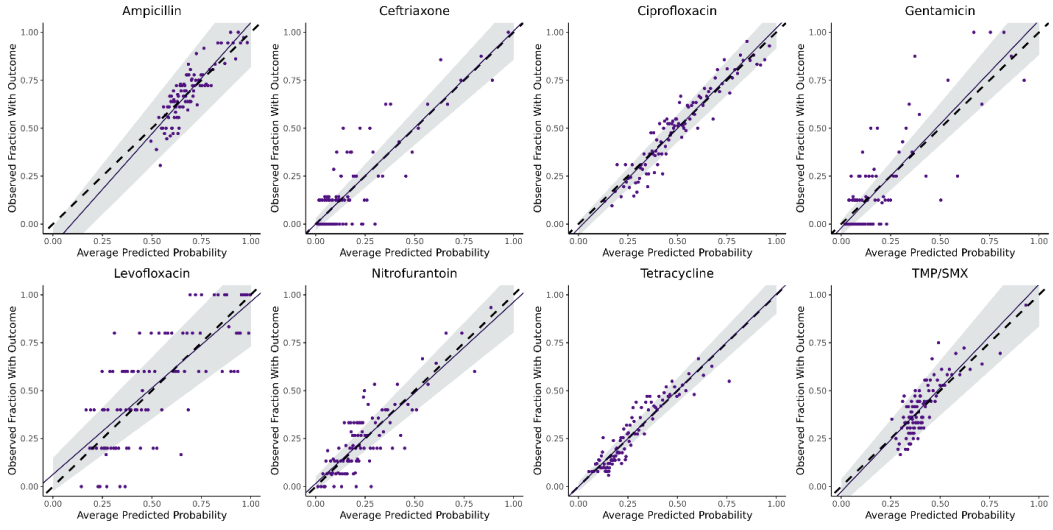

Figure 2 Calibration plots of antibiotic resistance prediction models which used lasso logistic regression

Conclusion

We developed machine-learning models for predicting antibiotic resistance with promising discrimination and calibration performances. It would contribute to the proper selection of empiric antibiotics susceptible to those causative pathogens in hospitalized patients with clinically suspected urinary tract infections. In addition, research on the development of a more delicate antibiotic resistance prediction model considering antibiogram by patient's visit type and length of stay is needed.

References

1. Ventola, C. L. (2015). The antibiotic resistance crisis: part 1: causes and threats. *P & T : a peer-reviewed journal for formulary management, 40*(4), 277-283.
2. Strich, J. R., Heil, E. L., & Masur, H. (2020). Considerations for Empiric Antimicrobial Therapy in Sepsis and Septic Shock in an Era of Antimicrobial Resistance. *The Journal of Infectious Diseases, 222*(Supplement_2), S119-S131. doi:10.1093/infdis/jiaa221
3. Barbieri, E., Bottigliengo, D., Tellini, M., Minotti, C., Marchiori, M., Cavicchioli, P., . . . Donà, D. (2021). Development of a Weighted-Incidence Syndromic Combination Antibiogram (WISCA) to guide the choice of the empiric antibiotic treatment for urinary tract infection in paediatric patients: a Bayesian approach. *Antimicrobial Resistance & Infection Control, 10*(1), 74. doi:10.1186/s13756-021-00939-2