The Rapid Prediction of Carbapenem Resistance in Patients With *Klebsiella pneumoniae* Bacteremia Using Electronic Medical Record Data

Timothy Sullivan, Osamu Ichikawa, Joel Dudley, Li Li, and Judith Aberg

Division of Infectious Diseases and Department of Genetics and Genomic Sciences, Institute of Next Generation Healthcare, Icahn School of Medicine at Mount Sinai, New York, New York

**Background.** The administration of active antibiotics is often delayed in cases of carbapenem-resistant gram-negative bacteremia. Using electronic medical record (EMR) data to rapidly predict carbapenem resistance in patients with *Klebsiella pneumoniae* bacteremia could help reduce the time to active therapy.

**Methods.** All cases of *Klebsiella pneumoniae* bacteremia at Mount Sinai Hospital from September 2012 through September 2016 were included. Cases were randomly divided into a "training set" and a "testing set." EMR data from the training set cases were reviewed, and significant risk factors for carbapenem resistance were entered into a multiple logistic regression model. Performance was assessed by repeated K-fold cross-validation and by applying the training set model to the testing set. All cases were also reviewed to determine the time to effective antibiotic therapy.

**Results.** A total of 613 cases of *Klebsiella pneumoniae* bacteremia were included, 61 (10%) of which were carbapenem-resistant. The training and testing sets consisted of 460 and 153 cases, respectively. The regression model derived from the training set correctly predicted 73% of carbapenem-resistant cases and 59% of carbapenem-susceptible cases in the testing set (sensitivity, 73%; specificity, 59%; positive predictive value, 16%; negative predictive value, 95%). The mean time between the expected and observed carbapenem-resistant cases was 0.731. Patients with carbapenem-resistant infections received active antibiotics significantly later than those with susceptible infections (40.4 hours vs 9.6 hours, \( P < .0001 \)).

**Conclusions.** A multiple logistic regression model using EMR data can generate rapid, sensitive predictions of carbapenem resistance in patients with *Klebsiella pneumoniae* bacteremia, which could help shorten the time to effective therapy in these cases.

**Keywords.** carbapenem; Klebsiella; prediction; resistance.

The early initiation of effective antibiotic therapy is a critical step in the successful treatment of gram-negative bacteremia. The use of inadequate initial antibiotics has been associated with an increased risk of death [1, 2]; therefore, patients with gram-negative bacteremia typically receive broad-spectrum antibiotics until susceptibilities are available. However, empiric antibiotic regimens may still be ineffective against carbapenem-resistant *Enterobacteriaceae* (CRE), and as a result, active treatment for CRE bacteremia is often delayed [3-5]. This delay may contribute to the high mortality associated with CRE infections [6].

Although treating CRE bacteremia with inappropriate antibiotics is associated with poor outcomes, the overuse of antibiotics active against CRE may also be harmful. Given the relatively low prevalence of CRE infections [7], routinely administering CRE therapy to all patients with gram-negative bacteremia may cause unnecessary toxicity and promote widespread resistance to critical last-line antibiotics [8, 9].

Identifying the bacterial species growing in a blood culture can help guide early antibiotic choices, and rapid species identification is now possible in many laboratories [10]. However, antibiotic susceptibility results may not be available for several days. Therefore, a method is needed to predict carbapenem resistance in patients with gram-negative bacteremia after the species is identified but before more definitive susceptibility results are available. Such a prediction tool would be especially useful when treating *Klebsiella pneumoniae* bacteremia, as this is the species of *Enterobacteriaceae* most likely to be carbapenem-resistant [3, 7].

We have developed a multiple logistic regression model using electronic medical record (EMR) data to generate immediate predictions of carbapenem resistance in patients with *Klebsiella pneumoniae* bacteremia.

**METHODS**

All cases of *Klebsiella pneumoniae* bacteremia at Mount Sinai Hospital from September 2012 through September 2016 were...
testing sets to evaluate for any introduced bias, and no significant differences were found.

Possible risk factors for carbapenem resistance were queried from the training set and analyzed via univariate methods, such as the Fisher exact test to compare categorical variables and the $t$ test to compare means. Potential independent variables were considered for inclusion in the regression model only if they (1) were biologically plausible and (2) existed within the EMR as objective data fields. Free text fields and data requiring a subjective opinion or manual input were not considered.

The hospital unit where each patient was located at the time of culture collection was classified as either high risk for carbapenem resistance, low risk, or the emergency department (ED). High-risk units were defined as intensive care units (ICUs) and acute medical and surgical floors. Acute floors and ICUs were grouped together because CRE rates in these units were similar. Low-risk units were all other hospital units excluding the ED.

Antibiotic exposures were quantified as “antibiotic-days.” One antibiotic-day was defined as each day that a distinct antibiotic was given. Patients receiving more than 1 distinct antibiotic on a given day could accrue more than 1 antibiotic-day per calendar day.

Colonization with imipenem-resistant *Klebsiella pneumoniae* was defined as growth of this organism in any prior culture, including blood, urine, respiratory, and wound cultures. Rectal swabs for multidrug-resistant gram-negative rods were collected sporadically during the study period, and these were also considered; however, all patients with positive surveillance cultures also had at least 1 other prior clinical specimen with growth of imipenem-resistant *Klebsiella pneumoniae*.

Factors that were significantly associated with imipenem resistance on univariate analysis ($P < .2$) were entered into a multiple logistic regression model in a forward stepwise approach using SPSS and R. The accuracy of potential models was assessed by comparing the area under the receiver operator characteristic curve (AUROC) of each model. Variables were
retained in the model if they resulted in a higher AUROC even if they were not associated with a P value of <.05. The final ROC curve for the model was also used to select a threshold for positivity of the regression output that emphasized sensitivity of the algorithm over specificity.

The logistic regression model created from the training set was applied to the testing set using the threshold for positivity defined above. A confusion matrix was generated from the classification of the testing set data, and performance characteristics were calculated.

Repeated K-fold Cross-Validation

Repeated K-fold cross-validation was performed on the training set using the same variables included in the model. The training set cases were divided into 10 equal folds, with 9 folds used to generate a logistic regression algorithm and the 10th fold used to test the predictive accuracy of the algorithm. This process was repeated for each of the 10 folds, and then the entire 10-fold cross-validation was repeated 500 times. An AUROC for each of the 5000 repeats was reported.

To compare the results of the cross-validation with random chance, a “random model” was created that consisted of random permutations of the dependent variable used in the logistic regression equation. This process was repeated 5000 times, and an AUROC was reported for each of these repeats. The mean AUROCs of the cross-validation model and random model were then compared.

RESULTS

A total of 613 cases of Klebsiella pneumoniae bacteremia were identified in 540 unique patients. The rate of imipenem resistance was 10% (61 cases in 53 unique patients). The training and testing sets consisted of 460 and 153 cases, respectively. Baseline demographic information and potential risk factors for imipenem resistance from the 460 training set cases are presented in Table 1.

Risk factors for imipenem resistance that were included in the logistic regression model were (1) prior colonization with imipenem-resistant Klebsiella pneumoniae, (2) hospital unit, (3) total inpatient days in the previous 5 years, (4) total days of oral or parenteral antibiotics in the past 2 years, and (5) age >60 years. The logistic regression coefficients and P values are listed in Table 2.

The logistic regression model created from the training set generated an ROC curve with an AUROC of 0.755. When applied to the testing set using a threshold of 0.08, the model correctly predicted 73% of the imipenem-resistant cases while incorrectly labeling 41% of the susceptible cases as resistant (sensitivity, 73%; specificity, 59%; positive predictive value [PPV], 16%; negative predictive value, 95%) (Table 3). The AUROC of the model when applied to the testing set was 0.721.

Ten-fold cross-validation was performed on the training set data and repeated 500 times. The mean AUROC of the 5000 cross-validation repeats was 0.731, whereas the mean AUROC of the random model repeats was 0.488 (difference in means, 0.243; P < .001) (Figure 2).

Time to Active Therapy

All 613 cases were reviewed to determine the time to active antibiotics. Of patients with imipenem-resistant Klebsiella pneumoniae bacteremia, 33% never received active antibiotics, compared with 3% of those with imipenem-susceptible infections (P < .0001). Among patients who did receive active antibiotics, the mean time to effective therapy was significantly greater for patients with imipenem-resistant infections (40.4 hours vs 9.6 hours, P < .0001). Only 44% of imipenem-resistant cases were appropriately treated within the first day, compared with 84% of susceptible infections (P < .0001).

DISCUSSION

This study demonstrates that a multiple logistic regression model using EMR data can generate immediate, sensitive predictions of carbapenem resistance in patients with Klebsiella pneumoniae bacteremia. Of patients who received appropriate antibiotics, the mean time to active therapy was more than 30 hours greater in patients with carbapenem-resistant infections, which underscores the need for rapid identification of carbapenem resistance in these cases.
Several new diagnostic tests may significantly shorten the time required to identify CRE infections [11–16]. However, some novel CRE diagnostics still require several hours after culture positivity for a result to return, whereas others have a faster turnaround time but only test for a limited number of resistance genes, which may not always correlate with a resistant phenotype [17]. In patients with septic shock, each hour without active antibiotic therapy has been associated with an increased risk of death [18]. Therefore, even when new molecular methods of CRE detection gain more widespread use, an algorithm that can immediately predict carbapenem resistance could still complement genotypic testing and impact patient outcomes.

In practice, clinicians may attempt to estimate a bacteremic patient’s risk for CRE infection by considering known risk factors, which include prior CRE colonization, previous antibiotic exposures, critical illness, renal failure, advanced age, recent surgery, hospital and ICU length of stay, and the presence of a central venous catheter [19–26].

Given the complexity of considering all these variables for each patient, prior studies have attempted to create clinical prediction models for gram-negative resistance. Logistic regression [24, 25, 27–31] and clinical decision trees [32, 33] have previously been used to predict patients at high risk for resistant gram-negative infections. However, some published models have been limited by the inclusion of few CRE cases, a focus on predicting colonization as opposed to infection, or a lack of robust validation methods. Others have incorporated variables that require manual data entry, such as travel history, or a subjective opinion, such as the suspected source of infection, which may impair the speed and reproducibility of predictions.

A major advantage of our model is that it includes only variables that exist as objective fields in the EMR. This approach could allow for the rapid generation of automated predictions of carbapenem resistance as soon as Klebsiella pneumoniae is identified in a blood culture, without the need for manual input or interpretation. Another benefit of this model is that it generates an automated review of prior microbiology data for each patient. Although clinicians would likely choose empiric antibiotics active against CRE for a bacteremic patient with a history of CRE infection, in practice, prior cultures may sometimes be overlooked. By automatically incorporating all prior culture data, our algorithm could help prevent this type of oversight.

### Table 2. Multiple Logistic Regression Model Variables and Coefficients

| Predicted Resistance | B  | OR  | SE  | Wald | df | P Value |
|----------------------|----|-----|-----|------|----|---------|
| Imi-R Klebsiella colonization | 2.873 | 17697 | 0.708 | 16.491 | 1 | <.001 |
| Location | 7.689 | 2 | .021 |
| Location (high risk) | 0.730 | 2.075 | 0.381 | 3.671 | 1 | .055 |
| Location (low risk) | −1.002 | 0.367 | 0.784 | 1.636 | 1 | .201 |
| Age >60 | 0.771 | 2.161 | 0.367 | 4.414 | 1 | .036 |
| ABX-days in past 2 y | 0.009 | 1.009 | 0.003 | 6.897 | 1 | .009 |
| Inpatient days in past 5 y | −0.004 | 0.996 | 0.002 | 3.544 | 1 | .060 |
| Constant | −3.144 | 0.043 | 0.430 | 53.543 | 1 | <.001 |

Abbreviations: ABX-days, days of antibiotics; B, regression coefficient; df, degrees of freedom; Imi-R, imipenem-resistant; OR, odds ratio.

*Emergency department was considered the reference location.

### Table 3. Performance of the Model on the Testing Set Data

| Predicted Resistance | Imi-R | Imi-S | % Correct |
|----------------------|-------|-------|-----------|
| Observed resistance  | Imi-R | 11    | 73.3      |
|                      | Imi-S | 56    | 59.4      |
| Overall % correct    |       |       | 60.8      |

Abbreviations: Imi-R, imipenem-resistant; Imi-S, imipenem-susceptible.

Figure 2. Histogram showing the area under the receiver operating characteristic curve for each K-fold cross-validation repeat and each random model repeat. Abbreviations: AUROC, area under the receiver operating characteristic curve.
This model could potentially be integrated directly into the EMR to provide real-time clinical predictions. Another strength of our model is that it is supported by 2 different validation methods. The K-fold cross-validation and the performance of the model when applied to the testing set both support its out-of-sample validity. Additionally, the significantly higher mean AUROC of the cross-validation repeats, when compared with the random model, indicates that our findings are not due to chance or sampling error.

The major limitation of this model is its relatively low PPV. The low PPV is likely driven by the low prevalence of carbapenem resistance in the study population, as well as our decision to choose a threshold for positivity that emphasizes sensitivity over specificity. Prospectively incorporating more data into the model in future years could potentially improve the specificity and PPV.

The single-center nature of this study is another drawback. Data were collected from only 1 institution, but patients may have been admitted to other hospitals and been exposed to other risks factors for CRE that were not included in the model.

Another limitation is that this model is not directly generalizable to other institutions. The inclusion of hospital unit as a variable precludes comparisons with other hospitals, and because the prevalence of CRE varies drastically between institutions and geographic locations, the model developed at our tertiary medical center in New York City may not be applicable to hospitals in other locations. Specifically, the PPV of the model might be lower if used in a population where CRE infections are less prevalent.

This study shows that a logistic regression model using EMR data can generate rapid, sensitive predictions of carbapenem resistance in patients with *Klebsiella pneumoniae* bacteremia. Because effective treatment is often delayed in cases of CRE bacteremia, and this delay may contribute to the high mortality associated with these infections, our model could potentially help improve outcomes by quickly identifying patients at risk for CRE infection. Further research is needed to determine how this model could best be applied to patient care.

**Acknowledgments**

Financial support. This work was supported by the New York State Department of Health Empire Clinical Research Investigator Program.

**References**

1. Gikas A, Samonis G, Christidou A, et al. Gram-negative bacteremia in non-neutropenic patients: a 3-year review. *Infection* 1998; 26:155–9.
2. Kang CI, Kim SH, Park WB, et al. Bloodstream infections caused by antibiotic-resistant gram-negative bacilli: risk factors for mortality and impact of inappropriate initial antimicrobial therapy on outcome. *Antimicrob Agents Chemother* 2005; 49:760–6.
3. Satlin MJ, Chen L, Patel G, et al. Bacteremia due to carbapenem-resistant *Enterobacteriaceae* (CRE): a multicenter clinical and molecular epidemiologic analysis in the nation’s epicenter for CRE. *Antimicrob Agents Chemother* 2017; 61(4).
4. Tumbarello M, Viale P, Viscoli C, et al. Predictors of mortality in bloodstream infections caused by *Klebsiella pneumoniae* carbapenemase-producing *K. pneumoniae*: importance of combination therapy. *Clin Infect Dis* 2012; 55:943–50.
5. Zarkotou O, Pouraras S, Tsolioti P, et al. Predictors of mortality in patients with bloodstream infections caused by KPC-producing *Klebsiella pneumoniae* and impact of appropriate antimicrobial treatment. *Clin Microbiol Infect* 2011; 17:798–803.
6. Lodise T, Berger A, Altincatal A, et al. Carbapenem-resistant *Enterobacteriaceae* (CRE) or delayed appropriate therapy (DAT)—does one affect outcomes more than the other among patients with serious infections due to *Enterobacteriaceae*? *Open Forum Infect Dis* 2017; 4(suppl 1): S14.
7. Guh AY, Bulens SN, Mu Y, et al. Epidemiology of carbapenem-resistant *Enterobacteriaceae* in 7 US communities, 2012–2013. *JAMA* 2015; 314:1479–87.
8. Ong DSY, Frencken JF, Klein Klouwenberg PMC, et al; MARS Consortium. Short-course adjunctive gentamicin as empirical therapy in patients with severe sepsis and septic shock: a prospective observational cohort study. *Clin Infect Dis* 2017; 64:1731–6.
9. van Duin D, Bonomo RA. Cefazidime/avibactam and ceftolozane/tazobactam: second-generation β-Lactam/β-Lactamase inhibitor combinations. *Clin Infect Dis* 2016; 63:234–41.
10. Patel R. Matrix-assisted laser desorption ionization-time of flight mass spectrometry in clinical microbiology. *Clin Infect Dis* 2013; 57:564–72.
11. Ledeboer NA, Lopansri BK, Dhiman N, et al. Identification of Gram-negative bacteria and genetic resistance determinants from positive blood culture broths by use of the verigene Gram-negative blood culture multiplex real-time molecular assay. *J Clin Microbiol* 2015; 53:2460–72.
12. Tato M, Ruiz-Garbajosa P, Tracienowski M, et al. Multisite evaluation of cepheid Xpert Carba R assay for detection of carbapenemase-producing organisms in rectal swabs. *J Clin Microbiol* 2016; 54:1814–9.
13. Lange C, Schubert S, Jung I, et al. Quantitative matrix-assisted laser desorption ionization-time of flight mass spectrometry for rapid resistance detection. *J Clin Microbiol* 2014; 52:4155–62.
14. Tijsen N, Boyd D, Patel SN, et al. Evaluation of the Carba NP test for rapid detection of carbapenemase-producing *Enterobacteriaceae* and *Pseudomonas aeruginosa*. *Antimicrob Agents Chemother* 2013; 57:4578–80.
15. Albon O, Almuhayawi M, Ullberg M, Ozenci V. Clinical evaluation of the FilmArray blood culture identification panel in identification of bacteria and yeasts from positive blood culture bottles. *J Clin Microbiol* 2013; 51:4130–6.
16. Marschal M, Bachmaier J, Autenrieth I, et al. Evaluation of the accelerated pheno system for fast identification and antimicrobial susceptibility testing from positive blood cultures in bloodstream infections caused by Gram-negative pathogens. *J Clin Microbiol* 2017; 55:2116–26.
17. Banerjee R, Humphries R. Clinical and laboratory considerations for the rapid detection of carbapenem-resistant *Enterobacteriaceae*. *Virulence* 2017; 8:427–39.
18. Kumar A, Roberts D, Wood KE, et al. Duration of hypotension before initiation of effective antimicrobial therapy as the critical determinant of survival in human septic shock. *Crit Care Med* 2006; 34:1589–96.
19. Falagas ME, Rafaillos PI, Kofteridis D, et al. Risk factors of carbapenem-resistant *Klebsiella pneumoniae* infections: a matched case-control study. *J Antimicrob Chemother* 2007; 60:1124–30.
20. Schwaber MJ, Klarfeld-Lidji S, Navon-Venezia S, et al. Predictors of carbapenem-resistant *Klebsiella pneumoniae* acquisition among hospitalized adults and effect of acquisition on mortality. *Antimicrob Agents Chemother* 2008; 52:1028–31.
21. Kofteridis DP, Valachis A, Dimopoulos D, et al. Risk factors for carbapenem-resistant *Klebsiella pneumoniae* infection/colonization: a case-case-control study. *J Infect Chemother* 2014; 20:293–7.
22. Ling ML, Tee YM, Tan SG, et al. Risk factors for acquisition of carbapenem-resistant *Enterobacteriaceae* in an acute tertiary care hospital in Singapore. *Antimicrob Resist Infect Control* 2015; 4:62.
23. Gasink LB, Edelstein PH, Lautenbach E, et al. Risk factors and clinical impact of *Klebsiella pneumoniae* carbapenemase-producing *K. pneumoniae*. *Infect Control Hosp Epidemiol* 2009; 30:1180–5.
24. Tumbarello M, Treccarichi EM, Turmiotto F, et al. Predictive models for identification of hospitalized patients harboring KPC-producing *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* 2014; 58:3514–20.
25. Yang D, Xie Z, Xin X, et al. A model for predicting nosocomial carbapenem-resistant *Klebsiella pneumoniae* infection. *Biomed Rep* 2016; 5:501–5.
26. Tischendorf J, de Avila RA, Safdar N. Risk of infection following colonization with carbapenem-resistant *Enterobacteriaceae*: a systematic review. *Am J Infect Control* 2016; 44:539–43.
27. Lin MY, Rezny S, Ray MJ, Jovanov D, Weinstein RA, Trick WE. Predicting carbapenem-resistant *Enterobacteriaceae* carriage at the time of admission.
using a state-wide hospital discharge database. Open Forum Infect Dis 2016; 3(suppl 1): 348.
28. Chiang WC, Chen SY, Chien KL, et al. Predictive model of antimicrobial-resistant gram-negative bacteremia at the ED. Am J Emerg Med 2007; 25:597–607.
29. Vasudevan A, Mukhopadhyay A, Li J, et al. A prediction tool for nosocomial multi-drug resistant Gram-negative bacilli infections in critically ill patients—prospective observational study. BMC Infect Dis 2014; 14:615.
30. Tseng WP, Chen YC, Yang BJ, et al. Predicting multidrug-resistant Gram-negative bacterial colonization and associated infection on hospital admission. Infect Control Hosp Epidemiol 2017; 38:1216–25.
31. Dan S, Shah A, Justo JA, et al. Prediction of fluoroquinolone resistance in Gram-negative bacteria causing bloodstream infections. Antimicrob Agents Chemother 2016; 60:2265–72.
32. Vazquez-Guillamet MC, Vazquez R, Micek ST, Kollef MH. Predicting resistance to piperacillin-tazobactam, ceftime and meropenem in septic patients with bloodstream infection due to Gram-negative bacteria. Clin Infect Dis 2017; 65:1607–14.
33. Goodman KE, Lessler J, Cosgrove SE, et al; Antibacterial Resistance Leadership Group. A clinical decision tree to predict whether a bacteremic patient is infected with an extended-spectrum β-lactamase-producing organism. Clin Infect Dis 2016; 63:896–903.