Infection Control & Hospital Epidemiology (2019), **40**, 400–407
doi:10.1017/ice.2019.17

Original Article

A methodological comparison of risk scores versus decision trees for predicting drug-resistant infections: A case study using extended-spectrum beta-lactamase (ESBL) bacteremia

Katherine E. Goodman[1], Justin Lessler[1], Anthony D. Harris[2], Aaron M. Milstone[1,3] and Pranita D. Tamma[3]

[1]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, Maryland, [2]Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore, Maryland and [3]Division of Infectious Diseases, Department of Pediatrics, Johns Hopkins University School of Medicine, Baltimore, Maryland

Abstract

Background: Timely identification of multidrug-resistant gram-negative infections remains an epidemiological challenge. Statistical models for predicting drug resistance can offer utility where rapid diagnostics are unavailable or resource-impractical. Logistic regression–derived risk scores are common in the healthcare epidemiology literature. Machine learning–derived decision trees are an alternative approach for developing decision support tools. Our group previously reported on a decision tree for predicting ESBL bloodstream infections. Our objective in the current study was to develop a risk score from the same ESBL dataset to compare these 2 methods and to offer general guiding principles for using each approach.

Methods: Using a dataset of 1,288 patients with *Escherichia coli* or *Klebsiella* spp bacteremia, we generated a risk score to predict the likelihood that a bacteremic patient was infected with an ESBL-producer. We evaluated discrimination (original and cross-validated models) using receiver operating characteristic curves and C statistics. We compared risk score and decision tree performance, and we reviewed their practical and methodological attributes.

Results: In total, 194 patients (15%) were infected with ESBL-producing bacteremia. The clinical risk score included 14 variables, compared to the 5 decision-tree variables. The positive and negative predictive values of the risk score and decision tree were similar (>90%), but the C statistic of the risk score (0.87) was 10% higher.

Conclusions: A decision tree and risk score performed similarly for predicting ESBL infection. The decision tree was more user-friendly, with fewer variables for the end user, whereas the risk score offered higher discrimination and greater flexibility for adjusting sensitivity and specificity.

(Received 14 October 2018; accepted 13 January 2019)

Multidrug-resistant gram-negative (MDRGN) organisms represent a growing clinical threat. These bacteria can spread rapidly among vulnerable hospitalized populations, and MDRGN infections are associated with significant morbidity and mortality.[1,2] Timely identification can limit nosocomial transmission and improve patient outcomes by facilitating prompt initiation of appropriate treatment.[3,4] However, rapid diagnostics that can be readily incorporated into routine laboratory workflows are limited or lacking for many MDRGNs, posing clinical and epidemiological challenges Extended-spectrum β-lactamase (ESBL)–producing bacteria, which can hydrolyze most β-lactam antibiotics other than carbapenems, are a representative example of these MDRGNs.

Currently, no phenotypic method has been endorsed by the Clinical and Laboratory Standards Institute (CLSI) for ESBL detection.[5] Although molecular methods for identifying ESBL genes are commercially available, these assays do not include a comprehensive list of known ESBL genes and would require frequent panel updates to detect emerging ESBLs.[6,7] Molecular diagnostics can also be resource-intensive and are often not cost-effective for laboratories in regions where ESBL prevalence is low, and they are cost-prohibitive for developing areas of the world where ESBL prevalence is high.

Statistical models for identifying MDRGN infections can provide important information in settings where rapid diagnostics are unavailable or are resource-impractical. One particular approach, generating a logistic regression–derived risk score, is common in the healthcare epidemiology literature. However, classification and regression tree (CART) analysis or "recursive partitioning," a form of machine learning, is an alternative approach for developing this type of decision support tool. Our group previously developed a CART decision tree for predicting ESBL bloodstream infections.[8] Since publication, there has been interest in whether a risk score derived from the same population could achieve greater predictive accuracy while remaining sufficiently simple to incorporate into practice.

We performed a case study of the development of a risk score from the same ESBL dataset as our original decision tree to

Author for correspondence: Katherine E. Goodman, Email: kgoodma7@jhu.edu and Pranita D. Tamma, Email: ptamma1@jhmi.edu

Cite this article: Goodman KE, *et al.* (2019). A methodological comparison of risk scores versus decision trees for predicting drug-resistant infections: A case study using extended-spectrum beta-lactamase (ESBL) bacteremia. *Infection Control & Hospital Epidemiology*, 40: 400–407, https://doi.org/10.1017/ice.2019.17

Table 1. Regression Model and Corresponding Points Scoring System[a] for Predicting Extended-Spectrum β-Lactamase (ESBL) Status in a Cohort of Adult Patients with *Escherichia coli* and *Klebsiella* spp Bacteremia

Variable	β Coefficient	Odds Ratio (95% CI)	Points
Intercept	−3.81		. . .
Orthopedic hardware (day of culture)	1.30	3.68 (1.21–11.17)	2
Chronic indwelling vascular hardware (day of culture)	0.60	1.82 (1.13–2.94)	1
Nephrostomy tube or Foley catheter (day of culture)	1.17	3.22 (1.87–5.57)	2
Gastrointestinal feeding tube (day of culture)	0.97	2.65 (1.35–5.18)	2
Presumptive infection source: central venous catheter	0.98	2.68 (1.56–4.60)	2
Presumptive infection source: pneumonia	1.12	2.98 (1.37–6.49)	2
Structural lung disease[b]	1.15	3.17 (1.43–7.00)	2
Self-identifies as Asian race	1.07	2.93 (1.23–6.94)	2
Post-acute care facility stay (prior 6 mo)	1.04	2.84 (1.12–7.27)	2
≥1 night of hospitalization in an international ESBL high-burden region[c] (prior 6 mo)	3.21	24.86 (10.99–56.24)	5
ESBL colonization or infection (prior 6 mo)	3.92	50.68 (25.97–98.92)	6
Carbapenem-resistant *Enterobacteriaceae* colonization or infection (prior 6 mo)	3.45	31.47 (2.52–393.30)	6
Multidrug-resistant *Pseudomonas* spp (prior 6 mo)	−2.42	0.09 (0.01–0.83)	−4
Weeks of active gram-negative therapy (per week, up to a maximum of 4, in prior 6 mo)	0.15	1.17 (1.02–1.34)	0.25/week; max of 1 pt

[a]To create points, the smallest model coefficient (0.15, per week of antibiotic therapy) was identified. To simplify end-user calculations, antibiotic therapy was scaled to receive 0.25 points per week, up to a maximum of 1 point or ≥4 weeks, by dividing by 0.60 (0.15/0.60 = 0.25). All other coefficients were also divided by 0.60 and rounded to the nearest whole integer. Patient scores were calculated by summing their respective points (risk score model).
[b]Chronic obstructive pulmonary disease, emphysema, or chronic ventilator dependency.
[c]Latin America (excluding the Caribbean), the Middle East (including Egypt), South Asia, China, and the Mediterranean.

compare the predictive accuracy of these 2 methods and to illustrate the advantages and disadvantages of logistic regression risk scores versus CART decision trees. Our objective is to offer general guiding principles for epidemiologists and researchers for when they might consider one prediction approach versus the other.

Methods

Cohort

The full description of the cohort has been previously reported.[8] Briefly, the study included adults hospitalized at the Johns Hopkins Hospital with bacteremia due to *Escherichia coli* or *Klebsiella* spp, from 2008 to 2015. Only the first episode of bacteremia per patient was included. *Escherichia coli or Klebsiella* spp with ceftriaxone minimum inhibitory concentrations (MICs) ≥2 μg/mL underwent testing for ESBL production. A decrease of ≥3 doubling dilutions in the MIC for a third-generation cephalosporin tested in combination with 4 μg/mL of clavulanic acid, versus its MIC when tested alone, was used to confirm ESBL status.

Patient data were collected via manual chart review from all available inpatient and outpatient medical records from facilities within the Johns Hopkins Health System, as well as from medical records for patients who previously received medical care at institutions in the Epic Care Everywhere Network (www.epic.com/CareEverywhere/). Patient data collected, which was based on the time period prior to day 1 of bacteremia (defined as the date the initial blood culture was collected), included the following: (1) demographic data; (2) preexisting medical conditions; (3) presumptive source of bacteremia (eg, catheter, pneumonia); (4) indwelling hardware; (5) multidrug-resistant organism (MDRO) colonization or infection (MDR *Pseudomonas aeruginosa*, MDR *Acinetobacter baumannii*, ESBL-producing *Enterobacteriaceae*, carbapenem-resistant *Enterobacteriaceae*, vancomycin-resistant *Enterococcus* species, and methicillin-resistant *Staphylococcus aureus*)[9] in the prior 6 months; (6) days of antibiotic therapy with gram-negative activity in the prior 6 months; (7) length of stay in any healthcare facility in the prior 6 months; (8) post-acute care facility stay in the prior 6 months; and (9) hospitalization in another country in the prior 6 months (assessed by standard nursing intake questionnaire upon Johns Hopkins Hospital admission). International hospitalizations in the following regions were classified as ESBL "high-burden": Latin America (excluding the Caribbean), the Middle East (including Egypt), South Asia, China, and the Mediterranean.[10,11]

Statistical methods

Descriptive statistics, univariable analyses, and decision tree derivation and validation have been described previously.[8] Briefly, a tree was derived using the following process: (1) identification of the single variable that, when used to split the dataset into 2 groups ("nodes"), best separated ESBL-positive from ESBL-negative patients, according to the Gini impurity criterion[12,13]; (2) repetition of this partitioning process in each daughter node and subsequent generations of nodes ("branching"); and (3) termination at "terminal" nodes ("leaves") when no additional variables in the data sufficiently distinguished patients by their ESBL status. Terminal nodes in binary recursive partitioning trees predict ESBL status categorically, but by evaluating the node impurity (eg, the mixture of ESBL-positive and ESBL-negative patients), they also offer associated probabilities.

We internally validated the performance of our tree using the leave-one-out cross-validation method,[12] in which a single observation is held out and a new model is derived from a dataset containing the remaining $n - 1$ observations. The resulting model is

https://doi.org/10.1017/ice.2019.17 Published online by Cambridge University Press

Printable Clinical Risk Score for Predicting Extended-Spectrum β-Lactamase (ESBL) Status of Adult Patients with *Escherichia coli* and *Klebsiella* Species Bacteremia

RISK FACTORS	POINTS
Indwelling Hardware (Day of Culture):	
1. Orthopedic hardware (2)	_____
2. Central vascular catheter (1)	_____
3. Nephrostomy tube or Foley catheter (2)	_____
4. Gastrointestinal feeding tube (2)	_____
Presumptive Source of Bloodstream Infection:	
5. Catheter-related (2)	_____
6. Pneumonia (2)	_____
Patient Characteristics:	
7. Structural lung disease[a] (2)	_____
8. Self-identifies as Asian race (2)	_____
Healthcare Exposure in Prior 6 Months:	
9. Post-acute care facility (2)	_____
10. ≥1 night of international hospitalization in an ESBL high-burden region[b] (5)	_____
MDRGN Colonization or Infection in Prior 6 Months:	
11. ESBL (6)	_____
12. Carbapenem-resistant *Enterobacteriaceae* (6)	_____
13. Multidrug-resistant *Pseudomonas aeruginosa* (-4; subtract 4 pts.)	_____
Antibiotic Exposure in Prior 6 Months:	
14. Weeks of active gram-negative therapy (0.25/week; max. of 1 point)	_____

POINTS SCORE (SUM TOTAL): _____

Patients with a score of 7.25 or more points have a 95% probability of being infected with an ESBL-producing organism.*

Fig. 1. A printable clinical risk score for bedside use to predict a bacteremic patient's likelihood of infection with an extended-spectrum β-lactamase (ESBL)–producing organism at the time of organism genus and species identification. Risk-factor points are noted in parentheses and summed among the 14 variables to produce a patient's risk score. Possible score cutoffs for ESBL-positive bacteremia, and associated sensitivities and specificities, are reflected in Table 2. [a]Chronic obstructive pulmonary disease, emphysema, or chronic ventilator-dependency. [b]Latin America (excluding the Caribbean), the Middle East (including Egypt), South Asia, China, and the Mediterranean.
*This statement reflects the positive predictive value of the score at a cutoff point of 7.25 and should be modified by the facility to account for local prevalence of ESBL bacteremia. Note. MDRGN, multidrug-resistant gram-negative organism; CRE, carbapenem-resistant *Enterobacteriaceae*. Drug-resistant organisms were defined in accordance with the Centers for Disease Control and Prevention guidelines.[9]

used to predict the value of the held-out observation. This process is repeated for all observations in the dataset, and performance metrics (eg, error) can be averaged across the *n* fitted models (in this case, decision trees) to produce a single estimate. We evaluated the discrimination of the original and cross-validated models through the generation of receiver operating characteristic (ROC) curves and calculation of C statistics. Decision tree analyses were performed using the RPART (Recursive Partitioning and Regression Trees) package in R Studio version 4.1–90.99.902 software (R Foundation for Statistical Computing, Vienna, Austria).

To develop a risk score, continuous variables (eg, age and antibiotic days) were first converted into ordinal categories to reduce complexity, given the score's anticipated manual application. A multivariable logistic regression model was derived using stepwise variable selection with backward elimination at an α level of 0.05. To create points, regression coefficients were rescaled by dividing by the smallest final model coefficient and rounding to the nearest integer (with the exception of antibiotic therapy, which received 0.25 points per week (up to a maximum of 1 point or ≥4 weeks), to simplify end-user calculations). Patient scores were calculated by summing their respective points (risk score model).

For both the multivariable regression model and the risk score model, discrimination was assessed with ROC curves and accompanying C statistics (ie, area under the curve). Risk score model

https://doi.org/10.1017/ice.2019.17 Published online by Cambridge University Press

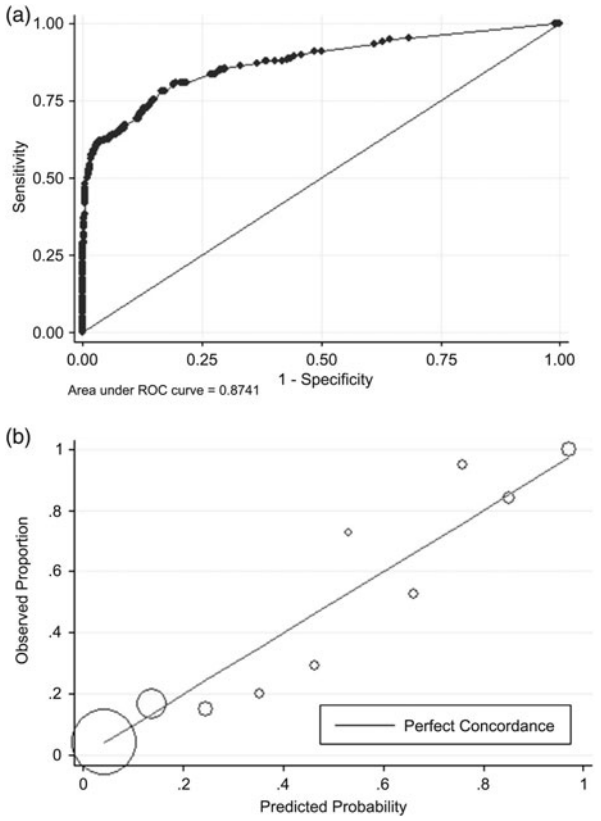

Fig. 2. Discrimination and calibration metrics for the multivariable logistic regression model and resulting risk score model. (A) Receiver operating characteristic (ROC) curve for the logistic regression model, prior to risk score transformation. The area under the curve (AUC) was 0.87 which, after rounding, was unchanged following conversion to a point-based risk-score model. See Table 2 for exact sensitivity and specificity values at different score cutoff points. (B) Calibration plot of observed proportion versus ESBL probabilities predicted by the risk score model, by decile groups.

calibration was evaluated using Hosmer-Lemeshow (HL) goodness-of-fit tests and graphical plots of observed proportion versus model-predicted ESBL probabilities by decile groups. Discrimination was internally validated with leave-one-out cross-validation. Risk score analyses were performed in Stata version 13.0 software (StataCorp, College Station, TX) and R Studio.

Results

Spanning the 2008 to 2015 time period, a total of 1,288 bacteremic patients met inclusion criteria, of whom 194 (15%) were ESBL positive. Patient and microbial characteristics have been reported previously.[8]

Risk score

The multivariable model and resulting risk score included 14 variables (Table 1), which were broadly categorizable into 6 groups (Fig. 1):

1. *Indwelling hardware on day of culture*. Orthopedic hardware (2 points); chronic indwelling vascular hardware (1 point); nephrostomy tube or Foley catheter (2 points); gastrointestinal feeding tube (2 points).
2. *Presumptive source of bloodstream infection*. central vascular catheter (2 points); pneumonia (2 points).
3. *Patient characteristics*. Structural lung disease (chronic obstructive pulmonary disease, emphysema, or tracheostomy dependency) (2 points); self-identification as Asian race (2 points).
4. *Healthcare exposure within the previous 6 months*. Post-acute care facility (2 points); ≥1 night of international hospitalization in an ESBL high-burden region (5 points).
5. MDRGN colonization or infection within the previous 6 months. ESBL (6 points); carbapenem-resistant *Enterobacteriaceae* (CRE) (6 points); MDR *Pseudomonas* spp (−4 points).
6. *Antibiotic exposure within the previous 6 months*. Weeks of therapy with gram-negative activity (0.25 points per week, up to a maximum of 1 point).

Patient scores ranged from −3 to 18.75, with a median score of 2 points (interquartile range: 0–3.25). The C statistic for the clinical risk score was 0.87 and 0.89 following cross-validation. The C statistic for the multivariable logistic regression model was also 0.87 (Fig. 2). The multivariable logistic regression model provided evidence of acceptable calibration (HL goodness-of-fit test $P = .13$). Following point conversion, however, the risk score model over- or underestimated the probability of ESBL infection at different points along the risk continuum, with the exception of very high-risk deciles (HL goodness-of-fit test $P < .001$) (Fig. 2). An ESBL-positive cutoff point of ≥7.25 maximized overall ESBL classification accuracy (92%). At this cutoff point, the risk score had a sensitivity of 49.5% and a specificity of 99.5%, and its positive and negative predictive values were 94.6% and 91.8%, respectively. Table 2 provides the risk score's sensitivity and specificity at each possible ESBL-positive cutoff point.

Decision tree

The final decision tree[8] included 5 predictors: central vascular catheter, age ≥43 years, and in the prior 6 months: history of ESBL colonization/infection, ≥1 night hospitalization in an ESBL high-burden region, and/or ≥1 week of gram-negative active antibiotic therapy (Fig. 3). The C statistic of the decision tree was 0.77 (unchanged in cross-validation); the sensitivity and specificity were 51.0% and 99.1%, and the positive and negative predictive values were 90.8% and 91.9%, respectively. Table 3 presents a comparison of the performance metrics of the risk score versus the decision tree.

Discussion

Despite advances in rapid diagnostics, timely identification of MDRGNs remains a clinical and epidemiological challenge. Diagnostic delays can prolong the period of ineffective antibiotic therapy and can increase the risk of nosocomial transmissions.[3,4] Statistical models for predicting drug resistance can play an important role in settings where rapid diagnostic tests are unavailable or are resource-impractical. This case study of ESBL bloodstream infections explores 2 approaches for developing predictive models: traditional logistic regression-derived risk scores and machine learning-derived decision trees.

The risk score included 14 independent predictors, broadly classifiable into 6 categories: indwelling hardware, bloodstream infection source, patient characteristics, recent gram-negative antibiotic exposure, healthcare exposure, and MDRO history. Many of these variables (eg, antibiotic use, prior ESBL colonization or infection) were retained in the decision tree. They are also consistent with other studies examining risk factors for MDRGN bloodstream infections[14] and recent scores for identifying

https://doi.org/10.1017/ice.2019.17 Published online by Cambridge University Press

Table 2. Risk Score Sensitivity, Specificity, and Overall Classification Accuracy at Select Cutoff Points for Predicting Extended-Spectrum β-Lactamase (ESBL) Status in a Cohort of Adult Patients with *Escherichia coli* and *Klebsiella* Species Bacteremia[a]

Risk Score Cutoff Point	Sensitivity, %	Specificity, %	Observations Correctly Classified, %
≥0	100.0	0.7	15.7
≥.25	95.4	31.5	41.2
≥.5	94.9	35.7	44.6
≥.75	93.8	37.0	45.6
≥1	93.3	38.8	47.0
≥1.25	90.7	51.3	57.2
≥1.5	89.7	54.1	59.5
≥1.75	89.2	55.6	60.6
≥2	88.7	56.7	61.5
≥2.25	85.6	70.2	72.5
≥2.5	84.0	71.6	73.5
≥2.75	83.5	72.5	74.2
≥3	83.5	73.1	74.7
≥3.25	77.8	83.4	82.5
≥3.5	74.2	86.8	84.9
≥3.75	71.7	87.7	85.3
≥4	70.6	88.3	85.6
≥4.25	65.5	92.6	88.5
≥4.5	64.4	92.8	88.5
≥4.75	63.9	93.2	88.8
≥5	63.9	93.4	89.0
≥5.25	61.9	95.7	90.6
≥5.5	61.3	96.2	90.9
≥5.75	60.8	96.6	91.2
≥6	60.8	97.0	91.5
≥6.25	55.2	98.2	91.7
≥6.5	54.6	98.4	91.8
≥6.75	54.6	98.5	91.9
≥7	54.1	98.5	91.9
≥7.25	49.5	99.5	91.9
≥7.5	46.9	99.5	91.5
≥7.75	46.4	99.5	91.5
≥8	45.9	99.5	91.4
≥8.25	40.2	99.5	90.6
≥8.5	38.7	99.7	90.5
≥8.75	38.1	99.8	90.5
≥9	37.6	99.8	90.5
≥9.25	31.4	100.0	89.7

Note. CI, confidence interval.[a]Cutoff points <0 and ≥9.5 were excluded because, respectively, they yielded equal sensitivity (100%) but inferior specificity, or inferior sensitivity but equal specificity (100%). Dark gray shading indicates the cutoff point that maximized overall classification accuracy (≥7.25 points).

Table 3. Comparative Performance Metrics of a Logistic Regression-Derived Clinical Risk Score and a Machine Learning-Derived Decision Tree to Predict Extended-Spectrum β-Lactamase (ESBL) Status

Variable	Risk Score	Decision Tree
No. of included variables	14	5
Sensitivity, %[a]	49.5	51.0
Specificity, %[a]	99.5	99.1
Positive predictive value (PPV), %[a]	94.6	90.8
Negative predictive value (NPV), %[a]	91.8	91.9
Naïve C statistic	0.87	0.77
Cross-validated C statistic	0.89	0.77

[a]Risk score values vary depending upon the selected cutoff point for dichotomization. Values reflected for the risk score are for the cutoff point of ≥7.25 points, which optimized overall classification accuracy.

community- and hospital-onset ESBL or third-generation cephalosporin-resistant bacteremia in other populations.[15,16] Taken together with the risk score's similar C statistic following cross-validation (0.89), this evidence suggests that despite the inclusion of a large number of variables, the risk score was not overfit.

Given that risk scores for binary predictions are dichotomized at a cutoff point, in practice the risk score and the decision tree performed similarly: sensitivities 49.5% and 51.0% and specificities 99.5% and 99.1%, respectively. However, the risk score had a ~10% higher area-under-the-curve (risk score and decision tree C statistics: 0.87 vs 0.77). This higher AUC offers users more latitude to prioritize sensitivity over specificity, or vice versa, by changing the cutoff point (as discussed in more detail below). In theory, a decision tree could also be developed to optimize a different balance of sensitivity and specificity, but this would require deriving an entirely new tree. The risk score's greater flexibility, however, came at a cost of low user-friendliness for manual application. Studies consistently demonstrate that incorporating decision support tools at the point of care is important to their success,[17] but manual tabulation of 14 variables would encounter significant bedside utilization barriers. In contrast, decision-tree branching logic does not require end-user calculations and, at least in this ESBL case study, the final decision tree included far fewer (ie, 5) predictors.

The potential tradeoff between flexibility and user friendliness is an important consideration when evaluating whether risk scores or decision trees are a more suitable decision support tool for a given application. Additional considerations, however, may also help to guide researchers in selecting one option versus the other. Below, we summarize the relative strengths of risk scores and decision trees for model development and fitting, implementation, and adaptability. Of note, the CART analysis is the tree-fitting process (approach), and a decision tree is the result (output), just as logistic regression is a common (but by no means the only or necessarily even the preferred) approach for developing a risk score. Approach and output can differ in their strengths and limitations, and we distinguish these concepts in our discussion.

Methodological differences between logistic regression and CART influence the data assumptions and exploratory analyses required for model development and fitting. In general, the more complex or challenging the underlying data, the more utility a machine learning approach can provide. Specifically, logistic regression imposes important data requirements, including minimal collinearity (ie, correlation) among independent variables and a sufficient ratio of cases to predictors (ie, sufficient sample size; a general, although debatable, guideline is 10 expected cases per

https://doi.org/10.1017/ice.2019.17 Published online by Cambridge University Press

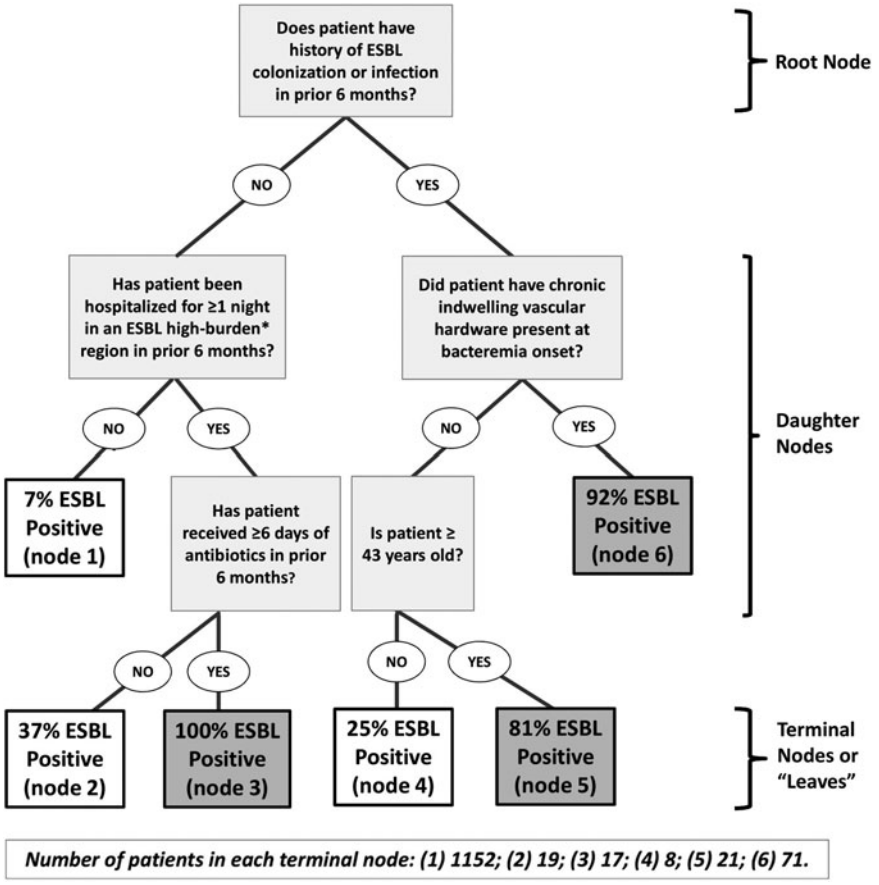

Fig. 3. A clinical decision tree to predict a bacteremic patient's likelihood of infection with an extended-spectrum β-lactamase (ESBL)–producing organism at the time of organism genus and species identification, adapted from Goodman et al (2016).[8] Gray-shaded terminal nodes indicate that the tree would classify patients as ESBL positive, and accompanying percentages (derived from terminal-node impurities) reflect the probability that patients assigned to a given terminal node are ESBL-positive. Terminal node numbering (1–6) is included in parentheses. *Latin America (excluding the Caribbean), the Middle East (including Egypt), South Asia, China, and the Mediterranean.

predictor evaluated).[18,19] In contrast, CART is nonparametric and makes fewer data assumptions,[13] and it can accommodate collinear independent variables. It is also less sensitive to outliers and more robust to high-dimensional data, which possess many independent variables relative to outcomes. These features are appealing in MDRGN research, given the abundance of predictors in patient medical records but the relative rarity of clinical outcomes. Moreover, logistic regression requires a priori specification and evaluation of variable interactions, whereas CART identifies interactions without user input,[13] a potentially helpful feature when the understanding of variable relationships is generally limited.

The benefits of CART, however, can come with a steep learning curve for researchers without prior experience with these methods. In particular, decision trees are prone to overfitting, in which they fit the data "too well" (including its idiosyncrasies and noise) and may consequently perform poorly on new data.[20] Sufficient expertise in pruning and/or stopping criteria during the tree-branching process is therefore critical to the utility and generalizability of the resulting tree, as is the use of internal validation methods (eg, cross-validation) when external testing datasets are unavailable. Although ensemble tree methods such as random forests analysis can address many of these challenges, these methods do not produce a single decision tree that can be used as a decision support tool (without automation).[21,22]

Decision tree branching logic does not require calculations, and decision trees are generally intuitive and user-friendly. When manual bedside use is anticipated, these features are especially beneficial. As facilities incorporate automated decision support tools and algorithms into electronic health records (EHRs), these benefits attenuate. In this ESBL case study, because important variables required clinical judgment (eg, source of infection) or were not hard-coded in the EHR (eg, foreign country of recent hospitalization was only entered as natural language), automating the decision support tool would have been challenging. As a result, the decision tree's simplicity for manual bedside use was highly valuable for this research application.

Finally, for applications in which decision support tool flexibility is paramount, risk scores are attractive because their cutoff points are modifiable by end users. Risk scores provide a range of score cutoffs, each with an associated sensitivity and specificity, which allow individual users to toggle the cutoff point to minimize the false-positive or false-negative rate (eg, depending upon infection severity or the clinical appearance of the patient). Using the current risk score, for example, a user seeking to increase sensitivity could choose a lower cutoff point of ≥3 points and reduce the risk of incorrectly classifying an ESBL infection as ESBL negative to <1 in 5 (sensitivity 83.5%, specificity 73.1%) (Table 1). This flexibility allows clinicians and hospital epidemiologists to maximize detection of cases (ie, ESBL-positive patients), though at the cost of attendant reductions in specificity and overall classification accuracy.

We caution, however, that although enhanced flexibility is generally beneficial, a risk score's utility depends upon users understanding the score and the implications of adjusting the cutoff point. Large score differences between patients may translate to minimal differences in risk, and vice versa. Moreover, cutoff-point positive and negative predictive values (ie, the probability that a patient does or does not have an ESBL-producing infection given a score that is respectively above or below the selected cutoff point) will vary by ESBL prevalence in the target population. It is

https://doi.org/10.1017/ice.2019.17 Published online by Cambridge University Press

Table 4. Comparative Strengths and Limitations of Logistic Regression-Derived Risk Scores and Classification and Regression Tree (CART) Analysis-Derived Decision Trees for Predicting Drug-Resistant Infections in Clinical Settings

	Risk Scores	Decision Trees	Notes
Data Characteristics			
High dimensionality	–	+++	Decision trees are well suited to high-dimensional data, which possess high predictor-to-outcome ratios. Logistic regression-derived risk scores impose more stringent sample size requirements (a general requirement is 10 expected cases per predictor).
Collinearity	–	+++	Logistic regression-derived risk scores require minimal collinearity among independent variables, unlike decision trees.
Interaction effects	+	+++	Logistic regression can accommodate interaction effects, but it requires moderately large sample sizes and a priori evaluation. CART decision trees can detect simple and higher-level interaction effects without user specification.
Rare outcome(s)	+	+	Rare outcomes pose challenges for both models. In logistic regression, rare outcomes limit the number of evaluable predictors. CART analysis may require parameter adjustment and/or case oversampling before model fitting and validation to improve sensitivity if outcomes are rare.
Model development			
Ease of development	++	+	Decision trees for standard applications are relatively straightforward to develop, but logistic regression-derived risk-score methodology is more well known in the infectious disease literature and more widely available on all common statistical computing platforms.
Robustness to overfitting	++	–	Both methods require validation, but decision trees are particularly prone to overfitting, in which they fit the data "too well" and may consequently perform poorly on new data. Methods to combat overfitting include imposing branching-stop criteria and "pruning" back terminal branches.
Implementation and usage			
Intuitiveness	+	+++	Decision-tree branching logic is highly intuitive.
Ease-of-use	+	+++	Decision trees generally do not require calculations, making them user-friendly for bedside application.
Adaptability			
End-user adjustment of sensitivity and specificity	+++	–	By changing the score cutoff point, individual users can tailor risk scores' sensitivity and specificity. A decision tree possesses a fixed sensitivity and specificity that, following model development, cannot be modified.
Addition of new variables over time	++	+	New variable(s) can be evaluated for risk score inclusion (eg, by comparing Akaike's information criterion (AIC) values of the original and expanded models).[27] Variable addition may change coefficient values and, accordingly, risk score points but will leave original score variables intact. Because decision trees are built "top-down," new variables require tree refitting and may substantially alter nodes and branching patterns.

imperative that the table of cutoff-point sensitivities and specificities, and an understanding that an institution's disease prevalence will affect the positive and negative predictive values, guides decisions about score thresholds for ESBL infection.

In contrast to risk scores, classification trees provide binary predictions (eg, "ESBL" or "not ESBL"), with a single sensitivity and specificity value for the tree as a whole. Terminal node percentages (eg, "37% probability of being ESBL positive") can quantify these predictions but do not provide a formal mechanism for prioritizing sensitivity versus specificity. For research applications in which sensitivity is the priority, methods are available to impose a greater "cost" for case misclassification during the tree-fitting process.[23] The limitation, however, is that these mechanisms are not adjustable by end users after a tree is built. In other words, whereas the CART approach provides flexibility to optimize sensitivity or specificity, once a single, final tree (output) is developed and provided to clinicians, the ability to adjust sensitivity and specificity is limited.

Although these considerations can help researchers to evaluate whether a risk score or a decision tree is preferable for a given research question (Table 4), a decision is rarely clear cut. In cases in which each model would at least partially meet stated goals, we encourage investigators to develop both support tools in parallel to compare their performance metrics. In particular, although model performance was comparable in this case study, other applications with more challenging data (eg, high-dimensionality, higher-order variable interactions) might more clearly favor a machine learning approach such as CART.

Our study has several limitations. This study was conducted in a single center, and although we internally validated our models, it lacked an external validation cohort. In addition, data may have been missing for patients treated outside of the Epic Care Everywhere network, although we do not expect such occurrences to have differed by ESBL status. As such, any resulting exposure misclassification would likely reduce predictive performance, and yet risk score discrimination remained robust, including in cross-validation. Nevertheless, we encourage others to evaluate and validate the risk score in their own patient populations, particularly for settings that differ from our academic, tertiary-care hospital cohort. Importantly, however, because study characteristics were constant across analyses, we expect decision tree and risk score comparisons to be unbiased. Finally, this case study intended to offer a practical, high-level introduction to a relatively simple machine learning approach, but we note that many machine learning methodologies (eg, random

https://doi.org/10.1017/ice.2019.17 Published online by Cambridge University Press

forests, Super Learner) offer potential healthcare epidemiology utility. We refer interested readers to additional resources that address these approaches and the underlying algorithms in greater technical detail.[22,24,26]

Overall, timely identification of MDRGN infections remains a clinical and epidemiological challenge. Rapid detection enables isolation of infected patients and prompt initiation of appropriate antibiotic treatment. Statistical models for predicting drug resistance can provide important information in settings when laboratory diagnostics are challenging to implement. This examination explored 2 alternative decision support tools, logistic regression–derived risk scores and machine learning–derived decision trees, in an inpatient cohort of bacteremic patients to predict ESBL infection. These methodologies offer different strengths and limitations, and we hope that their continued utilization in infectious disease research will assist with improving patient outcomes.

Author ORCIDs. Katherine E. Goodman, 0000-0003-2851-775X

Acknowledgments.

Financial support. This work was supported by funding from the Agency for Healthcare Research and Quality (grant no. R36HS025089 to K.E.G.), from the Sherrilyn and Ken Fisher Center for Environmental Infectious Diseases (grants to K.E.G. and A.M.M.), and from the National Institutes of Health (grant no. K23-AI127935 to P.D.T.).

Conflicts of interest. None of the authors report any conflicts of interest.

References

1. Livorsi DJ, Chorazy ML, Schweizer ML, *et al.* A systematic review of the epidemiology of carbapenem-resistant Enterobacteriaceae in the United States. *Antimicrob Resist Infect Control.* 2018;7:55.
2. McDanel J, Schweizer M, Crabb V, *et al.* Incidence of extended-spectrum beta-lactamase (ESBL)-producing *Escherichia coli* and *Klebsiella* infections in the United States: a systematic literature review. *Infect Control Hosp Epidemiol* 2017;38:1209–1215.
3. Micek ST, Hampton N, Kollef M. Risk factors and outcomes for ineffective empiric treatment of sepsis caused by gram-negative pathogens: stratification by onset of infection. *Antimicrob Agents Chemother* 2018;62(1): e01577–17.
4. Zhang D, Micek ST, Kollef MH. Time to appropriate antibiotic therapy is an independent determinant of postinfection ICU and hospital lengths of stay in patients with sepsis. *Crit Care Med* 2015;43:2133–2140.
5. Clinical and Laboratory Standards Institute. *Performance Standards for Antimicrobial Susceptibility Testing*, 28th ed. Supplement M100S. Wayne, PA: CLSI; 2018.
6. 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 microarray-based molecular assay. *J Clin Microbiol* 2015;53:2460–2472.
7. Ward C, Stocker K, Begum J, Wade P, Ebrahimsa U, Goldenberg SD. Performance evaluation of the Verigene (Nanosphere) and FilmArray (BioFire) molecular assays for identification of causative organisms in bacterial bloodstream infections. *Eur J Clin Microbiol Infect Dis* 2015;34:487–496.
8. Goodman KE, Lessler J, Cosgrove SE, *et al.* A clinical decision tree to predict whether a bacteremic patient is infected with an extended-spectrum beta-lactamase–producing organism. *Clin Infect Dis* 2016;63:896–903.
9. Antimicrobial resistant phenotype definitions. Centers for Disease Control and Prevention webstie. https://www.cdc.gov/nhsn/pdfs/ps-analysis-resources/phenotype_definitions.pdf. Published 2016. Accessed January 15, 2019.
10. Kantele A, Laaveri T, Mero S, *et al.* Antimicrobials increase travelers' risk of colonization by extended-spectrum beta-lactamase–producing Enterobacteriaceae. *Clin Infect Dis* 2015;60:837–846.
11. Ostholm-Balkhed A, Tarnberg M, Nilsson M, *et al.* Travel-associated faecal colonization with ESBL-producing Enterobacteriaceae: incidence and risk factors. *J Antimicrob Chemother* 2013;68:2144–2153.
12. Duda RO, Hart PE, Stork DG. *Pattern Classification*, 2nd ed. New York: Wiley-Interscience; 2001.
13. Breiman L, Friedman J, Stone C, Olshen R. *Classification and Regression Trees*. Boca Raton, FL: CRC/Chapman & Hall; 1984.
14. 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;38:1216–1225.
15. Rottier WC, van Werkhoven CH, Bamberg YRP, *et al.* Development of diagnostic prediction tools for bacteraemia caused by third-generation cephalosporin-resistant enterobacteria in suspected bacterial infections: a nested case-control study. *Clin Microbiol Infect* 2018;24: 1315–1321.
16. Augustine MR, Testerman TL, Justo JA, *et al.* Clinical risk score for prediction of extended-spectrum beta-lactamase–producing enterobacteriaceae in bloodstream isolates. *Infect Control Hosp Epidemiol* 2017; 38:266–272.
17. Kawamoto K, Houlihan CA, Balas EA, Lobach DF. Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success. *BMJ* 2005;330:765.
18. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 1996;49:1373–1379.
19. Vittinghoff E, McCulloch CE. Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007;165:710–718.
20. Dietterich T. Overfitting and undercomputing in machine learning. *ACM Comput Surv* 1995;27:326–327.
21. Chen X, Ishwaran H. Pathway hunting by random survival forests. *Bioinformatics* 2013;29:99–105.
22. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. *Psychol Method* 2009;14:323–348.
23. Drummond C, Holte RC. Exploiting the cost (in)sensitivity of decision tree splitting criteria. *Proceedings of the Seventeenth International Conference on Machine Learning*. Stanford, CA; 2000.
24. Tibshirani R. Regression shrinkage and selection via the lasso. *J Roy Stat Soc B* 1996;58:267–288.
25. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* 2007;6:article25. Epub 2007 Sep 16.
26. Song YY, Lu Y. Decision tree methods: applications for classification and prediction. *Shanghai Arch Psychiatr* 2015;27:130–135.
27. Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Control* 1974;19:716–723.

https://doi.org/10.1017/ice.2019.17 Published online by Cambridge University Press