

# A Clinical Decision Tree to Predict Whether a Bacteremic Patient Is Infected With an Extended-Spectrum $\beta$ -Lactamase–Producing Organism

Katherine E. Goodman,<sup>1</sup> Justin Lessler,<sup>1</sup> Sara E. Cosgrove,<sup>2</sup> Anthony D. Harris,<sup>3</sup> Ebbing Lautenbach,<sup>4</sup> Jennifer H. Han,<sup>4</sup> Aaron M. Milstone,<sup>5</sup> Colin J. Massey,<sup>6</sup> and Pranita D. Tamma<sup>5</sup>; for the Antibacterial Resistance Leadership Group

<sup>1</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, <sup>2</sup>Department of Medicine, Division of Infectious Diseases, Johns Hopkins University School of Medicine, and <sup>3</sup>Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore; <sup>4</sup>Department of Medicine, Division of Infectious Diseases, University of Pennsylvania School of Medicine, Philadelphia; <sup>5</sup>Department of Pediatrics, Division of Infectious Diseases, and <sup>6</sup>Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland

**Background.** Timely identification of extended-spectrum  $\beta$ -lactamase (ESBL) bacteremia can improve clinical outcomes while minimizing unnecessary use of broad-spectrum antibiotics, including carbapenems. However, most clinical microbiology laboratories currently require at least 24 additional hours from the time of microbial genus and species identification to confirm ESBL production. Our objective was to develop a user-friendly decision tree to predict which organisms are ESBL producing, to guide appropriate antibiotic therapy.

**Methods.** We included patients  $\geq 18$  years of age with bacteremia due to *Escherichia coli* or *Klebsiella* species from October 2008 to March 2015 at Johns Hopkins Hospital. Isolates with ceftriaxone minimum inhibitory concentrations  $\geq 2$   $\mu\text{g/mL}$  underwent ESBL confirmatory testing. Recursive partitioning was used to generate a decision tree to determine the likelihood that a bacteremic patient was infected with an ESBL producer. Discrimination of the original and cross-validated models was evaluated using receiver operating characteristic curves and by calculation of C-statistics.

**Results.** A total of 1288 patients with bacteremia met eligibility criteria. For 194 patients (15%), bacteremia was due to a confirmed ESBL producer. The final classification tree for predicting ESBL-positive bacteremia included 5 predictors: history of ESBL colonization/infection, chronic indwelling vascular hardware, age  $\geq 43$  years, recent hospitalization in an ESBL high-burden region, and  $\geq 6$  days of antibiotic exposure in the prior 6 months. The decision tree's positive and negative predictive values were 90.8% and 91.9%, respectively.

**Conclusions.** Our findings suggest that a clinical decision tree can be used to estimate a bacteremic patient's likelihood of infection with ESBL-producing bacteria. Recursive partitioning offers a practical, user-friendly approach for addressing important diagnostic questions.

**Keywords.** ESBL; bacteremia; carbapenem; machine learning; prediction.

Extended-spectrum  $\beta$ -lactamase (ESBL)–producing bacteria represent a serious clinical and public health challenge [1]. ESBL-producing bacteria can hydrolyze most broad-spectrum  $\beta$ -lactam antibiotics, with the exception of carbapenems [2]. Serious infections, including bacteremia, with ESBL-producing organisms are associated with higher morbidity and mortality relative to infections with more susceptible organisms [3, 4]. Existing data suggest that this disparity results at least in part from delayed initiation of appropriate therapy, as many empiric antibiotic regimens have limited activity against ESBL producers [5, 6]. While carbapenems remain effective against

ESBL-producing organisms, they should be used judiciously, because indiscriminate empiric carbapenem use may select for carbapenem-resistant Enterobacteriaceae [7, 8].

Rapid diagnostics to identify various  $\beta$ -lactamase genes are becoming increasingly available to reduce the time between Gram stain results and resistance mechanism identification, but such assays can be resource intensive and, thus, are currently not widely used in clinical microbiology laboratories. Additionally, commonly used molecular-based gram-negative panels do not include, or at most only identify one of, the ESBL gene groups [9]. Consequently, clinicians must select empiric antibiotic treatment for patients with gram-negative bacteremia without knowing whether the causative organism is ESBL producing, while balancing the risk of ineffective therapy against unnecessarily broad antibiotic treatment. This delay in selecting appropriate antibiotic treatment can lead to poor patient outcomes [4]. Statistical models for predicting ESBL-producing infections can help to address current diagnostic limitations.

Received 7 April 2016; accepted 20 June 2016; published online 28 June 2016.

Correspondence: P. D. Tamma, Department of Pediatrics, Division of Infectious Diseases, Johns Hopkins University School of Medicine, 200 N Wolfe St, Ste 3149, Baltimore, MD 21287 (ptamma1@jhmi.edu).

Clinical Infectious Diseases® 2016;63(7):896–903

© The Author 2016. Published by Oxford University Press for the Infectious Diseases Society of America. All rights reserved. For permissions, e-mail journals.permissions@oup.com. DOI: 10.1093/cid/ciw425

Numerous recent investigations have used multivariable logistic regression models to identify exposures independently associated with ESBLs (eg, previous antibiotic therapy, presence of an indwelling urinary catheter) [10–12]. Although valuable for exploring potential risk factors driving the emergence of ESBL-producing bacteria, these approaches do not help clinicians readily synthesize or decide how to prioritize multiple risk factors. Conversion of logistic regression coefficients into a risk score model addresses some of these concerns, but these models may be cumbersome to implement depending upon the number of included variables and complexity of end-user calculations.

Recursive partitioning is a form of machine learning rarely utilized in the clinical antibiotic resistance literature that may be helpful as a predictive modeling tool in these circumstances [13, 14]. Its output, a decision tree algorithm, has several practical advantages, including simplicity and intuitive interpretation. Our objective was to develop a user-friendly decision tree to predict, at the time of organism identification from a blood culture, which bacteremias are due to ESBL producers in order to guide appropriate antibiotic therapy.

## METHODS

### Setting and Participants

This study included patients aged  $\geq 18$  years hospitalized at Johns Hopkins Hospital with bloodstream isolates growing *Klebsiella pneumoniae*, *Klebsiella oxytoca*, or *Escherichia coli* from October 2008 to March 2015. Records were identified from the Johns Hopkins Hospital clinical microbiology laboratory database. Only first episodes of bacteremia with the above organisms for a given patient were included. This study was approved by the Johns Hopkins University School of Medicine Institutional Review Board, with a waiver of informed consent.

### Clinical Data Collection

Patient data were extracted 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 (<https://www.epic.com/CareEverywhere/>), into a REDCap database. The EPIC Care Everywhere Network is a secure health information exchange that allows clinicians to securely view previous patient medical information from a large number of inpatient and outpatient healthcare networks throughout the United States. The following patient data were collected, with all information based on the time period prior to day 1 of bacteremia, defined as the day the blood culture was obtained: (1) demographic data; (2) preexisting medical conditions; (3) source of bacteremia; (4) indwelling hardware (eg, orthopedic hardware, urology hardware, central venous catheters, grafts); (5) multidrug-resistant organism colonization or infection (multidrug-resistant *Pseudomonas aeruginosa*,

multidrug-resistant *Acinetobacter baumannii*, ESBLs, carbapenem-resistant Enterobacteriaceae, vancomycin-resistant *Enterococcus* species, and methicillin-resistant *Staphylococcus aureus*) within the previous 6 months [15]; (6) days of gram-negative active inpatient and outpatient antibiotic therapy (extended-spectrum penicillins, third- and fourth-generation cephalosporins, aztreonam, carbapenems, aminoglycosides, and fluoroquinolones) within the previous 6 months; (7) days of stay in any healthcare facility (outpatient procedures were assigned “1 day of stay”); (8) hospitalization in another country in the previous 6 months; and (9) residence in a long-term care facility or nursing home within the previous 6 months. Patients who were hospitalized in another country were separated into high-burden and low-burden ESBL regions. “High-burden” included the following regions: Latin America (excluding the Caribbean), the Middle East (including Egypt), South Asia, China, and the Mediterranean [16, 17].

### Microbiology Methods

Bloodstream isolates of *E. coli*, *K. pneumoniae*, and *K. oxytoca* were processed at the Johns Hopkins Hospital Microbiology Laboratory according to standard operating procedures. Antibiotic susceptibility data were determined by the BD Phoenix Automated System (BD Diagnostics, Sparks, Maryland). Organisms with minimum inhibitory concentrations (MICs)  $\geq 2$   $\mu\text{g/mL}$  for ceftriaxone underwent further confirmation for ESBL production. A decrease of  $>3$  doubling dilutions in the MIC for a third-generation cephalosporin tested in combination with 4  $\mu\text{g/mL}$  of clavulanic acid, vs its MIC when tested alone, was used to confirm ESBL status. There were no changes in the method of organism identification, antibiotic susceptibility testing, or ESBL confirmatory testing during the study period.

### Statistical Methods

#### Data Analysis and Logistic Regression

Descriptive statistics for patient variables were calculated using mean (standard deviation [SD]), median (range or interquartile range), or frequency count (percentage), as appropriate. The relationship between each study covariate and ESBL status was evaluated using univariable logistic regression, as summarized by odds ratios (ORs) and corresponding 95% confidence intervals (CIs). Final multivariable logistic regression models were derived using stepwise variable selection with backward elimination at an  $\alpha$  level of .05 (a common, though of debated validity, approach in the literature) and lasso regression at the value of the shrinkage parameter that minimized misclassification error in the cross-validated model [18]. Lasso regression was performed using the glmnet (Lasso and Elastic-Net Regularized Generalized Linear Models) package, version 2.0–2, in the R statistical package (version 3.0.3).

#### Decision Tree Derivation

We built a decision tree to predict whether a patient’s bacteremia was due to an ESBL producer applying the classification

and regression tree algorithm [14] on a dataset including all study variables using the rpart (Recursive Partitioning and Regression Trees) package, version 4.1–9, in R.

In brief, a tree was built using the following process: (1) identification of the single variable that, when used to split the dataset into 2 groups (“nodes”), best minimized impurity of ESBL status in each daughter node, according to the Gini impurity criterion [14, 19]; (2) repetition of the partitioning process within each daughter node and subsequent generations of nodes (“recursive partitioning” or “branching”); and (3) cessation at “terminal” nodes when no additional variables achieve further reductions in node impurity. Terminal nodes in binary recursive partitioning trees predict ESBL status categorically but, by evaluating the node impurity, also offer associated probabilities.

#### Decision Tree Validation

We internally validated the performance of our model using the leave-one-out cross-validation method [19]. 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 in R.

## RESULTS

#### Study Population

A total of 1288 Johns Hopkins Hospital patients with bacteremia due to *E. coli* (56%), *K. pneumoniae* (40%), or *K. oxytoca* (4%), spanning the period from October 2008 to March 2015, met eligibility criteria. For 194 patients (15%), bacteremia was due to a confirmed ESBL producer.

Patient and microbial characteristics are presented in Table 1. Evaluating the full cohort, patients had a mean age of 55 (SD, 16.4) years. Twenty-five percent of patients had a history of prior colonization or infection with a multidrug-resistant organism within the preceding 6 months. In the 6 months prior to bacteremia, patients had been hospitalized for a mean of 13.7 (SD, 20.3) days (excluding the current hospitalization) and had received a mean of 11.6 (SD, 20.2) days of antibiotic therapy. The majority of bacteremias originated from the urinary tract (37%), followed by intra-abdominal (24%), catheter-related (16%), and biliary (14%) sources.

Among patients with ESBL-positive bacteremia, 43% received chemotherapy within the prior 6 months, and the majority (68%) had chronic indwelling vascular hardware present at the time of bacteremia onset. Twenty-five percent had at least 1 overnight stay in a hospital in an ESBL high-burden region within the prior 6 months. Figure 1 reflects the distribution of ESBL-positive bacteremia cases by geographic region.

#### Logistic Regression

In univariable logistic regression analysis, a large proportion of collected data (25 study variables) were significantly associated with ESBL-positive bacteremia at an  $\alpha$  level of .05 (Table 1). The most strongly associated variables included prior history of

an ESBL (OR, 51.45 [95% CI, 29.11–90.93]) or carbapenem-resistant Enterobacteriaceae (OR, 23.01 [95% CI, 2.56–206.99]) colonization/infection, and recent international hospitalization in a high-burden region (OR, 30.47 [95% CI, 15.83–58.64]). Final multivariable models derived using stepwise variable selection and lasso regression included 14 and 16 variables, respectively (Table 1).

#### Decision Tree

Using binary recursive partitioning, the final classification tree for predicting ESBL-positive bacteremia included 5 study variables (Figure 2). The first question in the tree, also called the root node, asked: (1) Does the patient have a history of ESBL colonization or infection in the previous 6 months? In classification trees, positive or “yes” responses branch to the right. If “yes,” the second question queried: (2) Did the patient have chronic indwelling vascular hardware (defined as a dialysis or central venous catheter) at the time of bacteremia onset? Those patients meeting these criteria were classified as ESBL positive (terminal node 6) with an associated 92% probability. In patients with an ESBL history but lacking indwelling vascular hardware, the tree questioned: (3) Is the patient aged  $\geq 43$  years (based upon model-derived dichotomization at 43 years)? If “yes,” patients were classified as ESBL positive (terminal node 5, 81% probability) and if “no” were classified as ESBL negative (terminal node 4, 75% probability).

For those 1188 patients lacking a history of prior ESBL infection or colonization (question 1), the root node branched left. The tree then asked: (2) Has the patient been hospitalized in an ESBL high-burden region for 1 or more nights in the prior 6 months? If “yes”: (3) Has the patient received  $\geq 6$  days of antibiotics in the prior 6 months (based upon model-derived dichotomization at 6 days)? Those patients meeting these criteria were classified as ESBL positive (terminal node 3) with 100% probability. Patients who had been internationally hospitalized in a high-burden region but had not received at least 6 days of antibiotics were classified as ESBL negative (terminal node 2, 63% probability). Finally, patients who both lacked a prior ESBL history and recent high-risk international hospitalization were classified as ESBL negative, constituting the majority of the dataset (terminal node 1, 93% probability, 1152 patients).

The overall tree possessed a sensitivity of 51.0%, a specificity of 99.1%, and a  $\kappa$  value (reflecting chance-corrected agreement) of 0.61. The positive and negative predictive values were 90.8% and 91.9%, respectively. Incorporating outcome probabilities based on terminal node impurities, the C-statistic for the final tree trained on the full dataset was 0.77 and 0.77 following cross-validation.

Of the 194 patients with ESBL bacteremia, 35% (68) received empiric carbapenem therapy within 6 hours after genus and species identification. Utilization of the decision tree would have increased ESBL case detection during the empiric treatment window by approximately 50%. The decision tree identified

**Table 1. Description of Patient and Microbial Characteristics in a Cohort of Adult Patients With *Escherichia coli* and *Klebsiella* Species Bacteremia, by Extended-Spectrum  $\beta$ -Lactamase Status**

Variables on Day 1 of Bacteremia	ESBL Positive (n = 194)	ESBL Negative (n = 1094)	Odds Ratio (95% CI)	P Value
<b>Demographics</b>				
Age	51 $\pm$ 18.4	56 $\pm$ 15.9	0.98 (.97–.99)	<.001
Male sex	113 (58)	590 (54)	1.18 (.87–1.61)	.23
Race/ethnicity				
White	85 (44)	523 (48)	Reference	Reference
Black	49 (25)	458 (42)	0.66 (.45–.96)	.03
Latino	11 (6)	39 (4)	1.74 (.86–3.52)	.13
Asian	25 (13)	38 (3)	4.05 (2.33–7.05)	<.001 <sup>f,g</sup>
Middle Eastern	24 (12)	26 (2)	5.68 (3.12–10.35)	<.001
<b>Preexisting medical conditions</b>				
HIV infection	5 (3)	53 (5)	0.52 (.21–1.32)	.17
Chemotherapy within previous 6 mo	83 (43)	347 (32)	1.61 (1.18–2.20)	.003
Active immunosuppressant use <sup>a</sup>	8 (4)	65 (6)	0.68 (.32–1.44)	.32
Solid organ transplant	29 (15)	145 (13)	1.15 (.75–1.77)	.53
Hematopoietic stem cell transplant	12 (6)	48 (4)	1.44 (.75–2.76)	.28
End-stage liver disease	17 (9)	76 (7)	1.29 (.74–2.23)	.37
End-stage renal disease requiring dialysis	15 (8)	81 (7)	0.96 (.62–1.48)	.84
Congestive heart failure (ejection fraction <40)	16 (8)	81 (7)	1.12 (.64–1.97)	.68
Structural lung disease <sup>b</sup>	19 (10)	44 (4)	2.60 (1.48–4.54)	.001 <sup>f,g</sup>
<b>Indwelling hardware at the onset of bacteremia</b>				
Biliary stent	18 (9)	119 (11)	0.84 (.50–1.41)	.51
Gastrointestinal feeding tube	25 (13)	57 (5)	2.69 (1.64–4.43)	<.001 <sup>f,g</sup>
Nephrostomy tubes and/or Foley catheter	45 (23)	113 (10)	2.62 (1.78–3.86)	<.001 <sup>f,g</sup>
Chronic vascular hardware <sup>c</sup>	131 (68)	461 (42)	2.86 (2.07–3.95)	<.001 <sup>f,g</sup>
Orthopedic hardware	5 (3)	20 (2)	1.42 (.53–3.83)	.49 <sup>f,g</sup>
<b>Recent multidrug-resistant organism history (colonization or infection &lt;6 mo)</b>				
Vancomycin-resistant <i>Enterococcus</i> species	32 (17)	113 (10)	1.72 (1.12–2.63)	.01
Methicillin-resistant <i>Staphylococcus aureus</i>	8 (4)	45 (4)	1.00 (.47–2.16)	1.00
Extended-spectrum $\beta$ -lactamase producer	84 (43)	16 (2)	51.45 (29.11–90.93)	<.001 <sup>f,g</sup>
Carbapenem-resistant Enterobacteriaceae <sup>d</sup>	4 (2)	1 (<1)	23.01 (2.56–206.99)	.01 <sup>f,g</sup>
Multidrug-resistant <i>Pseudomonas</i> species <sup>d</sup>	4 (2)	14 (1)	1.62 (.53–4.99)	.40 <sup>f</sup>
Multidrug-resistant <i>Acinetobacter</i> species <sup>d</sup>	2 (1)	1 (<1)	11.39 (1.03–126.18)	.05
<b>Recent antibiotic exposure (&lt;6 mo)</b>				
Days of extended-spectrum penicillin therapy	6.6 $\pm$ 11.2	3.5 $\pm$ 8.0	1.03 (1.02–1.05)	<.001
Days of third- and fourth-generation cephalosporin therapy	4.9 $\pm$ 8.2	2.1 $\pm$ 4.8	1.07 (1.04–1.09)	<.001 <sup>g</sup>
Days of aztreonam therapy	0.3 $\pm$ 1.5	0.2 $\pm$ 1.9	1.02 (.95–1.10)	.61
Days of carbapenem therapy	5.0 $\pm$ 9.0	1.8 $\pm$ 6.2	1.05 (1.03–1.07)	<.001
Days of fluoroquinolone therapy	3.1 $\pm$ 6.8	2.2 $\pm$ 6.8	1.02 (1.00–1.04)	.10
Days of aminoglycoside therapy	1.3 $\pm$ 4.7	0.3 $\pm$ 1.9	1.11 (1.05–1.17)	<.001 <sup>f,g</sup>
Total days of antibiotics (combined)	21.0 $\pm$ 25.6	10.0 $\pm$ 18.6	1.02 (1.01–1.03)	<.001
Total days of hospitalization in the 6 mo prior to current hospitalization	23.1 $\pm$ 26.7	12.0 $\pm$ 18.5	1.02 (1.01–1.03)	<.001
Duration of time from hospital admission to positive blood culture, d	11.4 $\pm$ 41.0	5.7 $\pm$ 20.0	1.01 (1.002–1.01)	.01
<b>Recent international healthcare exposure (&lt;6 mo)</b>				
At least 1 overnight stay in a healthcare facility in an ESBL high-burden region <sup>e</sup>	49 (25)	12 (1)	30.47 (15.83–58.64)	<.001 <sup>f,g</sup>
<b>Other high-risk healthcare exposures (&lt;6 mo)</b>				
Long-term acute care facility residence	17 (9)	22 (2)	4.68 (2.44–8.99)	<.001 <sup>f,g</sup>
Nursing home residence	6 (3)	16 (2)	2.15 (.83–5.57)	.12
<b>Source of bacteremia</b>				
Urinary tract	65 (34)	407 (38)	Reference	Reference
Skin and soft tissue	4 (2)	43 (4)	0.59 (.21–1.71)	.33 <sup>g</sup>
Biliary	16 (8)	168 (15)	0.60 (.34–1.07)	.08
Intra-abdominal	35 (18)	271 (25)	0.80 (.52–1.25)	.33

Table 1 continued.

Variables on Day 1 of Bacteremia	ESBL Positive (n = 194)	ESBL Negative (n = 1094)	Odds Ratio (95% CI)	P Value
Catheter-related	57 (29)	143 (13)	2.48 (1.66–3.72)	<.001 <sup>f,g</sup>
Bone and/or joint	1 (<1)	10 (1)	0.62 (.08–4.95)	.66
Pneumonia	16 (8)	57 (5)	1.75 (.95–3.23)	.07 <sup>f,g</sup>

Data for ESBL status are presented as No. (%) or mean±SD.

Abbreviations: CI, confidence interval; ESBL, extended-spectrum  $\beta$ -lactamase; HIV, human immunodeficiency virus; SD, standard deviation.

<sup>a</sup> Excluding chemotherapy or immunosuppression for solid organ transplants.

<sup>b</sup> Chronic obstructive pulmonary disease, emphysema, tracheostomy dependent.

<sup>c</sup> Central venous catheter or dialysis catheter.

<sup>d</sup> [http://www.cdc.gov/nhsn/pdfs/ps-analysis-resources/phenotype\\_definitions.pdf](http://www.cdc.gov/nhsn/pdfs/ps-analysis-resources/phenotype_definitions.pdf).

<sup>e</sup> Colombia (1), Costa Rica (1), El Salvador (2), Honduras (4), Mexico (3), Panama (1), China (3), Iran (1), Jordan (1), Kuwait (4), Qatar (4), Saudi Arabia (10), United Arab Emirates (5), Bangladesh (2), India (7), Pakistan (5), Egypt (2), Greece (2), Turkey (3). An additional 8 and 9 ESBL-positive and ESBL-negative patients, respectively, were hospitalized internationally in a non-high-burden region in the 6 months preceding bacteremia.

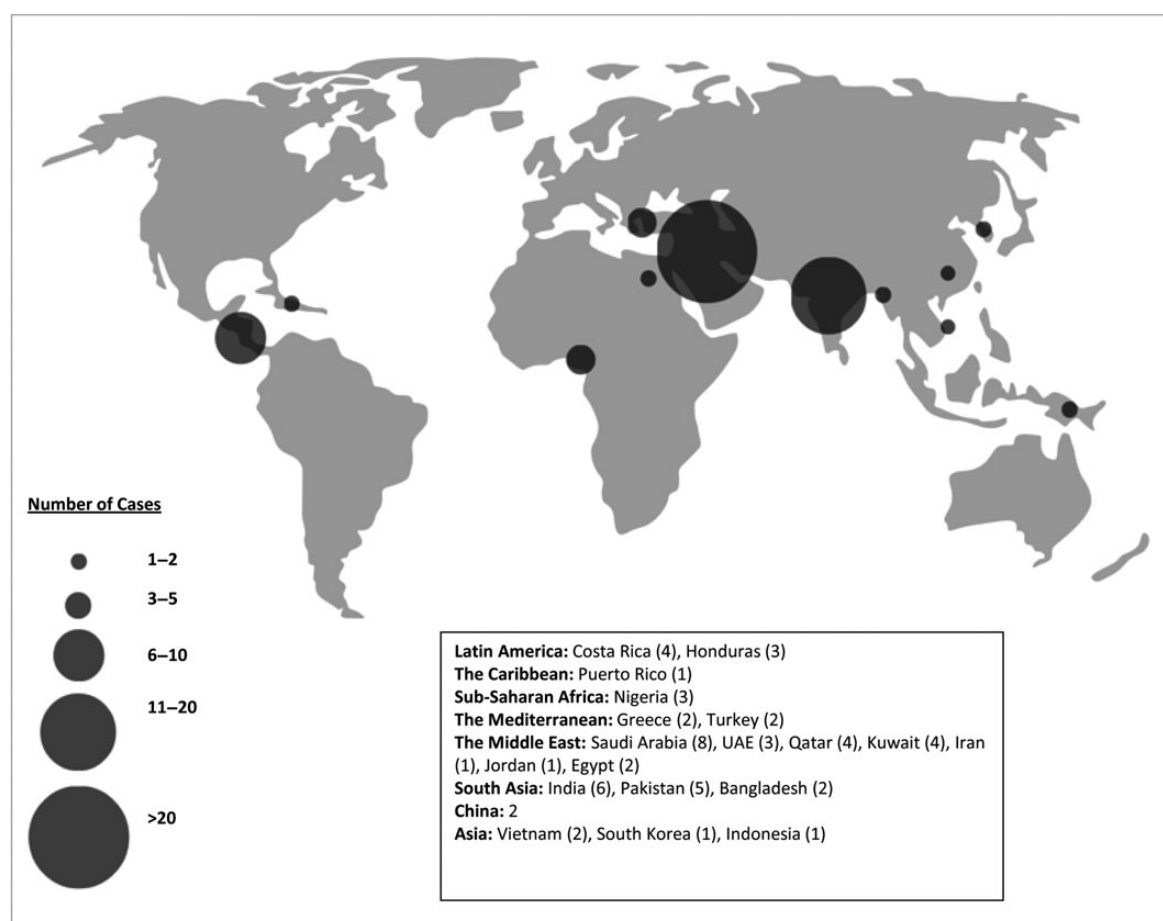
<sup>f</sup> Significant in multivariable analysis using stepwise selection with backwards elimination at an  $\alpha$  level of .05. Among variables that were significant in multivariable analysis, 1 variable, a history of multidrug-resistant *Pseudomonas* species, demonstrated qualitative confounding (univariable and multivariable odds ratios, 1.62 and 0.08, respectively).

<sup>g</sup> Retained in final multivariable model using lasso regression.

one-third of the original 68 patients, as well as an additional 78 ESBL cases, as “ESBL positive,” warranting empiric therapy with agents covering ESBL-producing bacteria.

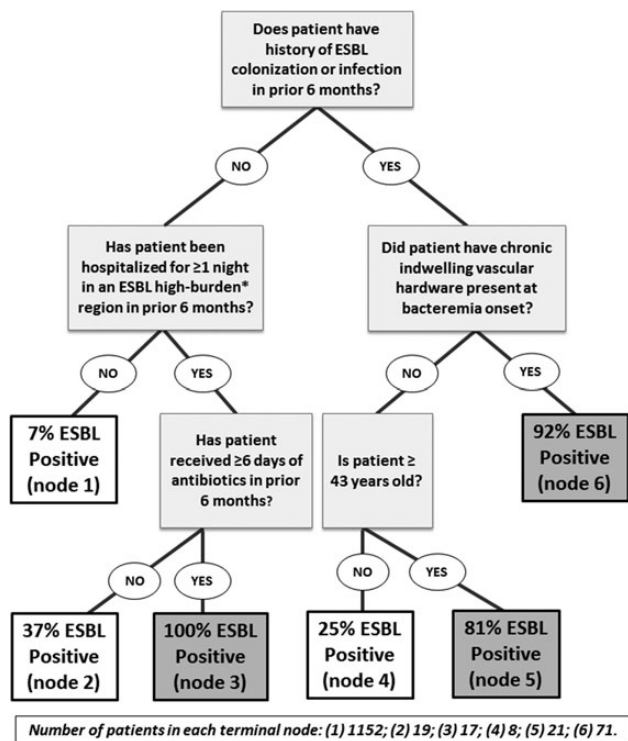
### Sensitivity Analyses

Approximately 45% (86/194) of patients with ESBL-positive bacteremia were classified in terminal node 1 as ESBL negative,



**Figure 1.** Distribution of recent international healthcare exposure among extended-spectrum  $\beta$ -lactamase (ESBL)-positive cases: 57 of 194 ESBL-positive patients had a recent international healthcare exposure, defined as hospitalization for 1 or more nights outside the United States in the 6 months preceding ESBL bacteremia. Abbreviation: UAE, United Arab Emirates.





**Figure 2.** Clinical decision tree to predict a bacteremic patient's likelihood of infection with an extended-spectrum  $\beta$ -lactamase (ESBL)-producing organism at the time of organism genus and species identification. 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.

compromising decision tree sensitivity. We performed sensitivity analyses on the subset of 1152 patients who lacked the 2 strongest study risk factors of prior ESBL infection or colonization history and recent international hospitalization in an ESBL high-burden region. We first refit a classification tree to this subset of data, and the resulting tree failed to branch (sensitivity 0%, C-statistic 0.50), consistent with truncation at terminal node 1 in the original tree. We also performed random forest analyses, a methodology that is less easily interpretable than binary recursive partitioning because it generates many bootstrapped classification trees, but that yields estimates of the most important classification variables [13, 20]. In random forest analysis on the data subset, no variables were strongly predictive of ESBL-positive bacteremia. An ROC curve generated from a logistic regression including the 3 most important variables yielded a C-statistic of 0.53. As definitions of “high burden” may reasonably differ, we also modeled international hospitalization to include all of Asia, as well as to include all countries without region restriction. Discriminatory performance remained similar to the original model in both analyses (C-statistics both 0.78).

## DISCUSSION

Timely identification of ESBL bacteremia can improve clinical outcomes while minimizing the unnecessary use of broad-spectrum antibiotics. Yet despite advances in rapid diagnostics, most clinical microbiology laboratories still require at least 24 additional hours from the time of organism identification to confirmation of ESBL production. Empirically treating serious gram-negative infections therefore remains a clinical challenge and leaves clinicians to balance the risks of ineffective agents against unnecessarily broad empiric antibiotic therapy on an ad hoc basis. A user-friendly clinical decision tree to determine a bacteremic patient's likelihood of infection with an ESBL-producing bacteria could assist clinicians with selecting appropriate empiric treatment at the time of organism identification.

From a dataset of >30 demographic and clinical variables, we developed a decision tree with 5 predictors: prior history of ESBL colonization or infection; presence of chronic indwelling vascular hardware; age (model-derived dichotomization at 43 years); recent hospitalization in an ESBL high-burden region; and total antibiotic exposure in the prior 6 months (model-derived dichotomization at 6 days). Patients classified as ESBL positive by the tree were 90.8% likely to be true ESBL cases (positive predictive value), and patients classified as negative were 91.9% likely to be true ESBL-negative cases (negative predictive value).

Our findings highlight the utility of recursive partitioning as a predictive modeling tool. In multivariable logistic regression, a high number of variables remained associated with ESBL-positive bacteremia, complicating efforts to translate statistical findings into practical application. Converting logistic regression coefficients into a risk score may have partially addressed this concern, but the resulting model would likely have been cumbersome to implement at the bedside. In contrast, a decision tree is generally intuitive and does not require tallying across variables. Moreover, recursive partitioning possesses attractive methodological features, including the ability to accommodate higher-order variable interactions and to generate automatic breakpoints for continuous variables [14, 21]. Perhaps most important, although decision trees yield categorical predictions (generally decided by majority rule in the terminal node), the strength of these predictions is quantifiable through terminal node impurities. As such, like risk scores, decision trees are flexible to differing risk-aversion attitudes, as well as to prioritizing sensitivity or specificity. For example, in a septic patient with a predicted 25% probability of ESBL-positive infection, it may be reasonable to prescribe empiric carbapenem therapy despite decision tree classification as ESBL negative. As with any methodological tool, classification trees can help to guide, but cannot replace, clinical judgment. The comfort level of clinicians, the clinical appearance of patients, and institutional treatment guidelines are necessary to fine-tune decisions.

Of note, a subset of ESBL-positive cases lacked a prior ESBL history and recent international hospitalization in an ESBL high-burden region and were classified by the tree as ESBL negative. Additional analyses suggested that no study variables were strongly discriminatory among this subset of patients. The poor predictive nature of healthcare-associated variables within this patient subset may suggest a high proportion of community-acquired ESBL infections. Indeed, although risk factors for ESBLs have traditionally focused on the healthcare setting, increasing reports describe the community as an important ESBL reservoir [22–26], with documented person-to-person transmission in the community and in households (predominantly *E. coli* sequence type 131) [27–29]. There is also evidence that livestock operations and food-supplying animals may be a source of ESBL-producing infections [30–33]. Additional information on community-associated exposures and isolate strain type were unavailable for these patients, unfortunately precluding further exploration of this hypothesis.

Our study has several limitations. First, this was a single-center study, and our results should be validated in other cohorts. Our results may not be generalizable to patients in other populations, particularly in regions with high ESBL prevalence. Second, recent international hospitalization was evaluated through a “yes/no” nursing intake questionnaire, which despite hospital policy to inquire of all patients may have been inconsistent during the study period. Selective questioning of patients perceived as higher risk could have artificially inflated the importance of this exposure. However, the association remained significant across calendar years, including later years when we expected greater policy compliance. Third, we recognize that individuals may define “high burden” international regions differently and that ESBL geographic prevalence changes over time. Sensitivity analyses yielded similar discriminatory performance under varied regional definitions, however, suggesting that the model was robust to these differences. Fourth, to reduce outcome misclassification, we restricted our study to *E. coli* and *Klebsiella* species, as the Centers for Disease Control and Prevention screening methodology to test for ESBL production is limited to these organisms [34]. As a result, our tree’s performance has only been validated from the point of genus and species identification of these common ESBL-producing organisms. If our tree is validated by others and evaluated in broader clinical practice, however, it may be reasonable at gram-negative confirmation to initiate carbapenem therapy in patients at high predicted risk of ESBL infection. Finally, despite our best attempt to gather detailed previous clinical data on all patients across health networks in the EPIC Care Everywhere network, due to the retrospective nature of this study there were likely missing data that could lead to exposure misclassification, although we would not expect this to be differential by ESBL status. In light of the decision tree’s intended real-world use, however, its performance under the practical constraints of missing data is arguably relevant. As the use of

electronic health records that interface across institutions becomes more widespread, these challenges may lessen.

Overall, our findings suggest that a clinical decision tree can be used to estimate, at the time of gram-negative organism identification, a bacteremic patient’s likelihood of infection with an ESBL-producing bacteria. These predictions may assist empiric treatment decisions, to optimize clinical outcomes while reducing administration of overly broad antibiotic agents that can select for further resistance emergence. The machine learning methodology relied upon in this study has been rarely utilized in the clinical infectious diseases literature but may offer a practical, user-friendly output for addressing important diagnostic questions.

## Notes

**Disclaimer.** The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH).

**Financial support.** This work was supported by the National Institute of Allergy and Infectious Diseases of the NIH (award number U1AI104681). The work was also supported in part by grants from the NIH (K24-AI079040 to A. D. H. and K24-AI080942 to E. L.).

**Potential conflicts of interest.** All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

## References

1. Spellberg B, Guidos R, Gilbert D, et al. The epidemic of antibiotic-resistant infections: a call to action for the medical community from the Infectious Diseases Society of America. *Clin Infect Dis* 2008; 46:155–64.
2. Paterson DL, Bonomo RA. Extended-spectrum beta-lactamases: a clinical update. *Clin Microbiol Rev* 2005; 18:657–86.
3. Menashe G, Borer A, Yagupsky P, et al. Clinical significance and impact on mortality of extended-spectrum beta lactamase-producing Enterobacteriaceae isolates in nosocomial bacteremia. *Scand J Infect Dis* 2001; 33:188–93.
4. Kang CI, Kim SH, Park WB, et al. Bloodstream infections due to extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae*: risk factors for mortality and treatment outcome, with special emphasis on antimicrobial therapy. *Antimicrob Agents Chemother* 2004; 48:4574–81.
5. 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.
6. Tamma PD, Han JH, Rock C, et al. Carbapenem therapy is associated with improved survival compared with piperacillin-tazobactam for patients with extended-spectrum beta-lactamase bacteremia. *Clin Infect Dis* 2015; 60:1319–25.
7. Armand-Lefevre L, Angebault C, Barbier F, et al. Emergence of imipenem-resistant gram-negative bacilli in intestinal flora of intensive care patients. *Antimicrob Agents Chemother* 2013; 57:1488–95.
8. McLaughlin M, Advincula MR, Malczynski M, et al. Correlations of antibiotic use and carbapenem resistance in Enterobacteriaceae. *Antimicrob Agents Chemother* 2013; 57:5131–3.
9. Dodemont M, De Mendonca R, Nonhoff S, Roisin S, Denis O. Performance of the Verigene gram-negative blood culture assay for rapid detection of bacteria and resistance determinants. *J Clin Microbiol* 2014; 52:3085–7.
10. Denis B, Lafaurie M, Donay JL, et al. Prevalence, risk factors, and impact on clinical outcome of extended-spectrum beta-lactamase-producing *Escherichia coli* bacteraemia: a five-year study. *Int J Infect Dis* 2015; 39:1–6.
11. Van Aken S, Lund N, Ahl J, Odenholt I, Tham J. Risk factors, outcome and impact of empirical antimicrobial treatment in extended-spectrum beta-lactamase-producing *Escherichia coli* bacteraemia. *Scand J Infect Dis* 2014; 46:753–62.
12. Nguyen ML, Toye B, Kanji S, Zvonar R. Risk factors for and outcomes of bacteremia caused by extended-spectrum  $\beta$ -lactamase-producing *Escherichia coli* and *Klebsiella* species at a Canadian tertiary care hospital. *Can J Hosp Pharm* 2015; 68:136–43.
13. 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 Methods* 2009; 14:323–48.
14. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. Boca Raton, Florida: CRC/Chapman & Hall, 1984.

15. Centers for Disease Control and Prevention. Antimicrobial resistant phenotype definitions. Available at: [http://www.cdc.gov/nhsn/pdfs/ps-analysis-resources/phenotype\\_definitions.pdf](http://www.cdc.gov/nhsn/pdfs/ps-analysis-resources/phenotype_definitions.pdf). Accessed 7 April 2016.
16. Ostholm-Balkhed A, Tarnberg M, Nilsson M, Nilsson LE, Hanberger H, Hallgren A. Travel-associated faecal colonization with ESBL-producing Enterobacteriaceae: incidence and risk factors. *J Antimicrob Chemother* **2013**; 68:2144–53.
17. 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**; 60:837–46.
18. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Series B* **1996**; 58:267–88.
19. Duda RO, Hart PE, Stork DG. Pattern classification. 2nd ed. Hoboken, NJ: Wiley Interscience, **2001**.
20. Chen S, Ishwaran H. Pathway hunting by random survival forests. *Bioinformatics* **2013**; 29:99–105.
21. Boulesteix AL, Janitza S, Hapfelmeier A, Van Steen K, Strobl C. On the term 'interaction' and related phrases in the literature on random forests. *Brief Bioinform* **2015**; 16:338–45.
22. Thaden JT, Fowler VG, Sexton DJ, Anderson DJ. Increasing incidence of extended-spectrum beta-lactamase-producing *Escherichia coli* in community hospitals throughout the southeastern United States. *Infect Control Hosp Epidemiol* **2016**; 37:49–54.
23. Hilty M, Betsch BY, Sogli-Stuber K, et al. Transmission dynamics of extended-spectrum beta-lactamase-producing Enterobacteriaceae in the tertiary care hospital and the household setting. *Clin Infect Dis* **2012**; 55:967–75.
24. Dayan N, Dabbah H, Weissman I, Aga I, Even L, Glikman D. Urinary tract infections caused by community-acquired extended-spectrum beta-lactamase-producing and nonproducing bacteria: a comparative study. *J Pediatr* **2013**; 163:1417–21.
25. Fan NC, Chen HH, Chen CL, et al. Rise of community-onset urinary tract infection caused by extended-spectrum beta-lactamase-producing *Escherichia coli* in children. *J Microbiol Immunol Infect* **2014**; 47:399–405.
26. Megged O. Extended-spectrum beta-lactamase-producing bacteria causing community-acquired urinary tract infections in children. *Pediatr Nephrol* **2014**; 29:1583–7.
27. Morgand M, Vimont S, Bleibtreu A, et al. Extended-spectrum beta-lactamase-producing *Escherichia coli* infections in children: are community-acquired strains different from nosocomial strains? *Int J Med Microbiol* **2014**; 304:970–6.
28. Madigan T, Johnson JH, Clabots C, et al. Extensive household outbreak of urinary tract infection and intestinal colonization due to extended-spectrum beta-lactamase-producing *Escherichia coli* sequence type 131. *Clin Infect Dis* **2015**; 61:e5–12.
29. Kaarme J, Molin Y, Olsen B, Melhus A, et al. Prevalence of extended-spectrum beta-lactamase-producing Enterobacteriaceae in healthy Swedish preschool children. *Acta Paediatr* **2013**; 102:655–60.
30. Lazarus B, Paterson DL, Mollinger JL, Rogers BA. Do human extraintestinal *Escherichia coli* infections resistant to expanded-spectrum cephalosporins originate from food-producing animals? A systematic review. *Clin Infect Dis* **2015**; 60:439–52.
31. Agerso Y, Jensen JD, Hasman H, Pedersen K. Spread of extended spectrum cephalosporinase-producing *Escherichia coli* clones and plasmids from parent animals to broilers and to broiler meat in a production without use of cephalosporins. *Foodborne Pathog Dis* **2014**; 11:740–6.
32. Olsen RH, Bisgaard M, Lohren U, Robineau B, Christensen H. Extended-spectrum beta-lactamase-producing *Escherichia coli* isolated from poultry: a review of current problems, illustrated with some laboratory findings. *Avian Pathol* **2014**; 43:199–208.
33. Voets GM, Fluit AC, Scharringa J, et al. Identical plasmid AmpC beta-lactamase genes and plasmid types in *E. coli* isolates from patients and poultry meat in the Netherlands. *Int J Food Microbiol* **2013**; 167:359–62.
34. Centers for Disease Control and Prevention. CDC laboratory detection of extended-spectrum  $\beta$ -lactamases (ESBLs). Available at: [http://www.cdc.gov/HAI/settings/lab/lab\\_esbl.html](http://www.cdc.gov/HAI/settings/lab/lab_esbl.html). Accessed 26 May 2016.