# PREDICTION MODELS FOR CARBAPENEM-RESISTANT *ENTEROBACTERIACEAE* (CRE) AND OTHER MULTIDRUG-RESISTANT GRAM-NEGATIVE (MDRGN) BACTERIA IN HEALTHCARE SETTINGS

by

Katherine E. Goodman, JD

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, MD

October, 2018

**ABSTRACT**

**Background.** Carbapenem-resistant *Enterobacteriaceae* (CRE) and other carbapenem-resistant organisms (CROs) pose urgent challenges to patient care. These bacteria are highly drug-resistant and are associated with significant attributable mortality. Current prevention strategies in United States (U.S.) healthcare facilities aim to reduce selective pressure from antibiotic exposure and to reduce patient-to-patient spread. These efforts are hampered by a lack of rapid and cost-effective diagnostics to identify these organisms. These diagnostic challenges leave basic epidemiological questions unanswered, including how many and which types of U.S. inpatients are asymptomatic carriers.

**Objectives.** We aimed to measure the prevalence of, and risk factors for, CRO colonization among high-risk U.S. hospitalized patients and to develop statistical and machine learning prediction models that could help to address existing diagnostic limitations.

**Methods.** To achieve these aims, we developed two study cohorts. The first, a one-year prospective cohort of Johns Hopkins Hospital (JHH) intensive care unit patients, screened patients for CRO carriage at unit admission. Isolates were speciated and molecularly characterized, and pre-admission exposure data were used to evaluate colonization risk factors and to develop predictive models of colonization with machine learning methodologies (Aim 1). The second, a retrospective cohort of JHH Gram-negative bacteremic patients, generated a

clinical decision tree (Aim 2) and a risk score (Aim 3) to predict whether infections were extended-spectrum B-lactamase (ESBL)-producing. ESBLs confer resistance to most antibiotics except carbapenems, and rapid identification can reduce unnecessary carbapenem administration. Through the lens of this real-world example, we methodologically compared these two prediction approaches (Aim 3).

**Results.** Aim 1 included 3,327 unit visits and 2,878 (87%) admission swabs. Our study found that 7.5% of patients were perirectally colonized with CROs and identified high organism and resistance mechanism diversity. Many variables were significantly associated with carriage, but resulting models were not highly predictive. Aims 2 and 3 analyzed 1,288 bacteremic patients and yielded higher performing prediction models for ESBL infection. We found that decision trees and risk scores performed similarly in our case study, but they offered different strengths and limitations.

**Conclusions.** Statistical and machine learning prediction models offer an important complement to microbiological diagnostics. They can circumvent existing resource and practical constraints, but high biological heterogeneity can compromise their performance. Increasing familiarity with these methods, as well as refining distinctions between causal inference and prediction, may improve statistical tools for identifying colonization or infection with CROs and other multidrug-resistant bacteria.

**ADVISORS AND READERS**

**Aaron M. Milstone, M.D., M.H.S. (Thesis Advisor)**
Associate Professor
Department of Pediatric Infectious Diseases
Johns Hopkins University School of Medicine

**Jonathan E. Golub, Ph.D. (Academic Advisor)**
Associate Professor
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health

**Justin Lessler, Ph.D. (Thesis Reader)**
Associate Professor
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health

**Christopher D. Heaney, Ph.D. (Thesis Reader)**
Associate Professor
Department of Environmental Health and Engineering
Johns Hopkins Bloomberg School of Public Health

**Bryan Lau, Ph.D. (Alternate Thesis Reader)**
Associate Professor
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health

**Alan L. Scott, Ph.D. (Alternate Thesis Reader)**
Professor
Department of Molecular Microbiology and Immunology
Johns Hopkins Bloomberg School of Public Health

**ACKNOWLEDGMENTS**

First and foremost, I would like to thank my thesis advisor, Dr. Aaron Milstone. Aaron has provided me with invaluable mentoring and practical guidance during his four years as my advisor, particularly from a clinician and healthcare epidemiologist's perspective. He has always made himself available to me for weekly (and more) meetings, phone calls and Skype, and many a late-night or weekend email with a research question. Aaron was not originally my advisor when I entered the program, but after taking his Healthcare Epidemiology class at the beginning of my second year, it reconfirmed for me that I knew I wanted to perform my dissertation research at the intersection of antibiotic resistance and healthcare epidemiology. He graciously agreed to serve as my thesis advisor even though he was already over-extended. I feel so fortunate that it was his Epicenter grant, a somewhat unexpected development at the beginning of my third year, that made it possible to perform the CRE and CRO colonization screening included in Aim 1 of my thesis — my dream goal when I entered the program, but for which I presumed I would never have the available data. I am grateful for Aaron's mentorship, and his friendship, and his continued support these past few years has been integral to my success.

I would also like to thank Dr. Jonathan Golub, who began as my advisor when I entered the Program and who has continued to serve as my academic advisor during the completion of my PhD studies. Jonathan has been an incredibly helpful mentor as I have navigated Departmental requirements and expectations, especially during the first two years of the program, and he graciously supported me in my pursuit of a healthcare epidemiology dissertation project even

though that is not his area of research. Moreover, I would like to thank Dr. Pranita Tamma and Dr. Justin Lessler, both members of my thesis committee. I am incredibly appreciative of their guidance and willingness to help me on projects, both dissertation-related and other papers, during my PhD studies, and I am proud of the variety of projects that we have collaborated on together. Our work has been highly interdisciplinary, drawing on their respective clinical and methodological expertise, and I am grateful for the opportunity to have had their input as my dissertation aims evolved. I would also like to thank Dr. Christopher Heaney and Dr. Amber D'Souza for their valuable feedback as members of my School-wide or Departmental oral exam committees.

Much of this work would not have been possible without the CDC Epicenter Study and the contributions of so many individuals at the Johns Hopkins Hospital and elsewhere in the University and School of Medicine. I am indebted to Dr. Trish Simner and other members of her laboratory for their microbiological expertise and swab processing, as well as Avi Gadala in the Healthcare Epidemiology and Infection Control Department for his many countless hours helping us to assemble our Aim 1 cohort. Many thanks as well to Verna Scheeler and her research assistant, Michael Anderson, for their help with swabbing compliance and many other aspects of study coordination; to Carrie Billman for her wisdom on clinical and nursing questions as I ascertained the location of data within the electronic medical record; and to Ani Voskertchian, Dina Khamash, and other members of the Department of Pediatrics, Division of Infectious Diseases, for their data assistance and research meeting feedback. In particular, I would also like to thank Dr. Eili Klein for his many, many months of work in extracting and

assembling much of our EMR data, as well as other individuals with whom he worked, including Matthew Toerper, Dr. Scott Levin, Dr. Diego Martinez, and Jiarui Cai. Finally, I would like to acknowledge other members of the CDC Epicenter study, including Dr. Clare Rock, Dr. Sara Cosgrove, and Dr. Lisa Maragakis, for their study involvement and very helpful feedback and guidance on Aim 1.

I was very fortunate to be well-funded during my studies and research. I owe a great amount of gratitude to the Department of Epidemiology, the Ellen B. Gold Fund Award, the AHRQ R36 Dissertation Research Award, the Sherrilyn and Ken Fisher Center for Environmental Infectious Diseases, and HRSA for a Public Health Trainee Award for providing funding for my academic pursuits and dissertation research.

In the Department of Epidemiology, I would like to thank my friends and colleagues, particularly members of my cohort, including Drs. Denali Boon, Marisa Hast, Fidel Desir, Megan Wallace, Talia Quandelacy, and Paul Sacamano. In addition, I would like to thank Dr. Becca Pierce for her guidance and insights, especially as one of the only other Epidemiology students to pursue a healthcare epidemiology dissertation. I would also like to thank the staff of the Department of Epidemiology, in particular Matt Miller for his guidance and patience in helping me successfully submit (and re-submit) my R36 application to AHRQ, and Fran Burman.

Last, but certainly not least, thank you to my friends and family for their unwavering support and encouragement in my pursuit and completion of my PhD. I am so fortunate to have parents

who have always been so supportive of my educational goals, even when they included the somewhat untraditional path of shifting from law to epidemiology, and without whom I have no doubt that I would not have had the courage to leave a good career behind in order to pursue my passion. I have also been blessed by a very close-knit extended family, including my maternal aunt and uncle, as well as my cousins, Sarah and Laura Claggett, who have been a source of great support and levity for more than three decades. Most importantly, I thank my husband, Simon Consedine, who I met in law school and who has now, for more than ten years, been my biggest champion and (incredibly) patient listener who lets me animatedly talk about health, science, and bacteria at every available opportunity. We were blessed to have a now toddler, Nora, during the middle of my PhD program, and his and my parents' support has been instrumental in helping to ensure that I completed my dissertation on time, for which I am eternally grateful.

**TABLE OF CONTENTS**

**LIST OF TABLES**

**LIST OF FIGURES**

## 1. INTRODUCTION

### 1.1. MultiDrug-Resistant Gram-Negative Bacteria in United States (U.S.) Healthcare Settings

Healthcare-associated infections (HAIs) represent a significant patient safety threat. Each day in United States (U.S.) hospitals, one of every 25 patients experiences an HAI [1], causing more than 90,000 patient deaths per year and in excess of $20 billion in direct healthcare costs [2]. Many of these infections are due to multidrug-resistant (MDR) bacteria, which patients can acquire from healthcare exposure, e.g., in a hospital, through multiple routes: environmental or medical device/equipment contamination, healthcare worker-mediated transmission, and/or direct contact with other patients [3].

Promisingly, in the last 15 years HAI rates due to MDR Gram-positive bacteria (e.g., methicillin-resistant *Staphylococcus aureus* (MRSA), vancomycin-resistant *Enterococci* (VRE)) have stabilized or declined in many healthcare populations and settings [4-6]. During this same time period, however, the threat from MDR Gram-negative (MDRGN) bacteria has increased. Of particular concern are carbapenem-resistant *Enterobacteriaceae* (CRE) and other carbapenem-resistant organisms (CROs), which are resistant to almost all commonly used antibiotics and can impose attributable mortality rates approaching 50 percent [7]. Informally termed the "Nightmare Bacteria" [8], CRE have been designated as one of only three urgent antibacterial threats by the U.S. Centers for Disease Control and Prevention (CDC), due to their increasing domestic prevalence, role in numerous nosocomial outbreaks, and poor patient outcomes [9].

Although still rare in the U.S., with a crude incidence of approximately 3 infections per 100,000 U.S. residents, the domestic incidence of CRE has more than tripled in the last 15 years [10].

## 1.2. Carbapenem-Resistant Organisms (CROs)

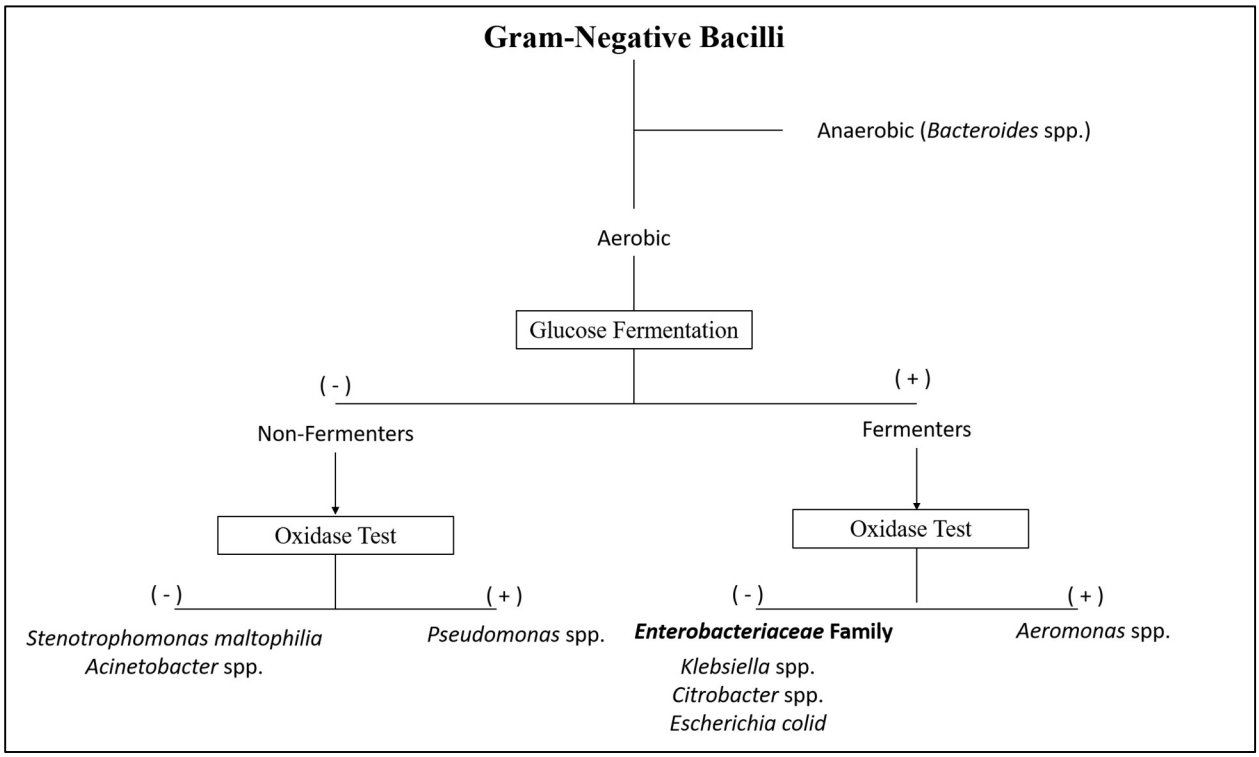
### *1.2.A. Carbapenem-Resistant Enterobacteriaceae (CRE)*

*Enterobacteriaceae* are a large family of Gram-negative bacteria that colonize the gastrointestinal (GI) tract of humans and animals (Figure 1.1) [11-13]. *Enterobacteriaceae* spread relatively easily among humans, including via environmental, food, and hand carriage routes [14]. Many *Enterobacteriaceae* species comprise important nosocomial pathogens, including *Escherichia coli*, *Klebsiella* spp., and *Enterobacter* spp. Although patients naturally carry, i.e., are colonized with, antibiotic-susceptible *Enterobacteriaceae* in their GI tracts, clinical (i.e., symptomatic) infections can occur in any body site, particularly the urinary tract or bloodstream [15].

The carbapenem antibacterial class includes four antibiotics: imipenem, meropenem, doripenem, and ertapenem. *Enterobacteriaceae* may be categorized as carbapenem-resistant based upon antimicrobial susceptibility test results following interpretative criteria established in the U.S. by the Clinical and Laboratory Standards Institute (CLSI). An *Enterobacteriaceae* currently qualifies as carbapenem-resistant with a minimum inhibitory concentration (MIC) of ≥4 ug/ml for imipenem/meropenem/doripenem, or ≥2 ug/ml for ertapenem [16, 17]. For surveillance purposes, the CDC defines a CRE as an *Enterobacteriaceae* that is resistant to any

carbapenem (or that has documented possession of a carbapenemase, *see* Section 1.2.A.1) [18]. Importantly, antimicrobial susceptibility testing does not reveal the molecular mechanism conferring carbapenem resistance.

**Figure 1.1. Classification of Gram-Negative Bacilli, with Select Genera and Species***




*Excludes Gram-negative bacilli with special growth requirements

### *1.2.A.1. CRE Sub-Types*

CRE are further distinguishable by their carbapenem resistance mechanisms. Carbapenem resistance may be achieved through production of a carbapenemase (CP-CRE) or through production of other β-lactamase enzymes coupled with altered membrane permeability (non-

CP-CRE) (Table 1.1). Carbapenemases are enzymes that hydrolyze carbapenems and are encoded by genes frequently carried on mobile genetic elements such as plasmids and transposons [19-22]. Multiple types and sub-types of carbapenemases exist, which are generally categorized by their molecular hydrolytic sites [23, 24]. Carbapenemases can transfer between bacteria, including across species and families. Epidemiologists fear that carbapenemase genes may become established in bacterial strains with heightened transmissibility and/or pathogenic potential. Due to the mobility of their plasmid-mediated resistance mechanisms and propensity for causing outbreaks, the literature generally expresses greater concern for CP-CRE [25]. CP-CRE are also associated with poorer clinical outcomes compared to non-CP-CRE [26].

Non-CP-CRE arise through mechanisms other than carbapenemase production. These mechanisms most commonly include production of extended-spectrum β-lactamases (ESBLs) and/or AmpC cephalosporinases (AmpCs), in combination with cell membrane alterations. ESBLs are generally plasmid-encoded, and AmpCs generally result from either deregulation or induction of a chromosomally-encoded *ampC* gene or from acquisition of a plasmid-encoded AmpC gene (e.g., $bla_{CMY}$) [23, 27]. ESBLs and AmpCs are sometimes capable of hydrolyzing carbapenems at very low levels [28, 29], but in combination with membrane impermeability (principally arising from mutated or absent porin channels) or increased drug efflux [30, 31], are effective at preventing carbapenems from reaching their binding targets at sufficient concentrations to exert their antibacterial effect [30-34].

**Table 1.1. Underlying Molecular Resistance Mechanisms in Carbapenem-Resistant *Enterobacteriaceae* (CRE)***

| **Carbapenemase-Producing CRE (CP-CRE)** | |
|---|---|
| **MOLECULAR RESISTANCE MECHANISM** | **EXAMPLES** |
| Ambler Class A: Serine Carbapenemases | KPC; SME; GES |
| Ampler Class B: Metallo-β-Lactamases | NDM; IMP; VIM |
| Ambler Class D: Serine Carbapenemases | OXA-23, -24, -48 |
| **Non-Carbapenemase-Producing CRE (non-CP-CRE)** | |
| Ambler Class A: ESBL + porin alterations or increased efflux pump activity | CTX-M-group; SHV-type; TEM-type |
| Ambler Class C: AmpC hyper-expression + porin alterations or increased efflux pump activity | Inducible AmpC β-lactamases (*Enterobacter* spp., *Citrobacter* spp., *Serratia* marcescens, *et cetera*.) or *E. coli* possessing chromosomal *ampC* promoter and/or attenuator mutations |
| Plasmid-mediated AmpC (*Ambler class C*) + porin alterations or increased efflux pump activity | CMY; ACT; DHA |

**Abbreviations**: *Klebsiella pneumoniae* carbapenemase (KPC); New Delhi metallo-β-lactamase (NDM); Imipenemase metallo-β-lactamase (IMP); Verona integron-encoded metallo-β-lactamase (VIM); Oxacillinase metallo-β-lactamase (OXA); Extended-spectrum β-lactamase (ESBL); AmpC cephalosporinase (AmpC)

*Adapted from Goodman KE, Simner PJ, Tamma PD, Milstone AM. Infection control implications of heterogeneous resistance mechanisms in carbapenem-resistant *Enterobacteriaceae* (CRE). Expert Rev Anti Infect Ther. 2016 Jan;14(1):95-108.

### *1.2.B. Other, Non-CRE Carbapenem-Resistant Organisms (CROs)*

Although CRE have received significant attention [9], other non-*Enterobacteriaceae* bacteria commonly found in the GI tract are an emerging reservoir of carbapenem resistance [35, 36]. Of these, the most notable are the glucose non-fermenting Gram-negative bacilli (glucose non-fermenters). This bacterial family encompasses important healthcare-associated pathogens, including *Pseudomonas aeruginosa* and *Acinetobacter baumannii* (**Figure 1.1**). In addition to the GI tract, these organisms frequently colonize additional body sites, including the respiratory

tract and nares/pharynx [37-39]. Glucose non-fermenters can cause serious clinical infections, particularly in patients with underlying comorbidities. The CDC has designated MDR glucose non-fermenters, including those which are carbapenem-resistant, as a “serious” public health threat [9].

Concurrent with the general increase in carbapenem resistance among glucose non-fermenters has been the continued dissemination of carbapenemases among these bacteria. All carbapenemase classes have now been detected in *Acinetobacter* and *Pseudomonas* species [40], with the Ambler class D carbapenemases (e.g, OXA-23) most common in the former and the class B carbapenemases (e.g., VIM) most common in the latter [41]. Carbapenemases can also migrate between *Enterobacteriaceae* and glucose non-fermenting bacteria, and carbapenemases traditionally observed in CRE (e.g., KPC, NDM) have been detected in glucose non-fermenters [42, 43]. Like CP-CRE, carbapenemase-producing glucose non-fermenters have caused nosocomial outbreaks with resulting serious clinical infections [44, 45]. Collectively, CP-CRE and carbapenemase-producing glucose non-fermenters constitute carbapenemase-producing organisms (CPOs).

### 1.3. Strategies for Reducing Carbapenem Resistance in Healthcare Settings

Strategies for minimizing carbapenem resistance in healthcare settings are complementary and multi-factorial. They include efforts to reduce selective pressure from antibiotic exposure, e.g.,

antibiotic stewardship programs, and infection control policies to reduce nosocomial spread following resistance emergence or CRO acquisition.

### *1.3.A. Reduce Carbapenem Use*

For a variety of nosocomial pathogens, including CRE and other CROs, antibiotic exposure is a risk factor for subsequent colonization or infection [46-50]. The processes underlying this association are complex. CROs may endogenously arise in patients receiving antibiotic therapy through a variety of mechanisms, including "classical" selective pressure but also through ecological effects on the microbiota.

First, CROs may arise in previously-susceptible patient microbiota during antibiotic exposure when *de novo* mutations confer resistance to select antibiotics or antibiotic classes. Carbapenemases, which are often encoded on mobile genetic elements such as transposons or plasmids, do not primarily arise from *de novo* mutation in bacterial cells and are not anticipated to endogenously arise during antibiotic therapy [41]. In contrast, non-CP-CROs can emerge if a patient with an infection receives carbapenem therapy, and the organism mutates to acquire a mutation in a membrane protein (i.e., porin), such that carbapenems can no longer cross into the bacterial cell [51-54]. Multiple studies and case reports have documented non-CP-CRE emergence following prolonged carbapenem treatment [55-59], and a large surveillance study identified non-CP-CRE acquisition as statistically associated with imipenem exposure [60].

Second, carbapenem therapy may exert selective pressure that enriches pre-existing CRE or other CROs. This selective pressure may allow CROs to flourish and may account for some of the extensively documented association between carbapenem exposure and subsequent carbapenem-resistant infection, particularly in endemic settings.

Third, carbapenem exposure may increase the risk of both CRO transmission and acquisition, although available data are more robust for CRE. Recent antibiotic exposure prior to CRE detection is associated with higher CRE bacterial load in fecal samples [61]. Higher bacterial burden may increase the risk of environmental contamination or transmission to others [62]. Conversely, antibiotic exposure may increase susceptibility to colonization among CRE-negative patients by disrupting the intestinal microbial ecology, leading to a loss of colonization resistance. Healthy commensal microbiota provide ecologic niche protection against externally encountered pathogens through competitive exclusion and inflammation-mediated mechanisms [63, 64]. Any antibiotic, including carbapenems, has the potential to cause gut dysbiosis [65].

Achieving reductions in carbapenem use is challenging. Currently, carbapenems are neither available orally nor first-line therapy for most infections, and are therefore already used more judiciously than many other antibacterial classes. Nevertheless, important opportunities for reducing carbapenem administration remain. Although less drug-resistant than CROs, ESBL-producing organisms are a much more common cause of healthcare-associated infections [66].

Importantly, because ESBLs can hydrolyze most broad-spectrum β-lactam antibiotics *except* carbapenems, carbapenems are generally the treatment of choice for serious ESBL infections [54, 67]. However, clinical microbiology laboratories currently require at least 24 additional hours from the time of microbial genus and species identification to confirm ESBL production; many other U.S. laboratories, including at the Johns Hopkins Hospital, have abandoned routine ESBL confirmatory testing due to resource constraints and because it is no longer required by CLSI guidance [16, 68]. Consequently, clinicians must generally select treatment for patients with serious Gram-negative infections without knowing whether the infection is ESBL-producing. These empirical treatment decisions must balance the risk of ineffective therapy against the resistance risks posed by overly broad antibiotic treatment. Erring towards caution in critically-ill patients and administering carbapenems understandably, but nevertheless concerningly, contributes to further carbapenem resistance development.

### *1.3.B. Reduce CRO Introduction and Transmission*

While reducing the emergence of carbapenem resistance is one important element of CRO prevention, mitigating the transmissibility of CROs that do emerge or are otherwise introduced into healthcare facilities is of equal importance. Due to the significant nosocomial outbreak potential and clinical harms posed by CRE, the CDC has published comprehensive guidance for controlling CRE in healthcare settings; many of these policies are also relevant to other CROs [69]. Recommendations include rigorous hand hygiene and environmental cleaning compliance, contact precautions for colonized or infected patients (e.g., use of disposable gowns and gloves

for all patient contact), and the consideration of staff and/or geographic cohorting of CP-CRE cases (e.g., physically separating patients or providing dedicated equipment or healthcare workers) [69].

The arguably missing element to this otherwise robust strategy is routine active surveillance for CRE or CRO colonization among high-risk populations (e.g., intensive care unit patients), which is a cornerstone of HAI prevention for many MDR Gram-positive organisms [70, 71]. Unlike infections, which are detected in the routine course of clinical care, identifying asymptomatic CRE colonization requires surveillance, generally with rectal or peri-rectal swabs. The CDC recommends limited CRE colonization surveillance, including patients with recent international hospitalization and contacts of confirmed CP-CRE cases [69]. Compliance with surveillance swabbing is often poor and wide-scale screening in U.S. hospitals remain uncommon. Unfortunately, recent studies suggest that existing policies and current recommendations may miss many CRE colonized patients [72]. Diagnostic challenges, including workflow and cost considerations, underlie the current impracticality of universal surveillance for CRE and other CROs. Moreover, due to the existing lack of screening, colonization prevalence among inpatient populations remains poorly quantified — i.e., the specific data necessary to inform evidence-based assessments about whether and how to expand screening programs. Better understanding the burden of CRO and CPO carriage and who to target for screening will be critical for justifying increased use of these surveillance assays.

### 1.4. Diagnostics for CROs and Other MDRGNs

Successful carbapenem resistance prevention strategies, whether targeted at reducing antibiotic use or at interrupting transmission, hinge on the ability to rapidly, and cost-effectively, identify patients harboring CROs and other MDRGNs, as well as to distinguish their underlying resistance mechanisms. While diagnostics for CROs have remained relatively unchanged, CPO diagnostics are advancing rapidly. Multiple assays for carbapenemase detection became clinically available in recent years, and others are in late-stage research testing [73, 74]. Each available method, either for CROs more broadly or CPOs more specifically, however, has limitations. These diagnostic limitations threaten our ability to effectively characterize and respond to evolving CRO and CPO epidemiology.

Well-established methods are available for identifying CRE or other CROs, whether from clinical specimens or surveillance swabs. Because CROs, including CRE, are defined in reference to their antibiotic resistance profile, however, identifying them requires organism culture and subsequent antibiotic susceptibility testing (AST). In the case of surveillance swabs, for example, from the time of laboratory receipt, culturing (e.g., direct plating or broth enrichment) requires ~24 hours, followed by an additional 24 or more hours to obtain AST results [75, 76]. This process is relatively inexpensive but has a long turnaround time, depending on the method may impose significant workflow and personnel constraints, and may exhibit reduced sensitivity in swabs with low organism burden. After two days, it also remains unknown whether organisms are carbapenemase-producing.

Various phenotypic tests are available to identify carbapenemases, but these tests have historically performed poorly or had other important limitations. Newer assays for identifying carbapenemase production have achieved high sensitivity and specificity, and the CDC now encourages laboratories with sufficient institutional capacity to test CRE isolates for carbapenemases for epidemiological purposes [69]. For example, the Carba NP, which produces a color change in the presence of carbapenemase-induced imipenem hydrolysis, is highly accurate in CRE (with the exception of identifying OXA-48-like carbapenemases) and has a fast turnaround time of approximately two hours [75, 77]. However, the Carba NP test uses reagents with short shelf-lives and may be challenging to incorporate into routine laboratory workflows. Commercial versions of this and similar tests resolve this problem but, at a cost of $2-10 per test, may be prohibitively expensive for wide-scale screening [73]. The modified carbapenem inactivation method (mCIM), another phenotypic assay for carbapenemase production, is accurate, relatively simple to implement, and only costs ~$1 per test. However, it requires overnight incubation, thereby lengthening processing time [78, 79].

Importantly, each of the phenotypic tests discussed above require cultured isolates. In theory, one could perform these tests following organism identification without waiting for AST results (e.g., ~12-24 hours after specimen collection), but given the presumed low CPO prevalence in U.S. facilities, this approach would be high-resource, low-yield, especially for screening. On the other hand, waiting for carbapenem resistance confirmation before performing subsequent

phenotypic testing produces an approximately 2.5-3 day turnaround time. Particularly for shorter ICU admissions, patients could be discharged before isolate processing is complete. Alternative methods (e.g., carbapenemase chromogenic media) can shorten this window to approximately 24 hours from laboratory receipt, but they are labor-intensive and with poorer sensitivity/specificity profiles they would likely require subsequent confirmatory testing [75, 77].

Molecular methods identify carbapenemase genes through nucleic acid amplification and can be performed directly from specimens or swabs without the need to culture. These methods can produce results within a few hours from swab collection, but at potentially prohibitive expense. DNA isolation and extraction can also be challenging for some assays (e.g., Check-Points), and although faster methods such as the CARBA-R are available, they currently cost ~$50 per test and run fewer samples per batch [80, 81]. Moreover, because amplification (without prior organism enrichment) can fail if bioburden is insufficient, real-time molecular methods for surveillance screening may perform better with rectal rather than perirectal swabs [82]. This shift would entail attendant practical challenges and need for floor staff re-education.

Of note, although some of the preceding limitations will likely lessen as technology advances, others are likely to persist. As experience with older MDRGNs, including ESBLs, illustrates, most confirmatory testing still requires an additional 24 hours after AST results are known. Moreover, the various available tests are either resource-intensive or expensive, or both; and

despite the availability of various molecular assays, they are not routinely utilized. As healthcare facilities consider whether and how to implement CRO screening programs, they have and will continue to grapple with these challenges.

### 1.5. Statistical and Machine Learning Approaches for Circumventing MDRGN Diagnostic Limitations

Statistical models for identifying CRO, CRE and other MDRGNs can help to address existing diagnostic limitations. Logistic regression-derived risk prediction scores are common in the clinical infectious disease and healthcare epidemiology literature. They have been used, for example, to predict risk of drug-resistant pneumonia in patients presenting to emergency departments [83] and risk of VRE bloodstream infection following hematopoietic stem cell transplantation [84]. These models are attractive because they are relatively simply to develop and allow individual users to alter score cut-points in order to prioritize sensitivity or specificity for a given application. However, risk scores may also be cumbersome to implement manually depending upon the number of included variables and complexity of end-user calculations. Of note, significance-based variable selection procedures, as are commonly used to build multivariable logistic regression models and risk scores, may perform poorly for prediction [85]. Alternative methods more suited to prediction goals (e.g., lasso regression [86]) are available.

Other approaches for developing clinical prediction tools include tree-based machine learning methodologies. Classification and regression tree (CART) analysis or “recursive partitioning,” for

example, is an approach rarely utilized in the clinical antibiotic resistance literature that may be helpful as a practical predictive modeling tool in these circumstances [87]. Decision trees use a series of continuous and/or categorical input variables to predict an outcome. Classification trees predict categorical outcomes, and regression trees predict continuous outcomes. These trees, which rely on branching logic, are simple to interpret and generally do not require calculations [87-89]. Decision trees are prone to overfitting, however, potentially limiting their generalizability without appropriate methodological corrections and internal or external validation. Random forests analysis, an ensemble approach that produces many bootstrapped decision trees, addresses many of these challenges although it does not produce a single, easily interpretable tree [90, 91]. In hospitalized settings, random forests have commonly been used in emergency departments to build electronic triage models for risk-stratifying patients [92, 93].

### 1.6. Summary of Current Challenges and Gaps in Knowledge

Current prevention strategies for CROs and CPOs are hampered by a lack of rapid and cost-effective diagnostics to identify these and other MDRGN organisms. Statistical approaches can help fill knowledge gaps that are critical to reducing CRO acquisition and spread in U.S. healthcare facilities.

First, available data on the prevalence of carbapenemase-producing and non-carbapenemase-producing CROs among U.S. inpatients are extremely limited. Standard antimicrobial

susceptibility testing (AST) cannot distinguish carbapenem resistance mechanisms, and current CLSI guidance does not recommend routine confirmation of carbapenemase production among CRO isolates. Many U.S. diagnostic laboratories therefore do not distinguish CROs from CPOs for therapeutic decision-making, rendering data from clinical specimens largely uninformative [68, 94, 95]. Data on CRO and CPO colonization prevalence are even more limited, because unlike clinical infections, colonization detection requires active surveillance. With the exception of specific targeted screening (e.g., patients recently hospitalized internationally), most facilities do not perform routine CRO or CPO active surveillance due to existing diagnostic limitations. The net result of these challenges are missed opportunities for intervening to reduce CRO introduction and healthcare-associated transmission. Better understanding the burden of CRO and CPO carriage, and risk factors and predictors for colonization (which may differ from clinical infection), may improve algorithms for identifying and screening patients at highest carriage risk.

Second, diagnostic challenges persist even for older, less resistant MDRGNs (e.g. ESBLs), which contributes to ongoing CRO emergence by facilitating unnecessary carbapenem use. Carbapenems remain the treatment of choice for serious ESBL infections, but specimen collection-to-ESBL-confirmation still requires approximately three days; many other laboratories have abandoned ESBL confirmatory testing entirely. Additionally, although molecular methods are faster, commonly used Gram-negative molecular panels do not include, or generally at most only identify one of, the ESBL gene groups [96]. Consequently, clinicians

must select empirical antibiotic treatment for patients with serious Gram-negative infections without knowing whether the causative organism is ESBL-producing, while balancing the risk of ineffective therapy against overly broad antibiotic treatment. The ability to discriminate rapidly between ESBL-positive and ESBL-negative infections at the point of treatment initiation could reduce inappropriate carbapenem administration.

Third, prediction models can help to address these diagnostic challenges by complementing microbiological methods. Recognizing that wide-scale CRO screening remains impractical, these models can help to identify who to target for screening in order to allocate resources efficiently. Clinically, they can also help to guide who can be more conservatively treated with carbapenem-sparing antibiotic regimens in the absence of, or while awaiting, microbiological confirmation. Yet, these approaches remain under, or at a minimum non-optimally, utilized in the clinical infectious disease and healthcare epidemiology literature. Older, more traditional methods (e.g., logistic regression-derived risk scores) are common, with desirable attributes. However, they frequently rely on significance-based variable selection procedures, which may perform poorly for prediction. Newer methods and machine learning-based approaches, e.g., decision trees, have practical utility and avoid significance-based, “risk factor” variable selection pitfalls. But, despite these benefits, they remain uncommon. Increasing familiarity with these methods, as well as refining distinctions between causal inference and prediction, may improve statistical tools for identifying CROs and other MDRGNs.

### 1.7. Specific Aims

To address these challenges and knowledge gaps, we aim to:

1. **Estimate the prevalence of and risk factors for, and develop predictive models of, colonization with carbapenem-resistant organisms (CROs) and carbapenemase-producing organisms (CPOs) at hospital unit admission.** This study will utilize a prospective cohort of approximately 3,000 patients admitted to the medical intensive care unit (MICU) and comprehensive transplant unit (CTU) at The Johns Hopkins Hospital (JHH). Patients will have peri-rectal swabs screened for CROs and CPOs at unit admission. Comprehensive data on pre-admission exposures will be collected through automated extraction from electronic medical records (EMRs). We will develop predictive models for colonization through classification and regression tree (CART) analysis, a machine learning methodology that yields a user-friendly decision tree algorithm, in order to identify patients at high risk of CRO carriage (composite and stratified by bacterial class and carbapenemase status). Alternative approaches (e.g., random forests analysis) will also be investigated as necessary.

2. **Develop a user-friendly clinical decision tree to predict infection with an extended-spectrum β-lactamase (ESBL)–producing organism, in order to reduce unnecessary carbapenem administration.** This study includes patients ≥18 years of age with bacteremia due to *Escherichia coli* or *Klebsiella* species, from October 2008 to March

2015 at JHH. The retrospective cohort comprises a total of 1288 bacteremic patients, of whom 194 (15%) are infected with ESBL-producers (all study isolates underwent ESBL confirmatory testing). Pre-infection demographic and clinical data will be collected through manual chart review. Recursive partitioning will be used to generate a practical, user-friendly decision tree to determine the likelihood that a bacteremic patient is infected with an ESBL-producer.

3. **Methodologically compare decision trees versus risk scores for predicting drug-resistant colonization or infection.** This case study will use the same cohort as Aim 2 in order to derive a clinical risk score to predict the likelihood that a bacteremic patient is infected with an ESBL-producer. Risk score and decision tree performance will be compared, and we will review their practical and methodological attributes. By comparing these methods through the lens of a real example, we aim to offer highly accessible, general guiding principles for when clinicians or hospital epidemiologists might consider each approach to complement existing microbiological diagnostics.

## REFERENCES


1. HAI Data and Statistics [https://www.cdc.gov/hai/surveillance/index.html]
2. Stone PW: Economic burden of healthcare-associated infections: an American perspective. *Expert review of pharmacoeconomics & outcomes research* 2009, 9(5):417-422.
3. Goodman KE, Simner PJ, Tamma PD, Milstone AM: Infection control implications of heterogeneous resistance mechanisms in carbapenem-resistant Enterobacteriaceae (CRE). *Expert review of anti-infective therapy* 2016, 14(1):95-108.
4. Kanjilal S, Sater MRA, Thayer M, Lagoudas GK, Kim S, Blainey PC, Grad YH: Trends in Antibiotic Susceptibility in Staphylococcus aureus in Boston, Massachusetts, from 2000 to 2014. *J Clin Microbiol* 2018, 56(1).
5. Kallen AJ, Mu Y, Bulens S, Reingold A, Petit S, Gershman K, Ray SM, Harrison LH, Lynfield R, Dumyati G *et al*: Health care-associated invasive MRSA infections, 2005-2008. *Jama* 2010, 304(6):641-648.
6. Mendes RE, Castanheira M, Farrell DJ, Flamm RK, Sader HS, Jones RN: Longitudinal (2001–14) analysis of enterococci and VRE causing invasive infections in European and US hospitals, including a contemporary (2010–13) analysis of oritavancin in vitro potency. *Journal of Antimicrobial Chemotherapy* 2016, 71(12):3453-3458.
7. Falagas ME, Tansarli GS, Karageorgopoulos DE, Vardakas KZ: Deaths attributable to carbapenem-resistant Enterobacteriaceae infections. *Emerging infectious diseases* 2014, 20(7):1170-1175.
8. Maggie Fox: 'Nightmare bacteria' are trying to spread in the U.S., CDC says. In: *NBC News.* 2018.
9. CDC: Antibiotic resistance threats in the United States, 2013. In.; 2014: 1-114.
10. Guh AY, Bulens SN, Mu Y, et al.: Epidemiology of carbapenem-resistant enterobacteriaceae in 7 us communities, 2012-2013. *JAMA* 2015, 314(14):1479-1487.

11. Muscarella LF: Risk of transmission of carbapenem-resistant Enterobacteriaceae and related "superbugs" during gastrointestinal endoscopy. *World journal of gastrointestinal endoscopy* 2014, 6(10):457-474.

12. Laurent C, Rodriguez-Villalobos H, Rost F, Strale H, Vincent JL, Deplano A, Struelens MJ, Byl B: Intensive care unit outbreak of extended-spectrum beta-lactamase-producing Klebsiella pneumoniae controlled by cohorting patients and reinforcing infection control measures. *Infection control and hospital epidemiology* 2008, 29(6):517-524.

13. Piednoir E, Thibon P, Borderan GC, Godde F, Borgey F, Le Coutour X, Parienti JJ: Long-term clinical and economic benefits associated with the management of a nosocomial outbreak resulting from extended-spectrum beta-lactamase-producing Klebsiella pneumoniae. *Critical care medicine* 2011, 39(12):2672-2677.

14. Oteo J, Perez-Vazquez M, Campos J: Extended-spectrum [beta]-lactamase producing Escherichia coli: changing epidemiology and clinical impact. *Current opinion in infectious diseases* 2010, 23(4):320-326.

15. Tischendorf J, de Avila RA, Safdar N: Risk of infection following colonization with carbapenem-resistant Enterobactericeae: A systematic review. *American journal of infection control* 2016, 44(5):539-543.

16. (CLSI) CLSI: Performance Standards for Antimicrobial Susceptibility Testing: Twentieth Informational Supplement (M100-S20). In. Wayne, PA: CLSI; 2010.

17. (CLSI) CLSI: Performance Standards for Antimicrobial Susceptibility Testing: Twenty-Second Informational Supplement (M100-S22). In. Wayne, PA: CLSI; 2012.

18. FAQs about Choosing and Implementing a CRE Definition [http://www.cdc.gov/hai/organisms/cre/definition.html]

19. Rasmussen BA, Bush K, Keeney D, Yang Y, Hare R, O'Gara C, Medeiros AA: Characterization of IMI-1 beta-lactamase, a class A carbapenem-hydrolyzing enzyme from Enterobacter cloacae. *Antimicrobial agents and chemotherapy* 1996, 40(9):2080-2086.

20. Conlan S, Deming C, Tsai YC, Lau AF, Dekker JP, Korlach J, Segre JA: Complete Genome Sequence of a Klebsiella pneumoniae Isolate with Chromosomally Encoded Carbapenem Resistance and Colibactin Synthesis Loci. *Genome announcements* 2014, 2(6).

21. Espedido B, Iredell J, Thomas L, Zelynski A: Wide dissemination of a carbapenemase plasmid among gram-negative bacteria: implications of the variable phenotype. *Journal of clinical microbiology* 2005, 43(9):4918-4919.

22. Hu L, Zhong Q, Shang Y, Wang H, Ning C, Li Y, Hang Y, Xiong J, Wang X, Xu Y *et al*: The prevalence of carbapenemase genes and plasmid-mediated quinolone resistance determinants in carbapenem-resistant Enterobacteriaceae from five teaching hospitals in central China. *Epidemiology and infection* 2014, 142(9):1972-1977.

23. Bush K, Jacoby GA: Updated functional classification of beta-lactamases. *Antimicrobial agents and chemotherapy* 2010, 54(3):969-976.

24. Ambler RP: The structure of beta-lactamases. *Philosophical transactions of the Royal Society of London Series B, Biological sciences* 1980, 289(1036):321-331.

25. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Group NCSP, Henderson DK, Palmore TN, Segre JA: Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med* 2012, 4(148):148ra116.

26. Tamma PD: Comparing the Outcomes of Patients With Carbapenemase-Producing and Non-Carbapenemase-Producing Carbapenem-Resistant Enterobacteriaceae Bacteremia. 2017, 64(3):257-264.

27. Woodford N, Reddy S, Fagan EJ, Hill RL, Hopkins KL, Kaufmann ME, Kistler J, Palepou MF, Pike R, Ward ME *et al*: Wide geographic spread of diverse acquired AmpC beta-lactamases among Escherichia coli and Klebsiella spp. in the UK and Ireland. *The Journal of antimicrobial chemotherapy* 2007, 59(1):102-105.

28. Livermore DM, Yang YJ: Beta-lactamase lability and inducer power of newer beta-lactam antibiotics in relation to their activity against beta-lactamase-inducibility mutants of Pseudomonas aeruginosa. *The Journal of infectious diseases* 1987, 155(4):775-782.

29. Girlich D, Poirel L, Nordmann P: Do CTX-M beta-lactamases hydrolyse ertapenem? *The Journal of antimicrobial chemotherapy* 2008, 62(5):1155-1156.

30. Livermore DM, Woodford N: The beta-lactamase threat in Enterobacteriaceae, Pseudomonas and Acinetobacter. *Trends in microbiology* 2006, 14(9):413-420.

31. Doumith M, Ellington MJ, Livermore DM, Woodford N: Molecular mechanisms disrupting porin expression in ertapenem-resistant Klebsiella and Enterobacter spp. clinical isolates from the UK. *The Journal of antimicrobial chemotherapy* 2009, 63(4):659-667.

32. Livingstone D, Gill MJ, Wise R: Mechanisms of resistance to the carbapenems. *The Journal of antimicrobial chemotherapy* 1995, 35(1):1-5.

33. Poirel L, Heritier C, Spicq C, Nordmann P: In vivo acquisition of high-level resistance to imipenem in Escherichia coli. *Journal of clinical microbiology* 2004, 42(8):3831-3833.

34. Oteo J, Delgado-Iribarren A, Vega D, Bautista V, Rodriguez MC, Velasco M, Saavedra JM, Perez-Vazquez M, Garcia-Cobos S, Martinez-Martinez L *et al*: Emergence of imipenem resistance in clinical Escherichia coli during therapy. *International journal of antimicrobial agents* 2008, 32(6):534-537.

35. Hong DJ, Bae IK, Jang I-H, Jeong SH, Kang H-K, Lee K: Epidemiology and Characteristics of Metallo-β-Lactamase-Producing Pseudomonas aeruginosa. *Infect Chemother* 2015, 47(2):81-97.

36. Kim UJ, Kim HK, An JH, Cho SK, Park K-H, Jang H-C: Update on the Epidemiology, Treatment, and Outcomes of Carbapenem-resistant Acinetobacter infections. *Chonnam Med J* 2014, 50(2):37-44.

37. Cobos-Trigueros N, Sole M, Castro P, Torres JL, Hernandez C, Rinaudo M, Fernandez S, Soriano A, Nicolas JM, Mensa J *et al*: Acquisition of Pseudomonas aeruginosa and its resistance phenotypes in critically ill medical patients: role of colonization pressure and antibiotic exposure. *Critical care (London, England)* 2015, 19:218.

38. Gniadek TJ, Carroll KC, Simner PJ: Carbapenem-Resistant Non-Glucose-Fermenting Gram-Negative Bacilli: the Missing Piece to the Puzzle. *Journal of Clinical Microbiology* 2016, 54(7):1700-1710.

39. Marchaim D, Navon-Venezia S, Schwartz D, Tarabeia J, Fefer I, Schwaber MJ, Carmeli Y: Surveillance cultures and duration of carriage of multidrug-resistant Acinetobacter baumannii. *J Clin Microbiol* 2007, 45(5):1551-1555.

40. Viau R, Frank KM, Jacobs MR, Wilson B, Kaye K, Donskey CJ, Perez F, Endimiani A, Bonomo RA: Intestinal Carriage of Carbapenemase-Producing Organisms: Current Status of Surveillance Methods. *Clinical microbiology reviews* 2016, 29(1):1-27.

41. Queenan AM, Bush K: Carbapenemases: the versatile beta-lactamases. *Clinical microbiology reviews* 2007, 20(3):440-458, table of contents.

42. Robledo IE, Aquino EE, Vazquez GJ: Detection of the KPC gene in Escherichia coli, Klebsiella pneumoniae, Pseudomonas aeruginosa, and Acinetobacter baumannii during a PCR-based nosocomial surveillance study in Puerto Rico. *Antimicrob Agents Chemother* 2011, 55(6):2968-2970.

43. Hong DJ, Bae IK, Jang IH, Jeong SH, Kang HK, Lee K: Epidemiology and Characteristics of Metallo-beta-Lactamase-Producing Pseudomonas aeruginosa. *Infect Chemother* 2015, 47(2):81-97.

44. Zavascki AP, Gaspareto PB, Martins AF, Goncalves AL, Barth AL: Outbreak of carbapenem-resistant Pseudomonas aeruginosa producing SPM-1 metallo-{beta}-lactamase in a teaching hospital in southern Brazil. *J Antimicrob Chemother* 2005, 56(6):1148-1151.

45. Yang HY, Lee HJ, Suh JT, Lee KM: Outbreaks of imipenem resistant Acinetobacter baumannii producing OXA-23 beta-lactamase in a tertiary care hospital in Korea. *Yonsei medical journal* 2009, 50(6):764-770.

46. Ben-David D, Masarwa S, Navon-Venezia S, Mishali H, Fridental I, Rubinovitch B, Smollan G, Carmeli Y, Schwaber MJ, Israel PCWG: Carbapenem-resistant Klebsiella pneumoniae in post-acute-care facilities in Israel. *Infection control and hospital epidemiology* 2011, 32(9):845-853.

47. Swaminathan M, Sharma S, Poliansky Blash S, Patel G, Banach DB, Phillips M, LaBombardi V, Anderson KF, Kitchel B, Srinivasan A *et al*: Prevalence and risk factors for acquisition of carbapenem-resistant Enterobacteriaceae in the setting of endemicity. *Infection control and hospital epidemiology* 2013, 34(8):809-817.

48. Wiener-Well Y, Rudensky B, Yinnon AM, Kopuit P, Schlesinger Y, Broide E, Lachish T, Raveh D: Carriage rate of carbapenem-resistant Klebsiella pneumoniae in hospitalised patients during a national outbreak. *The Journal of hospital infection* 2010, 74(4):344-349.

49. Marchaim D, Chopra T, Bhargava A, Bogan C, Dhar S, Hayakawa K, Pogue JM, Bheemreddy S, Blunden C, Shango M *et al*: Recent exposure to antimicrobials and carbapenem-resistant Enterobacteriaceae: the role of antimicrobial stewardship. *Infection control and hospital epidemiology* 2012, 33(8):817-830.

50. Chitnis AS, Caruthers PS, Rao AK, Lamb J, Lurvey R, Beau De Rochars V, Kitchel B, Cancio M, Torok TJ, Guh AY *et al*: Outbreak of carbapenem-resistant enterobacteriaceae at a long-term acute care hospital: sustained reductions in transmission through active surveillance and targeted interventions. *Infection control and hospital epidemiology* 2012, 33(10):984-992.

51. Warner DM, Yang Q, Duval V, Chen M, Xu Y, Levy SB: Involvement of MarR and YedS in carbapenem resistance in a clinical isolate of Escherichia coli from China. *Antimicrobial agents and chemotherapy* 2013, 57(4):1935-1937.

52. Chia JH, Siu LK, Su LH, Lin HS, Kuo AJ, Lee MH, Wu TL: Emergence of carbapenem-resistant Escherichia coli in Taiwan: resistance due to combined CMY-2 production and porin deficiency. *Journal of chemotherapy* 2009, 21(6):621-626.

53. Yang Q, Wang H, Sun H, Chen H, Xu Y, Chen M: Phenotypic and genotypic characterization of Enterobacteriaceae with decreased susceptibility to carbapenems: results from large hospital-based surveillance studies in China. *Antimicrobial agents and chemotherapy* 2010, 54(1):573-577.

54. Tamma PD, Han JH, Rock C, Harris AD, Lautenbach E, Hsu AJ, Avdic E, Cosgrove SE, Antibacterial Resistance Leadership G: Carbapenem therapy is associated with improved survival compared with piperacillin-tazobactam for patients with extended-spectrum beta-lactamase bacteremia. *Clin Infect Dis* 2015, 60(9):1319-1325.

55. Trevino M, Moldes L, Martinez-Lamas L, Varon C, Regueiro BJ: Carbapenem-resistant Enterobacter cloacae and the emergence of metallo-beta-lactamase-producing strains in a third-level hospital (Santiago de Compostela, NW Spain). *European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology* 2009, 28(10):1253-1258.

56. Kim J, Lee JY, Kim SI, Song W, Kim JS, Jung S, Yu JK, Park KG, Park YJ: Rates of fecal transmission of extended-spectrum beta-lactamase-producing and carbapenem-resistant Enterobacteriaceae among patients in intensive care units in Korea. *Annals of laboratory medicine* 2014, 34(1):20-25.

57. Grobner S, Linke D, Schutz W, Fladerer C, Madlung J, Autenrieth IB, Witte W, Pfeifer Y: Emergence of carbapenem-non-susceptible extended-spectrum beta-lactamase-producing Klebsiella pneumoniae isolates at the university hospital of Tubingen, Germany. *Journal of medical microbiology* 2009, 58(Pt 7):912-922.

58. Garcia-Fernandez A, Miriagou V, Papagiannitsis CC, Giordano A, Venditti M, Mancini C, Carattoli A: An ertapenem-resistant extended-spectrum-beta-lactamase-producing Klebsiella pneumoniae clone carries a novel OmpK36 porin variant. *Antimicrobial agents and chemotherapy* 2010, 54(10):4178-4184.

59. Lartigue MF, Poirel L, Poyart C, Reglier-Poupet H, Nordmann P: Ertapenem resistance of Escherichia coli. *Emerging infectious diseases* 2007, 13(2):315-317.

60. Armand-Lefevre L, Angebault C, Barbier F, Hamelet E, Defrance G, Ruppe E, Bronchard R, Lepeule R, Lucet JC, El Mniai A *et al*: Emergence of imipenem-resistant gram-negative bacilli in intestinal flora of intensive care patients. *Antimicrob Agents Chemother* 2013, 57(3):1488-1495.

61. Cheng VC, Chan JF, Wong SC, Chen JH, Tai JW, Yan MK, Kwan GS, Tse H, To KK, Ho PL *et al*: Proactive infection control measures to prevent nosocomial transmission of carbapenem-resistant Enterobacteriaceae in a non-endemic area. *Chin Med J (Engl)* 2013, 126(23):4504-4509.

62. Rock C, Thom KA, Masnick M, Johnson JK, Harris AD, Morgan DJ: Frequency of Klebsiella pneumoniae carbapenemase (KPC)-producing and non-KPC-producing Klebsiella species contamination of healthcare workers and the environment. *Infection control and hospital epidemiology* 2014, 35(4):426-429.

63. Sassone-Corsi M, Raffatellu M: No vacancy: how beneficial microbes cooperate with immunity to provide colonization resistance to pathogens. *Journal of immunology* 2015, 194(9):4081-4087.

64. Spees AM, Lopez CA, Kingsbury DD, Winter SE, Baumler AJ: Colonization resistance: battle of the bugs or Menage a Trois with the host? *PLoS pathogens* 2013, 9(11):e1003730.

65. Nord CE, Sillerstrom E, Wahlund E: Effect of tigecycline on normal oropharyngeal and intestinal microflora. *Antimicrobial agents and chemotherapy* 2006, 50(10):3375-3380.

66. Woodworth KR, Walters MS, Weiner LM, Edwards J, Brown AC, Huang JY, Malik S, Slayton RB, Paul P, Capers C *et al*: Vital Signs: Containment of Novel Multidrug-Resistant Organisms and Resistance Mechanisms - United States, 2006-2017. *MMWR Morb Mortal Wkly Rep* 2018, 67(13):396-401.

67. Paterson DL, Bonomo RA: Extended-spectrum beta-lactamases: a clinical update. *Clinical microbiology reviews* 2005, 18(4):657-686.

68. Clinical and Laboratory Standards Institute: Performance standards for antimicrobial susceptibility testing, 28th ed. Supplement M100S. In. Wayne, PA: CLSI; 2018.

69. CDC: Facility Guidance for Control of Carbapenem-resistant Enterobacteriaceae (CRE) – November 2015 Update CRE Toolkit. In. Atlanta, Georgia: CDC; 2015.

70. Lin MY, Hayden MK, Lyles RD, Lolans K, Fogg LF, Kallen AJ, Weber SG, Weinstein RA, Trick WE: Regional Epidemiology of Methicillin-Resistant Staphylococcus aureus Among Adult Intensive Care Unit Patients Following State-Mandated Active Surveillance. *Clin Infect Dis* 2018, 66(10):1535-1539.

71. Isenman H, Michaels J, Fisher D: Global variances in infection control practices for vancomycin resistant Enterococcus - results of an electronic survey. *Antimicrob Resist Infect Control* 2016, 5:41.

72. Shimasaki T, Segreti J, Tomich A, Kim J, Hayden MK, Lin MY: Active screening and interfacility communication of carbapenem-resistant Enterobacteriaceae (CRE) in a tertiary-care hospital. *Infect Control Hosp Epidemiol* 2018:1-5.

73. Tamma PD, Opene BN, Gluck A, Chambers KK, Carroll KC, Simner PJ: Comparison of 11 Phenotypic Assays for Accurate Detection of Carbapenemase-Producing Enterobacteriaceae. *J Clin Microbiol* 2017, 55(4):1046-1055.

74. Boutal H, Vogel A, Bernabeu S, Devilliers K, Creton E, Cotellon G, Plaisance M, Oueslati S, Dortet L, Jousset A *et al*: A multiplex lateral flow immunoassay for the rapid identification of NDM-, KPC-, IMP- and VIM-type and OXA-48-like carbapenemase-producing Enterobacteriaceae. *J Antimicrob Chemother* 2018, 73(4):909-915.

75. Simner PJ, Gilmour MW, DeGagne P, Nichol K, Karlowsky JA: Evaluation of five chromogenic agar media and the Rosco Rapid Carb screen kit for detection and confirmation of carbapenemase production in Gram-negative bacilli. *J Clin Microbiol* 2015, 53(1):105-112.

76. Simner PJ, Martin I, Opene B, Tamma PD, Carroll KC, Milstone AM: Evaluation of Multiple Methods for the Detection of Gastrointestinal Colonization of Carbapenem-Resistant Organisms from Rectal Swabs. *J Clin Microbiol* 2016.

77. Huang TD, Berhin C, Bogaerts P, Glupczynski Y: Comparative evaluation of two chromogenic tests for rapid detection of carbapenemase in Enterobacteriaceae and in Pseudomonas aeruginosa isolates. *J Clin Microbiol* 2014, 52(8):3060-3063.

78. Pierce VM, Simner PJ, Lonsway DR, Roe-Carpenter DE, Johnson JK, Brasso WB, Bobenchik AM, Lockett ZC, Charnot-Katsikas A, Ferraro MJ *et al*: Modified Carbapenem Inactivation Method for Phenotypic Detection of Carbapenemase Production among Enterobacteriaceae. *J Clin Microbiol* 2017, 55(8):2321-2333.

79. Tamma PD, Opene BNA, Gluck A, Chambers KK, Carroll KC, Simner PJ: Comparison of 11 Phenotypic Assays for Accurate Detection of Carbapenemase-Producing Enterobacteriaceae. *Journal of Clinical Microbiology* 2017, 55(4):1046-1055.

80. Moore NM, Canton R, Carretto E, Peterson LR, Sautter RL, Traczewski MM: Rapid Identification of Five Classes of Carbapenem Resistance Genes Directly from Rectal Swabs by Use of the Xpert Carba-R Assay. *J Clin Microbiol* 2017, 55(7):2268-2275.

81. Tato M, Ruiz-Garbajosa P, Traczewski M, Dodgson A, McEwan A, Humphries R, Hindler J, Veltman J, Wang H, Canton R: Multisite Evaluation of Cepheid Xpert Carba-R Assay for Detection of Carbapenemase-Producing Organisms in Rectal Swabs. *J Clin Microbiol* 2016, 54(7):1814-1819.

82. O'Connor C, Kiernan MG, Finnegan C, O'Hara M, Power L, O'Connell NH, Dunne CP: An optimized work-flow to reduce time-to-detection of carbapenemase-producing Enterobacteriaceae (CPE) using direct testing from rectal swabs. *Bioengineered* 2017, 8(3):217-224.

83. Shorr AF, Zilberberg MD, Reichley R, Kan J, Hoban A, Hoffman J, Micek ST, Kollef MH: Validation of a clinical score for assessing the risk of resistant pathogens in patients with pneumonia presenting to the emergency department. *Clin Infect Dis* 2012, 54(2):193-198.

84. Webb BJ, Healy R, Majers J, Burr Z, Gazdik M, Lopansri B, Hoda D, Petersen FB, Ford C: Prediction of Bloodstream Infection Due to Vancomycin-Resistant Enterococcus in Patients Undergoing Leukemia Induction or Hematopoietic Stem-Cell Transplantation. *Clin Infect Dis* 2017, 64(12):1753-1759.

85. Lo A, Chernoff H, Zheng T, Lo SH: Why significant variables aren't automatically good predictors. *Proceedings of the National Academy of Sciences of the United States of America* 2015, 112(45):13892-13897.

86. Tibshirani R: Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B* 1996, 58:267-288.

87. Breiman L, Friedman J, Stone C, Olshen R: Classification and regression trees. Boca Raton, Florida: CRC/Chapman & Hall; 1984.

88. Duda RO, Hart PE, Stork DG: Pattern Classification, 2nd edn. United States: Wiley-Interscience; 2001.

89. Breiman L: Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statist Sci* 2001, 16(3):199-231.

90. Casanova R, Saldana S, Chew EY, Danis RP, Greven CM, Ambrosius WT: Application of random forests methods to diabetic retinopathy classification analyses. *PloS one* 2014, 9(6):e98587.

91. 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(4):323-348.

92. Levin S, Toerper M, Hamrock E, Hinson JS, Barnes S, Gardner H, Dugas A, Linton B, Kirsch T, Kelen G: Machine-Learning-Based Electronic Triage More Accurately Differentiates Patients With Respect to Clinical Outcomes Compared With the Emergency Severity Index. *Annals of emergency medicine* 2018, 71(5):565-574.e562.

93. Demsar J, Zupan B, Aoki N, Wall MJ, Granchi TH, Robert Beck J: Feature mining and predictive model construction from severe trauma patient's data. *International journal of medical informatics* 2001, 63(1-2):41-50.

94. CLSI: Performance Standards for Antimicrobial Susceptibility Testing; Twenty-Fifth Informational Supplement. In., vol. M100-S25; 2015.

95. FAQs about Choosing and Implementing a CRE Definition [http://www.cdc.gov/hai/organisms/cre/definition.html]

96. Dodemont M, De Mendonca R, Nonhoff C, 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(8):3085-3087.

**4.**

**2. PREVALENCE, RISK FACTORS, AND PREDICTIVE MODELS OF PERIRECTAL COLONIZATION WITH CARBAPENEM-RESISTANT *ENTEROBACTERIACEAE* (CRE) AND OTHER CARBAPENEM-RESISTANT ORGANISMS (CROS) AT HOSPITAL UNIT ADMISSION**


**Authors:**

Katherine E. Goodman[1], Patricia J. Simner[2], Eili Y. Klein[1,3-4], Abida Kazmi[2], Avinash Gadala[5], Clare Rock[5-6], Pranita D. Tamma[5,7], Sara E. Cosgrove[5-6], Lisa L. Maragakis[5-6], and Aaron M. Milstone[1,5,7], for the CDC Prevention Epicenters Program

**Institutions:**

1 – Johns Hopkins University Bloomberg School of Public Health, Department of Epidemiology, Baltimore, MD

2 – Johns Hopkins University School of Medicine, Department of Pathology, Division of Medical Microbiology, Baltimore, MD

3 – Johns Hopkins University School of Medicine, Department of Emergency Medicine, Baltimore, MD

4 – The Center for Disease Dynamics, Economics & Policy, Washington, D.C.

5 – The Johns Hopkins Hospital, Department of Hospital Epidemiology and Infection Control, Baltimore, MD

6 – Johns Hopkins University School of Medicine, Department of Medicine, Division of Infectious Diseases, Baltimore, MD

7 – Johns Hopkins University School of Medicine, Department of Pediatrics, Division of Infectious Diseases, Baltimore, MD

**ABSTRACT**

**Background.** Carbapenem-resistant organisms (CROs), and the subset of carbapenemase-producing organisms (CPOs), pose urgent challenges to patient care in United States facilities. Data on inpatient colonization prevalence are limited. Admission screening for carriage remains impractical for many acute care settings. Our objective was to measure the prevalence of, and risk factors for, CRO and CPO perirectal colonization at hospital unit admission, and to develop predictive models for organism carriage to inform targeted screening programs.

**Methods.** We performed a prospective cohort study of all patients admitted to the medical intensive care unit or comprehensive transplant unit at The Johns Hopkins Hospital between July 1, 2016 and July 1, 2017. Using an existing vancomycin-resistant *Enterococcus* surveillance program, admission perirectal swabs were screened for CROs and CPOs. All CPOs were also molecularly analyzed. Pre-admission exposure data were collected through automated extraction from electronic medical records. We analyzed colonization risk factors with logistic regression and developed predictive models with machine learning decision tree methodologies.

**Results.** The study included 3,327 unit visits, with 2,878 (87%) admission swabs available for processing. We found that 7.5% and 1.3% of patients had perirectal colonization with CROs and

CPOs, respectively. There was high organism and carbapenemase gene diversity among CPO isolates. Many clinical characteristics were associated with both CRO and CPO carriage, but resulting models did not highly predict colonization. High-risk sub-groups included patients with recent CRO-positive cultures who use proton-pump inhibitors.

**Conclusions.** In this high-risk inpatient population, CRO carriage was infrequent but higher than previously published estimates. Colonization was characterized by significant microbiological and risk factor heterogeneity, making prediction challenging.

**INTRODUCTION**

Carbapenem-resistant organisms (CROs) are often resistant to nearly all routinely used antibiotics with Gram-negative coverage, and resulting infections are associated with high morbidity and mortality [1-6]. Historically, carbapenem-resistant *Enterobacteriaceae* (CRE) have received significant attention [7], but non-*Enterobacteriaceae* organisms, most notably the glucose non-fermenters (e.g., *Acinetobacter baumannii*, *Pseudomonas aeruginosa*), are an additional and increasingly recognized reservoir of carbapenem resistance genes [8, 9]. Of particular concern among CROs are the subset of carbapenamase-producing organisms (CPOs), for which carbapenem resistance is generally plasmid-mediated and can transfer between organisms, including among different bacterial species and families. CPOs have been implicated in high-profile nosocomial outbreaks [10] and, at least among CRE, may be associated with poorer clinical outcomes [11].

CRO colonization is an important risk factor for subsequent infection. Admission screening for CRO and/or CPO carriage enables rapid isolation of colonized patients and may provide an opportunity for individualized care, such as targeted empiric antibiotic therapy or fecal transplantation [12-15]. The Centers for Disease Control and Prevention (CDC) recommends CRE colonization screening in limited instances [16], but most U.S. hospitals do not perform routine CRE or CRO screening. Given limited data on inpatient colonization prevalence in non-outbreak periods and current limitations of CRO and CPO diagnostics (e.g., cost, turn-around time), universal screening remains impractical for many acute care settings. Yet, recent CRE

data suggests that existing targeted screening policies may miss many colonized patients [17]. Better understanding the burden of CRO and CPO carriage, and risk factors and predictors for colonization (which may differ from clinical infection), may improve algorithms for identifying and screening patients at highest carriage risk. Our objective was to measure the prevalence of, and risk factors for, CRO and CPO perirectal colonization at hospital unit admission, and to develop predictive models for organism carriage to inform targeted screening programs.

## METHODS

### Study Setting and Population

This study included patients aged ≥ 16 years admitted to the Johns Hopkins Hospital (JHH) medical intensive care unit (MICU) or solid organ transplant unit (Transplant Unit), from July 1, 2016 – July 1, 2017. The MICU provides intensive care for patients with complex and multi-system medical illnesses, and the Transplant Unit provides intensive and intermediate care for adult abdominal organ and reconstructive transplant recipients, as well management of pre-transplant patients and patients with transplantation medical complications. Both units have a longstanding vancomycin-resistant *Enterococcus* (VRE) surveillance program and collect patient peri-rectal Eswabs (Copan) at unit admission (defined as ≤ 2 calendar days from unit entry) and weekly thereafter. This study was approved by the Johns Hopkins University School of Medicine Institutional Review Board, with a waiver of consent.

### Microbiology Methods and Outcome Definitions

Residual Amies media from Eswab collection vials was stored at 4˚C and, within 4 days of swab collection, directly plated onto MacConkey agar with ertapenem and meropenem disks [18]. Colonies growing within 27 mm of ertapenem and 32 mm of meropenem were identified by matrix-assisted laser-desorption ionization time-of-flight mass spectrometry (MALDI-TOF; Bruker Daltonics), and carbapenem antimicrobial susceptibility testing (ertapenem, meropenem and imipenem) was performed by disk diffusion applying Clinical and Laboratory Standards Institute guidelines [19].

*Enterobacteriaceae* resistant to ertapenem, meropenem, and/or imipenem were categorized as CRE. Glucose non-fermenting (NF) Gram-negative bacilli resistant to meropenem and/or imipenem were categorized as NFCROs, with the exception of: 1) *Stenotrophomonas maltophilia,* which was excluded due to intrinsic carbapenem resistance, and 2) *Aeromonas* spp., a non-*Enterobacteriaceae* glucose fermenter, was included for lack of an alternative category. CRE and NFCRO isolates were tested for carbapenemase production by the modified carbapenem inactivation method (mCIM) [20]. CRE and NFCROs positive for carbapenemase production by the mCIM test were defined as CP-CREs and CP-NFCROs, respectively (collectively, carbapenemase-producing organisms or “CPOs”). mCIM-negative isolates were defined as non-CP-CRE and non-CP-NFCROs (Figure 2.1). CPOs underwent molecular carbapenemase genotype testing by the Check-MDR CT103XL assay (Check-Points). All study laboratory data were coded without direct patient identifiers until six months after sample

collection, and results were not entered in the medical record, so neither infection control staff nor direct care providers knew patients' colonization status.

**Clinical Data Collection**

Patient data were collected retrospectively through automated extraction from JHH electronic medical record (EMR), infection control, and administrative databases. EMR data were available for inpatient and outpatient encounters across the Johns Hopkins Healthcare system, which includes seven hospitals across Maryland and the District of Columbia. Briefly, an experienced data user (E.Y.K.) extracted all encounter-level information from a relational database that underlies the JHH EMR (Epic). Data relating to demographics, procedures and medication administration, laboratory results, and clinical diagnoses were extracted with the same structured query language for every patient. Infection control variables (e.g., contact precautions status, National Healthcare Safety Network (NHSN)-reportable indwelling hardware) were extracted by an experienced data user (A.G.) from Healthcare Epidemiology and Infection Control (HEIC) operational databases.

Patient data was based on the day of, or days/weeks prior to, unit admission, and included the following: (a) demographic data; (b) hospitalization encounter-level data (e.g., admission type, pre-admission location); (c) pre-existing medical conditions and co-morbidities; (d) multidrug-resistant organism (MDRO) colonization or infection in the prior six months, and contact precautions and VRE colonization status at unit admission; (e) days of inpatient and/or ICU

hospitalization in the prior six months; (f) discharge to a post-acute care facility (long-term acute care hospital, skilled nursing or rehabilitation facility) in the prior six months; (g) indwelling hardware in the prior three months; (h) days and total defined daily dose (DDD)-adjusted doses [21] of immunosuppressive therapy, proton-pump inhibitors (PPIs), or histamine H2-receptor antagonists (H2-blockers) in the prior three months; (i) days and total DDD-adjusted doses of Gram-negative active antibiotic therapy in the prior three months; (j) invasive abdominal procedures or surgeries in the prior three months; and (k) recent international exposure, including hospitalization in another country in the prior six months and/or foreign travel in the prior 21 days, both assessed by standard nursing intake questionnaire at admission. Consistent with CDC guidance [22], Johns Hopkins Healthcare system policy refers patients with recent international hospitalization (<6 mos.) for CRE peri-rectal colonization surveillance screening at admission. We also calculated a 30-category, unweighted Elixhauser Comorbidity Index score (including both primary diagnoses and comorbidities) from the EMR [23].

**Statistical Methods**

***Data Analysis and Logistic Regression.*** Descriptive statistics for patient variables were calculated using mean (standard deviation [SD]), median (range or interquartile range (IQR)), or frequency count (percentage), as appropriate. The relationship between each study covariate and the study outcomes was evaluated using univariable logistic regression with general estimating equations and robust standard errors to account for patient-clustering due to repeat

unit admissions. Results were summarized by odds ratios (ORs) and corresponding 95% confidence intervals (CIs). All tests were 2-tailed, and *P* values ≤ 0.05 were used for statistical significance testing in primary analyses and ≤ 0.025 in sensitivity analyses in order to adjust for multiple outcomes (CRO and CPO colonization). Due to the large number of collected variables and lack of strong *a priori* causal hypotheses, we did not perform multivariable logistic regression. Descriptive and logistic regression analyses were performed in Stata, version 13.0 (StataCorp, College Station, TX).

***Machine Learning-Derived Predictive Models.*** We developed predictive models for the outcomes of CRO, CRE, and CPO colonization at unit admission that evaluated all available collected variables. We built decision trees applying the classification and regression tree (CART) algorithm [24] using the rpart (Recursive Partitioning and Regression Trees) package, version 4.1–13. To fit our trees, we employed the Gini impurity criterion for splitting rules [25]. We also performed random forests analyses to fit 1,000 bootstrapped classification trees, using the randomForest package (version 4.6-14). In sensitivity analyses, we tuned model parameters for both CART decision trees and random forests to incorporate CRO and CPO outcome prior probabilities of 0.50 and 0.10, respectively, in order to address the large imbalances (i.e., rare outcomes) in our data and increase model sensitivity. All machine learning predictive models were developed in the R statistical package (version 3.0.5).

***Predictive Model Validation.*** We internally validated the performance of our CART decision trees using leave-one-out cross-validation [25]. We evaluated the discrimination of all models, both original and cross-validated, through the generation of receiver operating characteristic (ROC) curves and calculation of C-statistics (i.e., area under the curve) in R.

## RESULTS

### Study Population

There were 3,327 unit admissions during the study period: 1,796 (54%) in the MICU and 1,531 (46%) in the Transplant Unit. Of these encounters, 2,878 (87%), representing 2,165 unique patients, had stored perirectal admission screening swabs that were processed for CROs (Figure 2.2).

Patient characteristics are presented in Table 2.1. Evaluating the cohort of swabbed patients (n=2878), patients had a mean age of 55 (SD, 15.4) years and 54.0% were male. Twenty-nine (1.0%) had documented foreign permanent residence. The majority of hospitalizations were emergency/urgent (91.9%), with direct or rapid transfer to the one of the two study units (median time-to-unit admission from hospital admission: 0 days, IQR: 0 – 1 days). The median unweighted Elixhauser Comorbidity Index score was 4 (IQR, 2-7). The most common principal diagnoses or co-morbidities were deficiency anemia (41.8%), renal failure (40.4%), and/or liver disease (29.6%).

In the six months preceding unit admission, 54% of patients were hospitalized and 17.5% had a prior ICU stay. A combined 173 (6.0%) patients had documented discharge to a long-term acute care hospital (1.2%) and/or skilled nursing or rehabilitation facility (5.3%), and 2.6% of patients were directly admitted to the hospital from a long-term care facility during the current study encounter (excludes long-term acute care hospitals). Thirty patients (1.0%) had documented overnight hospitalization in a foreign country. Seventy-four (2.6%) of patient encounters had a prior CRO-positive clinical or surveillance culture: 11 CREs and 63 NFCROs. One hundred-and-seven patients (3.7%) had a prior extended-spectrum β-lactamase (ESBL)-producing or ceftriaxone-resistant *Enterobacteriaceae* culture.

In the three months preceding unit admission, 330 patients (11.5%) had an endoscopic procedure: 93 patients (3.2%) had lower endoscopies and 302 (10.5%) had upper endoscopies (sum >330 due to co-receipt of both procedures). Three hundred-and-three patients (10.6%) had abdominal, colorectal, or urologic surgery. Six hundred-and-seven (21.1%) of patients received antibiotics with Gram-negative coverage; 128 (4.5%) received carbapenems. Six hundred-and-eleven (21.2%) patients received PPIs or H-2 blockers.

At or on the day preceding unit admission, overall 887 (30.8%) of patients had indwelling hardware due to: urinary catheters (21.9%), central lines (13.6%), mechanical ventilation (7.2%), gastrointestinal (oragastric, nasogastric, G-, or GJ-) tubes (4.2%), fecal management

systems (0.3%), and/or ostomy (0.03%). Three hundred-and-fifteen (10.9%) of patients were positive for VRE on their unit admission surveillance swabs, i.e., on the same swab that was processed for CROs.

**CRO and CPO Colonization Admission Prevalence**

Overall, 217 of 2,878 admission swabs (7.5%; 95% CI: 6.6 – 8.5%), from 192 unique patients, tested positive for one or more CROs (Figure 2.2). The CRO prevalence was higher among MICU admissions than among Transplant Unit admissions (9.4% vs. 5.1%, p <0.001). Of the 217 CRO-positive swabs, 36 (16.7%) were positive for carbapenemase production, from 32 unique patients, yielding a CPO colonization admission prevalence of 1.3% (95% CI: 0.9 – 1.7%). The prevalence of CPOs was similar in both units (1.3% in the MICU vs. 1.1% in the Transplant Unit, p=0.64).

**Colonization Admission Prevalence by Bacterial Class: CRE versus NFCROs**

Of the 217 CRO-positive admission swabs, 121 swabs (56%) possessed one or more CREs. These 121 CRE-positive swabs, from 113 unique patients, comprised 97 non-CP-CRE swabs (from 91 unique patients) and 24 CP-CRE swabs (from 22 unique patients). The overall prevalence of CRE and CP-CRE perirectal colonization at admission was 4.2% (95% CI: 3.5 – 5.0%) and 0.8% (95% CI: 0.5 – 1.2%), respectively. Approximately 20% of CRE were carbapenemase-producers. The majority of CP-CRE organisms (n=26 due to CP-CRE co-colonization on two swabs) were

*Klebsiella pneumoniae* (50%), followed by *Enterobacter cloacae* (27%), *Citrobacter amalonaticus* (11%), *Escherichia coli* (8%), and *Citrobacter freundii* (4%).

One hundred and seven admission swabs, from 92 unique patients, were positive for one or more NFCROs, yielding a NFCRO perirectal colonization admission prevalence of 3.7% (95% CI: 3.0 – 4.4%). Ninety-five of the 107 swabs (88.8%) were non-CP-NFCROs, from 83 unique patients, and 12 (11.2%) were CP-NFCROs, from 10 unique patients. Overall, the admission prevalence of carbapenemase-producing NFCRO colonization was 0.4% (95% CI: 0.2 – 0.7%), relative to a 0.8% point prevalence of CP-CRE colonization. Of the 12 carbapenemase-producing NFCROs (no swabs were co-colonized with multiple CP-NFCROs), the majority were *Acinetobacter baumannii* (58%), followed by *Aeromonas* spp. (17%), and one organism (8%) each of: *Acinetobacter radioresistens*, *Achromobacter xylosoxidans*, and *Pseudomonas aeruginosa*.

Eleven admission swabs (0.4%), all from unique patients, were co-colonized with both CRE(s) and NFCRO(s). Most *Enterobacteriaceae* in these swabs were *K. pneumoniae* or *E. coli*, and the majority of the non-fermenters were *P. aeruginosa*. Three of 11 swabs possessed a CPO, but no admission swabs possessed CP-CRE and CP-NFCRO co-colonization.

**Molecular Carbapenemase Genotypes**

Thirty-two of 36 CPO-positive swabs, as defined by a positive mCIM test, underwent molecular Checkpoints processing. Twenty-three of the 33 (70%) mCIM-positive organisms were confirmed to have carbapemenase genes by the Checkpoints assay (one processed swab was co-colonized with two CP-CREs). The distribution of organisms and carbapenemase results are presented in Table 2.2.

Among CP-CREs, the identified carbapenemases in order of frequency were: *bla*$_{KPC}$ (67%), *bla*$_{NDM + OXA-48}$ (17%), *bla*$_{NDM}$ (11%), and *bla*$_{OXA-48}$ (5%). Among the CP-NFCROs, *bla*$_{OXA-23}$ and *bla*$_{OXA-24}$ were the most common carbapenemases (40% each), followed by *bla*$_{VIM}$ (20%). Only two patients with molecularly-confirmed carbapenemases, a *bla*$_{KPC}$-producing *K. pneumoniae* and *bla*$_{OXA-48}$-producing *K. pneumoniae*, reported recent foreign travel; none reported recent international hospitalization. Of the three patients with *bla*$_{NDM + OXA-48}$-producing CRE, one was directly transferred from a rehabilitation facility, had a recent endoscopy, and was already flagged for CRE in the infection control system. The two remaining patients lacked these and other traditional high-risk exposures, but both were immunosuppressed at admission.

**Risk Factors for CRO and CPO Colonization at Unit Admission**

In univariable logistic regression analysis accounting for repeat-patient admissions, a large proportion of variables were significantly associated with CRO colonization at an α level of .05 (Table 2.1). These variables were broadly classifiable into the following categories: MDRO history, invasive abdominal procedures, healthcare or international exposures, indwelling

hardware, and medications (e.g., antibiotics, PPIs or H2-blockers). The most strongly associated variables included recent or at-admission presence of ostomy hardware (OR 23.07, 95% CI: 7.06 – 75.40); recent history of a CRO (OR 12.13, 95% CI: 7.36 – 19.97), CRE (OR 17.72, 95% CI: 4.74 – 66.24), NFCRO (OR 10.21, 95% CI: 6.03 – 17.28), or ESBL-producing and/or ceftriaxone-resistant *Enterobacteriaceae* culture (OR 5.40, 95% CI: 3.18 – 9.14); recent carbapenem (OR 4.32, 95% CI: 2.72 – 6.88) or PPI/H2-blocker use (OR 2.16, 95% CI: 1.56 – 2.99); and admit from a skilled nursing or rehabilitation facility, relative to a community or Emergency Department admission (OR 3.53, 95% CI: 1.84 – 6.77), or recent discharge to a long-term care facility (long-term acute care hospital, skilled nursing or rehabilitation facility) (OR 3.04, 95% CI: 1.90 - 4.87).

Restricting to the subset of CPOs, each of the preceding variables remained significantly associated with CPO colonization, with the exception of admission from a skilled nursing or rehabilitation facility due to model non-convergence/non-estimability (Table 2.1). In addition, some variables that were not significantly associated with CRO colonization were significant for CPO colonization. These included recent colorectal surgery (OR 26.70, 95% CI: 2.75 – 259.24) and recent foreign travel by the patient or a partner (OR 10.80, 95% CI: 2.52 – 46.30).

Both CRO and CPO colonized patients were significantly more likely to test positive for VRE colonization at unit admission, i.e., on the same admission swab that was processed for CROs (respective ORs: 2.62, 95% CI: 1.85 – 3.73; and 3.39, 95% CI: 1.63 – 7.07). These VRE colonized patients accounted for 51 of 217 (24%) and 12 of 36 (33%) of CRO and CPO colonization

outcomes, respectively. In the two study units, patients who test VRE-positive are placed on contact precautions, although not until VRE results become available approximately two days after swabbing. However, thirty-five of the 51 CRO-VRE co-colonized patients were already on contact precautions for other indications (including prior VRE-positive cultures), before VRE results became available. Similarly, 8 of 12 CPO-VRE co-colonized patients were on pre-existing contact precautions at unit admission.

**Predictive Models of Colonization at Unit Admission**

We evaluated all collected study variables, including permutations (e.g., longer or shorter lookback periods, composite and individual variable categories), for inclusion in machine learning models in order to optimize predictive potential. Unlike logistic regression, the CART and random forests approaches used in these analyses can accommodate high predictor-to-outcome ratios (i.e., high-dimensional data), variable collinearities, and simple and higher-order interaction effects without *a priori* specification [24]. Because institutions may differ in their screening objectives and resources, as well as underlying prevalence and organism distribution, we derived models for three alternate outcomes: CRO, CPO, and CRE (included in supplementary material) colonization.

***CART Decision Trees.*** The final decision tree for predicting CRO colonization at unit admission included three study variables (Figure 2.3). The first question in the tree, also called the root node, asked (1) Did the patient have a CRO-positive culture in the previous six months? In

classification trees, positive or “yes” responses branch to the right. If “yes,” the second question queried (2) Did the patient receive 26 or more days (model-derived cut-point) of PPIs in the prior three months? Those patients meeting these criteria were classified as CRO-positive (Terminal node 4) with an associated 93% probability. In patients with a CRO history but lacking this PPI exposure, the tree questioned (3) Has the patient been hospitalized for 51 or more days in the prior six months? If “yes,” patients were classified as CRO-positive (Terminal node 3, 80% probability) and if “no” were classified as CRO-negative (Terminal node 2, 74% probability).

For those 2804 patients lacking a recent CRO history (question 1), the root node branched left and terminated. Patients lacking this history were classified as CRO-negative (Terminal node 1, 93% probability).

The overall tree possessed a specificity of 99.9% and a sensitivity of 9.8%. The positive and negative predictive values were 87.5% and 93.1%, respectively. Incorporating outcome probabilities based on terminal node impurities, the C-statistic for the final tree trained on the full dataset was 0.57 and unchanged following cross-validation. Thirty-four CRO-colonized patients had a prior CRO history, of whom 29 (85%) were already on contact precautions at unit admission. Of the five remaining patients who were not on contact precautions, two additional patients (40%) would have been captured by the decision tree for screening. One of these

patients was colonized with a non-CP-NFCRO and the other was colonized with a VIM-positive CP-NFCRO.

The CPO decision tree truncated at a single variable, history of a CRE-positive culture in the prior six months. Its specificity was 99.8%, and its sensitivity was 16.7%. In other words, 16.7% (6/36) of CPO-colonized patients had a recent CRE culture preceding admission. The CPO tree's discrimination was 0.58 (unchanged following cross-validation).

***Random Forests.*** Random forests (RF) analysis is similar to CART analysis, or other tree-fitting algorithms, except that it generates many bootstrapped decision trees [26, 27]. Its output is less easily interpretable because it does not produce a singular, final tree, but as an ensemble method it generally improves accuracy and reduces model overfitting, i.e., increases generalizability. It also estimates the most important variables for predicting a given outcome [27-29]. The C-statistic for CRO colonization in RF analysis was 0.65, a 14% increase from the single decision tree (C-statistic 0.57). Consistent with the single decision tree's placement of a recent CRO culture in the root node, which is reserved for the most discriminatory variable, RF analysis also identified this exposure as the most important for predicting CRO colonization status at admission (Figure 2.4). Using the RF variable rankings, we selected the five most important, non-nested variables for inclusion in a multivariable logistic regression model (nested variables, e.g., prior CRO culture and prior MDRGN culture, would violate logistic regression's requirement for variable non-collinearity). These variables in order of importance

were: CRO-positive culture in the prior six months, total days of hospitalization in the prior six months, Elixhauser severity of illness score, and total DDD-standardized doses of antibiotics with Gram-negative coverage and immunosuppressive therapy in the prior three months. The resulting model's C-statistic was 0.62 (Figures 2.4 and 2.5**)**.

In the RF model for CPO colonization, discrimination also rose (C-statistic 0.70) relative to the original CART decision tree. Unlike the single decision tree (which placed a recent CRE history at its root node), but consistent with the other models, a recent CRO-positive culture was ranked as the most important value for predicting CPO colonization (Figure 2.6). The top five, non-nested variables in order of importance were: CRO-positive culture in the prior six months, Elixhauser severity of illness score, total DDD-standardized doses of antibiotics with Gram-negative activity in the prior three months, time from hospital-to-unit admission (days), and total number of days of immunosuppressive therapy in the three months preceding unit admission. In logistic regression incorporating these variables, the C-statistic was 0.70 (Figures 2.4 and 2.6).

**Sensitivity Analyses**

To optimize model performance and address possible outcome misclassification, we performed three sensitivity analyses: 1) Refitting decision trees with adjusted prior outcome probabilities in order to increase sensitivity by imposing a greater "cost" during tree-building for misclassifying colonized patients as negative; 2) Refitting decision trees restricting to first,

unique-patient encounters (n=2165); and 3) Refitting a decision tree and performing random forests analysis for predicting CPO colonization, after restricting the CPO outcome to isolates with molecularly-confirmed carbapenemases. These results are discussed in the Supplementary Material (Supplement).

**DISCUSSION**

Identifying CRO- and CPO-colonized patients at hospital unit admission can facilitate timely infection control interventions, such as placing colonized patients on contact precautions, in order to limit healthcare-associated transmission. The CDC recommends CRE colonization screening in limited instances (e.g., patients with recent international hospitalization [16]), but most U.S. hospitals do not perform routine colonization screening for CRE or other CROs. Evaluating patients admitted to a medical intensive care unit and a solid organ transplant unit, we found that 7.5% and 1.3% of patients were perirectally colonized with CROs and CPOs, respectively. Among CROs, the distribution of CRE versus carbapenem-resistant glucose non-fermenters was roughly similar (54% vs. 46%), with a CRE admission prevalence of 4.2% (95% CI: 3.5 – 5.0%). This estimate is somewhat higher than the proportion of CRE (3.1%) among clinical isolates reported to the National Healthcare Safety Network in 2015 [30], and considerably higher than the 0.5% CRE admission prevalence at a Chicago tertiary-care hospital recently reported by Shimasaki *et al.* (2013 data from universal screening at ICU admission) [17].

CPO colonization reflected substantial organism and carbapenemase diversity. Approximately 20% of CRE were carbapenemase-producers, and only half of CP-CREs were *K. pneumoniae*. Consistent with other publicly available data, KPC was the most common carbapenemase among CP-CRE [30], but 1/3 of CP-CREs encoded other carbapenemases, including co-possession of NDM + OXA-48. The proportion of carbapenemase-producers among NFCROs was lower (11.2%) and included a variety of organisms; only 1 of 12 CP-NFCRO isolates was *P. aeruginosa* (VIM-producing). Of note, three CP-CREs were *Citrobacter amalonaticus*, a traditionally less common healthcare pathogen. These isolates carried different carbapenemases (1 NDM, 1 KPC, and 1 Checkpoints-negative), suggesting there were unlikely to be part of an undetected cluster. Recent data document carbapenemase dissemination among less common *Enterobacteriaceae* genera [31], and this finding reinforces that any Gram-negative bacteria of the gut can serve as a reservoir for further spread of carbapenemases.

In univariable analysis, a large number of study variables were significantly associated with both CRO and CPO colonization, even when instituting a more conservative, global significance threshold of $p \leq 0.025$. Most variables' odds ratio point-estimates also strengthened in CPO analysis. All variables reflecting recent MDRO history (MRSA, VRE, CRO, CRE, and ESBL) were significantly associated with colonization, with prior CRO or CRE cultures being the strongest. These findings comport with the overlap in risk factors (e.g., antibiotic use, exposure to high-risk healthcare facilities) among drug-resistant organisms [32-34], and in the case of CRO or CRE histories, with data suggesting strong relationships between infection and underlying

colonization [12-14] and colonization persistence [35]. Other significant variables were consistent with published risk factors for CRE/CRO carriage or infection in other populations, including but not limited to, antibiotic exposure overall [36-38] or carbapenems specifically [39, 40]; long-term care facility stay [41, 42]; immunosuppression [36]; endoscopic procedures [38, 39, 43]; and indwelling hardware [36, 41, 44]. We did not observe strong associations for diabetes mellitus, as other studies have identified [45, 46], or for recent international hospitalization. In fact, no CPO-colonized patients had documented recent international hospitalization, the current CDC-recommended exposure for targeted CRE screening [22]. However, recent foreign travel was strongly associated with CPO colonization.

Despite the large number of significant variables, our models did not highly predict colonization at unit admission. A decision tree to predict CRO colonization possessed an overall C-statistic of 0.57, driven largely by very low (9.8%) sensitivity. Even though its global performance was sub-optimal, however, the decision tree did reveal high-risk sub-groups. Patients with a recent CRO-positive culture (≤ 6 mos.) who had also been on PPIs for 26 or more days of the last three months were 93% likely to be CRO-colonized (13 of 14 patients). Recent studies have identified PPI use as a significant risk factor for admission carriage of other MDRGNs (e.g., ESBLs) [47]. Among patients with a recent CRO history but without this duration of PPI usage, those who had been hospitalized for 51 or more days in the prior six months were 80% likely to be colonized (8 of 10 patients). Using these criteria for targeted surveillance would capture 21 colonized patients while only producing three “unnecessary” (false-positive) screening referrals.

As a practical note, we recognize that algorithm-derived cut-points of 26 and 51 days may be difficult to capture without EMR-automated extraction. We encourage future evaluation after converting these variables to more user-friendly metrics (e.g., “4 weeks” versus 26 days).

Random forests (RF) models were less readily interpretable than the single decision trees, but they improved classifier discrimination, particularly for CPO colonization (RF C-statistic for CPO colonization, 0.70). An appealing feature of RF analysis is that it ranks variables by their predictive strength, and these variables may be subsequently evaluated in other, more user-friendly models (e.g., logistic regression). We evaluated the top five variables in a multivariable logistic regression model for CPO colonization, with similar model performance (C-statistic, 0.70). Of note, some variables clustered closely in terms of their predictive power, as reflected in the variable importance plots. This information may guide practical variable selection decisions when designing targeted screening or other prediction tools — e.g., if variables’ predictive strength is roughly equivalent, use the exposure that is simplest to ascertain.

The sub-optimal sensitivity of our predictive models, in particular our primary decision trees, was likely driven by a combination of factors. These lessons may be informative to future research. First, risk factors measure relative effects, and strong risk factors can account for few absolute numbers of cases. For example, one of the strongest risk factors in univariable regression and the most discriminatory predictor in most of our decision trees – a recent CRO history – was only present in 34% of CRO-colonized patients. This discrepancy between risk

factors and predictors is further underscored by the fact that of the five most predictive variables from our best-performing RF model, most were not independently significant in the corresponding multivariable logistic regression model. These variables would be missed by traditional stepwise variable selection procedures that rely upon significance testing [48]. Second, high predictor heterogeneity, with few dominant exposures, can compromise predictive accuracy. Because organism identity would not inform differential infection control interventions, we did not restrict to specific species (e.g., *K. pneumoniae*, *P. aeruginosa*), as is common in other published analyses [30, 49]. Moreover, many existing models predict clinical infection, not colonization (e.g., [49]). Infection models may differentially capture predictors of colonization-to-infection progression (a potentially more predictable, homogenous process) rather than risk factors for acquisition. In addition, given the long duration of CRE carriage documented in the literature, acquisition risk factors may predate the time-scales captured in the EMR. Third, although we collected extensive information on pre-admission healthcare-associated exposures, the poor model sensitivity among patients lacking a recent CRO history may suggest that important predictors were missing for this population. Indeed, although risk factors for CRE and other CROs in U.S. patients have traditionally focused on the healthcare setting, increasing reports describe the community (e.g., domestic porcine farms, retail seafood) as a carbapenem resistance reservoir [50, 51]. These challenges are likely to intensify as CRE and CRO endemicity in the U.S. rises and community-acquired colonization increases [52].

Notwithstanding these challenges, our results offer some actionable conclusions. Across all models evaluated, recent CRO or CRE culture was consistently the strongest predictor of colonization at admission. Many infection control programs already capture and flag these cultures, suggesting that existing policies — however imperfect — may be performing equivalently to a targeted screening program. We note, however, that we manually coded our 'CRO' culture definition to synchronize with our CRO study outcome; some NFCRO cultures would not qualify as MDRGNs (e.g., resistant to 4/5 antibiotic classes evaluated), as customarily defined. Moreover, 24% and 33% of CRO- and CPO-colonized patients, respectively, were co-colonized with VRE detected during routine admission screening. These patients would be placed on contact precautions even without dedicated CRO surveillance. These findings suggest that existing screening policies for other organisms may have unrecognized, off-target benefits, of arguable relevance to ongoing national conversations about the utility of continued VRE rectal surveillance [53].

Our study has several limitations. First, this was a single-center study, and although we internally validated our models, our results should be validated in other cohorts. Our results may not be generalizable to other, lower-risk hospitalized populations or areas with higher endemicity (e.g., New York City). Second, we used culture-based laboratory methods to isolate CROs. Culture-based methods can produce false-negative results, particularly for surveillance swabs with low bioburden [54, 55]. However, our screening methodology was selected to maximize sensitivity, based upon a 5-arm comparator pilot study [18]. Moreover, our CRO

recovery rates were consistent with, or higher, than other published estimates. Third, concordance between phenotypic and molecular carbapenemase assays was lower than expected. The mCIM demonstrates very high sensitivity and specificity in published data [20], but our experience suggests that its specificity may be lower than previously reported, particularly for *E. cloacae*. We repeated mCIM tests on the four mCIM+, Checkpoints-negative *E. cloacae* isolates with identical results, and whole genome sequencing suggested these isolates did not encode novel carbapenemases (data unpublished). Nevertheless, in sensitivity analyses we restricted the CPO outcome to molecularly-confirmed isolates, with similar findings, suggesting that our models were robust to this possible misclassification. Finally, despite gathering extensive demographic and clinical information on all study patients, due to our restriction to Johns Hopkins Healthcare system data, there was inevitably missing data that could lead to exposure misclassification. However, because we would not expect this missingness to be differential by colonization status, any associations would likely attenuate towards the null; yet, many exposures remained strongly associated with CRO and CPO colonization in regression analyses. More importantly, because the prediction models were designed to inform practical, real-world screening decisions, their performance under the operational constraints of incomplete data is arguably relevant.

Overall, in this high-risk inpatient population CRO and CPO carriage was infrequent but higher than previously published estimates. Colonization was characterized by significant organism and resistance mechanism heterogeneity, suggesting that colonization did not originate from a

single or few clonal strains. We identified carbapenemases in many organism types, including glucose non-fermenters other than *A. baumannii* and *P. aeruginosa* and organisms not commonly associated with healthcare infections (e.g., *C. amalonaticus*), providing an important reminder that many different GI-colonizing organisms can serve as carbapenemase gene reservoirs. Many pre-admission exposures were strongly associated with CRO and CPO colonization, consistent with risk factors identified in other published studies, but organism and resistance mechanism heterogeneity made prediction challenging. Active surveillance for CROs and CPOs is cost and/or resource-intensive, and although were unable to develop high-performing targeted screening algorithms, ongoing efforts should continue to use available EMR and other data to identify ways to limit resource utilization.

**ACKNOWLEDGMENTS**

We would like to thank Verna Scheeler, Michael Anderson, Dina Khamash, and Sean Thompson for their assistance with study coordination, and data collection and validation, as well as Belita Opene, Shawna Lewis, Yehudit Bergman, and Krizia Chambers for their work processing surveillance cultures. We would also like to thank Matthew Toerper and Scott Levin for their assistance with electronic medical record data extraction, and members of the JHU Clinical Microbiology Laboratory staff for helping with collection of surveillance swabs for the study.

***Financial Support*****:** This work was supported by the CDC Prevention Epicenters Program CDC 1U54CK000447, AHRQ R36HS025089, and The Sherrilyn and Ken Fisher Center for Environmental Infectious Diseases.

**REFERENCES**


1. Amit S, Mishali H, Kotlovsky T, Schwaber MJ, Carmeli Y: Bloodstream infections among carriers of carbapenem-resistant Klebsiella pneumoniae: etiology, incidence and predictors. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 2015, 21(1):30-34.
2. Balkan, II, Aygun G, Aydin S, Mutcali SI, Kara Z, Kuskucu M, Midilli K, Semen V, Aras S, Yemisen M *et al*: Blood stream infections due to OXA-48-like carbapenemase-producing Enterobacteriaceae: treatment and survival. *International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases* 2014, 26:51-56.
3. Ben-David D, Kordevani R, Keller N, Tal I, Marzel A, Gal-Mor O, Maor Y, Rahav G: Outcome of carbapenem resistant Klebsiella pneumoniae bloodstream infections. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases* 2012, 18(1):54-60.
4. Falagas ME, Tansarli GS, Karageorgopoulos DE, Vardakas KZ: Deaths attributable to carbapenem-resistant Enterobacteriaceae infections. *Emerging infectious diseases* 2014, 20(7):1170-1175.
5. Jamulitrat S, Arunpan P, Phainuphong P: Attributable mortality of imipenem-resistant nosocomial Acinetobacter baumannii bloodstream infection. *Journal of the Medical Association of Thailand = Chotmaihet thangphaet* 2009, 92(3):413-419.
6. Kumar SH, De AS, Baveja SM, Gore MA: Prevalence and risk factors of Metallo beta-lactamase producing Pseudomonas aeruginosa and Acinetobacter species in burns and surgical wards in a tertiary care hospital. *Journal of laboratory physicians* 2012, 4(1):39-42.
7. CDC: Antibiotic resistance threats in the United States, 2013. In.; 2014: 1-114.
8. Hong DJ, Bae IK, Jang I-H, Jeong SH, Kang H-K, Lee K: Epidemiology and Characteristics of Metallo-β-Lactamase-Producing Pseudomonas aeruginosa. *Infect Chemother* 2015, 47(2):81-97.

9. Kim UJ, Kim HK, An JH, Cho SK, Park K-H, Jang H-C: Update on the Epidemiology, Treatment, and Outcomes of Carbapenem-resistant Acinetobacter infections. *Chonnam Med J* 2014, 50(2):37-44.

10. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Group NCSP, Henderson DK, Palmore TN, Segre JA: Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med* 2012, 4(148):148ra116.

11. Tamma PD: Comparing the Outcomes of Patients With Carbapenemase-Producing and Non-Carbapenemase-Producing Carbapenem-Resistant Enterobacteriaceae Bacteremia. 2017, 64(3):257-264.

12. Dickstein Y, Edelman R, Dror T, Hussein K, Bar-Lavie Y, Paul M: Carbapenem-resistant Enterobacteriaceae colonization and infection in critically ill patients: a retrospective matched cohort comparison with non-carriers. *The Journal of hospital infection* 2016, 94(1):54-59.

13. Giannella M, Bartoletti M, Morelli MC, Tedeschi S, Cristini F, Tumietto F, Pasqualini E, Danese I, Campoli C, Lauria ND *et al*: Risk factors for infection with carbapenem-resistant Klebsiella pneumoniae after liver transplantation: the importance of pre- and posttransplant colonization. *American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons* 2015, 15(6):1708-1715.

14. Macesic N, Gomez-Simmonds A, Sullivan SB, Giddins MJ, Ferguson SA, Korakavi G, Leeds D, Park S, Shim K, Sowash MG *et al*: Genomic Surveillance Reveals Diversity of Multidrug-Resistant Organism Colonization and Infection: A Prospective Cohort Study in Liver Transplant Recipients. *Clin Infect Dis* 2018.

15. Davido B, Batista R, Michelon H, Lepainteur M, Bouchand F, Lepeule R, Salomon J, Vittecoq D, Duran C, Escaut L *et al*: Is faecal microbiota transplantation an option to eradicate highly drug-resistant enteric bacteria carriage? *The Journal of hospital infection* 2017, 95(4):433-437.

16. Centers for Disease Control and Prevention: Facility Guidance for Control of Carbapenem-resistant Enterobacteriaceae (CRE): November 2015 Update. In.; 2015.

17. Shimasaki T, Segreti J, Tomich A, Kim J, Hayden MK, Lin MY: Active screening and interfacility communication of carbapenem-resistant Enterobacteriaceae (CRE) in a tertiary-care hospital. *Infect Control Hosp Epidemiol* 2018:1-5.

18. Simner PJ, Martin I, Opene B, Tamma PD, Carroll KC, Milstone AM: Evaluation of Multiple Methods for the Detection of Gastrointestinal Colonization of Carbapenem-Resistant Organisms from Rectal Swabs. *J Clin Microbiol* 2016.

19. Clinical and Laboratory Standards Institute: Performance standards for antimicrobial susceptibility testing, 28th ed. Supplement M100S. In. Wayne, PA: CLSI; 2018.

20. Pierce VM, Simner PJ, Lonsway DR, Roe-Carpenter DE, Johnson JK, Brasso WB, Bobenchik AM, Lockett ZC, Charnot-Katsikas A, Ferraro MJ *et al*: Modified Carbapenem Inactivation Method for Phenotypic Detection of Carbapenemase Production among Enterobacteriaceae. *J Clin Microbiol* 2017, 55(8):2321-2333.

21. Defined Daily Dose (DDD) [http://www.who.int/medicines/regulation/medicines-safety/toolkit_ddd/en/]

22. CDC: Facility Guidance for Control of Carbapenem-resistant Enterobacteriaceae (CRE) – November 2015 Update CRE Toolkit. In. Atlanta, Georgia: CDC; 2015.

23. Elixhauser A, Steiner C, Harris DR, Coffey RM: Comorbidity measures for use with administrative data. *Medical care* 1998, 36(1):8-27.

24. Breiman L, Friedman J, Stone C, Olshen R: Classification and regression trees. Boca Raton, Florida: CRC/Chapman & Hall; 1984.

25. Duda RO, Hart PE, Stork DG: Pattern Classification, 2nd edn. United States: Wiley-Interscience; 2001.

26. Chen X, Ishwaran H: Pathway hunting by random survival forests. *Bioinformatics* 2013, 29(1):99-105.

27. 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(4):323-348.

28. Casanova R, Saldana S, Chew EY, Danis RP, Greven CM, Ambrosius WT: Application of random forests methods to diabetic retinopathy classification analyses. *PloS one* 2014, 9(6):e98587.

29. Rose S: Mortality risk score prediction in an elderly population using machine learning. *Am J Epidemiol* 2013, 177(5):443-452.

30. Woodworth KR, Walters MS, Weiner LM, Edwards J, Brown AC, Huang JY, Malik S, Slayton RB, Paul P, Capers C *et al*: Vital Signs: Containment of Novel Multidrug-Resistant Organisms and Resistance Mechanisms - United States, 2006-2017. *MMWR Morb Mortal Wkly Rep* 2018, 67(13):396-401.

31. Walters MS, Witwer M, Lee YK, Albrecht V, Lonsway D, Rasheed JK, Anacker M, Snippes-Vagnone P, Lynfield R, Kallen AJ: Notes from the Field: Carbapenemase-Producing Carbapenem-Resistant Enterobacteriaceae from Less Common Enterobacteriaceae Genera - United States, 2014-2017. *MMWR Morb Mortal Wkly Rep* 2018, 67(23):668-669.

32. Logan LK, Weinstein RA: The Epidemiology of Carbapenem-Resistant Enterobacteriaceae: The Impact and Evolution of a Global Menace. *J Infect Dis* 2017, 215(suppl_1):S28-S36.

33. McKinnell JA, Miller LG, Eells SJ, Cui E, Huang SS: A systematic literature review and meta-analysis of factors associated with methicillin-resistant Staphylococcus aureus colonization at time of hospital or intensive care unit admission. *Infect Control Hosp Epidemiol* 2013, 34(10):1077-1086.

34. Simner PJ, Goodman KE, Carroll KC, Harris AD, Han JH, Tamma PD: Using Patient Risk Factors to Identify Whether Carbapenem-Resistant Enterobacteriaceae Infections Are Caused by Carbapenemase-Producing Organisms. *Open Forum Infectious Diseases* 2018, 5(5):ofy094-ofy094.

35. Bar-Yoseph H, Hussein K, Braun E, Paul M: Natural history and decolonization strategies for ESBL/carbapenem-resistant Enterobacteriaceae carriage: systematic review and meta-analysis. *J Antimicrob Chemother* 2016, 71(10):2729-2739.

36. Bhargava A, Hayakawa K, Silverman E, Haider S, Alluri KC, Datla S, Diviti S, Kuchipudi V, Muppavarapu KS, Lephart PR *et al*: Risk factors for colonization due to carbapenem-resistant Enterobacteriaceae among patients exposed to long-term acute care and acute care facilities. *Infect Control Hosp Epidemiol* 2014, 35(4):398-405.

37. Marchaim D, Chopra T, Bhargava A, Bogan C, Dhar S, Hayakawa K, Pogue JM, Bheemreddy S, Blunden C, Shango M *et al*: Recent exposure to antimicrobials and carbapenem-resistant Enterobacteriaceae: the role of antimicrobial stewardship. *Infect Control Hosp Epidemiol* 2012, 33(8):817-830.

38. Predic M, Delano JP, Tremblay E, Iovine N, Brown S, Prins C: Risk Factors for Carbapenem-Resistant Enterobacteriaceae Infection. *American journal of infection control* 2017, 45(6, Supplement):S14.

39. Torres-Gonzalez P, Cervera-Hernandez ME, Niembro-Ortega MD, Leal-Vega F, Cruz-Hervert LP, Garcia-Garcia L, Galindo-Fraga A, Martinez-Gamboa A, Bobadilla-Del Valle M, Sifuentes-Osornio J *et al*: Factors Associated to Prevalence and Incidence of Carbapenem-Resistant Enterobacteriaceae Fecal Carriage: A Cohort Study in a Mexican Tertiary Care Hospital. *PloS one* 2015, 10(10):e0139883.

40. Armand-Lefevre L, Angebault C, Barbier F, Hamelet E, Defrance G, Ruppe E, Bronchard R, Lepeule R, Lucet JC, El Mniai A *et al*: Emergence of imipenem-resistant gram-negative bacilli in intestinal flora of intensive care patients. *Antimicrob Agents Chemother* 2013, 57(3):1488-1495.

41. Cunha CB, Kassakian SZ, Chan R, Tenover FC, Ziakas P, Chapin KC, Mermel LA: Screening of nursing home residents for colonization with carbapenem-resistant Enterobacteriaceae admitted to acute care hospitals: Incidence and risk factors. *American journal of infection control* 2016, 44(2):126-130.

42. Prabaker K, Lin MY, McNally M, Cherabuddi K, Ahmed S, Norris A, Lolans K, Odeh R, Chundi V, Weinstein RA *et al*: Transfer from high-acuity long-term care facilities is associated with carriage of Klebsiella pneumoniae carbapenemase-producing Enterobacteriaceae: a multihospital study. *Infect Control Hosp Epidemiol* 2012, 33(12):1193-1199.

43. Muscarella LF: Risk of transmission of carbapenem-resistant Enterobacteriaceae and related "superbugs" during gastrointestinal endoscopy. *World J Gastrointest Endosc* 2014, 6(10):457-474.

44. Prasad N, Labaze G, Kopacz J, Chwa S, Platis D, Pan CX, Russo D, LaBombardi VJ, Osorio G, Pollack S *et al*: Asymptomatic rectal colonization with carbapenem-resistant Enterobacteriaceae and Clostridium difficile among residents of a long-term care facility in New York City. *American journal of infection control* 2016, 44(5):525-532.

45. Ciobotaro P, Flaks-Manov N, Oved M, Schattner A, Hoshen M, Ben-Yosef E, Balicer RD, Zimhony O: Predictors of Persistent Carbapenem-Resistant Enterobacteriaceae Carriage upon Readmission and Score Development. *Infect Control Hosp Epidemiol* 2016, 37(2):188-196.

46. Meng X, Liu S, Duan J, Huang X, Zhou P, Xiong X, Gong R, Zhang Y, Liu Y, Fu C *et al*: Risk factors and medical costs for healthcare-associated carbapenem-resistant Escherichia coli infection

among hospitalized patients in a Chinese teaching hospital. *BMC infectious diseases* 2017, 17(1):82.

47. Huizinga P, van den Bergh MK, van Rijen M, Willemsen I, van 't Veer N, Kluytmans J: Proton Pump Inhibitor Use Is Associated With Extended-Spectrum beta-Lactamase-Producing Enterobacteriaceae Rectal Carriage at Hospital Admission: A Cross-Sectional Study. *Clin Infect Dis* 2017, 64(3):361-363.

48. Wiegand RE: Performance of using multiple stepwise algorithms for variable selection. *Statistics in medicine* 2010, 29(15):1647-1659.

49. Sullivan T, Ichikawa O, Dudley J, Li L, Aberg J: The Rapid Prediction of Carbapenem Resistance in Patients With Klebsiella pneumoniae Bacteremia Using Electronic Medical Record Data. *Open Forum Infect Dis* 2018, 5(5):ofy091.

50. Mollenkopf DF, Stull JW, Mathys DA, Bowman AS, Feicht SM, Grooters SV, Daniels JB, Wittum TE: Carbapenemase-producing Enterobacteriaceae recovered from the environment of a swine farrow-to-finish operation in the United States. *Antimicrobial Agents and Chemotherapy* 2016.

51. Janecko N, Martz SL, Avery BP, Daignault D, Desruisseau A, Boyd D, Irwin RJ, Mulvey MR, Reid-Smith RJ: Carbapenem-Resistant Enterobacter spp. in Retail Seafood Imported from Southeast Asia to Canada. *Emerging infectious diseases* 2016, 22(9):1675-1677.

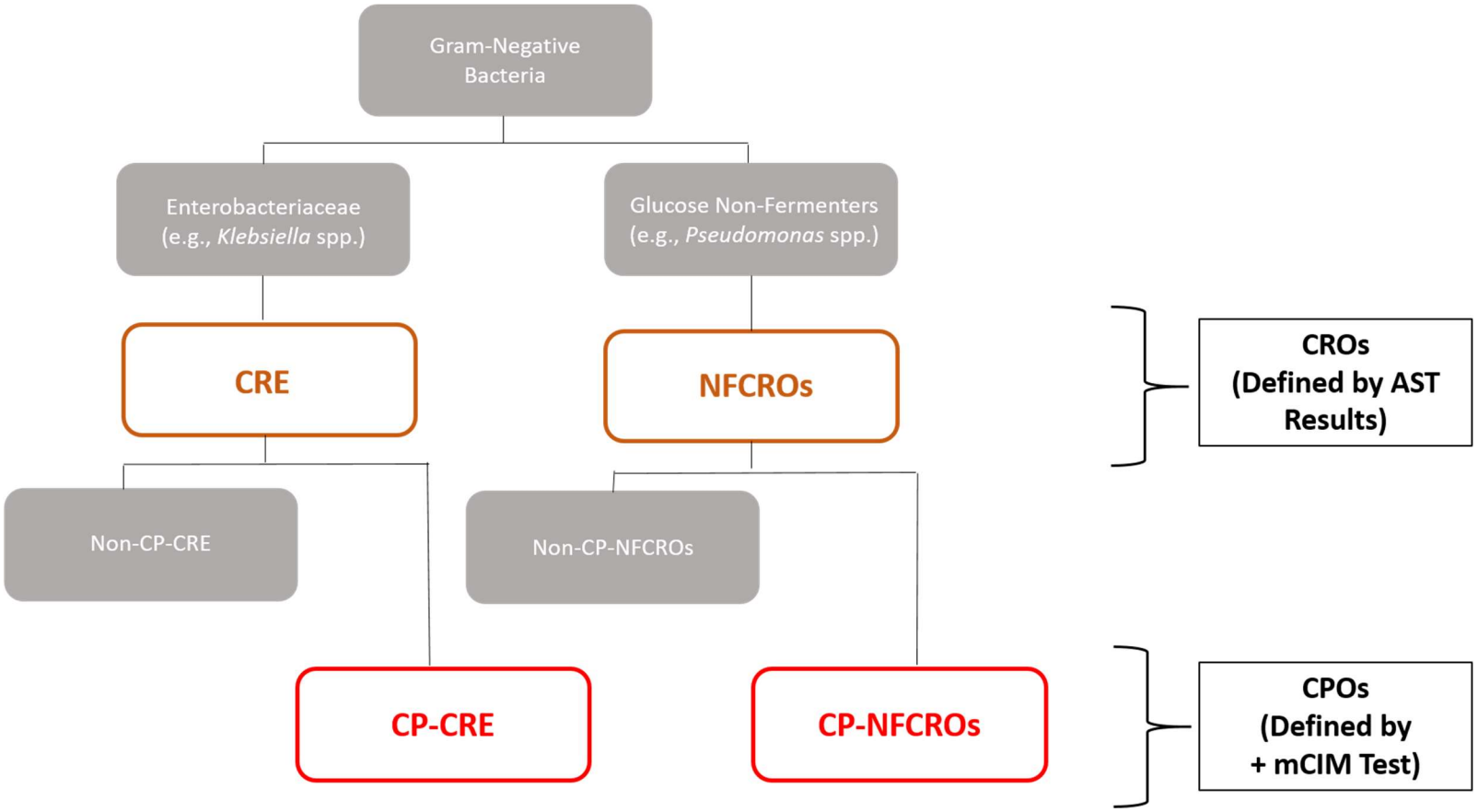
52. Janelle SJ, Kallen A, de Man T, Limbago B, Walters M, Halpin A, Xavier K, Knutsen J, Badolato E, Bamberg WM: Notes from the Field: New Delhi Metallo-beta-Lactamase-Producing Carbapenem-Resistant Enterobacteriaceae Identified in Patients Without Known Health Care Risk Factors - Colorado, 2014-2016. *MMWR Morb Mortal Wkly Rep* 2016, 65(49):1414-1415.

53. Martin EM, Bryant B, Grogan TR, Rubin ZA, Russell DL, Elashoff D, Uslan DZ: Noninfectious Hospital Adverse Events Decline After Elimination of Contact Precautions for MRSA and VRE. *Infect Control Hosp Epidemiol* 2018, 39(7):788-796.

54. Snyder GM, D'Agata EM: Diagnostic accuracy of surveillance cultures to detect gastrointestinal colonization with multidrug-resistant gram-negative bacteria. *American journal of infection control* 2012, 40(5):474-476.

55. Amar M, Adler A: Validation of quality assurance criteria for rectal surveillance cultures. *American journal of infection control* 2017, 45(9):1041-1042.

**Figure 2.1. Study Outcome Definitions by Bacterial Class**




*Glucose-fermenting *Aeromonas* spp. included in glucose non-fermenting category, as a non-*Enterobacteriaceae*.

Abbreviations – Carbapenem-resistant organisms (CROs); Carbapenemase-producing organisms (CPOs); Carbapenem-resistant *Enterobacteriaceae* (CRE); Carbapenemase-producing CRE (CP-CRE); Non-carbapenemase-producing CRE (Non-CP-CRE); Glucose non-fermenting carbapenem-resistant organisms (NFCROs); Carbapenemase-producing NFCROs (CP-NFCROs); Non-carbapenemase-producing NFCROs (Non-CP-NFCROs); Antimicrobial susceptibility testing (AST); Modified carbapenem inactivation method (mCIM).

**Figure 2.2. Study Flowchart of Carbapenem-Resistant Organism (CRO) and Carbapenemase-Producing Organism (CPO) Colonization at Hospital Unit Admission**

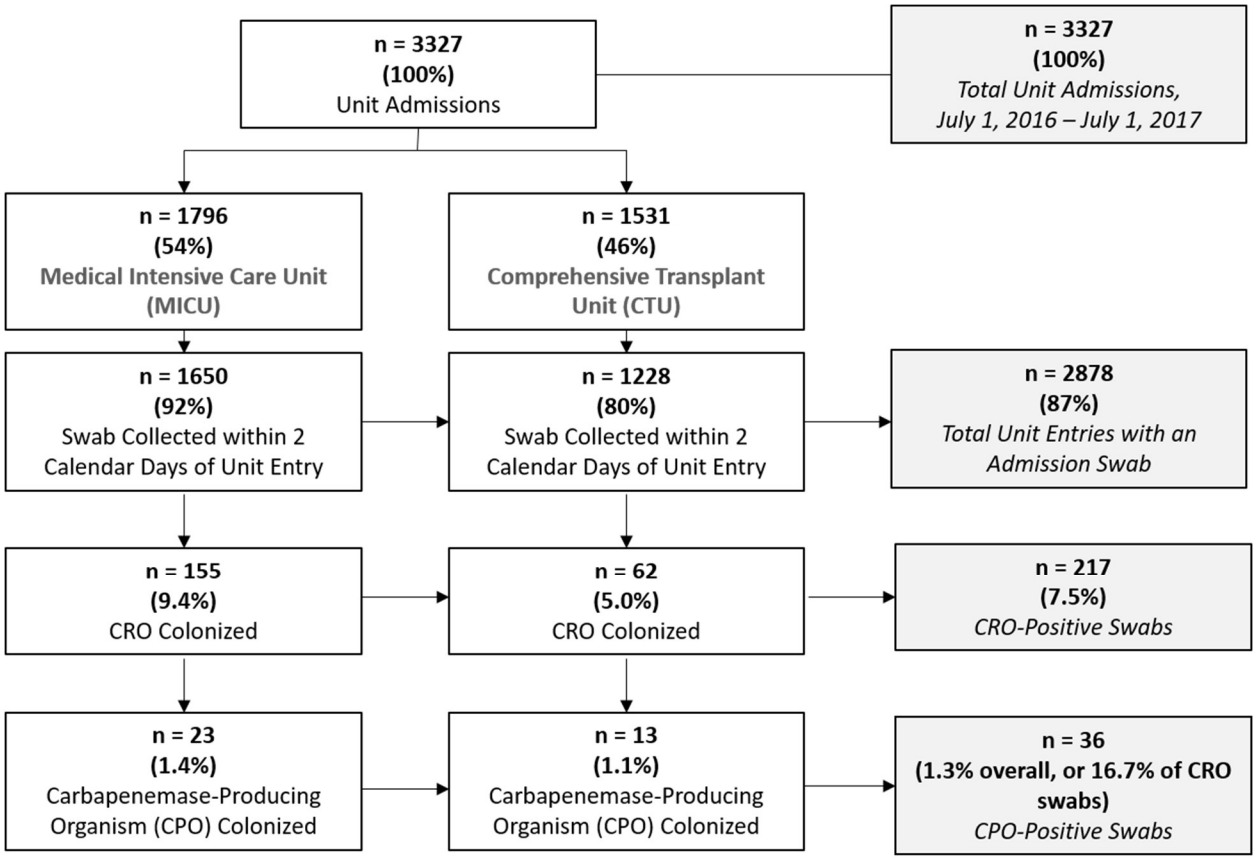


**Table 2.1. Description of Patient Characteristics in a Cohort of Medical Intensive Care Unit (MICU) and Comprehensive Transplant Unit (Transplant Unit) Patients, by Carbapenem-Resistant Organism (CRO) and Carbapenemase-Producing Organism (CPO) Colonization Status at Unit Admission**

| **Variables at or Preceding Unit Admission** | **Total Swabbed Cohort** | **CRO Colonized** Odds Ratio (95% Confidence Interval) | **CPO Colonized** Odds Ratio (95% Confidence Interval) | **P-values** CRO ; CPO |
|---|---|---|---|---|
| | n = 2878 | n = 217 | n = 36 | |
| **Demographics** | | | | |
| Age | 55 ± 15.4 | 1.01 (1.00 – 1.03) | 1.01 (0.98 – 1.04) | 0.01*; 0.54 |
| Female sex | 1325 (46%) | 0.99 (0.74 – 1.34) | 0.76 (0.36 – 1.60) | 0.97; 0.47 |
| Race | | | | |
| White | 1317 (46%) | Reference | Reference | Reference |
| Black | 1272 (44%) | NA | NA | NA |
| Asian | 61 (2%) | NA | NA | NA |
| American Indian, Alaska Native or Native Hawaiian, or Pacific Islander | 13 (0.5%) | NA | NA | NA |
| Other | 215 (7%) | NA | NA | NA |
| Foreign Permanent Residence | 29 (1%) | 2.10 (0.75 – 5.92) | 3.77 (0.49 – 29.01) | 0.16; 0.20 |
| **Encounter-Level Characteristics** | | | | |
| Admission type | | | | |
| Emergency/urgent (non-trauma) | 2646 (92%) | Reference | Reference | Reference |
| Trauma | 26 (1%) | NA | NA | NA |
| Non-urgent/elective | 206 (7%) | NA | NA | NA |
| Admission source | | | | |
| ER/Community | 2353 (82%) | Reference | Reference | Reference |
| Acute care hospital, direct transfer | 434 (15%) | 1.63 (1.13 – 2.35) | NA | 0.01* |
| Long-term care facility (non-acute), direct transfer | 74 (3%) | 3.53 (1.84 – 6.77) | NA | <0.001* |
| Other/unknown | 17 (0.6%) | 0.93 (0.12 – 7.12) | NA | 0.95 |
| **Elixhauser Comorbidity Score and Select Pre-Existing Medical Conditions** | | | | |
| Elixhauser Score, median (IQR) | 4 (2 – 7) | 1.02 (0.98 – 1.07) | 1.04 (0.95 – 1.12) | 0.26; 0.40 |
| Chronic peptic ulcer disease | 81 (3%) | 1.47 (0.71 – 3.08) | 1.19 (0.23 – 6.07) | 0.30; 0.83 |
| Solid tumor without metastasis | 468 (16%) | 1.24 (0.85 – 1.82) | 1.40 (0.59 – 3.36) | 0.27; 0.45 |
| Metastatic cancer | 197 (7%) | 1.50 (0.90 – 2.50) | 1.13 (0.33 – 3.90) | 0.12; 0.84 |
| Renal failure | 1164 (40%) | 0.94 (0.69 – 1.29) | 0.81 (0.37 – 1.77) | 0.72; 0.61 |

| | | | | |
|---|---|---|---|---|
| Liver disease | 852 (30%) | 0.78 (0.56 – 1.10) | 1.20 (0.57 – 2.50) | 0.15; 0.64 |
| Diabetes | 912 (32%) | 1.36 (1.00 – 1.85) | 1.12 (0.54 – 2.34) | 0.05; 0.76 |
| Iron-deficiency anemia | 1203 (42%) | 1.22 (0.91 – 1.64) | 1.57 (0.75 – 3.28) | 0.18; 0.23 |
| Chronic pulmonary disease | 630 (22) | 0.99 (0.68 – 1.43) | 1.06 (0.45 – 2.51) | 0.95; 0.90 |
| Paralysis | 68 (2%) | 3.00 (1.58 – 5.68) | 5.09 (1.50 – 17.25) | 0.001*; 0.01* |
| Human Immunodeficiency Virus positive | 159 (6%) | 0.68 (0.33 – 1.40) | 0.48 (0.07 – 3.53) | 0.30; 0.48 |
| Immunosuppressed[1] | 772 (27%) | 1.31 (0.94 – 1.83) | 1.81 (0.82 – 4.01) | 0.11; 0.14 |
| **Indwelling Hardware at Admission** | 887 (31%) | 1.60 (1.19 – 2.16) | 2.12 (0.95 – 4.73) | 0.002*; 0.07 |
| Central line | 393 (14%) | 1.54 (1.05 – 2.25) | 2.56 (1.08 – 6.06) | 0.03*; 0.03* |
| Urologic catheter | 631 (22%) | 1.20 (0.86 – 1.66) | 1.73 (0.77 – 3.88) | 0.28; 0.19 |
| Mechanical ventilation | 207 (7%) | 1.33 (0.79 – 2.25) | 0.32 (0.03 – 4.03) | 0.28; 0.38 |
| Gastrointestinal upper or lower tube | 122 (4%) | 1.04 (0.50 – 2.18) | NA | 0.91 |
| Fecal management device | 8 (0.3%) | 2.94 (0.43 – 20.10) | NA | 0.27 |
| Ostomy | 1 (0.03%) | NA | NA | NA |
| **Indwelling Hardware (< 3 Months)** | 1148 (40%) | 1.74 (1.31 – 2.31) | 3.51 (1.56 – 7.94) | <0.001*; 0.003* |
| Central line | 569 (20%) | 1.60 (1.15 – 2.22) | 3.66 (1.72 – 7.81) | 0.01*; 0.001* |
| Urologic catheter | 876 (30%) | 1.35 (1.01 – 1.82) | 2.32 (1.06 – 5.10) | 0.04; 0.04 |
| Mechanical ventilation | 324 (11%) | 1.71 (1.13 – 2.60) | 2.31 (1.09 – 4.90) | 0.012*; 0.03 |
| Gastrointestinal upper or lower tube | 189 (7%) | 2.09 (1.28 – 3.40) | 1.21 (0.26 – 5.65) | 0.003*; 0.81 |
| Fecal management device | 0 (0%) | NA | NA | NA |
| Ostomy | 13 (0.5%) | 23.07 (7.06 – 75.40) | 24.12 (4.77 – 121.79) | <0.001*; <0.001* |
| **On Contact Precautions at Admission (Prior to VRE Screening)** | 796 (28%) | 2.28 (1.68 – 3.09) | 3.19 (1.49 – 6.83) | <0.001*; 0.03 |
| **Admission Swab Positive for VRE Colonization** | 315 (11%) | 2.62 (1.85 – 3.73) | 3.39 (1.63 – 7.07) | <0.001*; 0.001* |
| **Recent Multidrug-resistant Organism History (Colonization or Infection <6 Months)** | | | | |
| Vancomycin-resistant *Enterococcus* species. | 274 (10%) | 2.09 (1.35 – 3.25) | 4.36 (1.87 – 10.14) | 0.001*; 0.001* |
| Methicillin-resistant *Staphylococcus aureus* | 168 (6%) | 2.47 (1.55 – 3.93) | 4.84 (1.94 – 12.01) | <0.001*; 0.001* |
| Extended-spectrum β-lactamase (ESBL) or ceftriaxone-resistant Enterobacteriaceae | 107 (4%) | 3.84 (2.41 – 6.14) | 8.40 (3.48 – 20.30) | <0.001*; 0.001* |

| | | | | |
|---|---|---|---|---|
| Carbapenem-resistant organism (CRO) | 74 (3%) | 12.13 (7.36 – 19.97) | 29.41 (13.24 – 65.33) | <0.001*; <0.001* |
| Carbapenem-resistant Enterobacteriaceae (CRE) | 11 (0.4%) | 17.72 (4.74 – 66.24) | 77.30 (17.02 – 351.12) | <0.001*; <0.001* |
| Carbapenem-resistant glucose non-fermenting bacilli (NFCRO) | 64 (2%) | 10.21 (6.03 – 17.28) | 13.59 (4.90 – 37.71) | <0.001*; <0.001* |
| Multidrug-resistant *Pseudomonas* species[2] | 28 (1%) | 9.18 (3.33 – 25.25) | 10.43 (2.35 – 46.35) | <0.001*; <0.001* |
| Multidrug-resistant *Acinetobacter* species[2] | 41 (1%) | 12.86 (6.01 – 27.54) | 16.36 (5.45 – 49.13) | <0.001*; <0.001* |
| **Recent Medication Exposure (< 3 Months)** | | | | |
| Immunosuppressive therapy[3] | 620 (22%) | 1.54 (1.09 – 2.19) | 2.07 (0.86 – 4.97) | 0.02*; 0.10 |
| Proton-pump inhibitors (PPIs) or H2-Blockers | 611 (21%) | 2.16 (1.56 – 2.99) | 3.08 (1.37 – 6.93) | <0.001*; 0.007* |
| **Recent Antibiotic Exposure (<3 Months)** | | | | |
| Extended-spectrum penicillin therapy | 313 (11%) | 2.13 (1.42 – 3.19) | 3.80 (1.49 – 9.70) | <0.001*; 0.01* |
| Third and fourth-generation cephalosporin therapy | 379 (13%) | 1.33 (0.86 – 2.03) | 1.87 (0.72 – 4.85) | 0.20; 0.20 |
| Aztreonam therapy | 21 (0.7%) | 4.86 (1.78 – 13.25) | 5.83 (0.92 – 36.81) | 0.002*; 0.06 |
| Carbapenems | 128 (4%) | 4.32 (2.72 – 6.88) | 5.74 (2.23 – 14.77) | <0.001*; <0.001* |
| Fluoroquinolone therapy | 144 (5%) | 2.20 (1.27 – 3.82) | 1.69 (0.35 – 8.15) | 0.01*; 0.51 |
| Aminoglycoside therapy | 49 (1.7%) | 4.91 (2.26 – 10.64) | 3.02 (0.48 – 18.97) | <0.001*; 0.24 |
| Any antibiotics (combined) | 607 (21%) | 1.92 (1.39 – 2.64) | 2.07 (0.88 – 4.92) | <0.001*; 0.10 |
| **Duration of Time from Hospital Admission to Unit Admission (Days), median (IQR)** | 0 (0 – 1) | 1.01 (1.01 – 1.02) | 1.01 (1.01 – 1.02) | <0.001*; <0.001* |
| **Recent International Exposure** | | | | |
| International Hospitalization (1+ nights, < 6 Months) | 30 (1%) | 1.72 (0.58 – 5.12) | 3.34 (0.46 – 24.32) | 0.33; 0.23 |
| International travel, patient or spouse (< 21 days) | 18 (0.6%) | 2.62 (0.79 – 8.71) | 10.80 (2.52 – 46.30) | 0.12; 0.001* |
| **Other High-Risk Healthcare Exposures (<6 Months)** | | | | |
| Inpatient hospitalization | 1553 (54%) | 1.37 (1.03 -1.83) | 1.24 (0.61 – 2.51) | 0.03; 0.55 |

| | | | | |
|---|---|---|---|---|
| Intensive care unit | 503 (17%) | 1.93 (1.38 – 2.69) | 2.66 (1.20 – 5.88) | <0.001*; 0.02* |
| Long-term care facility | 173 (6%) | 3.04 (1.90 – 4.87) | 4.59 (1.63 – 12.96) | <0.001*; 0.004* |
| Long-term acute care hospital | 34 (1%) | 3.70 (1.42 – 9.64) | NA | 0.01* |
| Skilled nursing or rehabilitation facility | 153 (5%) | 3.14 (1.92 – 5.16) | 5.95 (2.40 – 14.75) | <0.001*; <0.001* |
| **Invasive Procedures (< 3 months)** | | | | |
| Endoscopy | 330 (11%) | 1.89 (1.27 – 2.83) | 1.40 (0.42 – 4.63) | 0.002*; 0.59 |
| Lower endoscopy | 93 (3%) | 1.56 (0.67 – 3.65) | 1.13 (0.16 – 8.09) | 0.31; 0.90 |
| Upper endoscopy | 302 (11%) | 1.64 (1.07 – 2.51) | 1.72 (0.50 – 5.95) | 0.02*; 0.39 |
| Bronchoscopy | 56 (2%) | 0.66 (0.23 – 1.92) | NA | 0.45 |
| Surgery | 306 (11%) | 0.61 (0.35 – 1.08) | 1.32 (0.44 – 4.00) | 0.09; 0.62 |
| Colorectal surgery | 6 (0.2%) | 2.46 (0.32 – 18.97) | 26.70 (2.75 – 259.24) | 0.39; 0.01* |
| Abdominal surgery | 282 (10%) | 0.59 (0.32 – 1.08) | 1.01 (0.29 – 3.49) | 0.09; 1.00 |
| Urologic surgery | 22 (0.7%) | 0.44 (0.08 – 2.46) | NA | 0.35 |

* Significant at a globally-corrected P-value of ≤ 0.025 (α/2), accounting for two study outcomes (CRO and CPO colonization).

[1] Receipt of chemotherapy or immunosuppressive therapy in the prior 3 months, human immunodeficiency virus (HIV)-positive, and/or documented CBC immunosuppressive abnormalities within 24 hours preceding unit admission (defined as absolute neutrophil counts or total WBC counts less than 500 cells/$mm^3$).

[2] Pan-resistant to 4 of 5 antibiotic classes tested.

[3] Immunosuppressant or non-topical glucocorticoid.

Abbreviations – Not applicable due to model non-convergence or non-estimability (NA).

**Table 2.2. Carbapenemase Genes in Carbapenemase-Producing Admission Isolates**

| **Ambler Class Carbapenemase and Overall Percentage (%) Among Identified Carbapenemases** | **Species** | **No. of Isolates (n=38)** |
|---|---|---|
| **Class A** – *bla*$_{KPC}$ (52%) | *Klebsiella pneumoniae* | 8 |
| | *Enterobacter cloacae* | 2 |
| | *Escherichia coli* | 1 |
| | *Citrobacter amalonaticus* | 1 |
| **Class B** – *bla*$_{NDM}$ (9%) | *Klebsiella pneumoniae* | 1 |
| | *Citrobacter amalonaticus* | 1 |
| – *bla*$_{VIM}$ (4%) | *Pseudomonas aeruginosa* | 1 |
| **Class D** – *bla*$_{OXA-23}$ (9%) | *Acinetobacter baumannii* | 1 |
| | *Acinetobacter radioresistans* | 1 |
| – *bla*$_{OXA-24}$ (9%) | *Acinetobacter baumannii* | 2 |
| – *bla*$_{OXA-48}$ (4%) | *Klebsiella pneumoniae* | 1 |
| **Classes B and D** – bla$_{NDM + OXA-48}$ (13%) | *Klebsiella pneumoniae* | 3 |
| **Disposition of Remaining Isolates** | | |
| **Carbapenemase-Gene Negative on Check-Points Assay** | *Enterobacter cloacae (AmpC)* | 4 |
| | *Acinetobacter baumannii (Negative)* | 3 |
| | *Citrobacter amalonaticus (Negative)* | 1 |
| | *Achromobacter xylosoxidans (ESBL)* | 1 |
| **Check-Points Assay Failed Extraction** (Multiple attempts) | *Acinetobacter baumannii* | 1 |
| **Not Evaluated** – Chromosomally-encoded carbapenemase | *Aeromonas spp.* | 2 |
| **Not Evaluated** - Received whole genome sequencing (Results pending) | *Citrobacter freundii, Escherichia coli (co-colonized swab); Enterobacter cloacae* | 3 |

**Figure 2.3. Decision tree for Predicting CRO Perirectal Colonization at Hospital Unit Admission.** Gray-shaded terminal nodes indicate that the tree would classify patients as CRO-colonized, and accompanying percentages reflect the probability that patients assigned to a given terminal node are CRO-positive. Terminal node numbering, 1 through 4, is included in parentheses.

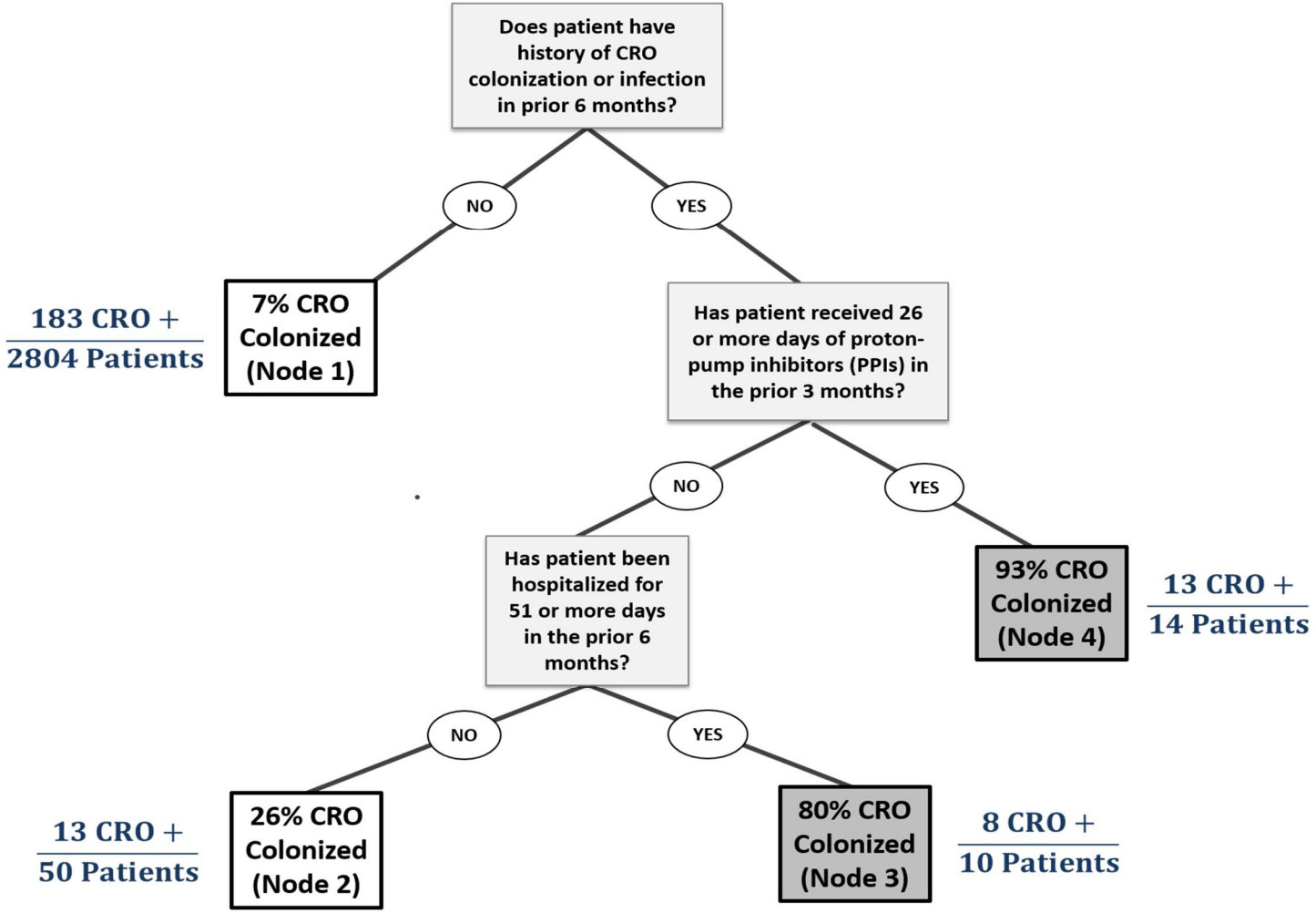
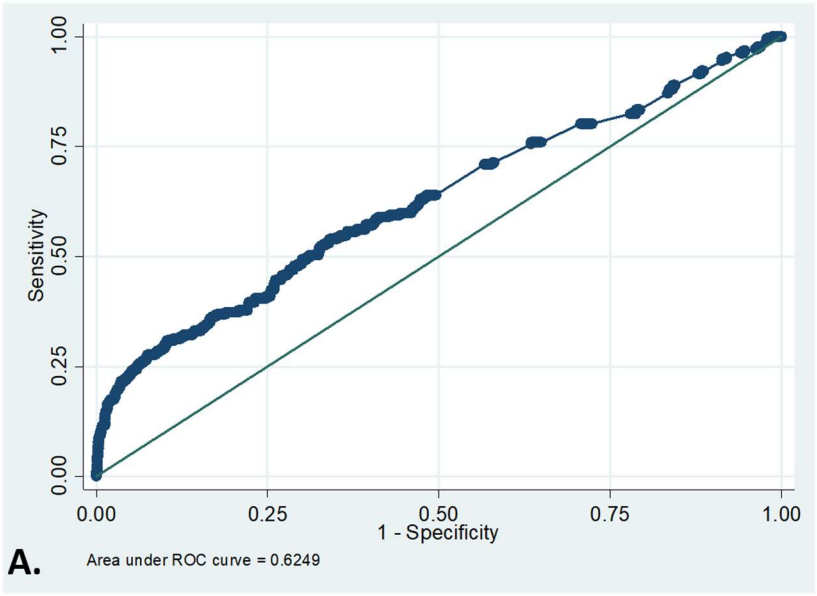
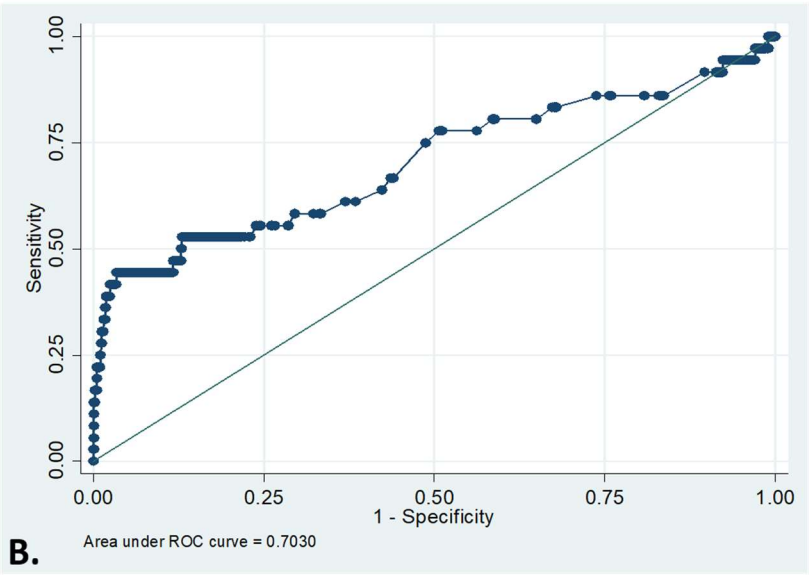


**Figure 2.4. Receiver Operating Characteristic (ROC) Curves for Multivariable Logistic Regression Models Predicting CRO and CPO Colonization at Unit Admission, Incorporating the Five Most Predictive Variables Identified in Random Forests Analysis.**

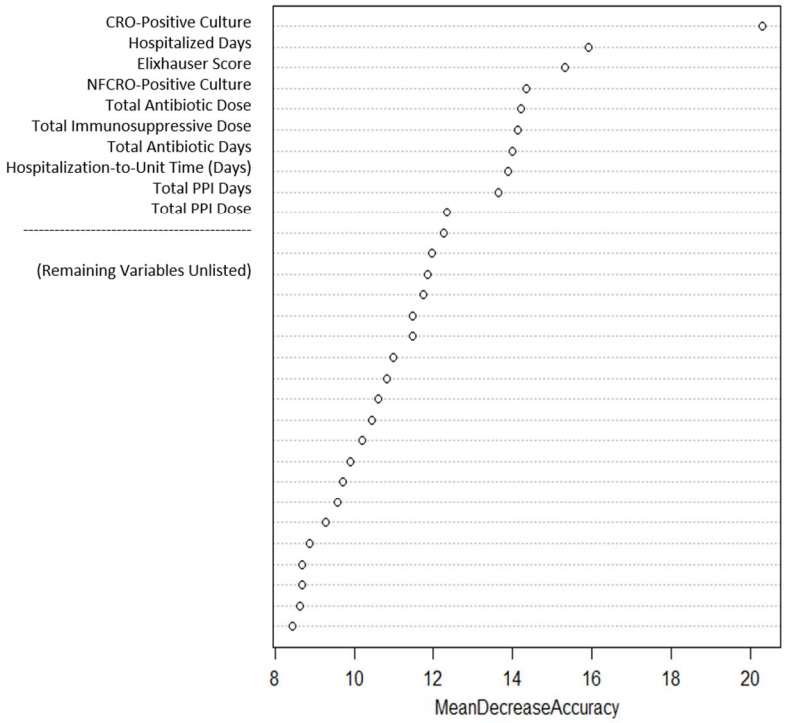



**A).** Receiver operating characteristic (ROC) curve for a multivariable logistic regression model with the outcome of carbapenem-resistant organism (CRO) colonization at unit admission and five independent variables: CRO-positive culture in the prior six months, total days of hospitalization in the prior six months, Elixhauser severity of illness score, total defined daily dose (DDD)-standardized doses of antibiotics with Gram-negative coverage in the prior three months, and total DDD-standardized doses of immunosuppressive therapy in the prior three months. Area under the curve (AUC) of 0.62; **B).** Corresponding ROC curve for the outcome of CPO colonization and five independent variables: CRO-positive culture in the prior six months, Elixhauser severity of illness score, total DDD-standardized doses of antibiotics with Gram-negative activity in the prior three months, time from hospital-to-unit admission (days), and total number of days of immunosuppressive therapy in the prior three months. AUC of 0.70.

**Figure 2.5. Variable Importance Plot of Most Predictive Variables for CRO Colonization and Corresponding Logistic Regression Output**

**A. Variable Importance Plot for Predicting CRO Colonization**



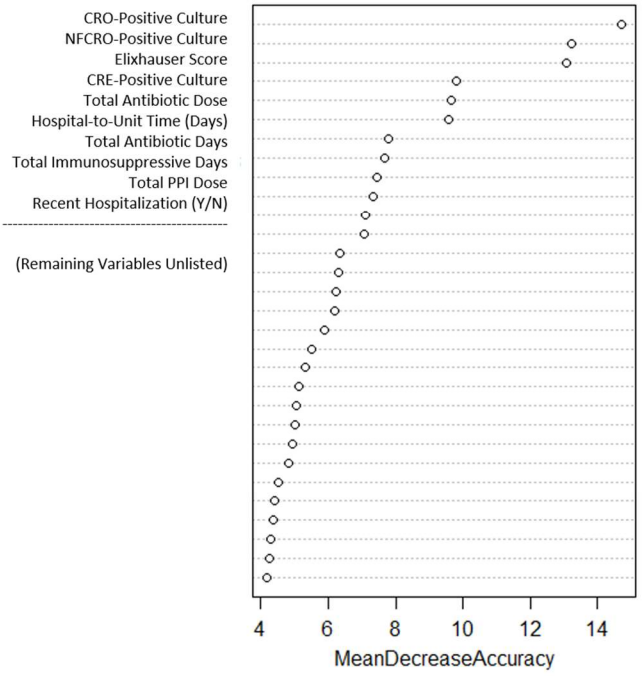

**B. Logistic Regression Results for CRO Colonization**

| cro | Odds Ratio | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| CRO-Positive Culture | 8.457535 | 2.216038 | 8.15 | 0.000 | 5.060722 | 14.13433 |
| Hospitalized Days | 1.011652 | .0035189 | 3.33 | 0.001 | 1.004779 | 1.018573 |
| Elixhauser Score | .9972176 | .0232808 | -0.12 | 0.905 | .9526163 | 1.043907 |
| Total Antibiotic Dose | 1.011914 | .0057096 | 2.10 | 0.036 | 1.000785 | 1.023167 |
| Total Immunosuppressive Dose | .9997978 | .0005066 | -0.40 | 0.690 | .9988053 | 1.000791 |
| Constant | .0619809 | .0083559 | -20.63 | 0.000 | .0475887 | .0807255 |

**A).** Random forests-generated variable importance plot ranking the ten most predictive variables for CRO colonization; **B).** Output from multivariable logistic regression analysis incorporating the top five, non-collinear variables from (A).

**Figure 2.6. Variable Importance Plot of Most Predictive Variables for CPO Colonization and Corresponding Logistic Regression Output**

### A. Variable Importance Plot for Predicting CPO Colonization




### B. Logistic Regression Results for CPO Colonization

| cpo | Odds Ratio | Std. Err. | z | P>\|z\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| CRO-Positive Culture | 21.63641 | 8.743855 | 7.61 | 0.000 | 9.79923 | 47.77257 |
| Hospital-to-Unit Time (Days) | .99983 | .0042456 | -0.04 | 0.968 | .9915433 | 1.008186 |
| Elixhauser Score | 1.001056 | .0581716 | 0.02 | 0.986 | .8932946 | 1.121816 |
| Total Antibiotic Dose | .9848818 | .0185744 | -0.81 | 0.419 | .9491413 | 1.021968 |
| Total Immunosuppressive Days | 1.035216 | .0144368 | 2.48 | 0.013 | 1.007304 | 1.063902 |
| Constant | .007044 | .0025195 | -13.86 | 0.000 | .0034944 | .0141995 |

**A).** Random forests-generated variable importance plot ranking the ten most predictive variables for CPO colonization; **B).** Output from multivariable logistic regression analysis incorporating the top five, non-collinear variables from (A).

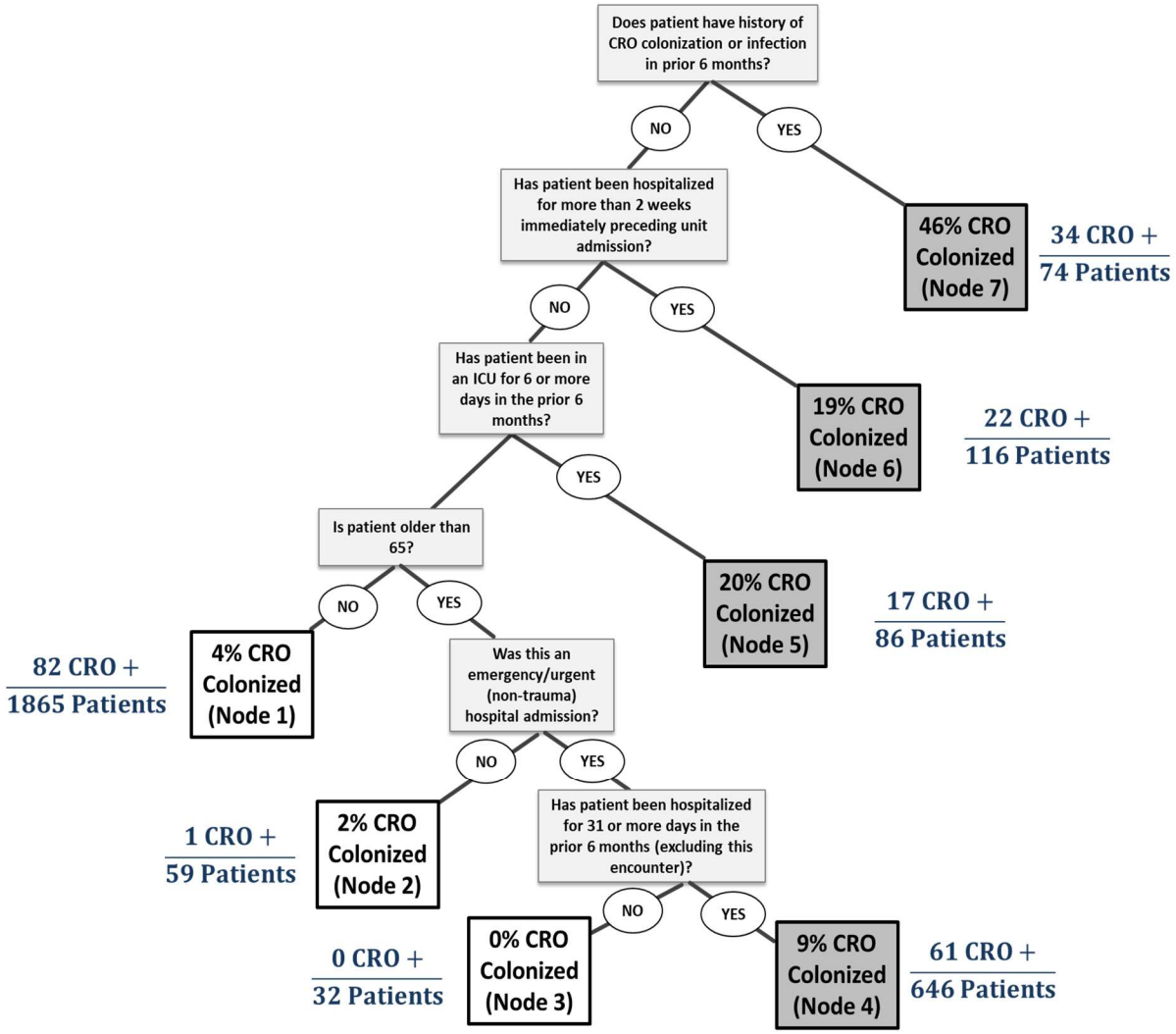
**SUPPLEMENTARY MATERIAL**

**Supplementary Table S.2.1.** Performance Metrics of CART Decision Trees for Predicting CRE Colonization at Unit Admission

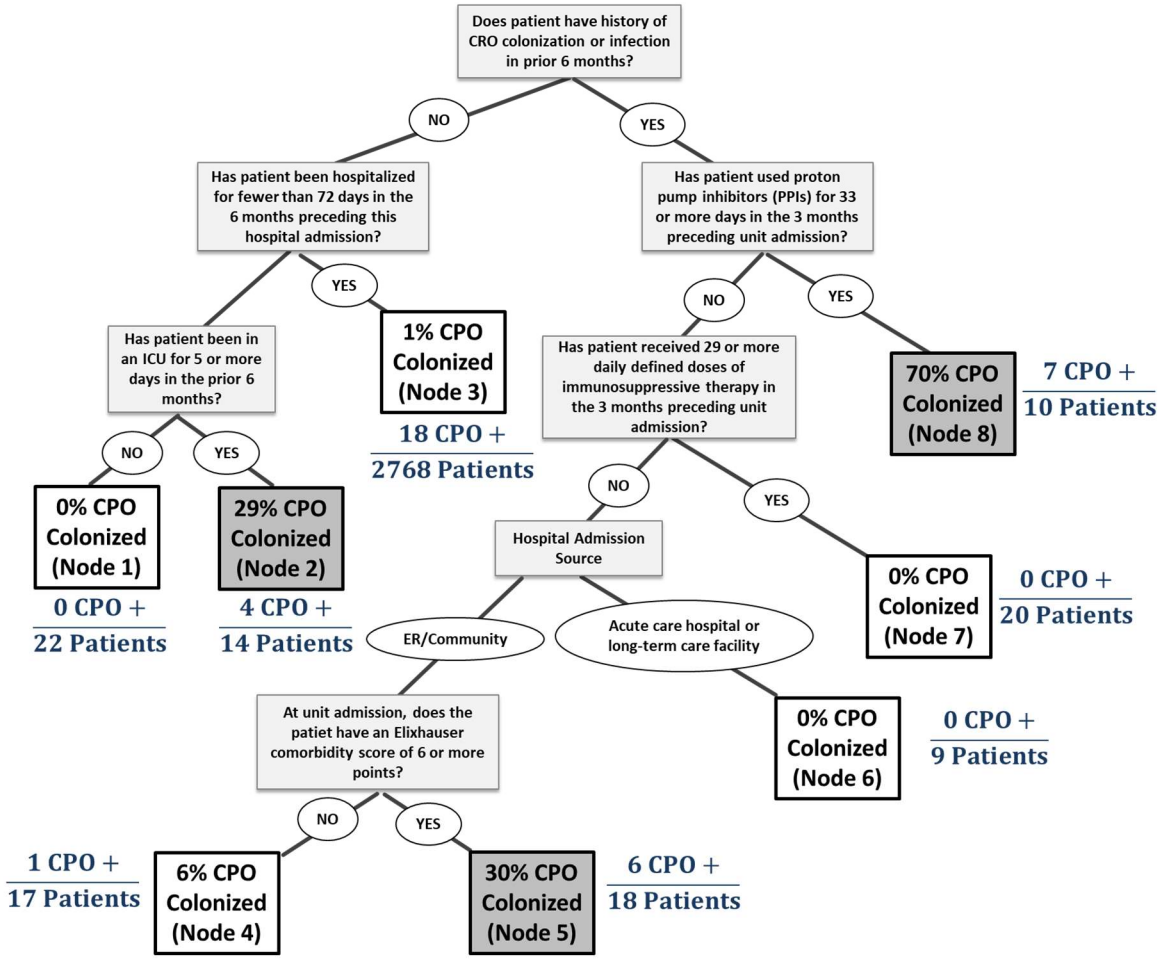
| | **Raw (No Tuning)** | **Tuned to Increase Sensitivity** |
|---|---|---|
| No. of Included Variables | 1 | 12 |
| Sensitivity | 5.8% | 70.2% |
| Specificity | 99.9% | 72.3% |
| Positive Predictive Value (PPV) | 64% | 10% |
| Negative Predictive Value (NPV) | 96% | 98% |
| C-Statistic | 0.53 | 0.76 |
| C-Statistic in Random Forests (RF) Analysis | 0.60 | 0.60 |

**Sensitivity Analyses**

**Supplementary Figure S.2.1. Decision tree for Predicting CRO Perirectal Colonization at Hospital Unit Admission, Fit with 50% Priors to Up-Weight Tree Sensitivity.** Gray-shaded terminal nodes indicate that the tree would classify patients as CRO-colonized, and accompanying percentages reflect the probability that patients assigned to a given terminal node are CRO-positive. Terminal node numbering is included in parentheses. The tree possessed a sensitivity of 61.8%, a specificity of 70.4%, a positive predictive value of 14.5%, and a negative predictive value of 95.8%. Its C-statistic was 0.70, and its C-statistic following random forest analysis was 0.65.



**Supplementary Figure S.2.2. Decision tree for Predicting CPO Perirectal Colonization at Hospital Unit Admission, Fit with 10% Priors to Up-Weight Tree Sensitivity.** Gray-shaded terminal nodes indicate that the tree would classify patients as CPO-colonized, and accompanying percentages reflect the probability that patients assigned to a given terminal node are CPO-positive. Terminal node numbering is included in parentheses. The tree possessed a sensitivity of 47.2%, a specificity of 99.1%, a positive predictive value of 40.5%, and a negative predictive value of 99.3%. Its C-statistic was 0.75, and its C-statistic following random forest analysis was 0.69.



**Supplementary Table S.2.2.** Performance Metrics of CART Decision Trees for Predicting CRO and CPO Colonization at Unit Admission, Restricted to First Unique-Patient Admission Encounters During the Study Period.*

| | **CRO Decision Tree** | **CPO Decision Tree** |
|---|---|---|
| No. of Outcomes (N=2165) | 141 | 20 |
| No. of Included Variables | 3 | 0 (Failed to Branch) |
| Sensitivity | 8.5% | NA |
| Specificity | 99.8% | NA |
| Positive Predictive Value (PPV) | 75% | NA |
| Negative Predictive Value (NPV) | 94% | NA |
| C-Statistic | 0.55 | 0.50 |

***** Number of first, unique-patient encounters during the study period equals 2165.
Abbreviations: “NA”, not applicable.

**Sensitivity Analysis Restricting to CPO Isolates with Molecularly-Confirmed Carbapenemases**

In order to address possible outcome misclassification, for the outcome of CPO colonization we refit a decision tree and performed random forests analysis restricting to CPO isolates with molecularly-confirmed carbapenemases. Of the 36 CPO-positive swabs, 22 were positive for one or more carbapenemases on the Checkpoints assay (representing 23 isolates, due to CP-CRE co-colonization on one swab). A decision tree fit to this data failed to branch, indicating that no variables were sufficiently predictive for this outcome. We also refit a decision tree with 10% priors in order to increase tree sensitivity. This tree possessed a C-statistic of 0.94 and included 14 variables, evidencing that it was overfit. In random forests analysis (no adjustment of priors), the C-statistic was 0.68, similar to the value for the primary CPO outcome (C-statistic 0.70). Taken together, the evidence indicates that restricting the CPO outcome to swabs possessing molecularly-confirmed carbapenemases did not materially improve, or in some cases reduced, model performance.

## 3. A Clinical Decision Tree to Predict Whether a Bacteremic Patient Is Infected With an Extended-Spectrum b-Lactamase–Producing Organism

**Authors:**

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

**Institutions:**

1 - Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

2 - Department of Medicine, Division of Infectious Diseases, Johns Hopkins University School of Medicine

3 - Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore

4 - Department of Medicine, Division of Infectious Diseases, University of Pennsylvania School of Medicine, Philadelphia

5 - Department of Pediatrics, Division of Infectious Diseases, Johns Hopkins University School of Medicine

6 - Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland1Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health

**Citation:**

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(7):896-903.

**ABSTRACT**

**Background.** Timely identification of extended-spectrum β-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 ≥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 ≥2 μg/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 ≥43 years, recent hospitalization in an ESBL high-burden

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

region, and ≥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.

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

## INTRODUCTION

Extended-spectrum β-lactamase (ESBL)-producing bacteria represent a serious clinical and public health challenge [1]. ESBL-producing bacteria can hydrolyze most broad-spectrum β-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 β-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 empirical 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].

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

Statistical models for predicting ESBL-producing infections can help to address current diagnostic limitations.

Numerous recent investigations have used multivariable logistic regression models to identify exposures independently associated with ESBLs (e.g., previous antibiotic therapy, presence of an indwelling urinary catheter, etc.) [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.

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

## METHODS

### Setting and Participants

This study included patients aged 18 years of age and older hospitalized at The 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 one of bacteremia, defined as the day the blood culture was obtained: (a) demographic data, (b) pre-existing medical conditions, (c) source of bacteremia, (d) indwelling hardware (e.g, orthopedic hardware, urology hardware,

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

central venous catheters, grafts, etc.), (d) 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], (e) 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, (f) days of stay in any healthcare facility (outpatient procedures were assigned "1 day of stay"), (g) hospitalization in another country in the previous 6 months, and (h) 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) ≥2 μg/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

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

μg/mL of clavulanic acid, versus 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 and corresponding 95% confidence intervals. Final multivariable logistic regression models were derived using stepwise variable selection with backward elimination at an alpha level of 0.05 (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 (CART) algorithm

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

Briefly, a tree was built using the following process: (1) identification of the single variable that, when used to split the dataset into two 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 (LOOCV) 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

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

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 3.1. Evaluating the full cohort, patients were a mean age of 55 years (SD, 16.4). Twenty-five percent of patients had a history of prior colonization or infection with a multidrug-resistant organism within the preceding six months. In the six 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 six months, and the majority (68%) had chronic indwelling vascular hardware present at the time of bacteremia onset. Twenty-five percent had at least one overnight stay in a hospital in an ESBL high-burden region within the prior six months. Figure 3.1 reflects the distribution of ESBL-positive bacteremia cases by geographic region.

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

**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 0.05, Table 1. The most strongly associated variables included prior history of an ESBL (odds ratio (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 3.1.

**Decision Tree**

Using binary recursive partitioning, the final classification tree for predicting ESBL-positive bacteremia included five study variables (Figure 3.2). 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 six 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 43 years or older? (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 one or more nights in the prior six months? If “yes,” (3) Has the patient received six or more days of antibiotics in the prior six months? (Based upon model-derived dichotomization at six 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 six 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 six hours after genus and species identification. Utilization of the

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

decision tree would have increased ESBL case detection during the empiric treatment window by approximately 50%. The decision tree identified one-third of the original 68 patients, as well as an additional 78 ESBL cases, as "ESBL-positive" and warranting empiric therapy with agents covering ESBL-producing bacteria.

**Sensitivity Analyses**

Approximately 45% (86/194) of ESBL-positive bacteremia patients were classified in Terminal node 1 as ESBL-negative, compromising decision tree sensitivity. We performed sensitivity analyses on the subset of 1152 patients who lacked the two strongest study risk factors of prior ESBL infection or colonization history and recent international hospitalization in an ESBL high-burden region. We first re-fit 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 three 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).

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

**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 more than 30 demographic and clinical variables, we developed a decision tree with five 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 six months (model-derived dichotomization at six days). Patients classified as ESBL-positive by the tree were 90.8% likely to be true ESBL cases (PPV), and patients classified as negative were 91.9% likely to be true ESBL-negative cases (NPV).

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

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 importantly, 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 remain flexible to differing risk-aversion attitudes. 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.

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

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 high ESBL prevalence regions. 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, in order to reduce outcome misclassification, we restricted our study to *E. coli* and *Klebsiella* spp., as Centers for Disease Control and Prevention screening methodology to test for ESBL production is limited to these organisms. 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 was likely missing data that could lead to exposure misclassification, although we would not expect it 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, in order 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

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

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.

**Funding**

This work was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under Award Number UM1AI104681. The work was also supported in part by a K24-AI079040 to A.D.H. and a K24-AI080942 to E.L, both from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

**Conflicts of Interest**

None of the authors report any conflicts of interest.

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

**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(2):155-64.

2. Paterson DL and Bonomo RA. Extended-spectrum beta-lactamases: a clinical update. Clin Microbiol Rev 2005;18(4): 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(3): 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(12):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(7):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(9):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(3):1488-95.

8. McLaughlin M, Advincula MR, Malczynski, et al. Correlations of antibiotic use and carbapenem resistance in enterobacteriaceae. Antimicrob Agents Chemother 2013; 57(10):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(8):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(11):753-62.

12. Nguyen ML, Toye B, Kanji S, Zvonar R. Risk Factors for and Outcomes of Bacteremia Caused by Extended-Spectrum ss-Lactamase-Producing Escherichia coli and Klebsiella Species at a Canadian Tertiary Care Hospital. Can J Hosp Pharm, 2015; 68(2):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(4):323-48.

14. Breiman L, Friedman J, Stone CJ, Olshen RA. Classification and regression trees. 1984, CRC: Chapman & Hall.

15. Antimicrobial Resistant Phenotype Definitions. http://www.cdc.gov/nhsn/pdfs/ps-analysis-resources/phenotype_definitions.pdf. Accessed April 7th, 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(9):2144-53.

17. Kantele A, Laaveri T, Mero S, et al. Antimicrobials increase travelers' risk of colonization by extended-spectrum betalactamase-producing Enterobacteriaceae. Clin Infect Dis 2015; 60(6):837-46.

18. Tibshirani R. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B 1996;**58**:267-288.

19. Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd ed. 2001, United States: Wiley-Interscience.

20. Chen S and Ishwaran H. Pathway hunting by random survival forests. Bioinformatics, 2013. 29(1): p. 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(2):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(1): 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(7): 967-75.

24. Dayan N, Dabbah, 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(5):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**(5): p. 399-405.

26. Megged O. Extended-spectrum beta-lactamase-producing bacteria causing community-acquired urinary tract infections in children. Pediatr Nephrol 2014; 29(9):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(8):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(1): p. 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(6): 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(3):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(9): 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(3):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(3): 359-62.

34. CDC Laboratory Detection of Extended-Spectrum B-Lactamases (ESBLs). http://www.cdc.gov/HAI/settings/lab/lab_esbl.html. Accessed May 26th, 2016.

**Table 3.1. 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 | ESBL Negative | Odds Ratio (95% Confidence Interval) | P-value |
|---|---|---|---|---|
| | N = 194 | N = 1094 | | |
| **Demographics** | | | | |
| Age | 51 ± 18.4 | 56 ± 15.9 | 0.98 (0.97 – 0.99) | <0.001 |
| Male sex | 113 (58%) | 590 (54%) | 1.18 (0.87 – 1.61) | 0.23 |
| Race/Ethnicity | | | | |
| White | 85 (44%) | 523 (48%) | Reference | Reference |
| Black | 49 (25%) | 458 (42%) | 0.66 (0.45 – 0.96) | 0.03 |
| Latino | 11 (6%) | 39 (4%) | 1.74 (0.86 – 3.52) | 0.13 |
| Asian | 25 (13%) | 38 (3%) | 4.05 (2.33 – 7.05) | <0.001¥θ |
| Middle Eastern | 24 (12%) | 26 (2%) | 5.68 (3.12 – 10.35) | <0.001 |
| **Pre-existing Medical Conditions** | | | | |
| Human Immunodeficiency Virus positive | 5 (3%) | 53 (5%) | 0.52 (0.21 – 1.32) | 0.17 |
| Chemotherapy within previous 6 months | 83 (43%) | 347 (32%) | 1.61 (1.18 – 2.20) | 0.003 |
| Active Immunosuppressant Use[1] | 8 (4%) | 65 (6%) | 0.68 (0.32 – 1.44) | 0.32 |
| Solid organ transplantation | 29 (15%) | 145 (13%) | 1.15 (0.75 – 1.77) | 0.53 |

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

| | | | | |
|---|---|---|---|---|
| Hematopoeitic stem cell transplantation | 12 (6%) | 48 (4%) | 1.44 (0.75 – 2.76) | 0.28 |
| End-stage liver disease | 17 (9%) | 76 (7%) | 1.29 (0.74 – 2.23) | 0.37 |
| End-stage renal disease requiring dialysis | 15 (8%) | 81 (7%) | 0.96 (0.62 – 1.48) | 0.84 |
| Congestive heart failure (ejection fraction <40) | 16 (8%) | 81 (7%) | 1.12 (0.64 – 1.97) | 0.68 |
| Structural lung disease[2] | 19 (10%) | 44 (4%) | 2.60 (1.48 – 4.54) | 0.001¥θ |
| **Indwelling Hardware at the Onset of Bacteremia** | | | | |
| Biliary stent | 18 (9%) | 119 (11%) | 0.84 (0.50 – 1.41) | 0.51 |
| Gastrointestinal feeding tube | 25 (13%) | 57 (5%) | 2.69 (1.64 – 4.43) | <0.001¥θ |
| Nephrostomy tubes and/or Foley catheter | 45 (23%) | 113 (10%) | 2.62 (1.78 – 3.86) | <0.001¥θ |
| Chronic vascular hardware[3] | 131 (68%) | 461 (42%) | 2.86 (2.07 – 3.95) | <0.001¥θ |
| Orthopedic hardware | 5 (3%) | 20 (2%) | 1.42 (0.53 – 3.83) | 0.49¥θ |
| **Recent Multidrug-resistant Organism History (Colonization or Infection <6 Months)** | | | | |
| Vancomycin-resistant *Enterococcus* species. | 32 (17%) | 113 (10%) | 1.72 (1.12 – 2.63) | 0.01 |
| Methicillin-resistant *Staphylococcus aureus* | 8 (4%) | 45 (4%) | 1.00 (0.47 – 2.16) | 1.00 |
| Extended-spectrum β-lactamase | 84 (43%) | 16 (2%) | 51.45 (29.11 – 90.93) | <0.001¥θ |
| Carbapenem-resistant Enterobacteriaceae[4] | 4 (2%) | 1 (<1%) | 23.01 (2.56 – 206.99) | 0.01¥θ |

| | | | | |
|---|---|---|---|---|
| Multidrug-resistant *Pseudomonas* species[4] | 4 (2%) | 14 (1%) | 1.62 (0.53 – 4.99) | 0.40[¥] |
| Multidrug-resistant *Acinetobacter* species[4] | 2 (1%) | 1 (<1%) | 11.39 (1.03 – 126.18) | 0.05 |
| **Recent Antibiotic Exposure (<6 Months)** | | | | |
| Days of extended-spectrum penicillin therapy | 6.6 ± 11.2 | 3.5 ± 8.0 | 1.03 (1.02 – 1.05) | <0.001 |
| Days of third and fourth-generation cephalosporin therapy | 4.9 ± 8.2 | 2.1 ± 4.8 | 1.07 (1.04 – 1.09) | <0.001[θ] |
| Days of aztreonam therapy | 0.3 ± 1.5 | 0.2 ± 1.9 | 1.02 (0.95 – 1.10) | 0.61 |
| Days of carbapenem therapy | 5.0 ± 9.0 | 1.8 ± 6.2 | 1.05 (1.03 – 1.07) | <0.001 |
| Days of fluoroquinolone therapy | 3.1 ± 6.8 | 2.2 ± 6.8 | 1.02 (1.00 – 1.04) | 0.10 |
| Days of aminoglycoside therapy | 1.3 ± 4.7 | 0.3 ± 1.9 | 1.11 (1.05 – 1.17) | <0.001[¥ θ] |
| Total days of antibiotics (combined) | 21.0 ± 25.6 | 10.0 ± 18.6 | 1.02 (1.01 – 1.03) | <0.001 |
| **Total Days of Hospitalization in the 6 Months Prior to Current Hospitalization** | 23.1 ± 26.7 | 12.0 ± 18.5 | 1.02 (1.01 – 1.03) | <0.001 |
| **Duration of Time from Hospital Admission to Positive Blood Culture (days)** | 11.4 ± 41.0 | 5.7 ± 20.0 | 1.01 (1.002 – 1.01) | 0.01 |
| **Recent International Healthcare Exposure (<6 Months)** | | | | |
| **At least one overnight stay in a healthcare facility in an ESBL high-burden region[5]** | 49 (25%) | 12 (1%) | 30.47 (15.83 – 58.64) | <0.001[¥ θ] |

| | | | | |
|---|---|---|---|---|
| **Other High-Risk Healthcare Exposures (<6 Months)** | | | | |
| Long-term acute care facility residence | 17 (9%) | 22 (2%) | 4.68 (2.44 – 8.99) | <0.001¥θ |
| Nursing home residence | 6 (3%) | 16 (2%) | 2.15 (0.83 – 5.57) | 0.12 |
| **Source of Bacteremia** | | | | |
| Urinary tract | 65 (34%) | 407 (38%) | Reference | Reference |
| Skin and soft tissue | 4 (2%) | 43 (4%) | 0.59 (0.21 – 1.71) | 0.33θ |
| Biliary | 16 (8%) | 168 (15%) | 0.60 (0.34 – 1.07) | 0.08 |
| Intra-abdominal | 35 (18%) | 271 (25%) | 0.80 (0.52 – 1.25) | 0.33 |
| Catheter-related | 57 (29%) | 143 (13%) | 2.48 (1.66 – 3.72) | <0.001¥θ |
| Bone and/or joint | 1(<1%) | 10(1%) | 0.62 (0.08 – 4.95) | 0.66 |
| Pneumonia | 16 (8%) | 57 (5%) | 1.75 (0.95 – 3.23) | 0.07¥θ |

[1]Excluding chemotherapy or immunosuppression for solid organ transplants; [2]Chronic obstructive pulmonary disease, emphysema, tracheostomy-dependent; [3]central venous catheter or dialysis catheter; [4]http://www.cdc.gov/nhsn/pdfs/ps-analysis-resources/phenotype_definitions.pdf;[5]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), UAE (5), Bangladesh (2), India (7), Pakistan (5), Egypt (2), Greece (2), Turkey (3). An additional eight and nine ESBL-positive and ESBL-negative patients, respectively, were hospitalized internationally in a non-high-burden region in the six months preceding bacteremia; ¥ - Significant in multivariable analysis using stepwise selection with backwards elimination at an alpha level 0.05. Among variables that were significant in multivariable analysis, one variable, a history of multidrug-resistant Pseudomonas species, demonstrated qualitative confounding (univariable and multivariable odds ratios 1.62 and 0.08, respectively); θ - Retained in final multivariable model using lasso regression.

**Figure 3.1. Distribution of recent international healthcare exposure among ESBL-positive cases**. 57 of 194 ESBL-positive patients had a recent international healthcare exposure, defined as hospitalization for one or more nights outside of the United States in the six months preceding ESBL bacteremia.

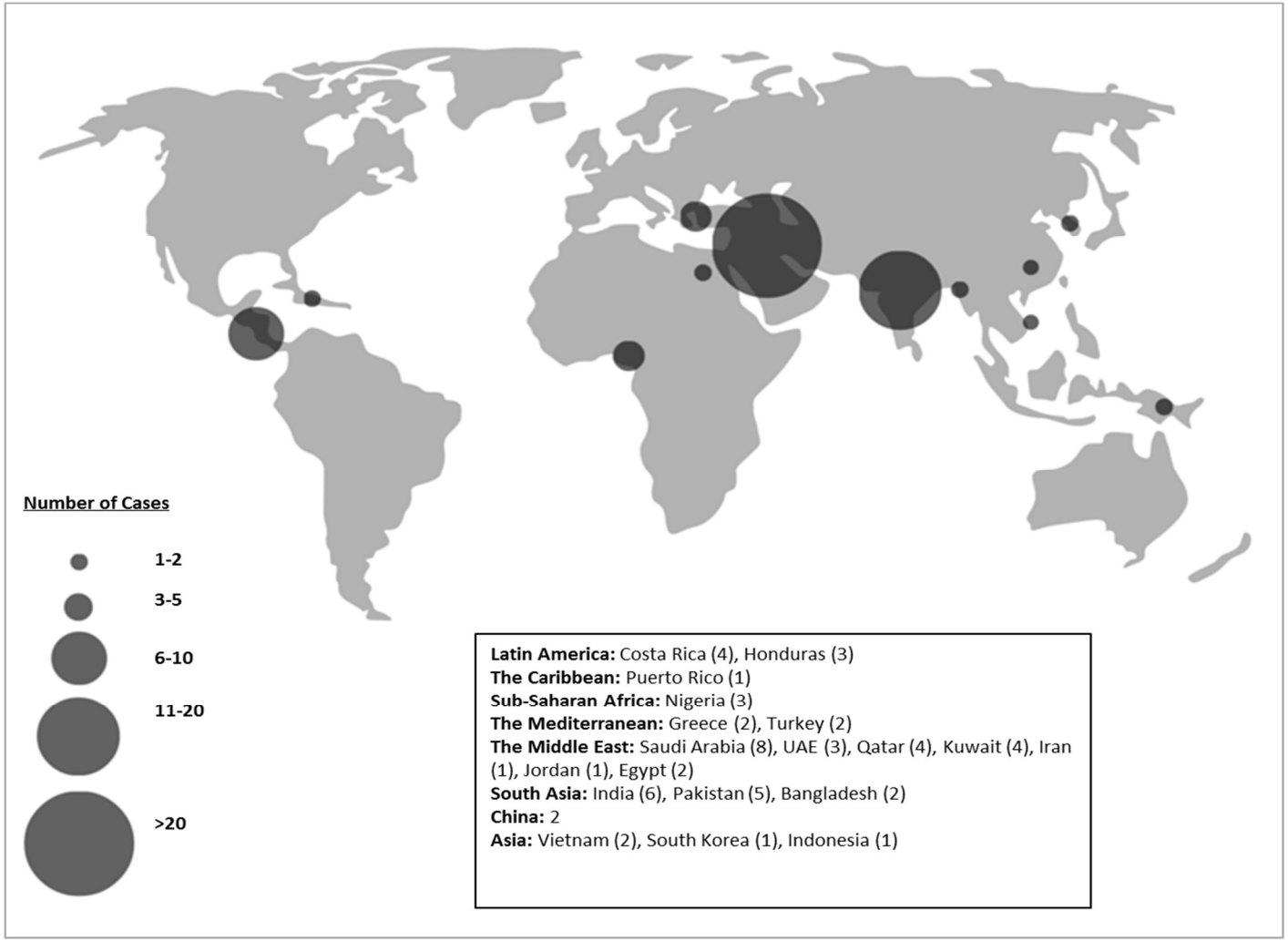
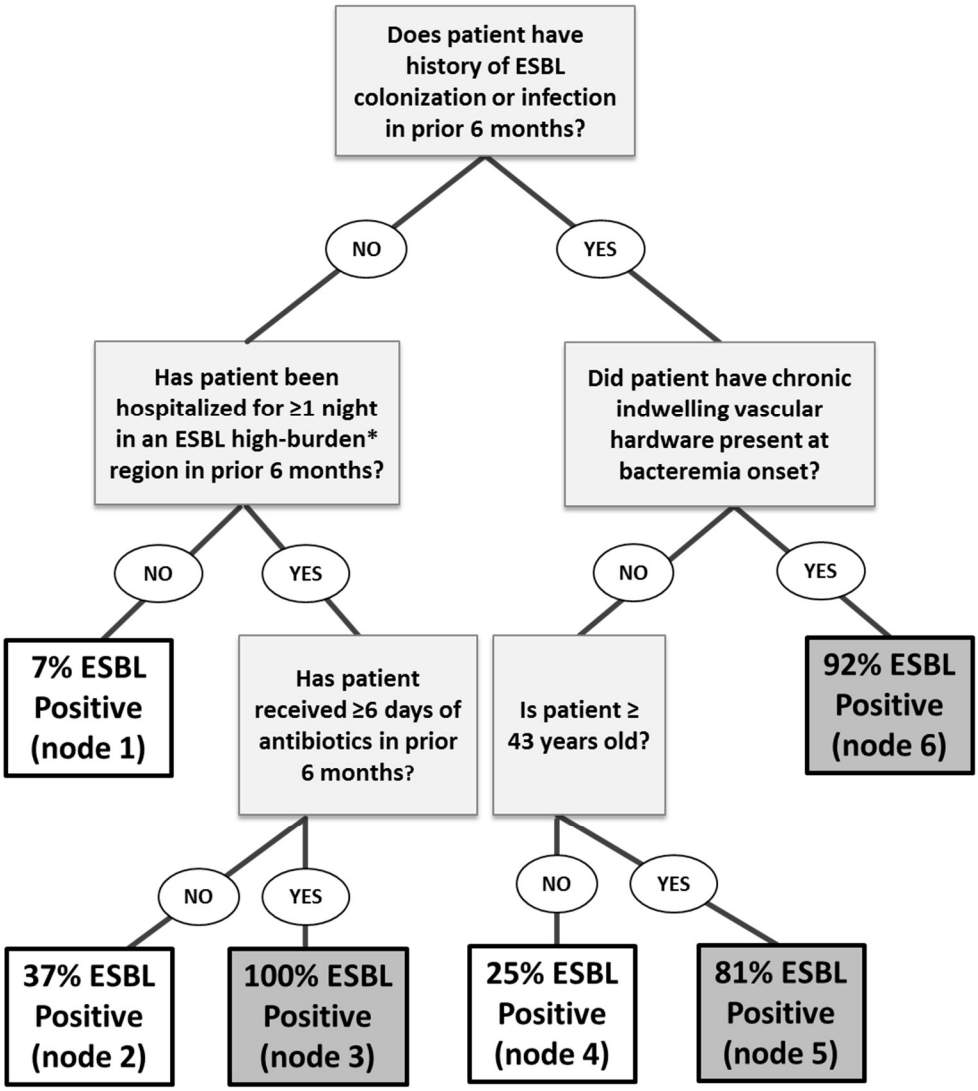


**Figure 3.2. Clinical decision tree to predict a bacteremic patient's likelihood of infection with an ESBL-producing bacteremia at the time of organism genus and species identification.** Gray-shaded terminal nodes indicate that the tree would classify patients as ESBL-positive, and accompanying percentages (derived from terminal node impurities) reflect the probability that patients assigned to a given terminal node are ESBL-positive. Terminal node numbering, 1 through 6, is included in parentheses.




***Number of patients in each terminal node: (1) 1152; (2) 19; (3) 17; (4) 8; (5) 21; (6) 71.***

* Latin America (excluding the Caribbean); the Middle East (including Egypt); South Asia; China; and the Mediterranean.

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

## 4. A Methodological Comparison of Risk Scores Versus Decision Trees for Predicting Drug-Resistant Infections: A Case Study using ESBL Bacteremia

**Authors:**

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

**Institutions:**

1 - Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health.

2 - Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore.

3 - Department of Pediatrics, Division of Infectious Diseases, Johns Hopkins University School of Medicine.

**ABSTRACT**

**Background.** Timely identification of multidrug-resistant Gram-negative infections, including extended-spectrum β-lactamase (ESBL)-producing organisms, remains a clinical 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 clinical infectious disease literature. Machine learning-derived decision trees are an alternative approach for developing clinical 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 two methods and to offer general guiding principles for when clinicians might consider each approach.

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

**Results.** 194 patients (15%) were infected with ESBL-producing bacteremia. The clinical risk score included 14 variables, compared to the decision tree's five. The score and decision tree's positive and negative predictive values were similar (>90%), but the score's C-statistic (0.87) was 10-percentage-points higher.

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

## INTRODUCTION

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. Currently there is no Clinical and Laboratory Standards Institute (CLSI)-endorsed phenotypic method 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 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 resource-impractical. Logistic regression-derived risk scores are common in the clinical infectious disease literature. Classification and

regression tree (CART) analysis or "recursive partitioning," a form of machine learning, is an alternative approach for developing this type of clinical decision support tool. Our group previously reported on a CART decision tree for predicting ESBL bloodstream infections to guide empiric therapy [8]. There has been interest in evaluating whether a risk score derived from the same population could achieve greater predictive accuracy while remaining sufficiently simple to incorporate into clinical practice.

We present a case study of the development of a risk score from the same ESBL dataset as our original decision tree to compare the predictive accuracy of these two methods, to illustrate the advantages and disadvantages of logistic regression risk scores versus CART decision trees, and to offer general guiding principles for when clinical researchers 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* species, 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 one of bacteremia (defined as the date the initial blood culture was collected), included the following: (a) demographic data; (b) pre-existing medical conditions; (c) presumptive source of bacteremia (e.g., catheter, pneumonia); (d) indwelling hardware; (e) multidrug-resistant organism (MDRO) colonization or infection (multidrug-resistant *Pseudomonas aeruginosa*, multidrug-resistant *Acinetobacter baumannii*, ESBL-producing *Enterobacteriaceae*, carbapenem-resistant *Enterobacteriaceae*, vancomycin-resistant *Enterococcus* species, and methicillin-resistant *Staphylococcus aureus*) [9] in the prior six months; (f) days of gram-negative active antibiotic therapy in the prior six months; (g) length of stay in any healthcare facility in the prior six months; (h) post-acute care facility stay in the prior six months; and (i) hospitalization in another country in the prior six 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 previously described [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 (e.g., the mixture of ESBL-positive and ESBL-negative patients), also offer associated probabilities. We internally validated the performance of our tree using the leave-one-out cross-validation method (LOOCV)[12] and 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, version 4.1–9, in R.

To develop a risk score, continuous variables (e.g., age, antibiotic days) were first converted into ordinal categories in order to reduce complexity, given the score's anticipated manual application. Multivariable logistic regression models were derived using stepwise variable

selection with backward elimination at an α level of 0.05 and lasso regression at the value of the shrinkage parameter that minimized misclassification error in the cross-validated lasso model. The most parsimonious model was selected for conversion to a risk score. 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), in order to simplify end-user calculations). Patient scores were calculated by summing their respective points (risk score model).

For both the final multivariable regression model and the risk score model, discrimination was assessed with ROC curves and accompanying C-statistics (i.e., area under the curve). Risk score model calibration was evaluated by 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 (LOOCV). Risk score analyses were performed in Stata, version 13.0 (StataCorp, College Station, TX) and R.

**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 previously reported [8].

**Risk Score**

Multivariable models derived using stepwise variable selection and lasso regression included 14 and 16 variables, respectively, with full agreement on the first 14 variables. The simpler stepwise-derived model was selected for risk score development. Therefore, the final clinical risk score for predicting ESBL-positive bloodstream infection at the time of organism and genus identification included 14 variables, broadly categorizable into six groups (Figure 4.1):

(A) 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).

(B) Presumptive source of bloodstream infection**:** Catheter (2 points); pneumonia (2 points).

(C) Patient characteristics**:** Structural lung disease (Chronic obstructive pulmonary disease, emphysema, or tracheostomy-dependency) (2 points); self-identification as Asian race (2 points).

(D) 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).

(E) MDRGN colonization or infection within the previous 6 months**:** ESBL (6 points); carbapenem-resistant *Enterobacteriaceae* (CRE) (6 points); MDR *Pseudomonas* species (-4 points).

(F) Antibiotic exposure within the previous 6 months**:** Weeks of active gram-negative therapy (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 to 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 (Supplement 4.1). The multivariable logistic regression model evidenced acceptable calibration (HL goodness-of-fit test $p$=0.13). Following points-conversion, however, the risk score model over- or under-estimated 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 $p$-value<0.001) (Supplement 4.2). An ESBL-positive cut-point of ≥ 7.25 points maximized overall ESBL-classification accuracy (92%). At this cut-off, 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 4.1 provides the risk score's sensitivity and specificity at each possible ESBL-positive cut-point.

**Decision Tree**

The decision tree using this cohort has been previously described [8] (Figure 4.2). The final tree included five predictors: central vascular catheter, age ≥43 years, and in the prior six 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. The tree's C-statistic was 0.77 (unchanged in cross-validation), its sensitivity and specificity were 51.0% and 99.1%, and its positive and negative predictive values were 90.8% and 91.9%, respectively. Table 4.2 compares the risk score and decision tree's performance metrics.

**DISCUSSION**

Despite advances in rapid diagnostics, timely identification of MDRGNs remains a clinical challenge. Diagnostic delays can prolong the period of ineffective antibiotic therapy and also 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 resource-impractical. This case study of ESBL bloodstream infections explores two approaches for developing predictive models: traditional logistic regression-derived risk scores and machine learning-derived decision trees.

Given that risk scores for binary predictions are dichotomized at a cut-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 did possess an approximately 10-percentage-point 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 cut-point. It should be noted that, 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 clinician application. Studies consistently demonstrate that incorporating decision support tools at the point of care is important to their success[14], but a risk score that requires 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 (i.e., five) predictors.

The potential tradeoff between flexibility and user-friendliness is one important consideration when evaluating whether risk scores or decision trees are a more suitable clinical decision support tool for a given application. Additional considerations at various stages of model development and implementation, however, may also help to guide clinical researchers in selecting one option versus the other. Below, we summarize risk scores' and decision trees' relative strengths for model development and fitting, implementation, and adaptability. We will examine the model algorithm and the outputted model (i.e., clinical decision support tool) in parallel because, in practice, this is how most clinical researchers are likely to implement these methods. These concepts, however, are not synonymous: CART analysis is the tree-fitting process (approach), and a decision tree is the result (output), just as logistic regression is one common approach for developing a risk score. Approach and output can differ in their strengths and limitations, and we attempt to distinguish these concepts in our discussion where relevant.

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 (i.e., correlation) among independent variables and a sufficient

ratio of cases-to-predictors (i.e., sufficient sample size; a general, although debatable, guideline is 10 expected cases per predictor evaluated) [15, 16]. In contrast, CART is non-parametric 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 clinical research, given the abundance of predictors in patient medical records but 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 where understanding of variable relationships is generally limited.

CART's benefits, 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[17]. 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 use of internal validation methods (e.g., cross-validation) when external testing datasets are unavailable. While 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) [18, 19].

Decision trees do not require calculations and are therefore intuitive and generally user-friendly for clinicians to apply. Where 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. Where an automated implementation is contemplated, a risk score might be equally easy to use. If automation is feasible, however, presumably technical expertise is also readily available, and ensemble machine learning algorithms (e.g., random forests or Super Learner) [19, 20] would likely outperform both logistic regression and CART (and yield scores or binary predictions as preferred). In this ESBL case study, because important variables required clinical judgment (e.g., source of infection) or were not hard-coded in the EHR (e.g., foreign country of recent hospitalization was only entered as natural language), automating the clinical decision support tool would have been challenging. As a result, the decision tree's simplicity was highly valuable for this research application.

Finally, for clinical applications where decision support tool flexibility is paramount, risk scores are attractive because their cut-points are modifiable by end-users. Risk scores provide a range of score cut-offs, each with an associated sensitivity and specificity, which allow individual users to toggle the cut-point in order to minimize the false-positive or false-negative rate (e.g., depending upon infection severity or the clinical appearance of the patient). Using the current clinical risk score, for example, a user seeking to increase sensitivity could choose a lower cut-point of ≥ 3 points and reduce the risk of incorrectly classifying an ESBL infection as ESBL-negative to less than 1 in 5 (sensitivity 83.5%, specificity 73.1%) (Table 4.1). This flexibility

allows clinicians and hospital epidemiologists to maximize detection of cases, i.e., 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 cut-point adjustment. Large score differences between patients may translate to minimal differences in risk, and vice versa. It is imperative that the table of cut-point sensitivities and specificities guides decisions about score thresholds for ESBL infection.

In contrast to risk scores, classification trees provide binary predictions (e.g., "ESBL" or "not ESBL"), with a single sensitivity and specificity value for the tree as a whole. Terminal node percentages (e.g. "37% probability that ESBL-positive") can quantify these predictions but do not provide a formal mechanism for prioritizing sensitivity versus specificity at earlier points in the tree's branching. For research applications where sensitivity is the priority, methods are available to impose a greater "cost" for case misclassification during the tree-fitting process [21]. The limitation, however, is that these mechanisms are not adjustable by end-users after a tree is built. In other words, while 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 the above considerations can help researchers to evaluate whether a risk score or a decision tree is preferable for a given research question (Table 4.3), a decision is rarely clear-

cut. Where each model would at least partially meet stated goals, we encourage investigators to develop both support tools in parallel in order to compare their performance metrics. In particular, although model performance was comparable in this case study, other applications with more challenging data (e.g., high-dimensionality, higher-order variable interactions) might more clearly favor a machine learning approach such as CART.

A hybrid analysis that combines methodological strengths — e.g., aspects of one approach with the output of the other — should also be considered. For example, even where a risk score is the optimal output, ensemble tree methods can rank the most predictive variables to compare to the variables retained in stepwise regression [18, 19]; preliminary data analysis with CART can also identify potentially meaningful variable interactions for modeling interaction terms during regression model development. Conversely, researchers fitting decision trees might still consider supplemental regression analysis if, for example, ascertainment of "independent" variables is important for generating causal hypotheses. Because most decision tree algorithms optimize predictive accuracy, variables retained in decision trees can be arbitrary among collinear predictors, and they do not necessarily reflect independent exposures [22].

Our study has several limitations. It was a single-center study, and although we internally validated our models, it lacked an external validation cohort. In addition, there was the potential for missing data in patients treated outside of the Epic Care Everywhere network, although we do not expect missingness to differ by ESBL status. As such, any resulting exposure

misclassification would likely reduce predictive performance, and yet risk score discrimination remained robust. Nevertheless, we would encourage future evaluation of the risk score in other cohorts. Importantly, however, because study characteristics were constant across analyses, we expect decision tree and risk score comparisons to be unbiased.

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 where laboratory diagnostics are challenging to implement. This examination explored two alternative decision support tools, logistic regression-derived risk scores and machine learning-derived decision trees, in an inpatient cohort of bacteremic patients in order to predict ESBL infection. These methodologies offer different strengths and limitations, and we hope that their continued utilization in clinical infectious disease research will assist with improving patient outcomes.

**Funding**

The work was supported by funding from the Agency for Healthcare Research and Quality [R36HS025089] awarded to KEG and from the National Institutes of Health [K23-AI127935] awarded to PDT.

**Conflicts of Interest**

None of the authors report any conflicts of interest.

**REFERENCES**


1. Livorsi DJ, Chorazy ML, Schweizer ML, Balkenende EC, Blevins AE, Nair R, Samore MH, Nelson RE, Khader K, Perencevich EN: 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, Nelson R, Samore M, Khader K, Blevins AE, Diekema D, Chiang HY, Nair R *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(10):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).
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(10):2133-2140.
5. Clinical and Laboratory Standards Institute: Performance standards for antimicrobial susceptibility testing, 28th ed. Supplement M100S. In. Wayne, PA: CLSI; 2018.
6. Ledeboer NA, Lopansri BK, Dhiman N, Cavagnolo R, Carroll KC, Granato P, Thomson R, Jr., Butler-Wu SM, Berger H, Samuel L *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(8):2460-2472.
7. Ward C, Stocker K, Begum J, Wade P, Ebrahimsa U, Goldenberg SD: Performance evaluation of the Verigene(R) (Nanosphere) and FilmArray(R) (BioFire(R)) molecular assays for identification of

causative organisms in bacterial bloodstream infections. *Eur J Clin Microbiol Infect Dis* 2015, 34(3):487-496.

8. Goodman KE, Lessler J, Cosgrove SE, Harris AD, Lautenbach E, Han JH, Milstone AM, Massey CJ, Tamma PD, Antibacterial Resistance Leadership G: 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(7):896-903.

9. Centers for Disease Control and Prevention: Antimicrobial resistant phenotype definitions. In.; 2016.

10. Kantele A, Laaveri T, Mero S, Vilkman K, Pakkanen SH, Ollgren J, Antikainen J, Kirveskari J: Antimicrobials increase travelers' risk of colonization by extended-spectrum betalactamase-producing Enterobacteriaceae. *Clin Infect Dis* 2015, 60(6):837-846.

11. Ostholm-Balkhed A, Tarnberg M, Nilsson M, Nilsson LE, Hanberger H, Hallgren A, Travel Study Group of Southeast S: Travel-associated faecal colonization with ESBL-producing Enterobacteriaceae: incidence and risk factors. *J Antimicrob Chemother* 2013, 68(9):2144-2153.

12. Duda RO, Hart PE, Stork DG: Pattern Classification, 2nd edn. United States: Wiley-Interscience; 2001.

13. Breiman L, Friedman J, Stone C, Olshen R: Classification and regression trees. Boca Raton, Florida: CRC/Chapman & Hall; 1984.

14. 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(7494):765.

15. 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(12):1373-1379.

16. Vittinghoff E, McCulloch CE: Relaxing the rule of ten events per variable in logistic and Cox regression. *Am J Epidemiol* 2007, 165(6):710-718.

17. Dietterich T: Overfitting and undercomputing in machine learning. *ACM Comput Surv* 1995, 27(3):326-327.

18. Chen X, Ishwaran H: Pathway hunting by random survival forests. *Bioinformatics* 2013, 29(1):99-105.

19. 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(4):323-348.

20. van der Laan MJ, Polley EC, Hubbard AE: Super learner. *Statistical applications in genetics and molecular biology* 2007, 6:Article25.

21. Drummond C, Holte RC: Exploiting the Cost (In)sensitivity of Decision Tree Splitting Criteria. In: *Proceedings of the Seventeenth International Conference on Machine Learning.* Morgan Kaufmann Publishers Inc.; 2000: 239-246.

22. Song YY, Lu Y: Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry* 2015, 27(2):130-135.

**Table 4.1. Risk Score Sensitivity, Specificity, and Overall Classification Accuracy at Select Cut-points for Predicting Extended-Spectrum β-Lactamase (ESBL) Status in a Cohort of Adult Patients with *Escherichia coli* and *Klebsiella* Species Bacteremia***

| Risk Score Cut-point | Sensitivity | Specificity | Percent of 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% |

* Cut-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 cut-point that maximized overall classification accuracy (≥ 7.25 points).

**Table 4.2. 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**

| | 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 cut-point for dichotomization. Values reflected for the risk score are for the cut-point of ≥ 7.25 points, which optimized overall classification accuracy.

**Table 4.3. 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 over-sampling before model fitting and validation in order 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 clinical 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 & Usage** | | | |
| Intuitiveness | + | +++ | Decision tree branching logic is highly intuitive. |
| Ease-of-Use | + | +++ | Decision trees do not require calculations, making them very user-friendly for bedside application. |
| **Adaptability** | | | |
| End-User Adjustment of Sensitivity and Specificity | +++ | - | By changing the score cut-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 (e.g., by comparing Akaike’s Information Criterion (AIC) values of the original and expanded models)[1]. 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. |

**Figure 4.1. 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 cut-offs for ESBL-positive bacteremia, and associated sensitivities and specificities, are reflected in Table 4.1.

**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.*

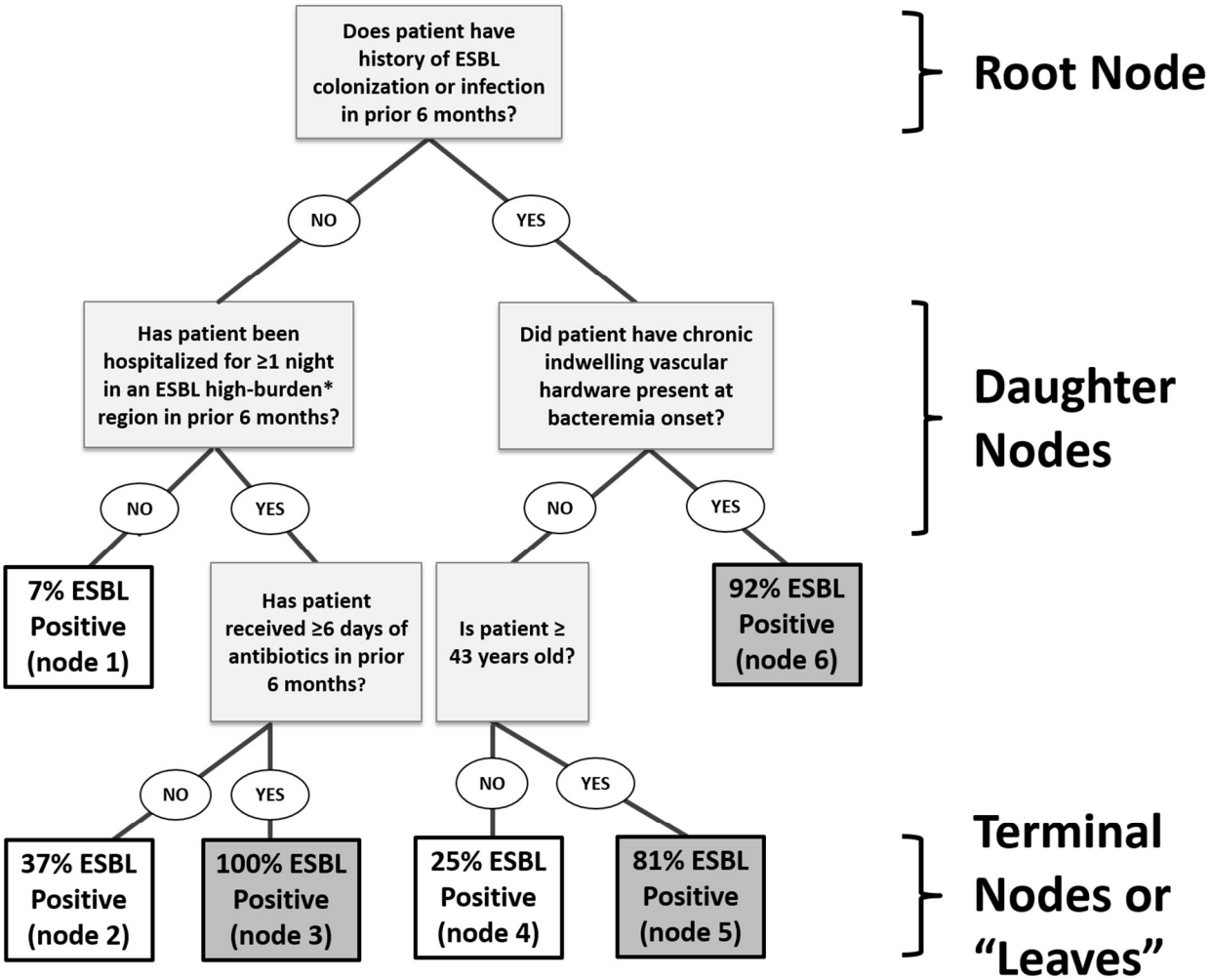
[a]Chronic obstructive pulmonary disease, emphysema, or tracheostomy-dependency.

[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 cut-point of 7.25 points and should be modified by the facility to account for local prevalence of ESBL bacteremia.

Abbreviations: MDRGN – Multidrug-resistant Gram-negative organism; CRE – Carbapenem-resistant *Enterobacteriaceae*; MDR – Multidrug-resistant. Drug-resistant organisms defined in accordance with The Centers for Disease Control and Prevention guidelines[2].

**Figure 4.2. 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[3].** 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.

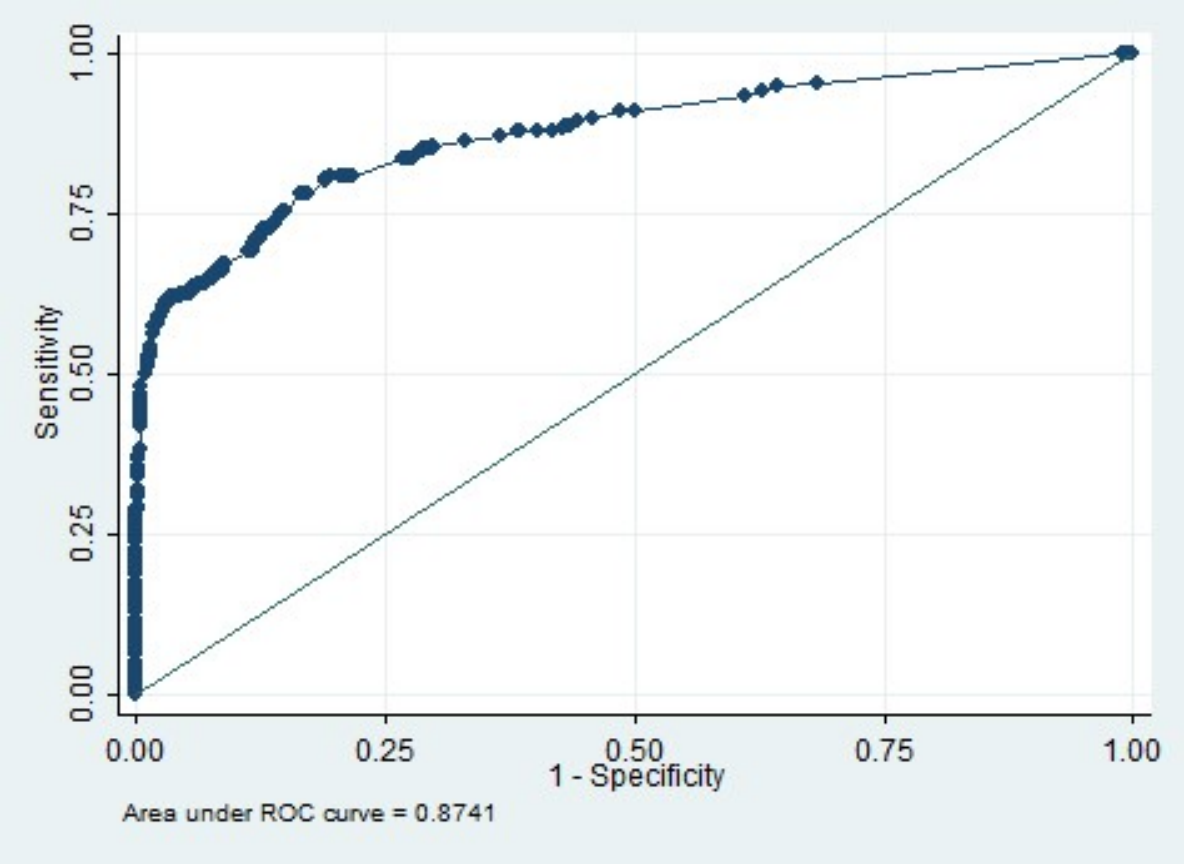



*Number of patients in each terminal node: (1) 1152; (2) 19; (3) 17; (4) 8; (5) 21; (6) 71.*

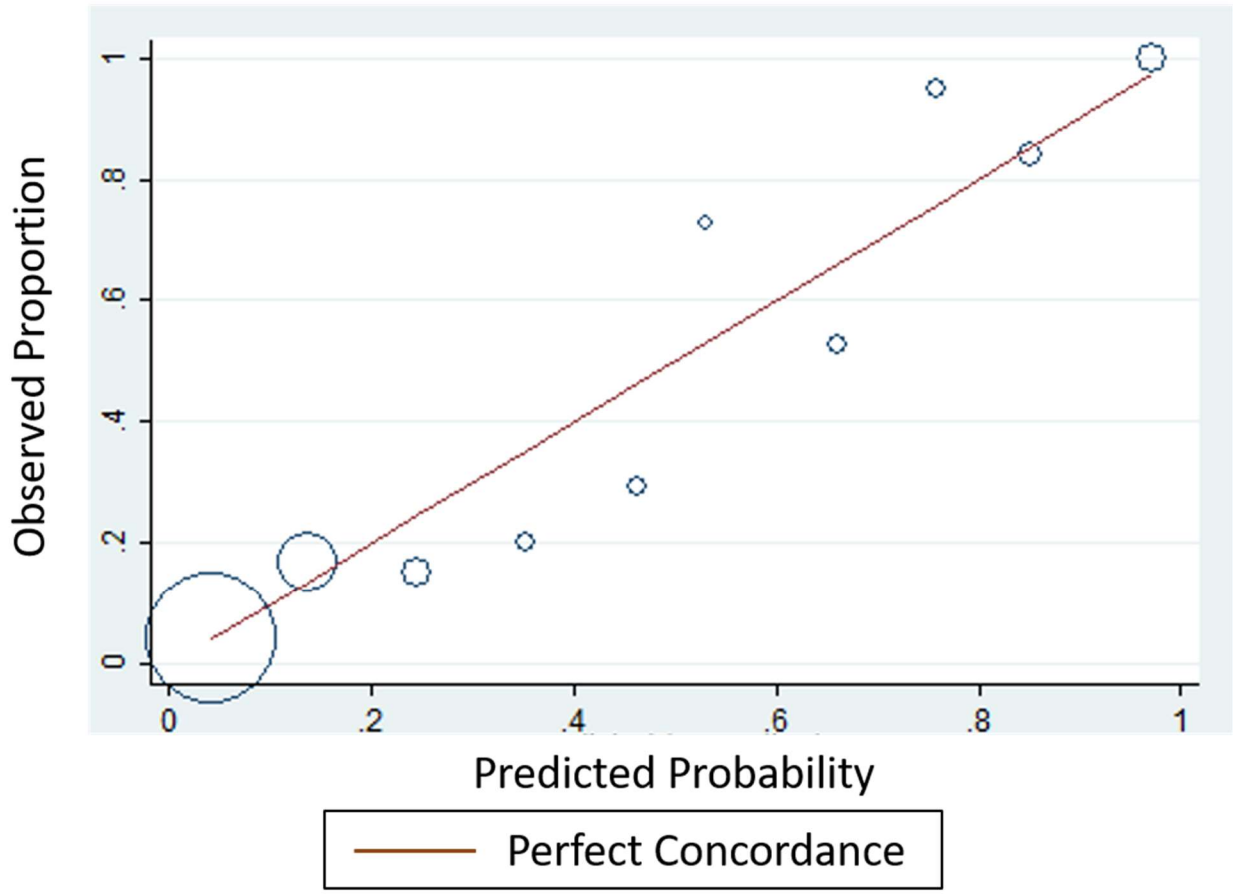
*Latin America (excluding the Caribbean), the Middle East (including Egypt), South Asia, China, and the Mediterranean.

**SUPPLEMENTAL MATERIAL, SECTION 4:**

**Supplemental Figure S.4.1. Receiver operating characteristic (ROC) curve for the logistic regression model, prior to risk score transformation.** Area under the curve (AUC), after rounding, was unchanged following conversion to a points-based risk score model.



**Supplemental Figure S.4.2. Calibration plot of observed proportion versus model-predicted ESBL probabilities by decile groups.**



## TABLE AND FIGURE REFERENCES


1. Akaike H: A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974, 19(6):716-723.
2. Centers for Disease Control and Prevention: Antimicrobial resistant phenotype definitions. In.; 2016.
3. Goodman KE, Lessler J, Cosgrove SE, Harris AD, Lautenbach E, Han JH, Milstone AM, Massey CJ, Tamma PD, Antibacterial Resistance Leadership G: 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(7):896-903.

# 5. CONCLUSION

## 5.1. Summary of Findings

Carbapenem-resistant Enterobacteriaceae (CRE) and other carbapenem-resistant organisms (CROs) pose important and increasingly urgent challenges to patient care. These bacteria are resistant to nearly all routinely used antibiotics and can impose mortality rates approaching 50 percent [1]. Of particular concern are the subset of CROs that are carbapenamase-producing (CPOs), for which carbapenem resistance is generally plasmid-mediated and can transfer between organisms. Carbapenemase genes have become established in some bacterial strains that are highly adept at clonal expansion, and CPOs have been implicated in numerous outbreaks in U.S. healthcare facilities [2].

Current prevention strategies for CROs and CPOs in healthcare facilities include efforts to reduce selective pressure from antibiotic exposure, e.g., antibiotic stewardship programs, and infection control policies to reduce patient-to-patient spread following resistance emergence or CRO/CPO introduction (e.g., active surveillance for CRO carriage at admission and placing colonized patients on contact precautions). These efforts are hampered by a lack of rapid and cost-effective diagnostics to identify CROs, CPOs, and other MDRGNs. We aimed to develop statistical and machine learning-based prediction models that could help to address existing diagnostic limitations.

In Aim 1, we screened nearly 3,000 patients admitted to the Johns Hopkins Hospital (JHH) medical intensive care unit (MICU) and comprehensive transplant unit (Transplant Unit) for perirectal colonization with CROs and CPOs. We found that 7.5% and 1.3% of patients were perirectally colonized with CROs and CPOs, respectively, and estimated a CRE admission prevalence of 4.2% (95% CI: 3.5 – 5.0%). We further tested all CPO isolates to reveal a diversity of carbapenemase gene-organism pairs. These data add important information to the epidemiological landscape of asymptomatic CRO and CPO carriage in U.S. inpatient populations. We collected comprehensive data (> 150 variables) on patient pre-admission exposures through automated extraction from electronic medical records (EMRs). These data were used to evaluate risk factors for colonization and to develop machine learning-derived predictive models to inform targeted screening. We identified a large number of variables significantly associated with colonization, many of which are consistent with published literature for CRO infection or for CRO colonization in other populations and settings. Despite this large number of significant variables, however, our models did not highly predict colonization, primarily due to low sensitivity. We were, however, able to identify specific high-risk sub-groups (e.g., patients with a recent CRO history and extended proton-pump inhibitor use). Our experience also yielded insights that may be informative to future research, including the role of heterogeneity in prediction models and differences between causal risk factors and predictors.

In Aim 2, we developed a clinical decision tree to predict infection with an extended-spectrum β-lactamase (ESBL)–producing organism, in order to reduce unnecessary carbapenem administration. We evaluated a total of 1,288 adult patients hospitalized at JHH between

October 2008 and March 2015 with bacteremia due to *Escherichia coli* or *Klebsiella* species. We collected pre-infection clinical and demographic information through manual chart review and used recursive partitioning to generate a decision tree to determine the likelihood that a bacteremic patient was infected with an ESBL-producer. The final classification tree for predicting ESBL-positive bacteremia included 5 predictors: history of ESBL colonization/infection, chronic indwelling vascular hardware, age ≥43 years, recent hospitalization in an ESBL high-burden region, and ≥6 days of antibiotic exposure in the prior 6 months. The decision tree's area under the curve (AUC) was 0.77 and unchanged in cross-validation, and its positive and negative predictive values were 90.8% and 91.9%, respectively. Our findings demonstrated that a user-friendly decision tree can reliably discriminate between ESBL-positive and –negative infections at the time of empirical treatment initiation in order to inform appropriate carbapenem usage.

In Aim 3, we developed a clinical risk score from the same ESBL dataset as Aim 2 in order to methodologically compare decision trees versus risk scores for predicting drug-resistant colonization or infection. The final clinical risk score for predicting ESBL-positive bloodstream infection at the time of organism and genus identification included 14 variables, broadly categorizable into six groups: Indwelling hardware on day of culture; presumptive source of bloodstream infection; patient characteristics; and, within the previous six months: healthcare exposure, MDRGN colonization or infection, or antibiotic usage. The risk score possessed higher discrimination than the decision tree (risk score C-statistic 0.87), but given that risk scores for binary predictions are dichotomized at a cut-point, in practice the risk score and the decision

tree performed similarly. However, with 14 variables, the risk score was less amenable to manual bedside use. In our discussion, we reviewed practical and methodological attributes of decision trees and risk scores and offered general guiding principles for when clinicians or hospital epidemiologists might consider each approach. We intended this examination to be highly practical and easily accessible, and believe that it addressed an unmet need in the clinical infectious disease and healthcare epidemiology literature.

### 5.2. Strengths and Limitations

Our studies had several strengths. The most important strength of Aims 1 and 2 (and Aim 3, by corollary) were our study populations and sample sizes. Aim 1's prospective cohort included thousands of patients, enabling robust colonization estimates despite outcome rarity, including stratified by bacterial class and resistance mechanism. The study also strongly benefitted from its ability to capitalize on an existing, longstanding VRE surveillance program. This infrastructure helped to maximize screening compliance and, in practice, made the study feasible. Without an existing program, individual informed consent would have been required, potentially biasing resulting estimates (e.g., sicker patients, such as those on ventilators, could be differentially excluded due to inability to consent). Similarly, to our knowledge Aim 2's retrospective cohort of 1,288 patients represents the largest existing cohort to examine risk factors and predictors for ESBL-positive bacteremia. Further, all aims benefitted from extensive clinical and demographic data collection, through a variety of both manual and automated methods, and built on the strengths of established clinical and molecular microbiology laboratories in order to characterize isolate resistance mechanisms. Finally, our aims utilized machine learning

methodology not routinely used in the infection control and clinical infectious disease literature. These approaches are methodologically well-suited to the high-dimensional data present in these aims, and decision trees in particular have practical advantages, including simplicity and intuitive interpretation.

Our studies also had several limitations. All aims involved single-center studies, and although we internally validated our models, our results should be validated in other cohorts. Our results may not be generalizable to other, in particular lower risk, patient populations. Aim 1 also had missing swabs for some unit admissions, although through the collective efforts of weekly real-time audits and floor re-education campaigns we were able to reduce missingness to less than 15 percent across the study period. Moreover, despite our best attempt to gather detailed previous clinical and demographic data on all patients, we were limited to Johns Hopkins Healthcare system (Aim 1) or Epic Care Everywhere (Aims 2 and 3) data. There was likely missing data on pre-admission or pre-infection exposures that could lead to exposure misclassification. Nevertheless, because of our prediction models' intended real-world use, their performance under the practical constraints of missing data was arguably relevant. As the use of electronic health records that interface across institutions becomes more widespread, these challenges will hopefully lessen. Finally, all of our tree-based predictive models demonstrated, to varying degrees, sub-optimal sensitivity. Because risk factor associations in our both our colonization and infection studies were numerous and strong, but sensitivity was very poor for specific sub-populations, we believe that these patient subsets may lack traditional healthcare-associated risk factors. Absent a prospective patient questionnaire, these

exposures would not be captured in medical records. Due to our study designs (de-linked patient identifiers in Aim 1 and a retrospective cohort in Aim 2), we were unable to explore these hypotheses further.

### 5.3. Public Health Implications and Recommendations for Future Research

In the U.S., CRO and CPO infections remain rare but devastating. Efforts to reduce carbapenem resistance in healthcare settings must maximize patient safety while simultaneously ensuring that policies are cost-effective and implementable. A lack of rapid, inexpensive microbiological diagnostics for many MDRGNs adds considerably to this challenge.

Our findings suggest that asymptomatic CRO and CPO carriage is infrequent among high-risk hospitalized patients, but nevertheless higher than the limited other published estimates available. With our data indicating that 1 of every 10 – 15 patients silently brings carbapenem-resistant bacteria into the ICU, the risk to other patients is potentially substantial. However, colonized patients' contribution to CRO spread once they enter hospital units remains poorly quantified. Despite circumstantial evidence linking CPOs to healthcare-associated outbreaks, it is also remains unknown whether propensity for intra-facility spread systematically differs between CROs and CPOS. These questions remain pressing as long as universal CRO screening remains impractical and accurate targeted screening algorithms are unavailable. Future research addressing these knowledge gaps is critical to ensure that policies target the highest-risk pathogens and those patients who pose the greatest risk of transmitting to other patients

and to the environment. Our Aim 1 study included weekly surveillance swabs, and we plan to investigate these questions in subsequent analyses.

We were unfortunately unable to develop highly predictive, targeted screening algorithms for CRO and CPO colonization despite extensive data collection on pre-admission risk factors. An important goal of artificial intelligence and other machine learning applications in healthcare is to capitalize on an abundance of 'Big Data,' despite its imperfections, to improve patient outcomes. This was a pragmatic study which demonstrated that the data currently available to us in the EMR, as extracted, did not meet these targets. Ongoing research and efforts to refine where, how, and which data we utilize will be important.

Notwithstanding these challenges, our results offer some actionable conclusions. A recent CRO- or CRE-positive culture was consistently the strongest predictor of colonization at admission, and many infection control programs already capture and flag these cultures. As such, existing policies — however imperfect — may be performing equivalently to a targeted screening program. Moreover, a sizable percentage of CRO- and CPO-colonized patients were co-colonized with VRE detected during routine admission screening. These patients would be placed on contact precautions even without dedicated CRO surveillance. These findings suggest that existing screening policies for other organisms may have unrecognized, off-target benefits, with important implications for current national conversations about the utility of continued

VRE rectal surveillance [3]. Cost-benefit analyses that account for these effects may be useful future investigations.

Finally and more generally, across our studies we encountered similar recurring challenges when developing prediction models for drug-resistant colonization or infection. Two of the most detrimental elements were outcome rarity, which compromised sensitivity in classification algorithms, and high heterogeneity or stochasticity underlying the biological process of acquisition or infection. As such, even strong risk factors accounted for only small absolute numbers of cases. Methodologically, we believe these findings would benefit from formal simulation studies in order to further dissect differences between risk factors and predictors and to better understand how to optimize EMR data retrieval when designing large, hospital-based cohort studies.

Practically, these findings also shed light on potential future model applications. In particular, although these limitations are likely to persist in U.S. patient populations, international, low-resource countries (e.g., India) generally have much higher CRO and CPO prevalence [4]. With a high proportion of community acquisition, they may also have more easily identifiable colonization or infection predictors (at least proportionally, although we note that EMR data is likely far more limited). Moreover, even as diagnostic improvements ease prediction challenges in the U.S., rapid molecular assays will remain cost-prohibitive for large scale screening in developed countries and even targeted screening in many developing regions. Taken together,

statistical models for predicting drug-resistant bacterial colonization or infection may achieve greater accuracy, and meet more pressing healthcare needs, in non-U.S. settings with higher prevalence. In both domestic and international settings, however, statistical models will continue to be an important tool in our armamentarium to tackle emerging antibiotic resistance.

**REFERENCES**


1. Falagas ME, Tansarli GS, Karageorgopoulos DE, Vardakas KZ: Deaths attributable to carbapenem-resistant Enterobacteriaceae infections. *Emerging infectious diseases* 2014, 20(7):1170-1175.
2. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Group NCSP, Henderson DK, Palmore TN, Segre JA: Tracking a hospital outbreak of carbapenem-resistant Klebsiella pneumoniae with whole-genome sequencing. *Sci Transl Med* 2012, 4(148):148ra116.
3. Martin EM, Bryant B, Grogan TR, Rubin ZA, Russell DL, Elashoff D, Uslan DZ: Noninfectious Hospital Adverse Events Decline After Elimination of Contact Precautions for MRSA and VRE. *Infect Control Hosp Epidemiol* 2018, 39(7):788-796.
4. Rousham EK, Unicomb L, Islam MA: Human, animal and environmental contributors to antibiotic resistance in low-resource settings: integrating behavioural, epidemiological and One Health approaches. *Proceedings Biological sciences* 2018, 285(1876).

# Katherine E. Goodman, JD

CURRICULUM VITAE

Johns Hopkins Bloomberg School of Public Health

## QUALIFICATIONS AND PERSONAL INFORMATION

### Personal Data

Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Street
Baltimore, MD 21205
**Tel:** (703) 217-3522
**E-mail:** kgoodma7@jhu.edu

### Education and Training

| | | |
|---|---|---|
| Expected Dec. 2018 | PhD | Johns Hopkins Bloomberg School of Public Health, Baltimore, MD<br>**Dissertation: Prediction Models for Carbapenem-Resistant *Enterobacteriaceae* and Other Multidrug-Resistant Gram-Negative Bacteria in Healthcare Settings** |
| 2013 | PGDiPH | University of Auckland, New Zealand |
| 2010 | JD | Columbia University School of Law, New York, NY |
| 2006 | BA | Dartmouth College (Moral Philosophy), Hanover, NH |

### Professional Experience

| | |
|---|---|
| 2014-Present | Research Assistant to Dr. Pranita Tamma, The Johns Hopkins Hospital, Baltimore, MD |
| 2015 | Lead Teaching Assistant, *Epidemiologic Methods II*, Johns Hopkins Bloomberg School of Public Health |
| 2015 | Health Resources and Services Administration Trainee Fellow, Office of Public Health Practice and Training, Johns Hopkins Bloomberg School of Public Health |

| | |
|---|---|
| 2014 | Teaching Assistant, *Epidemiologic Methods II*, Johns Hopkins Bloomberg School of Public Health |
| 2014 | Teaching Assistant, *Healthcare Epidemiology*, Johns Hopkins Bloomberg School of Public Health |
| 2010-2012 | Food & Drug Law Attorney, Covington & Burling LLP, Washington, DC |
| 2009 | Policy Intern, The American Association for the Advancement of Science, Washington, DC |
| 2008-2010 | Writer and Editor, The Myeloma Beacon™, New York, NY |
| 2008 | Legal Associate, Covington & Burling LLP, Washington, DC |
| 2007 | Legal Intern, U.S. Attorney's Office, Southern District of New York, New York, NY |
| 2005 | Summer Fellow, Bridgewater Associates, Westport, CT |
| 2004-2005 | Research Assistant, The Ethics Institute, Dartmouth College |
| 2003 | Research Intern, The President's Council on Bioethics, Washington, DC |

**Professional Training**

| | |
|---|---|
| 2015 | Certificate in Healthcare Epidemiology and Infection Control (HEIC), Johns Hopkins Bloomberg School of Public Health |

**Admissions to Legal Practice**

| | |
|---|---|
| 2011-Present | New York |
| 2011-Present | District of Columbia |

**RECOGNITION**

| | |
|---|---|
| 2018 | Student Travel Support Fund Award, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health |
| 2018 | "Top Poster Abstract Award," Society for Healthcare Epidemiology of America |
| 2016 | Ellen B. Gold Fund Award, Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health |
| 2013 | "Top Epidemiology Student" Award (highest score in an introductory epidemiology course), Australasian Epidemiological Association |

2013 "First-in-Course" Award, *Evidence for Best Practice*, University of Auckland

2012 "First-in-Course" Award, *Epidemiology*, University of Auckland

2012 "First-in-Course" Award, *Health and Society*, University of Auckland

2010 J.D. Awarded *cum Laude*, Columbia University School of Law

2008 Harlan Fiske Stone Scholar, Columbia University School of Law

2007 Harlan Fiske Stone Scholar, Columbia University School of Law

2007 "Best Orator" Award, Moot Court Debate, Columbia University School of Law

2006 Fred C. Scribner Jr. Fellowship Recipient, Dartmouth College

2006 B.A. Awarded *Summa cum Laude*, Dartmouth College

2006 Phi Beta Kappa Honor Society, Dartmouth College

**PUBLICATIONS**:

**Original Research:**

1. **Goodman KE**, Simner PJ, Klein EY, et al. How Frequently are Hospitalized Patients Colonized with Carbapenem-Resistant Enterobacteriaceae (CRE) Already on Contact Precautions for Other Indications? ICHE (In press).

2. Simner PJ, **Goodman KE**, Carroll KC, Harris AD, Han JH, & Tamma PD. Using Patient Risk Factors to Identify Whether Carbapenem-Resistant Enterobacteriaceae Infections Are Caused by Carbapenemase-Producing Organisms. OFID. 2018:5(5): Epub. ahead of print.

3. Tamma PD, **Goodman KE**, Harris AD, Tekle T, Robert A, Taiwo A, & Simner PJ. Comparing the Outcomes of Patients with Carbapenemase-Producing and Non-Carbapenemase-Producing Carbapenem-Resistant Enterobacteriaceae Bacteremia. Clin Infect Dis. 2017;64(3):257-264.

4. **Goodman KE**, Lessler J, Cosgrove SE, Harris AD, Lautenbach E, Han JH, Milstone AM, Masey C, Tamma PD; Antibacterial Resistance Leadership Group. A Clinical Decision Tree to Predict Whether a Bacteremic Patient is Infected With an ESBL-Producing Organism. Clin Infect Dis. 2016 Oct 1;63(7):896-903.

**Review Articles:**

1. **Goodman KE**, Simner PJ, Tamma PD, Milstone AM. Infection control implications of heterogeneous resistance mechanisms in carbapenem-resistant *Enterobacteriaceae* (CRE). Expert Rev Anti Infect Ther. 2016 Jan;14(1):95-108. Epub 2015 Nov 4.

**Other Publications**

**Original Legal Research:**

1. **Goodman KE**. Prosecution of Physicians as Drug Traffickers: The United States' Failed Protection of Legitimate Opioid Prescription Under the Controlled Substances Act and South Australia's Alternative Regulatory Approach, 47 Columbia Journal of Transnational Law. 210 (2008).
2. **Goodman KE.** Violent Pornography in the United States: An Argument for Content-Neutral Regulation Under the Jurisdiction of Workplace Safety Laws, 4(1) Dartmouth Law Journal. 22 (2006).

**Reports:**

1. The President's Council on Bioethics. Beyond Therapy: Biotechnology and the Pursuit of Happiness (Oct. 2003): Washington, DC (staff contributor).

**Journalism:**

1. **Goodman KE.** Stem Cell Research: Becoming Less Restrictive, 6(4) Scitech Lawyer 7 (2010).
2. **Goodman KE.** Supreme Court Rules Convicts Have No Constitutional Right to Test DNA Evidence, XXII(2) Professional Ethics Report. 5 (2009).
3. **Goodman KE.** User-Friendly Guide for Responding to Research Wrongdoing, XXII(2) Professional Ethics Report. 7 (2009).
4. **Goodman KE.** 2008 – 2010, The Myeloma Beacon™: Original Content and Medical Article Summaries available at: http://www.myelomabeacon.com/search/katherine+goodman

**PRESENTATIONS**

**Oral Presentations:**

1. **Goodman KE**, Opene BN, Voskertchian A, Simner PJ, Milstone AM, for the CDC Prevention Epicenters Program. Prevalence of Rectal Colonization with Carbapenem-

Resistant Organisms (CROs) at Hospital Unit Admission. Presented at the Society for Healthcare Epidemiology of America Spring 2017 Conference, St. Louis, MO.

**Poster Presentations:**

1. **Goodman KE**, Simner PJ, Kazmi A, Gadala A, Rock C, Maragakis L, Cosgrove SE, Milstone AM, for the CDC Prevention Epicenters Program. Prevalence and Contact Precautions Status of CRE Colonized Patients at Hospital Unit Admission. Society for Healthcare Epidemiology of America Spring 2018 Conference, Portland, OR.

**RESEARCH GRANT PARTICIPATION**

**Extramural – Completed:**

2/1/17- 6/31/18 **Infection Control Implications of Heterogeneous Resistance Mechanisms in Carbapenem-Resistant Enterobacteriaceae (CRE)**
R36HS025089
AHRQ
**Role**: Principal Investigator

**Intramural – Current:**

6/1/17 - 5/31/19 **Impact of Heterogeneous Resistance Mechanisms on Hospital Transmission of Carbapenem-Resistant Enterobacteriaceae (CRE)**
No assigned number
Sherrilyn and Ken Fisher Center for Environmental Infectious Diseases
**Role**: Co-Principal Investigator

**EDUCATIONAL ACTIVITIES**

2015 Lead Teaching Assistant, *Epidemiologic Methods II*, Johns Hopkins Bloomberg School of Public Health

2014 Teaching Assistant, *Epidemiologic Methods II*, Johns Hopkins Bloomberg School of Public Health

2014 Teaching Assistant, *Healthcare Epidemiology*, Johns Hopkins Bloomberg School of Public Health

2003 Teaching Assistant, *Human Biology*, Dartmouth College

## ORGANIZATIONAL ACTIVITIES

### Professional and Institutional

2016-2017 Research-in-Progress Coordinator, Department of Epidemiology (Infectious Disease Concentration), Johns Hopkins Bloomberg School of Public Health

2015-Present Student Outbreak Response Team, Johns Hopkins Bloomberg School of Public Health

2015-Present Infectious Disease Dynamics Group, Johns Hopkins Bloomberg School of Public Health

### Editorial

2007-2009 Editor, The Columbia Journal of Transnational Law

2004-2006 Executive Editor, The Dartmouth Law Journal

## RESEARCH ACTIVITIES

### Research Focus

As a PhD candidate in infectious disease epidemiology at the Johns Hopkins School of Public Health, my research interests include strategies for reducing healthcare-acquired infections, Gram-negative antimicrobial resistance in inpatient settings, and the development of evidence-based healthcare epidemiology and antimicrobial stewardship policies. Prior to starting my doctoral studies, I practiced FDA law in Washington, D.C.

### Research Contributions

I. Implications of Heterogeneous Resistance Mechanisms in CRE

My primary area of interest is multi-drug-resistant Gram-negative bacteria in hospitalized settings. In particular, I have focused on strategies for reducing carbapenem-resistant Enterobacteriaceae (CRE) emergence in domestic healthcare environments. There are considerable knowledge gaps in our understanding of the behavior of CRE in the healthcare environment, in particular the implications of different carbapenem resistance mechanisms to CRE control and treatment. To date, both policy approaches (e.g., how 'CRE' is defined) and the scientific literature have generally treated 'CRE' as a single composite category or as

synonymous with carbapenemase production, despite the fact that an unknown proportion — and possible majority — of CRE do not produce carbapenemases. My research aims to explore implications of this heterogeneity, which have become clinically and epidemiologically relevant as resistance diagnostics improve, as CRE prevalence increases, and with the development of new therapeutic agents that are designed to target specific resistance mechanisms (e.g., avibactam/AVYCAZ®). I led a review examining the available evidence and discussing many of the above knowledge gaps regarding potential differences in transmissibility and nosocomial acquisition pathways between CRE resistance types. I also closely collaborated with Dr. Tamma on a cohort of CRE bacteremia cases at JHH in order to investigate whether clinical outcomes differ by CRE resistance mechanism.

1. **Goodman KE**, Simner PJ, Tamma PD, Milstone AM (2016). Infection control implications of heterogeneous resistance mechanisms in carbapenem-resistant Enterobacteriaceae (CRE). *Expert Rev Anti Infect Ther.* Jan;14(1):95-108. PubMed PMID: 26535959.
2. Tamma PD, **Goodman KE**, Harris AD, Tekle T, Robert A, Taiwo A, & Simner PJ. (2017) Comparing the Outcomes of Patients with Carbapenemase-Producing and Non-Carbapenemase-Producing Carbapenem-Resistant *Enterobacteriaceae* Bacteremia. *Clin Infect Dis*. 64(3): 257-264. PubMED PMID: 28013264.
3. Simner PJ, **Goodman KE**, Carroll KC, Harris AD, Han JH, & Tamma PD. Using Patient Risk Factors to Identify Whether Carbapenem-Resistant Enterobacteriaceae Infections Are Caused by Carbapenemase-Producing Organisms. OFID. 2018:5(5): Epub. ahead of print.

II. Statistical and Machine Learning Approaches for Circumventing Diagnostic and Resource Limitations

As emerging pathogens, multi-drug-resistant gram-negative (MDRGN) bacteria frequently lack well-validated and readily-implementable rapid diagnostics, threatening their incorporation into routine laboratory workflows. These resource and performance limitations pose important hospital epidemiology and clinical challenges. My research has focused on deploying statistical and machine learning methodologies to circumvent these limitations. For example, timely identification of ESBL bacteremia can improve clinical outcomes while minimizing the unnecessary use of broad-spectrum antibiotics. Yet 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 leaves clinicians to balance the risks of ineffective agents against unnecessarily broad empiric antibiotic therapy on an ad-hoc basis. I led the development of a clinical algorithm to predict a bacteremic patient's likelihood of infection with an ESBL-producing bacteria in order to assist selection of empiric treatment. Dr. Justin Lessler, myself, and additional authors are also currently preparing a review for the *American Journal of Epidemiology* examining application of machine learning methodologies to public health. I have also submitted a manuscript that

methodological compares risk scores and decision trees, which is currently under review at *Clinical Infectious Diseases*.

1. **Goodman, K.**, Lessler, J., Cosgrove, S., Harris, A., Lautenbach, E., Han, J., Milstone, A., Massey, C., Tamma, P., for the Antibacterial Resistance Leadership Group. (2016) A Clinical Decision Tree to Predict Whether a Bacteremic Patient Is Infected With an Extended-Spectrum β-Lactamase–Producing Organism. *Clin Infect Dis*. 63 (7): 896-903. PubMed PMID: 27358356.
2. **Bi, Q.*, Goodman, K.***, Kaminsky, J., Lessler, J. "What is Machine Learning (and why Should Epidemiologists Care)?" (manuscript in preparation). ***Co-first authorship**
3. **Goodman, K.**, Lessler, J., Harris, A., Milstone, A., Tamma, P. (2018) A Methodological Comparison of Risk Scores Versus Decision Trees for Predicting Drug-Resistant Infections: A Case Study using ESBL Bacteremia (under review at *Clin Infect Dis*).

III. Legal Tools for Addressing Public Health Problems

Given my longstanding interest in science and bioethics, during law school I began to appreciate legal theory as an alternative methodology for conceptualizing and tackling pressing public health problems. In 2006 I published a peer-reviewed article advocating regulation of pornography under the blood-borne pathogens clause of workplace safety laws rather than under the traditional constitutional doctrine of freedom of speech. Six years later, spurred by an industry HIV outbreak the California Occupational Safety and Health Administration (OSHA) began requiring condom use on pornography sets under this same clause. Similarly, a 2008 article (peer-reviewed) examines standards for criminal conviction of opioid-prescribing physicians under the Controlled Substances Act (CSA), ultimately asserting that the CSA creates a unique regulatory vacuum in which federal agents and courts — entities arguably the least prepared to render such decisions — assume the predominant role of distinguishing legitimate treatment from criminal narcotic trafficking.

1. **Goodman, KE.** Prosecution of Physicians as Drug Traffickers: The United States' Failed Protection of Legitimate Opioid Prescription Under the Controlled Substances Act and South Australia's Alternative Regulatory Approach (2008), Columbia Journal of Transnational Law 47: 210–244.
2. **Goodman, KE**. Violent Pornography in the United States: An Argument for Content-Neutral Regulation Under the Jurisdiction of Workplace Safety Laws (2006), Dartmouth Law Journal. 4(1): 22–29.