ORIGINAL RESEARCH

Development of a Prediction Model for Antibiotic-Resistant Urinary Tract Infections Using Integrated Electronic Health Records from Multiple Clinics in North-Central Florida

Shannan N. Rich ⬤ · Inyoung Jun · Jiang Bian · Christina Boucher ·
Kartik Cherabuddi · J. Glenn Morris Jr. · Mattia Prosperi

ABSTRACT

Introduction: Urinary tract infections (UTIs) are common infections for which initial antibiotic treatment decisions are empirically based, often without antibiotic susceptibility testing to evaluate resistance, increasing the risk of inappropriate therapy. We hypothesized that models based on electronic health records (EHR) could assist in the identification of patients at higher risk for antibiotic-resistant UTIs and help guide the selection of antimicrobials in hospital and clinic settings.

Methods: EHR from multiple centers in North-Central Florida, including patient demographics, previous diagnoses, prescriptions, and antibiotic susceptibility tests, were obtained for 9990 patients diagnosed with a UTI during 2011–2019. Decision trees, boosted logistic regression (BLR), and random forest models were developed to predict resistance to common antibiotics used for UTI management [sulfamethoxazole-trimethoprim (SXT), nitrofurantoin (NIT), ciprofloxacin (CIP)] and multidrug resistance (MDR).

Results: There were 6307 (63.1%) individuals with a UTI caused by a resistant microorganism. Overall, the population was majority female, white, non-Hispanic, and older aged (mean = 60.7 years). The BLR models yielded the highest discriminative ability, as measured by the out-of-bag area under the receiver-operating curve (AUROC), for the resistance outcomes [AUROC = 0.58 (SXT), 0.62 (NIT), 0.64

S. N. Rich (✉) · I. Jun · M. Prosperi
Department of Epidemiology, College of Public Health and Health Professions and Medicine, University of Florida, PO Box 100009, Gainesville, FL 32610, USA
e-mail: shannanrich@ufl.edu

S. N. Rich · J. G. Morris Jr.
Emerging Pathogens Institute, University of Florida, Gainesville, FL, USA

J. Bian
Department of Health Outcomes and Biomedical Informatics, College of Medicine, University of Florida, Gainesville, FL, USA

C. Boucher
Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL, USA

K. Cherabuddi
Division of Infectious Diseases, Department of Medicine, University of Florida, Gainesville, FL, USA

J. G. Morris Jr.
Department of Medicine, College of Medicine, University of Florida, Gainesville, FL, USA

(CIP), and 0.66 (MDR)]. Variables in the best performing model were sex, history of UTIs, catheterization, renal disease, dementia, hemiplegia/paraplegia, and hypertension.

Conclusions: The discriminative ability of the prediction models was moderate. Nonetheless, these models based solely on EHR demonstrate utility for the identification of patients at higher risk for resistant infections. These models, in turn, may help guide clinical decision-making on the ordering of urine cultures and decisions regarding empiric therapy for these patients.

Keywords: Antimicrobial resistance; Clinical decision support; Electronic health records; Machine learning; Urinary tract infections

Key Summary Points
Why carry out this study?
UTIs are common bacterial infections that are often treated inappropriately, consequently resulting in treatment failures and the development of antimicrobial resistance.
Machine learning algorithms applied to large, curated electronic health records (EHR) data can permit the development of highly predictive models to aid diagnosis of resistant infections to guide treatment decisions more rapidly.
What was learned from the study?
Machine-learning algorithms predicted resistance to the antibiotics (SXT, NIT, and CIP) most prescribed for their treatment in this study population using variables easily accessible in patients' EHR.
Factors such as prior antibiotic use, renal disease, and diabetes were highly predictive of resistance phenotypes.
These models can inform decision-making in settings where resistance testing is not possible or rapid enough for patients with declining health status.

INTRODUCTION

Urinary tract infections (UTIs) are among the most frequent bacterial infections occurring in the USA [1]. UTI is an umbrella term for multiple syndromes; however, antibiotic therapy is typically only indicated for two syndromes, cystitis (UTIs confined to the bladder) and pyelonephritis, referring to UTIs that progress to involve the kidneys [2]. Among adult patients suffering acute pyelonephritis, hospitalization occurs in 10–30% [2]. As such, UTIs account for a significant economic burden, with direct and indirect costs projected at $2.9 billion annually [2, 3]. These infections are often treated inappropriately, consequently resulting in treatment failures and the development of antimicrobial resistance [4]. Levels of UTIs resistant to some of the most common antibiotics used for their treatment, including sulfatrimethoprim (SXT) and the fluoroquinolone ciprofloxacin (CIP), have risen substantially in the US [5]. Between 2000 and 2010, resistance levels rose from 17.9 to 24.2% for SXT and from 3.0 to 17.1% for CIP among outpatients suffering UTIs caused by *Escherichia coli*, the most prevalent UTI pathogen [5]. Another study reported similar levels of SXT resistance (25.2%) and even higher levels of resistance to fluoroquinolones (29.5%) in *E. coli* UTI isolates as of 2013 [6]. Resistance to nitrofurantoin (NIT) and fosfomycin (FOF), frontline antibiotics for treatment of uncomplicated UTI, has not noticeably increased since they were introduced [7]. The gold standard for testing for a resistant UTI is a urine culture and antibiotic susceptibility test. However, this test takes, on average, 24–48 h before results are available. Thus, decisions regarding diagnosis and treatment are typically empirically based on symptoms and results of simple assays such as a dipstick or urinalysis with microscopy [8, 9].

Under these circumstances, there is potential utility in the development of clinical prediction and decision support systems that can improve the culture-independent diagnosis and management of antimicrobial-resistant UTIs. Prior studies have identified several predictors of resistant UTI that are routinely collected and

captured in patients' electronic health records (EHR) including demographic factors, comorbidities (e.g., diabetes, immune deficiency), clinical history of UTI, and past antibiotic prescriptions [10–13]. However, many of these studies were conducted on relatively small datasets, in non-US populations, or using data not typically available at the point of care (e.g., infecting organism). The availability of large, structured, and curated datasets— in terms ofboth sample size and measurement space— from EHR, together with the ability of machine learning to fit models on big data and approximate complex outcome surfaces, permits the development of diagnostic models for a variety of diseases. The objective of this study was to develop a culture-independent diagnostic model of resistance to antimicrobial agents commonly used in the treatment of UTIs and multidrug-resistant (MDR) UTIs.

METHODS

As part of this study, we compared the performance of linear and nonlinear methods, using EHR data collected over a 2-decade-long period at a large, multi-center academic health system in the southeastern US. The models were fit to accommodate realistic data availability and aid clinicians in point-of-care decision-making in ambulatory and hospital settings. This article adheres to the Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD) statement (Supp. Fig. 1).

Ethics Statement

The study protocol for secondary data analysis and a waiver of informed consent were approved by the University of Florida (UF)'s Institutional Review Board (#IRB201900652).

Data Source

Deidentified EHR data were obtained from the UF Health and Shands Hospital (UF Health). The UF Health network includes two main

hospital systems and approximately 45 affiliated outpatient practices. The clinics are located primarily in Gainesville and Alachua County, Florida, with additional health centers in Jacksonville, Florida. UF Health EHRs have been collectively managed as an integrated data repository (IDR) since 2011. Clinical diagnoses and procedures are encoded using the International Classification of Diseases (ICD) codes, versions 9–11 depending on the year. Medications are encoded using RxNorm and laboratory tests via LOINC. In this work, diagnostic codes across all years were harmonized by converting to the ICD version nine (ICD-9). To be included in the study population patients had to be aged 18 years or older, diagnosed with a UTI (ICD-9: 595.*-'Cystitis' or 599.0-'Urinary tract infection, site not specified') at a UF health inpatient or outpatient center between January 1, 2011, and July 1, 2019, resided in a zip code sharing the same first three digits as the health network and had an antibiotic susceptibility test for their infection. Only the most recent UTI observation was considered for patients with multiple UTI diagnoses with an antibiotic susceptibility test present. The bacterial identification and antibiotic susceptibility test results were pulled from PDF reports through in-house scripts developed and maintained by IDR staff.

Antibiotic-resistant UTI

Infections were determined to be antibiotic-resistant based on the susceptibility test results. Antibiotic resistance was defined cumulatively as any infection with ≥ 1 "resistant" susceptibility findings for all antibiotics tested and by major drug (SXT, CIP, or NIT). Infections with resistance to all three major drug types were categorized as MDR. Antibacterial agents were abbreviated using standard American Society for Microbiology conventions (https://aac.asm.org/content/abbreviations-and-conventions).

Covariates

Data were collected on patient demographics, including age, sex at birth, race (categorized as Black, White, or other/unknown), ethnicity

(Hispanic or non-Hispanic), and three-digit zip code of residence. Prior diagnoses obtained from the EHR included components of the Charlson's comorbidity index, in addition to other factors that were previously identified as predictors of antimicrobial-resistant UTIs [11, 12, 14, 15] including diabetes, congestive heart failure, renal, peripheral vascular, cerebrovascular, chronic pulmonary, rheumatic, peptic ulcer, and liver diseases, hemiplegia or paraplegia, any malignancy except malignant neoplasm of skin, metastatic solid tumor, HIV/AIDS, and dementia. Additional comorbidities included hypertension, pregnancy, immune deficiency (non-HIV infection-related), nicotine dependence, birth control usage, history of UTI, vaginal infection, and urinary tract abnormalities. Symptoms/clinical features of the UTI that were collected included dysuria, frequency, urgency, hematuria, and pyelonephritis. ICD-9 code categorization schemes for all comorbidities were based on works by Glasheen (2019) [16] and Menendez (2014) [17] (Supplementary Material Table 1).

Dates of admission and discharge were obtained to determine hospitalization and intensive care unit (ICU) status at the time of the infection and urinary catheterization in the past year. Patients with at least two consecutive dates under hospital observation were considered hospitalized. Past antibiotic prescriptions for previous encounters were categorized by drug type, either as "cognate," defined as the same drug type to which the infection is resistant, or as "non-cognate," defined as a different drug type from the resistant infection as done in a previous study [11]. Organism identification from the susceptibility test was also extracted and categorized by the most common isolates (*Citrobacter* species, *Enterobacter* species, *Enterococcus faecalis*, *E. coli*, *Klebsiella pneumoniae*, *Proteus mirabilis*, *Pseudomonas aeruginosa*, *Staphylococcus aureus*) or other.

Statistical Analysis

All variables were considered for model inclusion. We did not attempt to distinguish between variables that were confounders (versus colliders) as the goal of this analysis was the prediction of resistance rather than causal inference [18]. We fit the following linear and nonlinear multivariable models: (1) main effects boosted logistic regression (BLR), (2) random forest (RF), and (3) decision tree (DT). The BLR models used 100 iterations; the RF models were fit with 500 trees, while different pruning strategies were used for the DTs. We assessed models' performance through bootstrap validation ($n = 25$) and assessed the out-of-bag average sensitivity, specificity, and area under the receiver-operating characteristic (AUROC) curve with corresponding standard deviations. To externally validate the models, we trained them using data from patients living in the predominant three-digit zip code ('326', accounting for 78.9% of the population) region where the main UF Health hospital is located and tested them on data from patients living in the greater-Jacksonville, Florida, region (three-digit zip code '322', accounting for 21.1% of the population) where a secondary UF Health hospital is located. Predictors' importance was assessed using odds ratios for the logistic model, with *p*-values, and of split-rank with a stratified *p*-value of node purity for the DTs. All analyses were performed in R [19], version 4.0.2, using the following packages: mboost, randomForest, rpart, party, caret, and ROCR [20–25].

RESULTS

Study Population

There were 9990 patients who met the study criteria. The overall study population was majority female (76.4%), white (67.7%), and non-Hispanic (96.1%) with a mean age of 60.7 (SD = 19.8) years (Table 1). The proportion of infections resistant to at least one antibiotic was 63.0%, ranging between 58.8 and 67.1% throughout the study period (2011–2019). The prevalence of all-cause resistant UTIs remained stable throughout the study period, with no clear upward/downward trend (Fig. 2). The most common uropathogens were *E. coli* (59.1%), *K. pneumoniae* (14.6%), and *E. faecalis* (5.5%), which was also consistent throughout

Table 1 Characteristics of the study population diagnosed with a urinary tract infection (UTI) stratified by antibiotic resistance status, 2011–2019

	Non-resistant UTI	Resistant UTI	Unadjusted OR
Age (years); mean (SD)	59.2 (20.3)	61.5 (19.5)	1.01 (1.00–1.01)
Age category (years)			
18–29	400 (10.9)	484 (7.7)	*Referent*
30–39	411 (11.2)	652 (10.3)	1.31 (1.09–1.57)
40–49	392 (10.6)	574 (9.1)	1.21 (1.01–1.46)
50–59	485 (13.2)	887 (14.1)	1.51 (1.27–1.80)
60–69	672 (18.2)	1272 (20.2)	1.56 (1.33–1.84)
70–79	678 (18.4)	1240 (19.7)	1.51 (1.29–1.78)
80+	645 (17.5)	1198 (19.0)	1.54 (1.30–1.81)
Sex			
Male	786 (21.3)	1568 (24.9)	*Referent*
Female	2897 (78.7)	4739 (75.1)	0.82 (0.74–0.90)
Race			
White	2508 (68.1)	4165 (66.0)	*Referent*
Black	953 (25.9)	1752 (27.8)	1.11 (1.01–1.22)
Other/unknown	222 (6.0)	390 (6.2)	1.06 (0.89–1.26)
Ethnicity			
Not Hispanic	3514 (95.4)	6024 (95.5)	*Referent*
Hispanic	148 (4.0)	243 (3.9)	0.96 (0.78–1.18)
Other/unknown	21 (0.6)	40 (0.6)	1.11 (0.66–1.92)
Diabetes	1255 (34.1)	2634 (41.8)	1.39 (1.28–1.51)
Hypertension	2108 (57.2)	4220 (66.9)	1.51 (1.39–1.64)
Renal disease	789 (21.4)	1762 (27.9)	1.42 (1.29–1.57)
Myocardial infarction	403 (10.9)	862 (13.7)	1.29 (1.14–1.46)
Congestive heart failure	221 (6.0)	464 (7.4)	1.24 (1.06–1.47)
Peripheral vascular disease	672 (18.2)	1302 (20.6)	1.17 (1.05–1.29)
Cerebrovascular disease	362 (9.8)	655 (10.4)	1.06 (0.93–1.22)
Chronic pulmonary disease	115 (3.1)	233 (3.7)	1.19 (0.95–1.50)
Rheumatic disease	130 (3.5)	205 (3.3)	0.92 (0.74–1.15)
Peptic ulcer disease	81 (2.2)	170 (2.7)	1.23 (0.95–1.62)
Liver disease	587 (15.9)	1115 (17.7)	1.13 (1.02–1.26)
Hemiplegia or paraplegia	196 (5.3)	497 (7.9)	1.52 (1.29–1.81)

Table 1 continued

	Non-resistant UTI	Resistant UTI	Unadjusted OR
Any malignancy (*except skin*)	371 (10.1)	649 (10.3)	1.02 (0.90–1.17)
Metastatic solid tumor	127 (3.4)	203 (3.2)	0.93 (0.74–1.17)
HIV/AIDS	53 (1.4)	125 (2.0)	1.38 (1.01–1.93)
Dementia	213 (5.8)	387 (6.1)	1.06 (0.90–1.27)
Pregnancy	60 (1.6)	80 (1.3)	0.78 (0.55–1.09)
History of UTI	679 (18.4)	1423 (22.6)	1.29 (1.16–1.43)
Immunodeficiency (*non-HIV*)	99 (2.7)	260 (4.1)	1.56 (1.23–1.98)
Vaginal infection	128 (3.5)	226 (3.6)	1.03 (0.83–1.29)
Catheterization	140 (3.8)	308 (4.9)	1.04 (0.81–1.34)
Hospitalization	2544 (69.1)	4513 (71.6)	1.13 (1.03–1.23)
ICU	243 (7.3)	631 (10.8)	1.57 (1.35–1.84)
Nicotine dependence	1197 (32.5)	2098 (33.3)	1.04 (0.95–1.13)
Birth control usage	496 (13.5)	680 (10.8)	0.78 (0.69–0.88)
Urinary tract abnormality	14 (0.4)	35 (0.6)	1.46 (0.80–2.81)
Dysuria	610 (16.6)	857 (13.6)	0.79 (0.71–0.89)
Frequency	202 (5.5)	321 (5.1)	0.92 (0.77–1.11)
Urgency	56 (1.5)	101 (1.6)	1.05 (0.76–1.47)
Hematuria	419 (11.4)	693 (11.0)	0.96 (0.85–1.09)
Pyelonephritis	412 (11.2)	742 (11.8)	1.06 (0.93–1.20)
Any antibiotic use	1398 (38.0)	2944 (46.7)	1.43 (1.32–1.55)
SXT use	138 (3.7)	460 (7.3)	2.02 (1.67–2.46)
NIT use	66 (1.8)	177 (2.8)	1.58 (1.20–2.12)
CIP use	271 (7.4)	733 (11.6)	1.66 (1.43–1.92)
Organism			
E. coli	2564 (69.6)	3340 (53.0)	*Referent*
Citrobacter spp.	111 (3.0)	93 (1.5)	0.64 (0.49–0.85)
Enterobacter spp.	11 (0.3)	319 (5.1)	22.26 (12.81–43.31)
E. faecalis	310 (8.4)	226 (3.6)	0.56 (0.47–0.67)
K. pneumoniae	121 (3.3)	1340 (21.2)	8.50 (7.04–10.36)
P. mirabilis	26 (0.7)	348 (5.5)	10.27 (7.02–15.73)
P. aeruginosa	251 (6.8)	99 (1.6)	0.30 (0.24–0.38)
S. aureus	74 (2.0)	120 (1.9)	1.24 (0.93–1.68)

Table 1 continued

	Non-resistant UTI	Resistant UTI	Unadjusted OR
Other	215 (5.8)	422 (6.7)	1.51 (1.27–1.79)

Values are presented as frequency (percentage) unless otherwise stated

UTI urinary tract infection, *OR* odds ratio, *SD* standard deviation, *HIV/AIDS* human immunodeficiency virus/acquired immune deficiency syndrome, *ICU* intensive care unit, *SXT* sulfa-trimethoprim, *NIT* nitrofurantoin, *CIP* ciprofloxacin, *Dx* diagnosis

the study period (Fig. 1). Resistance to CIP and SXT was highest for all-cause UTIs and UTIs caused by *E. coli* throughout the full study period, whereas resistance to NIT was highest for UTIs caused by *K. pneumoniae* in 2011–2015 only (Fig. 2). MDR UTIs were observed in 159 patients. Missing data were problematic for three of the outcome variables: status of SXT resistance (12.3% missing), NIT resistance (5.9% missing), and CIP resistance (6.3% missing) due to omission from the antibiotic susceptibility screening—likely deemed unnecessary upon organism identification. This was remedied by subsetting three new datasets for which the outcome status was non-missing, resulting in a final population count of 9072 for SXT, 9726 for NIT, 9688 for CIP, and 10,340 for MDR and 'any' resistance.

Strongest Predictors of Antibiotic-Resistant UTI

Increased age, presence of diabetes, hypertension, renal disease, myocardial infarction, congestive heart failure, peripheral vascular disease, liver disease, hemiplegia or paraplegia, HIV/AIDS, history of UTI, immunodeficiency (non-HIV), hospitalization, ICU status, and antibiotic use were all positively associated with having an all-cause resistant UTI (Table 1). Compared to infection with *E. coli*, infection with *Enterobacter* spp., *K. pneumoniae*, or *P. mirabilis* was associated with significantly greater odds of all-cause resistant UTI. Conversely, infection with *Citrobacter* spp., *E. faecalis*, or *P. aeruginosa* was associated with reduced odds of all-cause resistant UTI compared to infection with *E. coli*. Additionally, being female, using birth control, and having the symptom of dysuria at the time

of the encounter were also associated with reduced odds of all-cause-resistant UTI.

Model Performance

The BLR models yielded the highest discriminative performance as compared to the DT and RF models for all five outcomes: AUROC = 0.57 (SD = 0.01) for AMR-UTI, AUROC = 0.58 (SD = 0.01) for SXT-resistant UTI, AUROC = 0.62 (SD = 0.01) for NIT-resistant UTI, AUROC = 0.64 (SD = 0.01) for CIP-resistant UTI, and AUROC = 0.66 (SD = 0.02) for MDR UTI (Table 2, Fig. 3). The BLR model performed similarly on the external validation population (Supplementary Table 2). The best fit clinical decision support system was for MDR UTI and included the variables sex, history of UTI, history of catheterization, renal disease, dementia, hemiplegia or paraplegia, and hypertension (Supplementary Fig. 2).

Effect estimates for the remaining outcomes from the BLR models are in Supplementary Table 3. Given that the model performances were similar for each method, multiple decision paths may be considered. Based on the best-fit, pruned DTs, the most important feature for prediction of all-cause resistant (AMR)-UTI was past antibiotic use, followed by hypertension, sex, and birth control in the low antibiotic use category, and age and nicotine use in the high antibiotic use category (Supplementary Fig. 3). For SXT-resistant UTIs, the most important factors were cognate antibiotic use and renal disease ($\sim$ 60% resistant) or history of UTI and congestive heart failure ($\sim$ 40% resistant) (Supplementary Fig. 4). Female sex, renal disease, and other/unknown race/ethnicity were associated with NIT-resistant UTIs $\sim$ 30% of
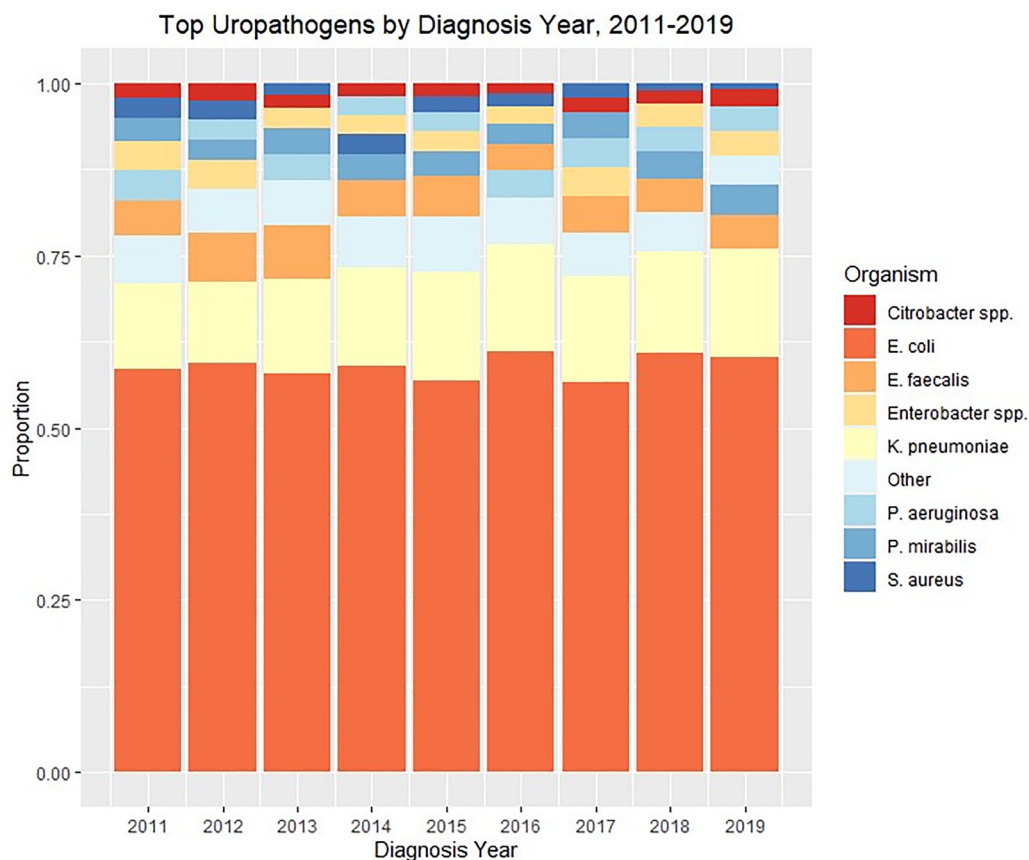
Fig. 1 Proportions of the most common uropathogens identified in this study population are plotted by year of diagnosis

the time, whereas male sex and catheter use were associated with NIT-resistant UTIs ∼ 26% of the time (Supplementary Fig. 5). For CIP-resistant infections, the most important predictor was age followed by renal disease (∼ 25% resistant) or HIV/AIDS (∼ 35% resistant) in the < 50 years old age group versus CIP use and hospitalization (∼ 50% resistant) or history of UTI (∼ 30% resistant) in the 50 + years old age group (Supplementary Fig. 6). The most important risk factors for MDR UTIs were age ≥ 60 years old, SXT use, and black race (∼ 60% resistant) (Supplementary Fig. 7).

DISCUSSION

Appropriate management of UTIs is a key component of antimicrobial stewardship in ambulatory and hospital settings. Despite this, there continues to be inappropriate selection of

antibiotics for empiric and definitive therapies for UTIs. The increasing availability of EHR, combined with the ability to quickly scan and utilize these data to guide clinical practice, motivated us to develop algorithms that would provide clinicians with an "early warning" that a UTI might be being caused by a microorganism with a single drug or multidrug-resistant phenotype. Our data suggest that, although not perfect, such algorithms can be of value to clinicians and persons involved with antimicrobial stewardship programs. Predictability of resistance and guidance toward a reasonable choice of first-line antibiotic could be valuable to a treating clinician. This predictability can prompt optimal resource utilization by consulting the infectious diseases specialist or stewardship team in guiding the use of non-standard oral or intravenous antibiotics as needed. Rather than asking clinicians to input multiple data with the use of practice alerts or
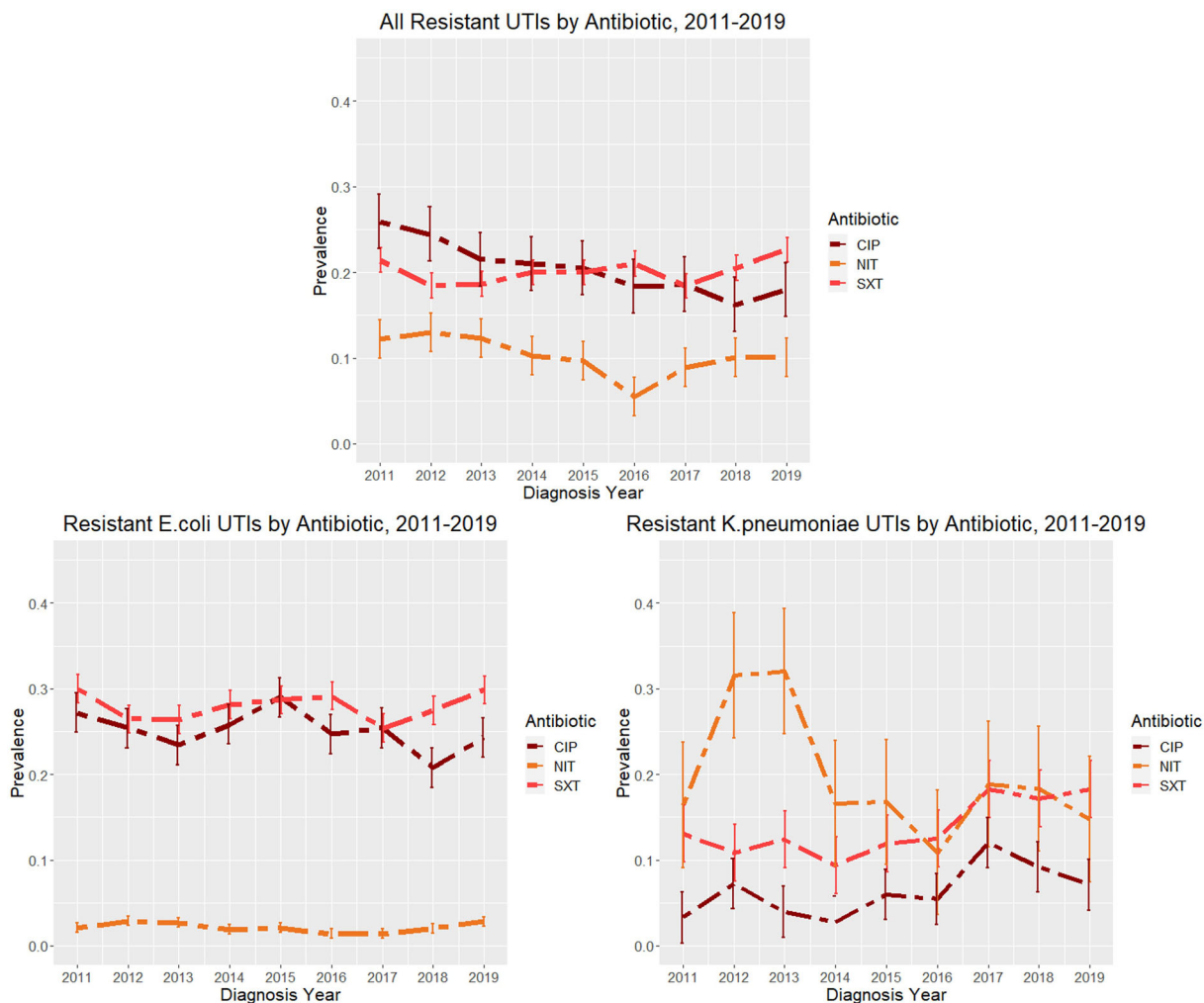
Fig. 2 Prevalence of (i) all-cause resistant urinary tract infections (UTIs) (top), (ii) resistant UTIs due to infection with *Escherichia coli* (bottom left), and (iii) *Klebsiella pneumoniae* (bottom right) by diagnosis year, grouped by major drug type: *SXT* sulfamethoxazole- trimethoprim, *NIT* nitrofurantoin, *CIP* ciprofloxacin, or multidrug resistance (MDR) to all three major drug types

order sets, these algorithms could automatically incorporate factors such as creatinine clearance or allergies, which preclude the use of certain antibiotics in addition to prediction of resistance.

In the development of such algorithms, this study was among the first to fit and compare the performance of both linear and nonlinear models developed to predict antibiotic-resistant UTIs using large, structured, and curated EHR data. The clinical decision support systems developed in this study were moderately predictive of antibiotic-resistant UTIs—with the highest performances (as measured by AUROC values) ranging from 0.57 to 0.66. These models performed similarly to those developed in another US-based population in the Northeast (AUROC = 0.56 for NIT, 0.59 for SXT, and 0.64 for CIP), although that study population was younger on average and excluded males [26]. In contrast, Yelin et al.'s algorithms using nationally representative data from Israel outperformed those in the current study (AUROC > 0.70 for all resistant outcomes) [11].

Table 2 Models' performances at predicting antibiotic-resistant urinary tract infections from bootstrap validation

Model	Any AMR (n = 6307)			SXT resistance (n = 2012)			NIT resistance (n = 1009)			CIP resistance (n = 1977)			MDR (n = 129)		
	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec	AUC	Sens	Spec
DT	0.55	0.55	0.54	0.57	0.53	0.58	0.60	0.57	0.57	0.62	0.58	0.61	0.63	0.62	0.59
	(0.01)	(0.08)	(0.08)	(0.01)	(0.05)	(0.05)	(0.02)	(0.10)	(0.11)	(0.01)	(0.05)	(0.05)	(0.02)	(0.11)	(0.10)
BLR	0.57	0.56	0.55	0.58	0.56	0.56	0.62	0.58	0.55	0.64	0.60	0.61	0.66	0.63	0.63
	(0.01)	(0.03)	(0.03)	(0.01)	(0.03)	(0.03)	(0.01)	(0.04)	(0.04)	(0.01)	(0.03)	(0.03)	(0.02)	(0.07)	(0.06)
RF	0.53	0.55	0.50	0.57	0.53	0.58	0.57	0.60	0.56	0.63	0.59	0.60	0.62	0.64	0.58
	(0.01)	(0.03)	(0.03)	(0.01)	(0.04)	(0.03)	(0.01)	(0.04)	(0.03)	(0.01)	(0.03)	(0.03)	(0.03)	(0.06)	(0.04)

Results are provided as area under the receiver-operating curve (AUC) and (standard deviation)

AMR antimicrobial resistance, *SXT* sulfa-trimethoprim, *NIT* nitrofurantoin, *CIP* ciprofloxacin, *MDR* multidrug-resistant, *DT* decision tree, *BLR* boosted logistic regression, *RF* random forest

Among the most important features selected for the prediction of resistant-UTI types in both studies were past antibiotic use and prior resistant infections. In the current study, variables most predictive of antibiotic-resistant UTIs were past cognate and non-cognate antibiotic use, renal disease, and diabetes. Previous antibiotic exposure was consistently proven to be one of the strongest predictors of future antibiotic-resistant UTIs, except for NIT-resistant UTIs. Prior cognate antibiotic use was not a significant predictor for NIT resistance. The effects of prior antibiotic use on the risk of CIP and SXT-resistant UTIs were stronger for cognate (same-drug) exposures as compared to non-cognate (different drug) exposures, also observed by Yelin et al. [11]. This finding underscores the importance of taking cumulative antibiotic exposures into consideration when assessing the risk for future resistance. Yelin and colleagues [11] also observed fewer UTIs resistant to NIT than to CIP and SXT as did we in the present study. NIT has consistently performed better than SXT and CIP in previous studies, including for the treatment of MDR *E. coli* UTIs [27].

The BLR (linear) models performed better than the RF and DT (non-linear) models in the current study, similarly observed by Kanjilal et al. [26]. AUROC values were 0.57–0.66 for the best performing models, indicating that these prediction systems may be a feasible option to support antibiotic decisions in outpatient settings, where the majority of antibiotics are prescribed in the US [28]. However, the AUROC values for these models are moderate, even if always > 0.55, indicating there is a portion of data variance that cannot be explained by the current covariate sets. One suggestion for future studies is to build separate models by age bracket to achieve better discrimination.

This study had limitations. Our population was relatively older with more comorbidities than may be prevalent in the general population at risk for UTI. Study of alternative antibiotics used in this population is needed. Additionally, information on prior healthcare exposures, such as in-patient stays, and previous drug-resistant infections were not included in this study but may be considered in future studies to improve models' predictive
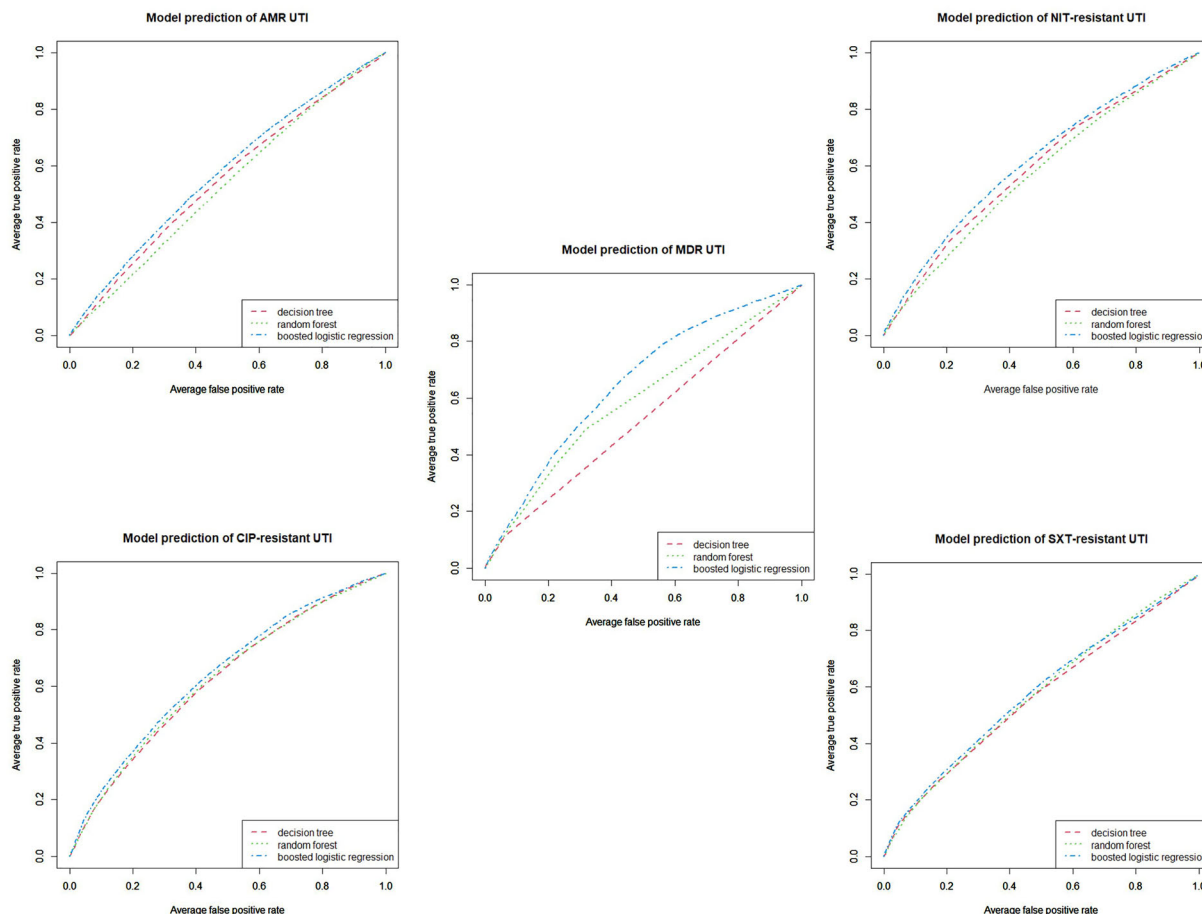
Fig. 3 Receiver-operator characteristic curves for model prediction of antibiotic-resistant urinary tract infections. *UTI* urinary tract infection, *AMR* antimicrobial resistance, *SXT* sulfa-trimethoprim, *NIT* nitrofurantoin, *CIP* ciprofloxacin, *MDR* multidrug-resistant

performances. Furthermore, EHR data from only one healthcare provider were used, which limits the generalizability of our findings to other populations; however, we did attempt to externally validate these models on a population that was geographically separate from the main study population (primarily located in Alachua County, Florida) by running the models using equivalent EHR data from the Jacksonville population (located in Duval County, Florida) with similar results. Application of these models to populations outside of Florida or the US will require additional studies. The use of variables such as race and ethnicity are readily available proxies to adjust for societal forces contributing to disease likelihood, but they by no means imply a biological/genetic

mechanism for these relationships, and future research could incorporate direct socioeconomic factors. Lastly, relying on ICD codes for comorbidity and symptom collection may have led to exposure misclassification and measurement error, which are common in studies using EHR.

Equipped with information on past antibiotic prescriptions, demographics, and comorbidities, the models presented in this study can better aid a clinician's decision-making to prevent a potential mismatched therapy in ambulatory and hospital settings. The models developed to predict antimicrobial-resistant UTIs in this study performed similarly to those published in previous studies [26]. Additionally, to enhance replicability and future work in this

area, we have proposed several computational phenotypes for each predictor used in this study, which will require validation in future studies. With the antibiotic development pipeline slowed—and no other clinically effective therapeutic options available [29]—improving the use of existing antibiotics to treat UTIs is of utmost importance in the battle against antimicrobial resistance.

CONCLUSION

In this study, we considered a variety of linear and nonlinear approaches to predict resistance to the top three antibiotics prescribed to treat UTIs in the US (SXT, NIT, and CIP) and MDR for use in both inpatient and outpatient settings. The variables included in these models are easily accessible in patients' EHR and can be used to inform the personalized prediction of antibiotic resistance in settings where phenotypic and/or genotypic resistance testing is not routine or prompt enough for patients with multiple risk factors who could have critical deterioration in health status. These data highlight the potential for use of EHR data in guiding clinical decision-making, including, in this instance, decisions regarding the selection of antimicrobial agents in UTIs.

ACKNOWLEDGEMENTS

to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES

1. Foxman B. The epidemiology of urinary tract infection. Nat Rev Urol. 2010;7:653–60.

2. Foxman B. Urinary tract infection syndromes: occurrence, recurrence, bacteriology, risk factors, and disease burden. Infect Dis Clin North Am. 2014;28:1–13.

3. Brown P, Ki M, Foxman B. Acute pyelonephritis among adults: cost of illness and considerations for the economic evaluation of therapy. Pharmacoeconomics. 2005;23:1123–42.

4. Hamilton KW, Fishman NO. Antimicrobial stewardship interventions. Infect Dis Clin North Am. 2014;28:301–13.

5. Sanchez GV, Master RN, Karlowsky JA, Bordon JM. In vitro antimicrobial resistance of urinary Escherichia coli isolates among U.S. outpatients from 2000 to 2010. Antimicrob Agents Chemother. 2012;56:2181–3.

6. Morrill HJ, Morton JB, Caffrey AR, Jiang L, Dosa D, Mermel LA, et al. Antimicrobial resistance of Escherichia coli urinary isolates in the veterans affairs health care system. Antimicrob Agents Chemother. 2017;61:e02236-e2316.

7. Grigoryan L, Trautner BW, Gupta K. Diagnosis and management of urinary tract infections in the outpatient setting: a review. JAMA. 2014;312:1677–84.

8. Taylor RA, Moore CL, Cheung K-H, Brandt C. Predicting urinary tract infections in the emergency department with machine learning. PLoS One. 2018;13: e0194085.

9. Didelot X, Pouwels KB. Machine-learning-assisted selection of antibiotic prescription. Nat Med. 2019;25:1033–4.

10. Shaikh N, Hoberman A, Keren R, Ivanova A, Gotman N, Chesney RW, et al. Predictors of antimicrobial resistance among pathogens causing urinary tract infection in children. J Pediatr. 2016;171:116–21.

11. Yelin I, Snitser O, Novich G, Katz R, Tal O, Parizade M, et al. Personal clinical history predicts antibiotic resistance of urinary tract infections. Nat Med. 2019;25:1143–52.

12. Ben Ayed H, Koubaa M, Hammami F, Marrakchi C, Rekik K, Ben Jemaa T, et al. Performance of an easy and simple new scoring model in predicting multidrug-resistant Enterobacteriaceae in community-acquired urinary tract infections. Open Forum Infect Dis [Internet]. 2019;6. https://doi.org/10.1093/ofid/ofz103/5369878 [cited 2020 Oct 3].

13. On behalf of COMBACTE-MAGNET WP5- RESCUING Study, Gomila A, Shaw E, Carratalà J, Leibovici L, Tebé C, et al. Predictive factors for multidrug-resistant gram-negative bacteria among hospitalised patients with complicated urinary tract infections. Antimicrob Resist Infect Control. 2018;7:111.

14. Sheikh K, Rangan S, Deshmukh D, Dholakia Y, Porter J. Urban private practitioners: potential partners in the care of patients with HIV/AIDS. Natl Med J India. 2005;18:32–6.

15. Laudisio A, Marinosci F, Gemma A, Bartoli IR, Montenegro N, Incalzi RA. The Burden of comorbidity is associated with antibiotic resistance among institutionalized elderly with urinary infection: a retrospective cohort study in a single Italian nursing home between 2009 and 2014. Microb Drug Resist. 2017;23:500–6.

16. Glasheen WP, Cordier T, Gumpina R, Haugh G, Davis J, Renda A. Charlson Comorbidity Index: ICD-9 update and ICD-10 translation. Am Health Drug Benefits. 2019;12:188–97.

17. Menendez ME, Neuhaus V, van Dijk CN, Ring D. The Elixhauser comorbidity method outperforms the Charlson index in predicting inpatient death after orthopaedic surgery. Clin Orthop Relat Res. 2014;472:2878–86.

18. Prosperi M, Guo Y, Sperrin M, Koopman JS, Min JS, He X, et al. Causal inference and counterfactual prediction in machine learning for actionable healthcare. Nat Mach Intell. 2020;2:369–75.

19. R Core Team. R: A language and environment for statistical computing [Internet]. Vienna: R

Foundation for Statistical Computing; 2021. https://www.R-project.org/.

20. Hothorn T, Buehlmann P, Kneib T, Schmit M, Hofner B. mboost: Model-Based Boosting [Internet]; 2020. https://CRAN.R-project.org/package=mboost.

21. Liaw A, Wiener M. Breiman and Cutler's random forests for classification and regression [Internet]; 2018. https://CRAN.R-project.org/doc/Rnews/.

22. Therneau T, Atkinson B. rpart: recursive partitioning and regression trees [Internet]; 2019. https://CRAN.R-project.org/package=rpart.

23. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework; 2006.

24. Kuhn M. caret: classification and regression training [Internet]; 2020. https://CRAN.R-project.org/package=caret.

25. Sing T, Sander O, Beerenwinkel N, Lengauer T. ROCR: visualizing classifier performance in R [Internet]. http://rocr.bioinf.mpi-sb.mpg.de.

26. Kanjilal S, Oberst M, Boominathan S, Zhou H, Hooper DC, Sontag D. A decision algorithm to promote outpatient antimicrobial stewardship for uncomplicated urinary tract infection. Sci Transl Med. 2020;12: eaay5067.

27. Sanchez GV, Baird AMG, Karlowsky JA, Master RN, Bordon JM. Nitrofurantoin retains antimicrobial activity against multidrug-resistant urinary *Escherichia coli* from US outpatients. J Antimicrob Chemother. 2014;69:3259–62.

28. Duffy E, Ritchie S, Metcalfe S, Van Bakel B, Thomas MG. Antibacterials dispensed in the community comprise 85%–95% of total human antibacterial consumption. J Clin Pharm Ther. 2018;43:59–64.

29. Wawrysiuk S, Naber K, Rechberger T, Miotla P. Prevention and treatment of uncomplicated lower urinary tract infections in the era of increasing antimicrobial resistance—non-antibiotic approaches: a systemic review. Arch Gynecol Obstet. 2019;300:821–8.

Publisher's Note