

# Predicting Risk of Multidrug-Resistant Enterobacterales Infections Among People With HIV

Heather I. Henderson,<sup>1,○</sup> Sonia Napravnik,<sup>1,○</sup> Michael R. Kosorok,<sup>2,○</sup> Emily W. Gower,<sup>1</sup> Alan C. Kinlaw,<sup>3,4,○</sup> Allison E. Aiello,<sup>1,○</sup> Billy Williams,<sup>5</sup> David A. Wohl,<sup>1,○</sup> and David van Duin<sup>1,○</sup>

<sup>1</sup>Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA, <sup>2</sup>Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA, <sup>3</sup>Division of Pharmaceutical Outcomes and Policy, University of North Carolina School of Pharmacy, Chapel Hill, North Carolina, USA, <sup>4</sup>Cecil G. Sheps Center for Health Services Research, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA, and <sup>5</sup>Clinical Microbiology Laboratory, University of North Carolina Hospitals, Chapel Hill, North Carolina, USA

**Background.** Medically vulnerable individuals are at increased risk of acquiring multidrug-resistant Enterobacterales (MDR-E) infections. People with HIV (PWH) experience a greater burden of comorbidities and may be more susceptible to MDR-E due to HIV-specific factors.

**Methods.** We performed an observational study of PWH participating in an HIV clinical cohort and engaged in care at a tertiary care center in the Southeastern United States from 2000 to 2018. We evaluated demographic and clinical predictors of MDR-E by estimating prevalence ratios (PRs) and employing machine learning classification algorithms. In addition, we created a predictive model to estimate risk of MDR-E among PWH using a machine learning approach.

**Results.** Among 4734 study participants, MDR-E was isolated from 1.6% (95% CI, 1.2%–2.1%). In unadjusted analyses, MDR-E was strongly associated with nadir CD4 cell count  $\leq 200$  cells/mm<sup>3</sup> (PR, 4.0; 95% CI, 2.3–7.4), history of an AIDS-defining clinical condition (PR, 3.7; 95% CI, 2.3–6.2), and hospital admission in the prior 12 months (PR, 5.0; 95% CI, 3.2–7.9). With all variables included in machine learning algorithms, the most important clinical predictors of MDR-E were hospitalization, history of renal disease, history of an AIDS-defining clinical condition, CD4 cell count nadir  $\leq 200$  cells/mm<sup>3</sup>, and current CD4 cell count 201–500 cells/mm<sup>3</sup>. Female gender was the most important demographic predictor.

**Conclusions.** PWH are at risk for MDR-E infection due to HIV-specific factors, in addition to established risk factors. Early HIV diagnosis, linkage to care, and antiretroviral therapy to prevent immunosuppression, comorbidities, and coinfections protect against antimicrobial-resistant bacterial infections.

**Keywords.** multidrug resistance; HIV; Enterobacterales; gram-negative; machine learning.

The World Health Organization identifies antimicrobial resistance as “a global health security threat that requires action across government sectors and society as a whole,” leading to greater morbidity and mortality [1]. Multidrug-resistant Enterobacterales (MDR-E) are especially important emerging pathogens, with carbapenem-resistant Enterobacterales (CRE) and extended-spectrum  $\beta$ -lactamase-producing Enterobacterales (ESBL-E) being classified by the US Centers for Disease Control and Prevention as “urgent” and “serious” threats, respectively [2, 3]. While CRE are mostly health care-associated infections in the United States, increasing numbers of community-associated CRE cases have been reported globally. Likewise,

ESBL-E infections due to community spread are becoming more common [3–5]. Identified risk factors for MDR-E infection include health care exposures, residence in long-term care facilities, antibiotic exposure, and immunosuppression [6–14]. Overall, medically vulnerable individuals are disproportionately at risk for antimicrobial-resistant bacterial infections, including MDR-E, as well as for adverse outcomes [3, 15].

It is well documented that people with HIV (PWH) are at increased risk of multidrug-resistant tuberculosis [16]. We and others have shown that PWH may also be at increased risk for MDR-E infection, compared with the general population [6, 17–21]. An excess risk of MDR-E among PWH is likely multifactorial. PWH have a greater burden of comorbidities, including cancer, metabolic disorders, cardiovascular disease, chronic kidney disease, liver disease, lung disease, and multimorbidity [22–28]. In addition, PWH may be more susceptible to MDR-E colonization and infection due to HIV-specific factors, including HIV-associated immunosuppression, dysbiosis, and antibiotic prophylaxis [29–31].

In this study, we assessed predictors of MDR-E in a clinical cohort of PWH receiving care at a large tertiary care center in the Southeastern United States from 2000 to 2018. We used machine learning classification algorithms to evaluate demographic and clinical factors that may be associated with MDR-E.

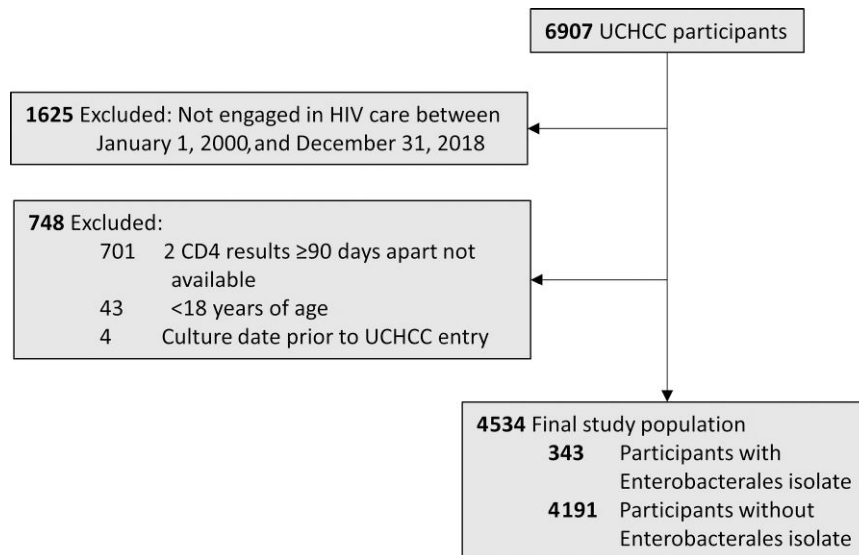
Received 27 May 2022; editorial decision 14 September 2022; accepted 15 September 2022; published online 17 September 2022

Correspondence: Heather I. Henderson, DVM, MPH, PhD, Bioinformatics Building, 130 Mason Farm Road, Chapel Hill, NC 27599 ([henderh@email.unc.edu](mailto:henderh@email.unc.edu)).

## Open Forum Infectious Diseases®

© The Author(s) 2022. Published by Oxford University Press on behalf of Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

<https://doi.org/10.1093/ofid/ofac487>



**Figure 1.** Diagram of patient flow, with inclusion and exclusion criteria. Abbreviation: UCHCC, University of North Carolina Center for AIDS Research HIV Clinical Cohort.

## METHODS

### Study Population

Our study population included all PWH receiving HIV primary care at the University of North Carolina at Chapel Hill (UNC) Infectious Diseases Clinic and participating in the UNC Center for AIDS Research HIV Clinical Cohort (UCHCC). The UNC Infectious Diseases Clinic is part of a not-for-profit integrated health care system serving PWH from all 100 North Carolina counties without regard to health insurance status. The clinic receives Ryan White funding, and the patients are representative of PWH in care across the state. Nearly all PWH seen in the clinic since 1996 participate in the UCHCC (>95%). The UCHCC study and participants have been described previously [32]. For this study, we included all adult UCHCC patients who received HIV care between January 1, 2000, and December 31, 2018. Patients were entered in the study on January 1, 2000, or the date of the first of 2 CD4 cell count measures obtained within a 12-month period and at least 90 days apart, whichever was later (baseline). They were followed until the culture date of the first MDR-E isolate, loss to follow-up, or administrative censoring (December 31, 2018), whichever was earlier. Patients were considered lost to follow-up if they did not have a CD4 cell count or HIV RNA level in >12 months. Patients <18 years of age at baseline were excluded. Inclusion and exclusion criteria are detailed in Figure 1.

### Patient Consent

Participants provided written informed consent to participate in the UCHCC. This study was approved by the UNC Office of Human Research Ethics/Institutional Review Board.

### Microbiology

Our primary outcome was an MDR-E isolate from a clinically obtained culture. We obtained microbiological data, including antibiotic susceptibility testing, from the UNC Hospitals Clinical Laboratory for bacterial cultures with growth of an Enterobacteriales species. The cultures for our analysis were obtained for clinical purposes from study participants during outpatient or inpatient visits at any UNC Health site and originated from any anatomical source (eg, blood, urine, wound). We selected the first Enterobacteriales isolate per patient, thereby excluding repeat cultures, and classified the anatomical source as “blood,” “respiratory,” “urine,” or “other.”

Susceptibility breakpoints for some antibiotics changed during the study period; therefore, we used numeric antibiotic susceptibility tests (ie, zone of inhibition measurements or minimum inhibitory concentrations) to standardize susceptibility interpretations to the current breakpoints published by the Clinical and Laboratory Standards Institute [33]. Species-antibiotic combinations were considered nonsusceptible to an antibiotic if the antibiotic susceptibility test interpretation was “intermediate” or “resistant” using current breakpoints. An isolate was considered nonsusceptible to an antibiotic class if it was nonsusceptible to any member of that class, after removing results corresponding to intrinsically resistant species-antibiotic combinations. MDR-E was defined as an Enterobacteriales that was nonsusceptible to at least 1 antibiotic from 3 or more separate antibiotic classes [34].

### Predictors of Interest

All predictive variables were obtained from the UCHCC, which includes all electronically available data from the institutional

electronic health record (EHR), including factors known to increase risk of antimicrobial resistance in the general population and factors specific to PWH. Demographic characteristics were self-reported date of birth (evaluated as age at the date of culture or censoring), gender, race, and ethnicity. Hospital admissions were evaluated as a measure of health care exposure, and CD4 cell counts and HIV RNA levels were evaluated as HIV-specific characteristics. A comorbidity was included if it was documented in the EHR at any time between baseline and the end of follow-up. We assessed comorbidities using International Classification of Diseases, 9th/10th Revision, Clinical Modification (ICD-CM), diagnosis codes. We categorized the diagnosis codes into clinically relevant groups using Healthcare Cost and Utilization Project Clinical Classifications Software [35] and aggregated the groups by disease type. The comorbidities included in our analysis were AIDS-defining clinical conditions, asthma, cerebrovascular disease, chronic obstructive pulmonary disease, diabetes mellitus, heart disease, hypertension, lipid disorders, liver disease, psychiatric disorders, renal disease, sexually transmitted infections, and substance use disorders. Their corresponding ICD-CM codes are listed in [Supplementary Table 1](#). For time-varying characteristics (age, CD4 cell count, and HIV RNA level), we included the value most proximal to the first date of a culture with an MDR-E isolate; for those without an MDR-E isolate, we used the value most proximal to censoring. In a secondary analysis, we also included calendar year of study entry.

For each of the predictors, we estimated the unadjusted prevalence ratio (PR) and 95% CI for having an Enterobacteriales isolate obtained, as well as for presence of MDR-E, our primary outcome of interest. For the analysis of predictors, we transformed continuous variables into binary or categorical form. Given the limits of detection of HIV RNA assays at the beginning of our study period, we dichotomized HIV RNA levels as  $\leq 400$  or  $> 400$  copies/mL. We categorized current CD4 cell count into 3 clinically relevant categories:  $\leq 200$ , 201–500, or  $> 500$  cells/mm<sup>3</sup>. Since only 1 patient with the outcome of interest had a nadir CD4 cell count  $> 500$  cells/mm<sup>3</sup>, we dichotomized nadir CD4 cell count as  $\leq 200$  or  $> 200$  cells/mm<sup>3</sup>. Based on the functional form of age and MDR-E prevalence, we dichotomized age as  $\leq 50$  or  $> 50$  years. Calendar year of study entry was categorized into the following time periods: 2000–2004, 2005–2009, 2010–2014, or 2015–2018.

### Machine Learning Approaches

We first fit a penalized logistic regression model (PLR [elastic-net]) to assess the relative importance of our candidate predictors in the prediction of MDR-E. In a secondary analysis, we fit an identical PLR model, except with calendar year of study entry as an additional variable.

For our predictive model, we employed super learning, a stacked ensemble machine learning approach that fits a

meta-learner on combined predictions from multiple base learners [36]. Rather than prespecifying a parametric model, machine learning algorithms exploit associations in data to model the outcome as a complex function of the covariates. The algorithms in our library of base learners were PLR, naïve Bayes, gradient boosting, support vector machines, and random forest. These base learner algorithms were selected to provide a diverse set of parametric and nonparametric models. We included each algorithm in multiple configurations to test an array of hyperparameters. Because the outcome of interest was rare in our study population, most algorithms would preferentially predict the majority class (ie, patients without MDR-E); therefore, we tested the algorithms using various combinations of class weighting, random under-sampling of the majority class, and oversampling of the minority class (synthetic minority oversampling technique [37]).

To train the super learner, we first split the data into 70% training and 30% test sets, stratified so the outcome of interest was present in an equal proportion in each set. We used 10-fold cross-validation to train the base learners, evaluating the algorithms based on their cross-validated area under the receiver operating curve (cv-AUC). For the meta-learner, we used logistic regression, regressing the actual outcome against the predictions of the base learners. The resulting coefficients then weighted the base learners to create the optimal combination for prediction. After training the super learner (with 10 repeats), we evaluated its discrimination on the held-out test set using the mean AUC. We also evaluated the relative importance of the features in the super learner's predictions using Shapley additive explanations (SHAP), a model-agnostic approach to explain how each feature contributes to predictions [38].

In each model, we included the full set of predictors, detailed in [Supplementary Table 2](#). For the PLR models, all input variables were binary. For the super learner, current and nadir CD4 cell counts, current HIV RNA level, age, and hospital length of stay were continuous (scaled to the range of 0–1). All other variables were binary. Current CD4 cell count and HIV RNA level were missing for 1% and 9% of patients, respectively. For these patients, we imputed the median value of the study population for the machine learning models. We also assessed estimates based on multiple imputation of missing values in unadjusted association analyses, and results were consistent with using the median for imputation.

Analyses were performed using R (version 4.0.2; R Foundation for Statistical Computing, Vienna, Austria) and Python (version 3.9.6; Python Software Foundation, Beaverton, OR 97008 USA).

## RESULTS

### Study Population

The characteristics of 4534 PWH in HIV care from January 2000 to December 2018 are detailed in [Tables 1](#) and [2](#).

**Table 1. Unadjusted Prevalence Ratios of Patient Characteristics and the Probability of Having a Clinically Obtained Enterobacterales Isolate, UCHCC 2000–2018**

Characteristic <sup>a</sup>	Isolate Obtained			Prevalence Ratio (95% CI)
	No. (%)	Yes, No. (%)	No, No. (%)	
Total No.	4534	343	4191	
Gender				
Male (referent)	3175 (70.0)	147 (42.9)	3028 (72.3)	—
Female	1359 (30.0)	196 (57.1)	1163 (27.7)	3.12 (2.52–3.86)
Age				
18–49 y (referent)	2715 (59.9)	166 (48.4)	2549 (60.8)	—
≥50 y	1819 (40.1)	177 (51.6)	1642 (39.2)	1.59 (1.29–1.97)
Race				
Black	2721 (60.0)	226 (65.9)	2495 (59.5)	1.35 (1.05–1.74)
Other <sup>d</sup> /unknown	449 (9.9)	33 (9.6)	416 (9.9)	1.19 (0.79–1.77)
White (referent)	1364 (30.1)	84 (24.5)	1280 (30.5)	—
Ethnicity				
Hispanic or Latino	255 (5.6)	23 (6.7)	232 (5.5)	1.12 (0.71–1.68)
Not Hispanic or Latino	2858 (63.0)	230 (67.1)	2628 (62.7)	—
Unknown	1421 (31.3)	90 (26.2)	1331 (31.8)	0.79 (0.61–1.00)
Nadir CD4 count				
≤200 cells/mm <sup>3</sup>	2223 (49.0)	227 (66.2)	1996 (47.6)	2.03 (1.63–2.55)
>200 cells/mm <sup>3</sup> (referent)	2311 (51.0)	116 (33.8)	2195 (52.4)	—
Current CD4 count <sup>b</sup>				
≤200 cells/mm <sup>3</sup>	791 (17.7)	79 (23.5)	712 (17.2)	1.55 (1.18–2.03)
201–500 cells/mm <sup>3</sup>	1368 (30.6)	108 (32.1)	1260 (30.4)	1.23 (0.96–1.57)
>500 cells/mm <sup>3</sup> (referent)	2316 (51.8)	149 (44.3)	2167 (52.4)	—
HIV RNA level <sup>c</sup>				
≤400 copies/mL (referent)	3058 (74.5)	225 (75.8)	2833 (74.5)	—
>400 copies/mL	1044 (25.5)	72 (24.2)	972 (25.5)	0.94 (0.71–1.22)
Hospital admission in prior 12 mo	772 (17.0)	125 (36.4)	647 (15.4)	2.79 (2.24–3.47)
Comorbidities				
AIDS-defining condition	1823 (40.2)	222 (64.7)	1601 (38.2)	2.73 (2.19–3.41)
Asthma	469 (10.3)	42 (12.2)	427 (10.2)	1.21 (0.86–1.65)
Cerebrovascular disease	269 (5.9)	40 (11.7)	229 (5.5)	2.09 (1.48–2.87)
Chronic obstructive pulmonary disease	407 (9.0)	54 (15.7)	353 (8.4)	1.89 (1.40–2.51)
Diabetes mellitus	558 (12.3)	67 (19.5)	491 (11.7)	1.73 (1.31–2.24)
Heart disease	1427 (31.5)	173 (50.4)	1254 (29.9)	2.22 (1.79–2.74)
Hypertension	1831 (40.4)	192 (56.0)	1639 (39.1)	1.88 (1.52–2.33)
Lipid disorder	1309 (28.9)	130 (37.9)	1179 (28.1)	1.50 (1.21–1.87)
Liver disease	1559 (34.4)	167 (48.7)	1392 (33.2)	1.81 (1.46–2.24)
Psychiatric disorder	2456 (54.2)	240 (70.0)	2216 (52.9)	1.97 (1.57–2.49)
Renal disease	1237 (27.3)	191 (55.7)	1046 (25.0)	3.35 (2.71–4.15)
Sexually transmitted infection	1353 (29.8)	96 (28.0)	1257 (30.0)	0.91 (0.72–1.15)
Substance use disorder	2404 (53.0)	191 (55.7)	2213 (52.8)	1.11 (0.90–1.38)
Time period, study entry				
2000–2004	847 (18.7)	43 (12.5)	804 (19.2)	0.65 (0.46–0.89)
2005–2009	664 (14.6)	62 (18.1)	602 (14.4)	1.19 (0.88–1.58)
2010–2014	780 (17.2)	62 (18.1)	718 (17.1)	1.01 (0.75–1.34)
2015–2018 (referent)	2243 (49.5)	176 (51.3)	2067 (49.3)	—

Abbreviations: MDR-E, multidrug-resistant Enterobacterales; UCHCC, University of North Carolina Center for AIDS Research HIV Clinical Cohort.

<sup>a</sup>For time-varying characteristics, we included the value most proximal to the first date of a culture with an MDR-E isolate; for those without an MDR-E isolate we used the value most proximal to censoring.

<sup>b</sup>Fifty-nine missing values.

<sup>c</sup>Four hundred twenty-eight missing values.

<sup>d</sup>“Other” race includes American Indian (34), Asian (16).

Participants’ median age (interquartile range [IQR]) was 46 (36–55) years, median nadir CD4 cell count (IQR) was 209 (58–378) cells/mm<sup>3</sup>, median current CD4 cell count (IQR)

was 515 (281–760) cells/mm<sup>3</sup>, 75% had a current HIV RNA level ≤400 copies/mL, and 93% had received antiretroviral therapy (ART). Comorbidities were common, particularly

**Table 2. Unadjusted Associations Between Patient Characteristics and the Probability of Having a Multidrug-Resistant Enterobacteriales Isolate, UCHCC 2000–2018**

Characteristic <sup>a</sup>	No. (%)	MDR-E Isolate		Prevalence Ratio (95% CI)
		Yes, No. (%)	No, No. (%)	
Total No.	4534	73	4461	
Gender				
Male (referent)	3175 (70.0)	35 (47.9)	3140 (70.4)	—
Female	1359 (30.0)	38 (52.1)	1321 (29.6)	2.54 (1.60–4.03)
Age				
18–49 y (referent)	2715 (59.9)	46 (63.0)	2669 (59.8)	—
≥50 y	1819 (40.1)	27 (37.0)	1792 (40.2)	0.88 (0.54–1.40)
Race				
Black	2721 (60.0)	50 (68.5)	2671 (59.9)	1.57 (0.91–2.84)
Other <sup>d</sup> /unknown	449 (9.9)	7 (9.6)	442 (9.9)	1.33 (0.51–3.11)
White (referent)	1364 (30.1)	16 (21.9)	1348 (30.2)	—
Ethnicity				
Hispanic or Latino	255 (5.6)	6 (8.2)	249 (5.6)	1.53 (0.58–3.32)
Not Hispanic or Latino (referent)	2858 (63.0)	44 (60.3)	2814 (63.1)	—
Unknown	1421 (31.3)	23 (31.5)	1398 (31.3)	1.05 (0.62–1.72)
Nadir CD4 count				
≤200 cells/mm <sup>3</sup>	2223 (49.0)	58 (79.5)	2165 (48.5)	4.02 (2.34–7.35)
>200 cells/mm <sup>3</sup> (referent)	2311 (51.0)	15 (20.5)	2296 (51.5)	—
Current CD4 count <sup>b</sup>				
≤200 cells/mm <sup>3</sup>	791 (17.7)	23 (31.9)	768 (17.4)	3.21 (1.77–5.83)
201–500 cells/mm <sup>3</sup>	1368 (30.6)	28 (38.9)	1340 (30.4)	2.26 (1.29–4.02)
>500 cells/mm <sup>3</sup> (referent)	2316 (51.8)	21 (29.2)	2295 (52.1)	—
HIV RNA level <sup>c</sup>				
≤400 copies/mL (referent)	3058 (74.5)	35 (58.3)	3023 (74.8)	—
>400 copies/mL	1044 (25.5)	25 (41.7)	1019 (25.2)	2.09 (1.24–3.48)
Hospital admission in prior 12 mo	772 (17.0)	37 (50.7)	735 (16.5)	5.01 (3.16–7.94)
Comorbidities				
Asthma	469 (10.3)	7 (9.6)	462 (10.4)	0.92 (0.38–1.87)
AIDS-defining condition	1823 (40.2)	52 (71.2)	1771 (39.7)	3.68 (2.25–6.24)
Cerebrovascular disease	269 (5.9)	13 (17.8)	256 (5.7)	3.44 (1.80–6.05)
Chronic obstructive pulmonary disease	407 (9.0)	7 (9.6)	400 (9.0)	1.08 (0.45–2.18)
Diabetes mellitus	558 (12.3)	11 (15.1)	547 (12.3)	1.26 (0.63–2.30)
Heart disease	1427 (31.5)	39 (53.4)	1388 (31.1)	2.50 (1.58–3.97)
Hypertension	1831 (40.4)	39 (53.4)	1792 (40.2)	1.69 (1.07–2.69)
Lipid disorder	1309 (28.9)	21 (28.8)	1288 (28.9)	0.99 (0.59–1.63)
Liver disease	1559 (34.4)	32 (43.8)	1527 (34.2)	1.49 (0.93–2.36)
Psychiatric disorder	2456 (54.2)	52 (71.2)	2404 (53.9)	2.10 (1.28–3.55)
Renal disease	1237 (27.3)	44 (60.3)	1193 (26.7)	4.04 (2.54–6.53)
Sexually transmitted infection	1353 (29.8)	14 (19.2)	1339 (30.0)	0.56 (0.30–0.97)
Substance use disorder	2404 (53.0)	39 (53.4)	2365 (53.0)	1.02 (0.64–1.62)
Time period, study entry				
2000–2004	847 (18.7)	14 (19.2)	833 (18.7)	3.09 (2.65–11.00)
2005–2009	664 (14.6)	25 (34.2)	639 (14.3)	7.04 (3.61–14.51)
2010–2014	780 (17.2)	22 (30.1)	758 (17.0)	5.27 (2.65–6.80)
2015–2018 (referent)	2243 (49.5)	12 (16.4)	2231 (50.0)	—

Abbreviations: MDR-E, multidrug-resistant Enterobacteriales; UCHCC, University of North Carolina Center for AIDS Research HIV Clinical Cohort.

<sup>a</sup>For time-varying characteristics, we included the value most proximal to the first date of a culture with an MDR-E isolate; for those without an MDR-E isolate, we used the value most proximal to censoring.

<sup>b</sup>Fifty-nine missing values.

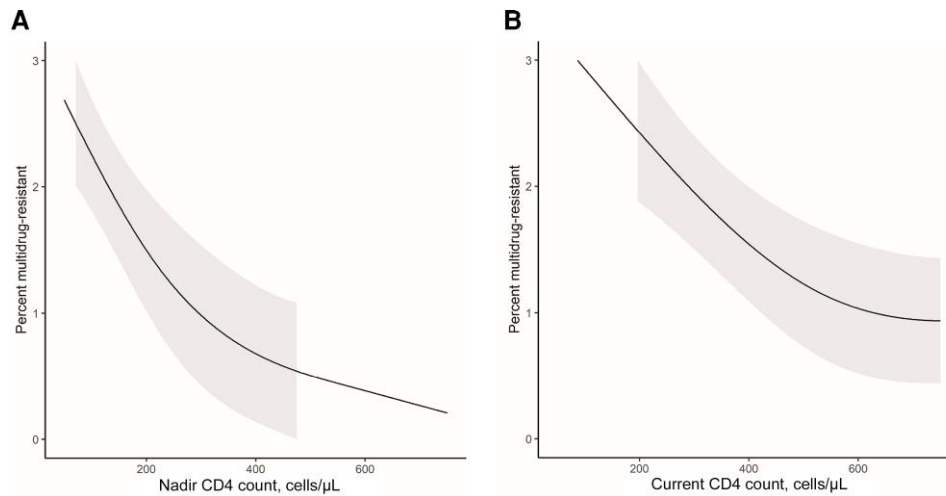
<sup>c</sup>Four hundred twenty-eight missing values.

<sup>d</sup>“Other” race includes American Indian (34), Asian (16).

psychiatric and substance use disorders and hypertension. An AIDS-defining clinical condition had been diagnosed in 40% of participants. Within the 12 months before the end of follow-

up, 17% of participants had been hospitalized at least once. The median follow-up time (IQR) was 5 (3–10) years until the first MDR-E isolate, loss to follow-up, or administrative censoring.





**Figure 2.** Predicted percentage of patients with multidrug-resistant Enterobacteriales infections by CD4 count. *A*, Nadir CD4 cell count. *B*, Current CD4 cell count. Percentages were estimated using restricted cubic spline models. Bands represent 95% confidence intervals (bands do not extend the full range of estimates due to lack of precision at the extremes).

### Enterobacteriales Isolates

Among the 4534 participants, 343 (8%) had an Enterobacteriales isolate obtained (Table 1). The majority of isolates were urinary (80%), followed by blood (12%) and respiratory (2%), with the remaining 6% originating from other anatomical sites. Participants who were female or  $\geq 50$  years of age were more likely to have had an isolate obtained (PR, 3.1; 95% CI, 2.5–3.9; and PR, 1.6; 95% CI, 1.3–2.0; respectively). A nadir CD4 cell count  $\leq 200$  cells/mm<sup>3</sup>, compared with  $>200$  cells/mm<sup>3</sup>, was also associated with having had an isolate obtained (PR, 2.0; 95% CI, 1.6–2.6), as were hospital admission within the prior 12 months (PR, 2.8; 95% CI, 2.2–3.5) and a history of an AIDS-defining clinical condition (PR, 2.7; 95% CI, 2.2–3.4). Several comorbidities were associated with having an Enterobacteriales isolate, most notably renal disease (PR, 3.4; 95% CI, 2.7–4.2). Participants who entered the study during 2000–2004 were less likely to have had an isolate obtained, compared with those who entered later (PR, 0.7; 95% CI, 0.5–0.9).

### Predictors of MDR-E

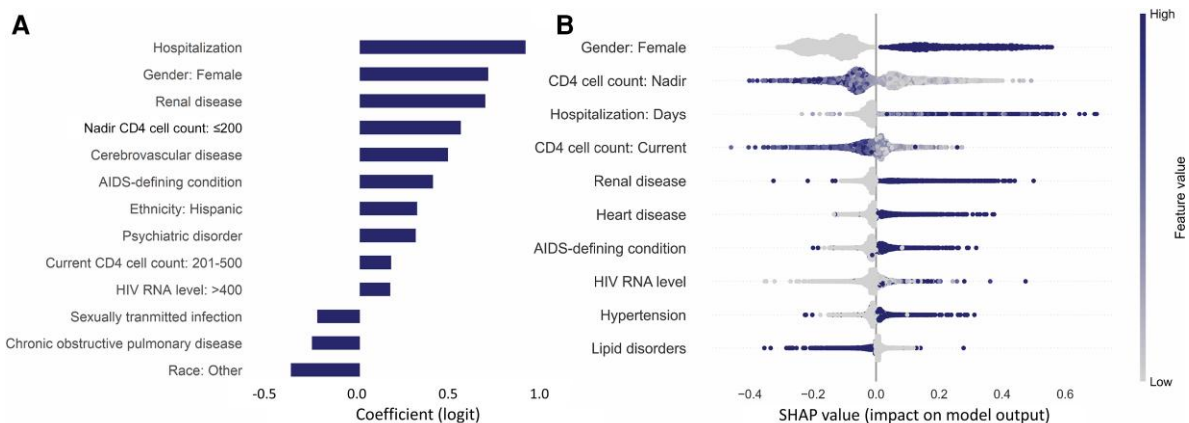
Overall, 1.6% of participants ( $n = 73$ ) had an MDR-E isolate (95% CI, 1.2%–2.1%). Unadjusted associations of demographic and clinical characteristics with MDR-E are presented in Table 2. Females had a higher prevalence of MDR-E than males, with a PR of 2.5 (95% CI, 1.6–4.0). Participants with a nadir CD4 cell count  $\leq 200$ , vs  $>200$  cells/mm<sup>3</sup>, had a higher prevalence of MDR-E (PR, 4.0; 95% CI, 2.3–7.4). Similarly, a current CD4 cell count of  $\leq 200$  or 201–500 cells/mm<sup>3</sup>, vs  $>500$  cells/mm<sup>3</sup>, was associated with higher MDR-E prevalence (PR, 3.2; 95% CI, 1.8–5.8; and PR, 2.3; 95% CI, 1.3–4.0; respectively). Restricted cubic spline models estimating the predicted percentage of patients with MDR-E by continuous CD4 cell count confirmed the

increasing MDR-E prevalence at lower nadir, as well as current, CD4 cell counts (Figure 2). Other characteristics associated with greater MDR-E prevalence included hospital admission in the prior 12 months (PR, 5.0; 95% CI, 3.2–7.9) and having a history of an AIDS-defining clinical condition, cerebrovascular disease, or renal disease, with PRs of 3.7 (95% CI, 2.3–6.2), 3.4 (95% CI, 1.8–6.0), and 4.0 (95% CI, 2.5–6.5), respectively. Participants who entered the study during 2015–2018 were less likely to have had MDR-E, compared with those entering earlier.

### Machine Learning Algorithms

Our PLR model supported the findings of the unadjusted analyses, identifying female gender, low nadir and current CD4 cell count, hospitalization within the prior 12 months, renal disease, and an AIDS-defining clinical condition as the most important predictors of MDR-E. Results are presented in Figure 3A, with the coefficient indicating the relative influence of each patient characteristic in predicting MDR-E. Regularization adds a penalty term to the loss function to avoid overfitting and obtain better predictive performance. The penalty biases the coefficients toward the null, complicating interpretation of the coefficients and estimation of meaningful confidence intervals. Therefore, we do not report measures of association from the PLR models.

To fit the super learner, we included 18 configurations of the base learners. The base learner configurations and their cv-AUCs are detailed in Supplementary Table 3. After fitting the base learners to the training data and using the cv-AUC as an indicator of performance, the algorithms all performed similarly, with cv-AUCs of 65%–72%. The mean AUC of the super learner on the held-out test data was 71%, sensitivity was 67%, specificity was 76%, and positive and negative predictive values were 5% and 99%, respectively. The SHAP analysis



**Figure 3.** Relative influence of predictors on machine learning model output. *A*, Coefficients from penalized logistic regression (elastic-net) model. The model includes all demographic and clinical predictors of interest, with variables specified as binary input features. *B*, Shapley additive explanations for super learner model. The model includes all demographic and clinical predictors of interest, with variables specified as either continuous or binary input features. Continuous features vary from low to high values, whereas binary features are either present or absent. Each dot represents the impact of a feature on the prediction of a multidrug-resistant Enterobacteriales isolate for 1 patient. Abbreviation: SHAP, Shapley additive explanations.

(Figure 3B) revealed female gender to be the most influential predictor of MDR-E in our study participants, followed by lower values of nadir CD4 cell count, higher values of days hospitalized in the prior 12 months, and lower values of current CD4 cell count, respectively. A history of renal disease, heart disease, and an AIDS-defining clinical condition were also influential in prediction of MDR-E.

In our secondary PLR model with the addition of calendar time of study entry, the time periods 2005–2009 and 2010–2014 were the strongest predictors of MDR-E (Supplementary Figure 1). This model retained the same predictors as the original PLR model, identifying hospitalization within the prior 12 months, female gender, renal disease, and low nadir CD4 cell count as important predictors of MDR-E, albeit with decreased influence on the model output.

## DISCUSSION

In this well-characterized clinical cohort of PWH from 2000 to 2018, the prevalence of MDR-E was 1.6%. As no nationwide surveillance system monitors the prevalence of MDR-E, we are unable to compare our findings to national estimates. Notably, we found that HIV-specific factors, including lower current and nadir CD4 cell counts, were strongly associated with having an MDR-E isolate. Participants also had a greater likelihood of MDR-E if they had evidence of MDR-E risk factors that have been established in the general population, including recent hospitalization and certain comorbidities. Consistent with greater risk among those with more interactions with the health care system, our results underscore excess risk among medically vulnerable PWH.

In unadjusted analyses, participants with nadir CD4 cell counts  $\leq 200$  cells/mm<sup>3</sup> had substantially increased prevalence

of MDR-E, as did those with current CD4 cell counts  $\leq 500$  cells/mm<sup>3</sup>. Our results show that the predicted prevalence of MDR-E rose sharply as the nadir CD4 cell count decreased below  $\sim 300$  cells/mm<sup>3</sup>. A similar pattern was observed as the current CD4 cell count decreased below  $\sim 500$  cells/mm<sup>3</sup>. CD4 cell count is a marker of severity of HIV-induced immune deficiency, and low CD4 cell counts increase the risk of mortality and morbidity, including opportunistic infections.

Enterobacteriales are colonizers of the gut, and HIV infection disrupts the gut mucosal epithelia and microbiota, resulting in microbial translocation [39]. Therefore, HIV-associated gut microbiota changes and inflammatory responses may lead to colonization and subsequent infection with MDR-E [39, 40]. It is also possible that low CD4 cell counts are associated with increased MDR-E risk by increasing a patient's interactions with the health care system through hospitalizations or clinical management of AIDS-defining clinical conditions or other comorbidities. Having a history of an AIDS-defining clinical condition was also strongly associated with MDR-E, and this association remained when all other predictors, including current and nadir CD4 cell counts and hospitalization, were included in the models.

In addition, PWH had a higher prevalence of MDR-E if they had been hospitalized or had renal disease. This is consistent with risk factors documented in the general population. For example, in a large study of gram-negative rods isolated from hospitalized patients, renal disease was present in 46% of patients with ertapenem-resistant isolates, compared with 32% of patients with ertapenem-susceptible isolates. In that study, renal disease was the only comorbidity retained in the final model for prediction of ertapenem-resistant gram-negative infection [10].

In our PLR model with calendar period of study entry included as a predictor, calendar years early in the study period

were the strongest predictors of MDR-E. The predictors identified in our primary analyses were retained, though their influence was somewhat reduced. This is likely related to the use of more potent ART and other changes in the management of HIV infection over the study period [41].

In contrast to the general population, we did not find age to be associated with risk of MDR-E, either in unadjusted analyses or in our models. This may be related to the young age at which many PWH are diagnosed with HIV-related conditions, as well as the younger ages at which PWH develop comorbidities in comparison to the general population. For instance, PWH are diagnosed with cancers and cardiovascular conditions at younger ages compared with persons without HIV [26, 28, 42].

To identify predictors of MDR-E among PWH, we relied on a set of machine learning classification algorithms. Because no single algorithm invariably outperforms all others, we used super learning to create a predictive model. As a predictive model, the super learner will, in theory, perform at least as well as the optimal algorithm in a library of base learners [36]. We found that all algorithms in the super learner library performed similarly on the training data. The super learner model had low-to-moderate sensitivity and specificity for predicting MDR-E in our participants. Furthermore, given the low prevalence of MDR-E in the study population, the negative predictive value of the super learner was high at 99%, while the positive predictive value was just 5%. With a rare condition, positive predictive value is expected to be low, even when sensitivity is high. Improving the performance of a predictive model would therefore require a population with a much higher prevalence of MDR-E.

Our findings of an association between MDR-E and low CD4 cell counts and diagnosis of an AIDS-defining clinical condition underscore the continuing need for strategies that minimize the time from HIV infection to linkage to HIV care, as well as the time from HIV care initiation to ART initiation [43, 44]. Strategies that prevent HIV progression are relevant to reducing MDR-E risk among PWH, in addition to their established effects decreasing HIV-related morbidity, mortality, and non-HIV clinical outcomes such as myocardial infarction, end-stage renal disease, cancer, and hospitalization [43, 45, 46]. Further efforts are also needed to support PWH in maintaining consistent engagement with HIV care and remaining on efficacious ART to preserve and improve immune function, with programs spanning individual, health care system, and policy levels [47].

A strength of this study is the large clinical cohort of PWH, with longitudinal data collected throughout the 19-year study period. We were able to verify susceptibility results of isolates to all antibiotic classes that were tested, as well as correct for changes in susceptibility definitions by updating interpretations with current susceptibility breakpoints. However, the study is subject to several limitations. The study is based on a

clinical cohort of PWH receiving care at a single tertiary care center in North Carolina, and results need to be reproduced in larger cohorts and other geographical areas. The outcome of interest was rare in our study population, limiting the performance of the classification algorithms as well as the number of predictors that could be analyzed. We were unable to include some patient characteristics that may increase MDR-E risk, such as residence in long-term care facilities, invasive medical procedures, and use of antibiotics including prophylactic medications, due to limitations of the available data. In spite of these limitations, we were able to show that some HIV-specific characteristics, as well as general risk factors that are more prevalent in PWH, were associated with MDR-E.

In conclusion, we found that prevalence of MDR-E in PWH is associated with HIV-specific factors, including low CD4 cell counts, in addition to established risk factors, such as hospitalization and comorbidities. Early diagnosis of HIV, linkage to care, and provision of ART to preserve immune function, which is critical to reducing HIV-related morbidity and mortality, may also protect PWH from antimicrobial-resistant bacterial infections. Additionally, screening and treatment for comorbidities among PWH may reduce MDR-E infections in this population. This may be increasingly important as PWH live to older ages, given the efficacy of current ART.

### Supplementary Data

**Supplementary materials** are available at *Open Forum Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

### Acknowledgments

**Financial support.** This work was supported by the University of North Carolina at Chapel Hill Center for AIDS Research, a National Institutes of Health-funded program (grant number P30 AI50410); the National Center for Advancing Translational Sciences, National Institutes of Health (grant number TL1TR002491); and the National Institute of Allergy and Infectious Diseases (grant numbers T32 AI070114 and AI007001 to H.I.H.).

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

**Author contributions.** Heather I. Henderson: conceptualization, methodology, formal analysis, data curation, writing - original draft, visualization. Sonia Napravnik: conceptualization, methodology, supervision, writing - review & editing, project administration, funding acquisition. Michael R. Kosorok: methodology, supervision, writing - review & editing. Emily W. Gower: conceptualization, methodology, supervision, writing - review & editing. Alan C. Kinlaw: conceptualization, methodology, supervision, writing - review & editing. Allison E. Aiello: conceptualization, methodology, supervision, writing - review & editing. Billy Williams: investigation, data curation, writing - review & editing. David A. Wohl: conceptualization, writing - review & editing, funding acquisition. David van Duin: conceptualization, methodology, supervision, writing - review & editing, project administration, funding acquisition.



## References

1. World Health Organization. Antimicrobial resistance: global report on surveillance 2014. Available at: <http://www.who.int/drugresistance/documents/surveillancereport/en/>. Accessed January 17, 2020.
2. Laxminarayan R, Duse A, Wattal C, et al. Antibiotic resistance—the need for global solutions. *Lancet Infect Dis* **2013**; 13:1057–98.
3. Centers for Disease Control and Prevention. The biggest antibiotic-resistant threats in the U.S. **2019**. Available at: <https://www.cdc.gov/drugresistance/biggest-threats.html>. Accessed January 17, 2020.
4. Thaden JT, Fowler VG, Sexton DJ, Anderson DJ. Increasing incidence of extended-spectrum  $\beta$ -lactamase-producing *Escherichia coli* in community hospitals throughout the Southeastern United States. *Infect Control Hosp Epidemiol* **2016**; 37:49–54.
5. van Duin D, Paterson DL. Multidrug-resistant bacteria in the community: an update. *Infect Dis Clin North Am* **2020**; 34:709–22.
6. da Silva Winter J, dos Santos RP, de Azambuja AZ, Cechinel AB, Goldani LZ. Microbiologic isolates and risk factors associated with antimicrobial resistance in patients admitted to the intensive care unit in a tertiary care hospital. *Am J Infect Control* **2013**; 41:846–8.
7. Lim CJ, Cheng AC, Kong DC, Peleg AY. Community-onset bloodstream infection with multidrug-resistant organisms: a matched case-control study. *BMC Infect Dis* **2014**; 14:126.
8. Neuner EA, Yeh J-Y, Hall GS, et al. Treatment and outcomes in carbapenem-resistant *Klebsiella pneumoniae* bloodstream infections. *Diagn Microbiol Infect Dis* **2011**; 69:357–62.
9. Nounoune A, Ticinesi A, Lauretani F, et al. Comorbidities and disease severity as risk factors for carbapenem-resistant *Klebsiella pneumoniae* colonization: report of an experience in an internal medicine unit. *PLoS One* **2014**; 9:e110001.
10. Richter SE, Miller L, Needleman J, et al. Risk factors for development of carbapenem resistance among gram-negative rods. *Open Forum Infect Dis* **2019**; 6:XXX–XX.
11. van Duin D, Strassle PD, DiBiase LM, et al. Timeline of healthcare-associated infections and pathogens after burn injuries. *Am J Infect Control* **2016**; 44:1511–6.
12. Ben-Ami R, Rodríguez-Baño J, Arslan H, et al. A multinational survey of risk factors for infection with extended-spectrum beta-lactamase-producing Enterobacteriaceae in nonhospitalized patients. *Clin Infect Dis* **2009**; 49:682–90.
13. Gasink LB, Edelstein PH, Lautenbach E, Synnestvedt M, Fishman NO. Risk factors and clinical impact of *Klebsiella pneumoniae* carbapenemase-producing *K. pneumoniae*. *Infect Control Hosp Epidemiol* **2009**; 30:1180–5.
14. Patel G, Huprikar S, Factor SH, Jenkins SG, Calfee DP. Outcomes of carbapenem-resistant *Klebsiella pneumoniae* infection and the impact of antimicrobial and adjunctive therapies. *Infect Control Hosp Epidemiol* **2008**; 29:1099–106.
15. Perez F, van Duin D. Carbapenem-resistant Enterobacteriaceae: a menace to our most vulnerable patients. *Cleve Clin J Med* **2013**; 80:225–33.
16. Sultana ZZ, Hoque FU, Beyene J, et al. HIV infection and multidrug resistant tuberculosis: a systematic review and meta-analysis. *BMC Infect Dis* **2021**; 21:51.
17. Madhi SA, Petersen K, Madhi A, Khoosal M, Klugman KP. Increased disease burden and antibiotic resistance of bacteria causing severe community-acquired lower respiratory tract infections in human immunodeficiency virus type 1-infected children. *Clin Infect Dis* **2000**; 31:170–6.
18. Marbou WJT, Kuete V. Bacterial resistance and immunological profiles in HIV-infected and non-infected patients at Mbouda AD LUCEM hospital in Cameroon. *J Infect Public Health* **2017**; 10:269–76.
19. Reinheimer C, Keppler OT, Stephan C, Wichelhaus TA, Friedrichs I, Kempf VAJ. Elevated prevalence of multidrug-resistant gram-negative organisms in HIV positive men. *BMC Infect Dis* **2017**; 17:206.
20. Oлару ID, Ferrand RA, Chisenga M, et al. Prevalence of ESBL-producing *Escherichia coli* in adults with and without HIV presenting with urinary tract infections to primary care clinics in Zimbabwe. *JAC Antimicrob Resist* **2021**; 3:dlab082.
21. Henderson HI, Napravnik S, Gower EW, et al. Resistance in Enterobacteriales is higher among people with HIV. *Clin Infect Dis* **2022**; 75:28–34.
22. Kaplan-Lewis E, Aberg JA, Lee M. Aging with HIV in the ART era. *Semin Diagn Pathol* **2017**; 34:384–97.
23. Aberg JA. Aging, inflammation, and HIV infection. *Top Antivir Med* **2012**; 20:101–5.
24. Wada N, Jacobson LP, Cohen M, French A, Phair J, Muñoz A. Cause-specific life expectancies after 35 years of age for human immunodeficiency syndrome-infected and human immunodeficiency syndrome-negative individuals followed simultaneously in long-term cohort studies, 1984–2008. *Am J Epidemiol* **2013**; 177:116–25.
25. Antiretroviral Therapy Cohort Collaboration. Causes of death in HIV-1-infected patients treated with antiretroviral therapy, 1996–2006: collaborative analysis of 13 HIV cohort studies. *Clin Infect Dis* **2010**; 50:1387–96.
26. Guaraldi G, Orlando G, Zona S, et al. Premature age-related comorbidities among HIV-infected persons compared with the general population. *Clin Infect Dis* **2011**; 53:1120–6.
27. Serrano-Villar S, Gutiérrez F, Miralles C, et al. Human immunodeficiency virus as a chronic disease: evaluation and management of non-acquired immune deficiency syndrome-defining conditions. *Open Forum Infect Dis* **2016**; 3:XXX–XX.
28. Shiels MS, Pfeiffer RM, Engels EA. Age at cancer diagnosis among people with AIDS in the United States. *Ann Intern Med* **2010**; 153:452–60.
29. Egwuatu CC, Iwuafor AA, Egwuatu TO, et al. Effect of trimethoprim-sulfamethoxazole prophylaxis on faecal carriage rates of resistant isolates of *Escherichia coli* in HIV-infected adult patients in Lagos. *Afr J Infect Dis* **2016**; 10:156–63.
30. Martin JN, Rose DA, Hadley WK, Perdreau-Remington F, Lam PK, Gerberding JL. Emergence of trimethoprim-sulfamethoxazole resistance in the AIDS era. *J Infect Dis* **1999**; 180:1809–18.
31. Sibanda EL, Weller IVD, Hakim JG, Cowan FM. Does trimethoprim-sulfamethoxazole prophylaxis for HIV induce bacterial resistance to other antibiotic classes? Results of a systematic review. *Clin Infect Dis* **2011**; 52:1184–94.
32. Napravnik S, Eron JJ Jr, McKaig RG, Heine AD, Menezes P, Quinlivan E. Factors associated with fewer visits for HIV primary care at a tertiary care center in the Southeastern U.S. *AIDS Care* **2006**; 18:45–50.
33. Clinical and Laboratory Standards Institute. M100 | Performance Standards for Antimicrobial Susceptibility Testing. 30th ed. Clinical and Laboratory Standards Institute; **2018**.
34. Magiorakos A-P, Srinivasan A, Carey RB, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect Dis* **2012**; 18:268–81.
35. Agency for Healthcare Research and Quality. Clinical Classifications Software Refined (CCSR) for ICD-10-CM diagnoses. Available at: <https://www.hcup-us.ahrq.gov/toolssoftware/ccsr/dxcsr.jsp>. Accessed July 2, 2021.
36. van der Laan MJ, Polley EC, Hubbard AE. Super learner. *Stat Appl Genet Mol Biol* **2007**; 6:Article25.
37. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res* **2002**; 16:321–57.
38. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, et al, eds. *Advances in Neural Information Processing Systems*. Curran Associates, Inc.; **2017**. Available at: <https://papers.nips.cc/paper/2017/hash/8a20a8621978632d76c43dfd2867767-Abstract.html>. Accessed July 12, 2022.
39. Ribeiro ABDTM, Heimesaat MM, Bereswill S. Changes of the intestinal microbiome-host homeostasis in HIV-infected individuals – a focus on the bacterial gut microbiome. *Eur J Microbiol Immunol* **2017**; 7:158–67.
40. Bandera A, De Benedetto I, Bozzi G, Gori A. Altered gut microbiome composition in HIV infection: causes, effects and potential intervention. *Curr Opin HIV AIDS* **2018**; 13:73–80.
41. Nance RM, Delaney JAC, Simoni JM, et al. HIV viral suppression trends over time among HIV-infected patients receiving care in the United States, 1997 to 2015: a cohort study. *Ann Intern Med* **2018**; 169:376–84.
42. Schouten J, Wit FW, Stolte IG, et al. Cross-sectional comparison of the prevalence of age-associated comorbidities and their risk factors between HIV-infected and uninfected individuals: the AGEHIV cohort study. *Clin Infect Dis* **2014**; 59:1787–97.
43. Althoff KN, Gange SJ, Klein MB, et al. Late presentation for human immunodeficiency virus care in the United States and Canada. *Clin Infect Dis* **2010**; 50:1512–20.
44. Kitahata MM, Gange SJ, Abraham AG, et al. Effect of early versus deferred antiretroviral therapy for HIV on survival. *N Engl J Med* **2009**; 360:1815–26.
45. Antiretroviral Therapy Cohort Collaboration. Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies. *Lancet HIV* **2017**; 4:e349–56.
46. Davy-Mendez T, Napravnik S, Eron JJ, et al. Current and past immunodeficiency are associated with higher hospitalization rates among persons on virologically suppressive antiretroviral therapy for up to 11 years. *J Infect Dis* **2021**; 224:657–66.
47. Mugavero MJ, Norton WE, Saag MS. Health care system and policy factors influencing engagement in HIV medical care: piecing together the fragments of a fractured health care delivery system. *Clin Infect Dis* **2011**; 52(Suppl 2):S238–46.