2021

Development and validation of a multivariable prediction model for missed HIV health care provider visits in a large US clinical cohort

April C. Pettit

Aihua Bian

Cassandra O. Schember

Peter F. Rebeiro

Jeanne C. Keruly

See next page for additional authors

Authors

April C. Pettit, Aihua Bian, Cassandra O. Schember, Peter F. Rebeiro, Jeanne C. Keruly, Kenneth H. Mayer, W. Christopher Mathews, Richard D. Moore, Heidi M. Crane, Elvin Geng, Sonia Napravnik, Bryan E. Shepherd, and Michael J. Mugavero

Open Forum Infectious Diseases

MAJOR ARTICLE

Development and Validation of a Multivariable Prediction Model for Missed HIV Health Care Provider Visits in a Large US Clinical Cohort

April C. Pettit,[1,2] Aihua Bian,[3] Cassandra O. Schember,[2] Peter F. Rebeiro,[1,2,3] Jeanne C. Keruly,[4] Kenneth H. Mayer,[5] W. Christopher Mathews,[6] Richard D. Moore,[4] Heidi M. Crane,[7] Elvin Geng,[8] Sonia Napravnik,[9] Bryan E. Shepherd,[3] and Michael J. Mugavero[10]; for the Centers for AIDS Research Network of Integrated Clinical Systems (CNICS)

[1]Division of Infectious Diseases, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [2]Division of Epidemiology, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [3]Department of Biostatistics, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [4]Division of Infectious Diseases, Johns Hopkins University, Baltimore, Maryland, USA, [5]Fenway Health and Harvard Medical School, Boston, Massachusetts, USA, [6]Department of Medicine, University of California San Diego, San Diego, California, USA, [7]Division of Infectious Diseases, University of Washington School of Medicine, Seattle, Washington, USA, [8]Division of Infectious Diseases, Washington University School of Medicine, St. Louis, Missouri, USA, [9]Division of Infectious Diseases, University of North Carolina Chapel Hill School of Medicine, Chapel Hill, North Carolina, USA, and [10]Division of Infectious Diseases, University of Alabama at Birmingham School of Medicine, Birmingham, Alabama, USA

Background. Identifying individuals at high risk of missing HIV care provider visits could support proactive intervention. Previous prediction models for missed visits have not incorporated data beyond the individual level.

Methods. We developed prediction models for missed visits among people with HIV (PWH) with ≥1 follow-up visit in the Center for AIDS Research Network of Integrated Clinical Systems from 2010 to 2016. Individual-level (medical record data and patient-reported outcomes), community-level (American Community Survey), HIV care site–level (standardized clinic leadership survey), and structural-level (HIV criminalization laws, Medicaid expansion, and state AIDS Drug Assistance Program budget) predictors were included. Models were developed using random forests with 10-fold cross-validation; candidate models with the highest area under the curve (AUC) were identified.

Results. Data from 382 432 visits among 20 807 PWH followed for a median of 3.8 years were included; the median age was 44 years, 81% were male, 37% were Black, 15% reported injection drug use, and 57% reported male-to-male sexual contact. The highest AUC was 0.76, and the strongest predictors were at the individual level (prior visit adherence, age, CD4+ count) and community level (proportion living in poverty, unemployed, and of Black race). A simplified model, including readily accessible variables available in a web-based calculator, had a slightly lower AUC of .700.

Conclusions. Prediction models validated using multilevel data had a similar AUC to previous models developed using only individual-level data. The strongest predictors were individual-level variables, particularly prior visit adherence, though community-level variables were also predictive. Absent additional data, PWH with previous missed visits should be prioritized by interventions to improve visit adherence.

Keywords. HIV; missed visits; prediction model; random forests; retention in care.

INTRODUCTION

HIV infection remains a significant public health problem in the United States, with estimates of >1 million people with HIV (PWH) in 2018 [1]. In 2019 the US Department of Health and Human Services announced the *Ending the HIV Epidemic: A Plan for America* goals, which include reduction of incident HIV by 75% in 2025 and by 90% in 2030 [2]. Increasing antiretroviral therapy (ART) uptake and adherence by PWH is effective in preventing HIV transmission [3].

Retention in care (RIC) by attendance at HIV provider visits is critical to sustained ART receipt [4–6]. Among numerous RIC measures, missed visits are uniquely captured in real time, amenable to immediate intervention, and associated with deleterious HIV outcomes [7]. Many studies have identified individual characteristics (age, sex, race) as risk factors for missing HIV health care provider visits [8], although these factors are often fixed, immutable characteristics that identify at-risk groups but do not serve as modifiable intervention targets. These characteristics must be complemented by health system–, community-, and structural-level factors, as conceptualized in the Socio-ecological Model of HIV Behaviors [9]. Multilevel

Received 9 December 2020; editorial decision 11 March 2021; accepted 12 March 2021.
Correspondence: April Pettit, MD, MPH, 1161 21st Avenue South, Nashville, TN 37232 (april.pettit@vumc.org).

Open Forum Infectious Diseases®2021

DOI: 10.1093/ofid/ofab130

analyses are lacking but would allow a more comprehensive approach to understanding and improving RIC.

We sought to develop and validate a predictive model for missing scheduled HIV health care provider visits, which included determinants from multiple levels (individual, health system, community, and structural levels). We also sought to compare the performance of a full and simple model incorporating factors readily accessible at the point of care (POC) to proactively identify patients who are likely to miss their next scheduled visit and to permit prioritized resource utilization aimed to improve RIC among those most likely to benefit.

METHODS

Study Population

We developed and validated a prediction model using data from PWH ≥18 years of age who attended a new patient and ≥1 follow-up HIV health care provider visit at a participating Center for AIDS Research (CFAR) Network of Integrated Clinical Systems (CNICS) site from January 1, 2010, through December 31, 2015. CNICS is a prospective observational cohort study of adult PWH in routine clinical care at 8 academic institutions across the United States, which integrates clinical data from electronic medical records with other data sources such as patient-reported outcomes (PROs) [10]. Individuals contributed person-time from their new patient visit to date of death, end of study follow-up (December 31, 2016), administrative censoring, or 12 months after the last completed HIV health care provider visit. CNICS uses the National Death Index to verify vital status and dates of death.

Patient Consent Statement

The Vanderbilt University Medical Center Institutional Review Board approved this study with a waiver of informed consent.

Study Definitions

We identified potential predictors at the individual, HIV care site, community, and structural levels based on a literature review and consultation with HIV care experts. In all candidate models, the outcome was missing the next scheduled HIV provider visit (no-show vs kept visit). All visits scheduled during study follow-up were included, and each patient could contribute multiple outcomes. Visits canceled ahead of time were excluded.

Individual-Level Data

Individual-level data included demographic characteristics, patient-reported outcomes (PROs), laboratory values, medical diagnoses, health insurance, and prior visit adherence. Demographic data included age, birth sex, present gender, race/ethnicity, HIV transmission risk factor, and laboratory data (CD4+ count and HIV-1 viral load [VL]). If both sexual and injection drug use (IDU) risk factors were present, risk was attributed to IDU given its stronger ability to predict visit adherence [11]. Baseline laboratory values were measured 180 days before and up to 14 days after the initial visit; missing laboratory values were carried forward until a new value became available. We included a VL indicator for each visit (0 = undetectable, 1 = detectable). To be pragmatic, VLs were defined as undetectable using the lower limit of quantification of the assay used at each site at the time of specimen reporting. For visits with a detectable VL, we included a continuous VL variable.

PROs assessed current tobacco, alcohol, and drug use as well as depression, quality of life, and symptom burden. Alcohol use variables included continuous Alcohol Use Disorders Identification Test (AUDIT-C) scores, high-risk alcohol use (AUDIT-C score ≥4 for men and ≥3 for women), and binge drinking (≥5 drinks in 1 sitting for men and ≥4 for women) [12]. Current drug use was captured using a modified World Health Organization (WHO) Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST) tool [13]. Depression, quality of life, and symptom burden were measured using continuous scores from the Patient Health Questionnaire (PHQ)–9 [14], EuroQOL Health-Related Quality of Life-5D [15], and HIV Symptom Index [16], respectively. We excluded visits with missing PRO data when PRO-derived predictors were included in the model.

Medical diagnoses in CNICS are either (1) verified via the electronic health record/adjudicated or (2) confirmed via laboratory results, medications, or objective measurements [17]. Both verified and confirmed diagnoses were included.

Time-updated health insurance type was categorized as private, public (Medicare/Medicaid), Ryan White HIV/AIDS Program (RWHAP), or uninsured (self-pay or unknown/missing). If >1 insurance type was documented, we used the following hierarchy for attribution: private, public, RWHAP, uninsured. We defined prior HIV visit adherence with 6 variables: 5 that require calculation from medical record data (number of scheduled visits before the visit of interest, time from study entry, number of missed visits before the visit of interest, proportion of missing visits before the visit of interest, rate of missing visits [number of prior missed visits divided by time from study entry]) and 1 easily obtained at the POC (an indicator for missing the last scheduled visit).

HIV Care Site–Level Data

We collected HIV care site–level data using a Research Electronic Data Capture (REDCap) [18] survey completed by site leadership (Supplementary Material). Site-level data were time-updated yearly and applied equally to each person receiving care at that site. HIV care site variables that did not differ across sites were not included as predictors of interest.

Community-Level Data

We obtained ZIP Code Tabulation Area (ZCTA)–level data from the US Census Bureau American Community Survey (ACS) [19]. The 2008–2012 and 2012-2016 5-year estimates were used from 2010–2011 and 2012–2016, respectively. All sites except 1 had 5-digit ZCTA datum available; the remaining site had 3-digit ZCTA data. We extracted ACS data on proportion of the ZCTA with less than a high school education, in the workforce but without current employment (unemployed), living below the federal poverty level, and of Black race. ACS data were merged with individual-level data by ZCTA of patient residence for each year between study entry and exit.

Structural-Level Data

We collected structural-level data on HIV criminalization laws [20], Medicaid expansion status under the Patient Protection and Affordable Care Act of 2010 [21], and proportion of each state's ADAP budget contributed by the state government of each patient's residence [22]. State of residence was determined using ZCTA data; visits with missing ZTCA data were assigned to the state of their HIV care site.

CD4+ count and VL data were missing for 6% and 8% of individuals at baseline. ZCTA data were missing for 8.8% of visits for all 4 ZCTA variables. Missing data were imputed using single imputation with predictive mean matching [23].

Statistical Analysis

An individual-level person-period data set was constructed to allow for time-varying data. Exposures were attributed based on a data structure in which individuals were nested within clinics, ZCTAs, and states. Model development and validation followed the Transparent Reporting of a Multivariable Prediction Model (TRIPOD) guidelines [24].

Candidate models with different sets of predictors were developed using random decision forests. Models were internally validated using 10-fold cross-validation. We randomly split our data set into 10 subsets, with 9 subsets pooled together to train the model and 1 subset reserved for testing; the area under the receiver operating characteristic (ROC) curve (AUC) was calculated using the reserved subset of the data. This process was repeated 10 times, and we selected the model with the largest AUC across the 10 replicates. This model was refit to the entire data set to create a final, validated predictive model. The discriminatory power of the final model was estimated using the average of the 10 AUCs obtained via cross-validation. Confidence intervals for AUCs were generated via bootstrap re-sampling with 200 replicates. By utilizing all available data, we a priori estimated a margin of error of <0.01 for our AUC estimates.

We developed a full model including all potential predictors as well as a simplified model including the most predictive variables identified in the full model that were feasible to obtain at the POC or in advance of the scheduled visit and were not redundant. Prediction models are not affected by simultaneous inclusion of correlated variables, so collinearity was not assessed [25]. As we were particularly interested in race/ethnicity because of prior associations with missed visits, we ordered all predictors based on variable importance in the full model and considered all variables through race/ethnicity for inclusion in the simplified model. Variable importance was determined using the Gini coefficient, a measure of how much the variable improves classification [26]. Analyses were completed in R, version 3.6.2. The statistical code is available at https://biostat.app.vumc.org/ArchivedAnalyses.

We developed a web-based calculator for the model, available at https://statcomp2.app.vumc.org/APP1/, which allows computation of an individual's predicted probability of a missed visit based on characteristic input. This calculator was developed using Shiny, a web application framework for R (http://shiny.rstudio.com). We utilized the R package "tidycensus" in order to populate US Census Bureau ACS data variables into our web-based calculator by inputting only the patient's ZCTA of residence.

RESULTS

Study Population

We included 20 807 PWH followed for a median of 3.8 years. The median age was 44 years, 81% were male, 2% were transgender, 37% were Black, 15% reported IDU as an HIV transmission risk factor, 57% reported male-to-male sexual contact, and 42% had public health insurance. The median baseline CD4+ count was 449 cells/mm^3; baseline VL was undetectable for 45% (Table 1). Additional medical diagnoses were identified among 18 812 (90.4%) (Supplementary Data).

There were 382 432 scheduled HIV health care provider visits not canceled ahead of time during the study period; 312 085 (82%) were kept, and 70 338 (18%) were missed. PRO data were available for 200 543 (64%) kept visits among 13 303 (64%) unique patients (Table 2). Person-visits were also characterized by clinical site-level variables, the average proportion of ZCTA-level properties, and state-level properties (Table 2).

Predictive Models

The AUCs for the full model including all predictor variables (200 543 visits) and the model excluding PRO variables (382 423 visits) were similar (AUC, 0.743 and 0.759, respectively). When PRO and prior adherence data were excluded, the AUC dropped to 0.709, and when only PRO data were included, the AUC dropped even further to 0.585 (Figure 1).

In the full model, the most important predictor of missed visits was a measure of previous visit adherence. In fact, a model including only prior adherence variables had an AUC of 0.710.

Table 1. Individual Demographic and Clinical Characteristics of the Study Population

Characteristic	n = 20 807
Age at baseline, median (IQR), y	44 (34–50)
Male birth sex, No. (%)	16 941 (81)
Transgender, No. (%)	372 (2)
Race/ethnicity, No. (%)	
Black, non-Hispanic	7691 (37)
Hispanic	2909 (14)
Other/unknown	1249 (6)
White, non-Hispanic	8958 (43)
HIV risk factor, No. (%)	
Heterosexual	5050 (24)
IDU	3186 (15)
MSM	11 841 (57)
Other/unknown	730 (4)
Baseline CD4+ count	
Median (IQR)	449 (264–652)
Missing, No. (%)	1275 (6)
Baseline HIV, copies/mL	
Median if detectable (IQR)	17 898 (1009–86 292)
Undetectable, No. (%)	9389 (45)
Missing, No. (%)	1612 (8)
Baseline insurance type, No. (%)	
Private	4705 (23)
Public	8674 (42)
Ryan White	3613 (17)
Uninsured/missing	3815 (18)
Follow-up duration, median (IQR), y	3.8 (1.6–6.4)
Follow-up visits, median (IQR)	15 (8–25)
Site of care, No. (%)	
Fenway Health/Harvard University	1678 (8)
John Hopkins University	2539 (12)
University of Alabama at Birmingham	4225 (20)
University of California San Diego	4297 (21)
University of California San Francisco	2898 (14)
University of North Carolina Chapel Hill	2136 (10)
University of Washington	3034 (15)

Abbreviations: IDU, injection drug use; IQR, interquartile range; MSM, men who have sex with men.

Among the top 23 important variables in the full model, 13 variables were included in the simplified model as they were deemed readily available at the POC or in advance of a scheduled visit and not redundant (Figure 2).

The AUC from the simplified model was 0.700 (n = 382 423). The ROC curves for the full and simplified models can be found in Figure 3. The vast majority (98.4%) of patient visits had ≤50% predicted probability of missing their next visit using the simplified model (Figure 4). A calibration curve comparing the predicted and observed probability for missing the next visit shows a good fit for the simplified model, with the predicted probability of missing the next visit being slightly higher than the observed probability for those most likely to miss a visit (Figure 5).

Table 2. Predictor Variables by Person Visit

Patient-Reported Outcomes (n = Total Visits With Available Data)	Median (IQR) or No. (%)
Smoking status (n = 241 069)	
Current smoker	95 698 (40)
Former smoker	61 343 (25)
Never smoker	84 028 (35)
Alcohol use	
AUDIT-C score (n = 235 396)	1 (0–3)
Binge drinking (n = 238 625)	72 103 (30)
High-risk alcohol use by AUDIT-C score (n = 235 396)	37 362 (16)
Current drug use	
Methamphetamines (n = 219 924)	26 186 (12)
Cocaine (n = 220 728)	21 712 (10)
Marijuana—regardless of local laws on use (n = 217 255)	71 810 (33)
Opiates—illicit and not taken as prescribed (n = 206 399)	8102 (4)
Any drug use (n = 224 860)	92 602 (41)
Depression score (n = 196 992)	5 (1–10)
Quality of life score (n = 211 859)	0.83 (0.76–1.00)
HIV Symptom Index Score (n = 204 844)	2 (0–6)
Clinic-Level Variables (n = 382 432)	Median (IQR) or No. (%)
Patients/ART prescriber per year	46 (29–91)
Trainees per year	6 (4–10)
Messaging on retention in care	
Posters	106 512 (28)
Brochures	100 194 (26)
Peer navigation	
HIV-positive	249 162 (65)
HIV-negative	175 342 (46)
Stigma support services	205 383 (54)
Financial assistance services	151 848 (40)
Appointment reminders	
Text	32 097 (8)
Personal phone call	118 915 (31)
Email	167 040 (44)
Flexible scheduling	275 911 (72)
Laboratory services	
Before appointment	55 628 (15)
Same day as appointment	258 360 (68)
ZCTA-Level Variables	Median (IQR) or No. (%)
Proportion of ZTCA with less than a high school education (n = 348 696), %	13 (8–19)
Proportion of ZCTA unemployed (n = 348 706), %	8 (6–10)
Proportion of ZTCA living below the FPL (n = 348 667), %	16 (12–24)
Proportion of ZTCA of Black race (n = 348 706), %	12 (4–22)
State-Level Variables	Median (IQR) or No. (%)
Proportion of state's ADAP budget contributed by the state government of each patient's residence (n = 362 484), %	12 (7–25)
Expansion of Medicaid in state of residence (n = 382 423)	121 160 (32)
Residence in a state with HIV criminalization laws (n = 382 423)	291 230 (76)

Abbreviations: ADAP, AIDS Drug Assistance Program; ART, antiretroviral therapy; AUDIT-C, Alcohol Use Disorders Identification Test; IQR, interquartile range; ZCTA, zip code tabulation area.

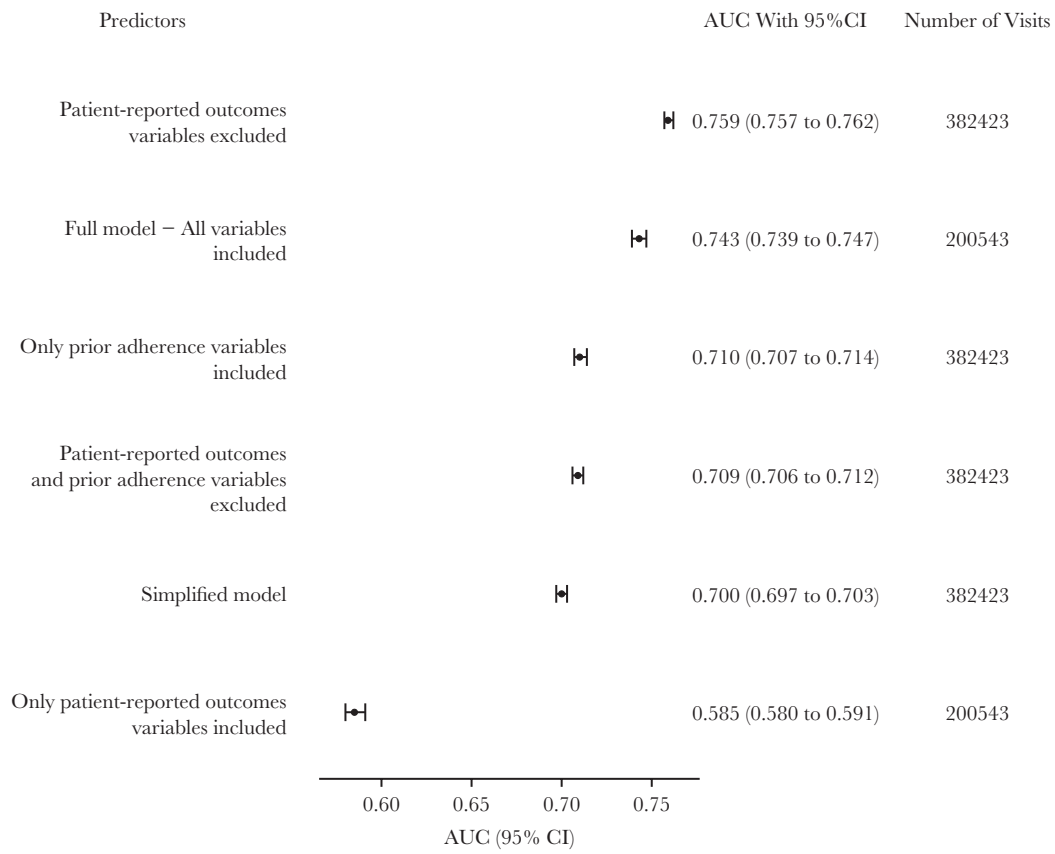

Figure 1. Discriminatory ability of candidate predictive models. Patient-reported outcomes data were only available for 200 543 of 382 423 (52%) visits. Abbreviation: AUC, area under the receiver operating characteristic curve.

DISCUSSION

We developed and internally validated a prediction model for missing HIV health care provider visits using individual-, HIV care site–, community-, and structural-level data collected in the CFAR Network of Integrated Clinical Systems (CNICS). Our full model had very good discriminatory ability between individuals who will miss vs attend their next HIV health care provider appointment (AUC, 0.743). Our simplified model, including only 13 variables readily accessible at POC or ahead of a visit from which we developed our web-based probability calculator, performed similarly (AUC, 0.700). In contrast to a previously published model [11], our model was developed and validated using random forest methods, incorporated multilevel data, and included data from over twice the number of patients and 3 times the number of HIV health care provider visits.

Similar to this previously published model, prior visit adherence data alone resulted in fairly high discriminatory power for identifying the individuals most likely to miss their next visit (AUC, 0.710). While some prior visit adherence measures are easily accessible (eg, last visit missed), others are difficult to obtain in the absence of an electronic health record that can provide these calculations (eg, proportion of previous visits missed). However, those difficult-to-obtain prior visit adherence measures were some of the strongest predictors of visit adherence. This highlights the importance of capturing accurate visit data and the potential utility of electronic health record tools that can quickly calculate variables from prior visits in order to correctly identify individuals at highest risk for missing their next visit.

Following previous visit adherence and selective individual-level predictors (age, CD4+ count, VL), 3 community-level characteristics (based on ZCTA) were highly predictive of missed visits. These findings are consistent with previous studies showing an association between community-level factors and RIC [27–29]. They also highlight the critical impact of contextual, structural factors in an individual's HIV health care provider visit attendance. While not readily modifiable, an individual's geographic place of residence can serve as a characteristic distinguishing them as someone who may benefit from a retention intervention, particularly when other highly predictive characteristics are present. Alternatively, these geographic areas may benefit from community health worker models, which have been utilized widely in low-resource settings but have also been shown to improve ART adherence in the United States [30].

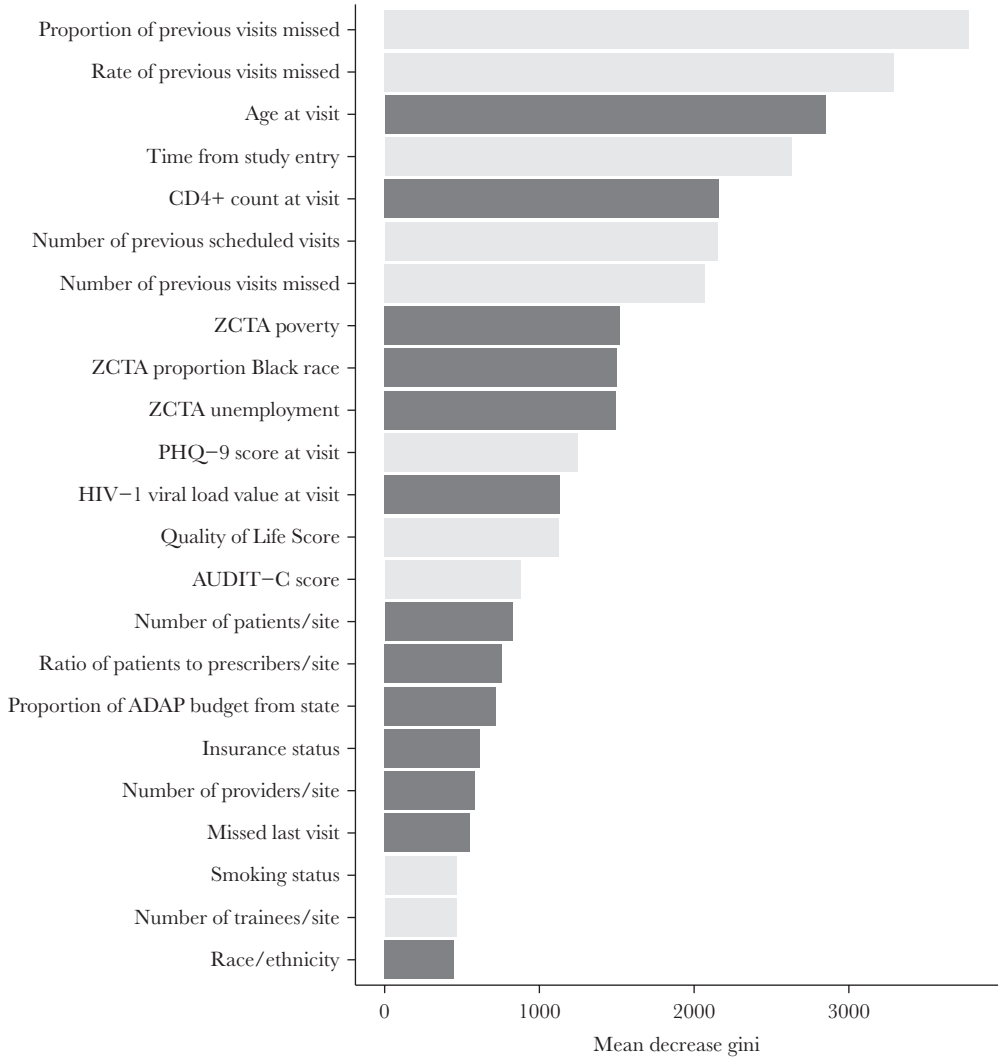

Figure 2. Random forest variable importance plot. Dark gray bars indicate the 13 predictors included in the simplified model from the top 23 most important predictors included in the full model. Abbreviations: ADAP, AIDS Drug Assistance Program; AUDIT-C, Alcohol Use Disorders Identification Test; PHQ-9, Patient Health Questionnaire-9; ZCTA, zip code tabulation area.

We hypothesized that patient-reported outcome (PRO) variables would add to our model's discriminatory performance. When only PRO data were included, the discriminatory power was not much better than chance alone (AUC, 0.585). During the study period, CNICS sites collected PRO data during in-person visits. In the context of the coronavirus disease 2019 (COVID-19) pandemic, CNICS has transitioned to asynchronous PRO data collection via web-based or mobile phone data capture. Therefore, future work can assess the impact of this change in PRO data collection methodology on their predictive ability. It is also possible that we did not collect data on the strongest PRO predictors of missed visits, such as housing stability, food insecurity, and transportation. CNICS has also begun collecting important social determinants of health data recently (eg, housing stability), allowing incorporation in future refinements of our prediction model.

While an individual's race/ethnicity has previously been identified as a strong predictor of missing HIV health care provider visits, this predictor was not among the top 20 most predictive in our full model. However, the racial composition of an individual's ZTCA of residence was a strong predictor of missed visits. This reflects the fact that race/ethnicity is a social construct [31]. An individual's race/ethnicity is likely a proxy for additional social determinants of health predictors (food insecurity, transportation, housing), and to the extent that these are correlated with ZTCA racial composition due to racial segregation, there is little additional predictive power of individual race. This also highlights the importance of measuring these important social determinants, modern-day manifestations of centuries-old structural racism [32].

Importantly, our model provided the discriminatory power to stratify a large number of HIV health care provider visits into

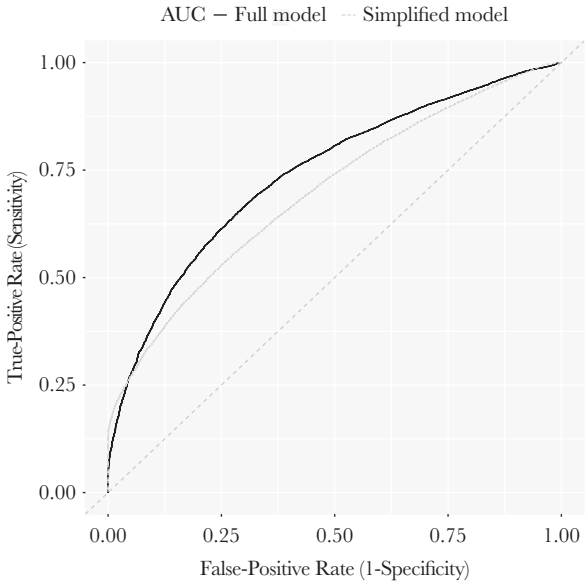

Figure 3. Receiver operating characteristic curves for the full model and simplified model. Abbreviation: AUC, area under the receiver operating characteristic curve.

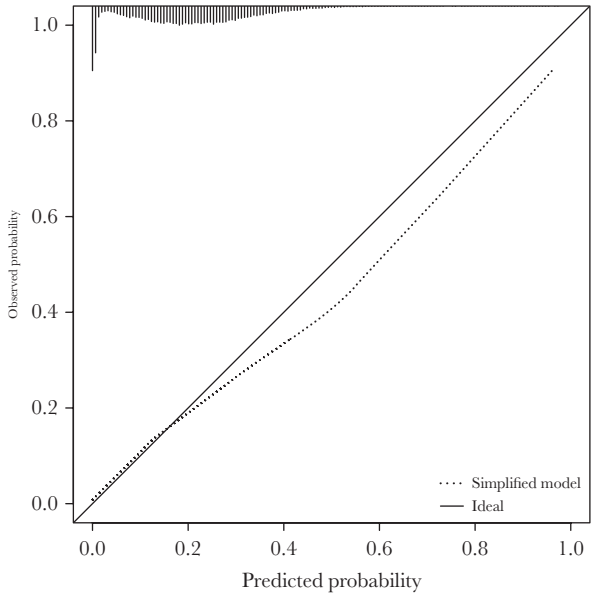

Figure 5. Calibration plot for simplified model.

smaller groups based on their predicted probability of missing the next visit. In fact, <2% of visits had >50% predicted probability for missing the next visit, allowing for potentially high-cost and high-resource-intensity interventions to be targeted to the smallest groups of patients who may benefit the most. Cost-effectiveness analyses based on specific interventions of interest will be needed to determine optimal economic cutoffs.

This model applies only to the prediction of missing in-person visits. With the COVID-19 pandemic, there has been a rapid rollout of telehealth. The 2020 Coronavirus Aid, Relief, and Economic Security (CARES) Act has allowed funds to be used by Health Resources and Service Administration (HRSA) RWHAP recipients for telehealth [33]. However, it is unknown if telehealth appointment adherence correlates with important

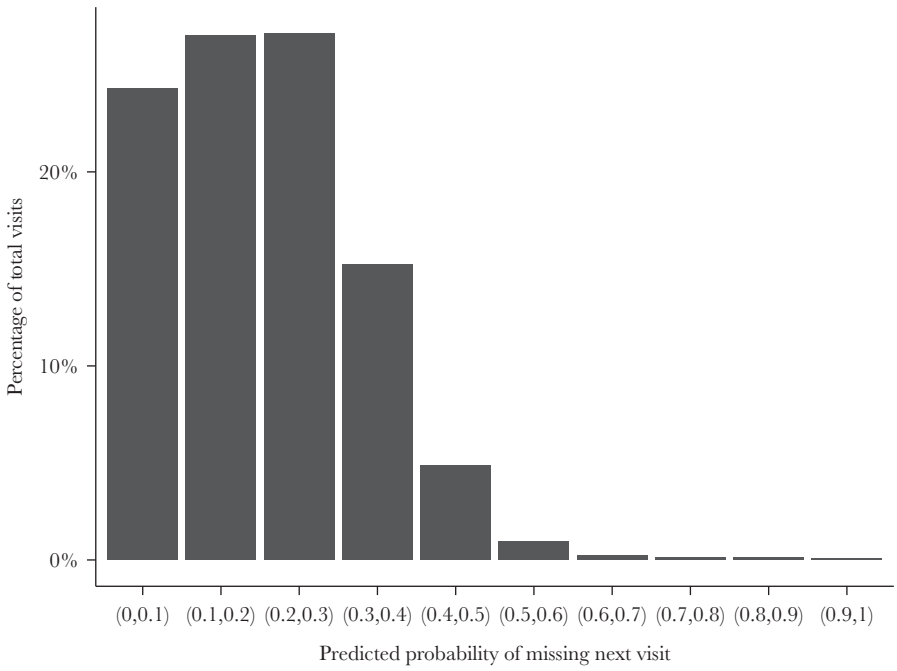

Figure 4. Proportion of visits by missed visit predicted probability deciles.

clinical outcomes, such as mortality, as has been shown for in-person appointments [7].

There were limitations to our study. First, the demographics of our study population differed slightly from those of PWH nationally [1]. Specifically, the White non-Hispanic population was overrepresented. Second, PRO data were only available for attended visits and focused on behavioral determinants (tobacco, alcohol, drug use) as opposed to social determinants of health (food insecurity, housing, transportation). Additionally, we did not have individual-level data on clinic-level resource access, so we applied the availability of the support resources equally to all persons in care at the site. Importantly, our model was developed and internally validated using data from a single US cohort of PWH who attended both a new patient visit and ≥1 follow-up visit, limiting its generalizability to other settings.

Our study also has several strengths. We used data from a large, geographically diverse US cohort of PWH, which allowed us to include a large number of predictor variables from multiple levels in our candidate models. We also used random forest methods, which do not assume linear relationships between variables, are flexible, and have excellent predictive performance [34]. Our simplified predictive model and web-based risk calculator allow for rapid, proactive assessment of an individual's risk of missing their next HIV health care provider visit in advance of a scheduled appointment. Therefore, the model could be integrated into routine clinical care in order to direct limited resources to those at the highest risk.

CONCLUSIONS

We developed a simple, point-of-care model, available via a web-based calculator with strong discriminatory power for predicting missed HIV care visits. Future refinements of this model should include data on additional important social determinants of health, external validation, and tailoring to varying clinical settings.

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. CNICS is an National Institutes of Health (NIH)–funded program (R24 AI067039) made possible by the National Institute of Allergy and Infectious Diseases (NIAID). The CFAR sites involved in CNICS include University of Alabama at Birmingham (P30 AI027767), University of Washington (P30 AI027757), University of California San Diego (P30 AI036214), University of California San Francisco (P30 AI027763), Johns Hopkins University (P30 AI094189, U01 DA036935), Fenway Health/Harvard (P30 AI060354), and University of North Carolina Chapel Hill (P30 AI50410). This work was also supported by NIH R01-MH113438 (Pettit), K01-AI131895 (Rebeiro), R01-AI093234 and R01-AI1311771 (Shepherd), the Tennessee Center for AIDS Research (P30 AI110527), Clinical Translational Science Award No. TL1TR002244 (Schember), and UL1TR000430 (REDCap) from the National Center for Advancing Translational Sciences.

Potential conflicts of interest. All authors report no conflicts of interest with respect to this work.

Author contributions. Study conception and design: A.C.P., P.F.R., B.E.S., M.M.; data collection: A.C.P., A.B., C.D.O., P.F.R., J.K., K.M., C.M., R.D.M., H.C., E.G., S.N., M.M.; data analysis: A.B., P.F.R., B.E.S.; data interpretation: A.C.P., A.B., C.D.O., P.F.R., J.K., K.M., C.M., R.D.M., H.C., E.G., S.N., B.E.S., M.M.; drafting of the initial manuscript: A.C.P., A.B., C.D.O., P.F.R., B.E.S.; critical review of the final draft of the manuscript: A.C.P., A.B., C.D.O., P.F.R., J.K., K.M., C.M., R.D.M., H.C., E.G., S.N., B.E.S., M.M.; access and verification of underlying data: A.C.P., A.B.

Prior presentation. Results were presented in part at the Conference on Retroviruses and Opportunistic Infections, Boston, Massachusetts, USA, March 8–11, 2020.

References

1. Centers for Disease Control and Prevention. HIV surveillance report, 2018 (updated); vol. 31. **2020**. Available at: http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html. Accessed 25 September 2020.
2. Fauci AS, Redfield RR, Sigounas G, et al. Ending the HIV epidemic: a plan for the United States. JAMA **2019**; 321:844–5.
3. Cohen MS, Chen YQ, McCauley M, et al; HPTN 052 Study Team. Prevention of HIV-1 infection with early antiretroviral therapy. N Engl J Med **2011**; 365:493–505.
4. Rebolledo P, Kourbatova E, Rothenberg R, Del Rio C. Factors associated with utilization of HAART amongst hard-to-reach HIV-infected individuals in Atlanta, Georgia. J AIDS HIV Res **2011**; 3:63–70.
5. Mugavero MJ, Lin HY, Allison JJ, et al. Racial disparities in HIV virologic failure: do missed visits matter? J Acquir Immune Defic Syndr **2009**; 50:100–8.
6. Berg MB, Safren SA, Mimiaga MJ, et al. Nonadherence to medical appointments is associated with increased plasma HIV RNA and decreased CD4 cell counts in a community-based HIV primary care clinic. AIDS Care **2005**; 17:902–7.
7. Mugavero MJ, Lin HY, Willig JH, et al. Missed visits and mortality among patients establishing initial outpatient HIV treatment. Clin Infect Dis **2009**; 48:248–56.
8. Bulsara SM, Wainberg ML, Newton-John TRO. Predictors of adult retention in HIV care: a systematic review. AIDS Behav **2018**; 22:752–64.
9. 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.
10. Kitahata MM, Rodriguez B, Haubrich R, et al. Cohort profile: the Centers for AIDS Research Network of Integrated Clinical Systems. Int J Epidemiol **2008**; 37:948–55.
11. Pence BW, Bengtson AM, Boswell S, et al. Who will show? Predicting missed visits among patients in routine HIV primary care in the United States. AIDS Behav **2019**; 23:418–26.
12. Bradley KA, DeBenedetti AF, Volk RJ, et al. AUDIT-C as a brief screen for alcohol misuse in primary care. Alcohol Clin Exp Res **2007**; 31:1208–17.
13. Humeniuk Rachel, Henry-Edwards S, Ali Robert, et al. *The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): Manual for Use in Primary Care*. Geneva: World Health Organization; **2010**.
14. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med **2001**; 16:606–13.
15. Delate T, Coons SJ. The use of 2 health-related quality-of-life measures in a sample of persons infected with human immunodeficiency virus. Clin Infect Dis **2001**; 32:E47–52.
16. Justice AC, Holmes W, Gifford AL, et al; Adult AIDS Clinical Trials Unit Outcomes Committee. Development and validation of a self-completed HIV symptom index. J Clin Epidemiol **2001**; 54(Suppl 1):S77–90.
17. CFAR Network of Integrated Clinical Systems (CNICS). CNICS data elements. Available at: https://sites.uab.edu/cnics/cnics-data-elements/. Accessed 4 November 2020.
18. Harris PA, Taylor R, Thielke R, et al. Research Electronic Data Capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform **2009**; 42:377–81.
19. US Census Bureau. A Compass for Understanding and Using American Community Survey Data: What Researchers Need to Know. Washington, DC: US Government Printing Office; **2009**.
20. Horvath KJ, Meyer C, Rosser BR. Men who have sex with men who believe that their state has a HIV criminal law report higher condomless anal sex than those who are unsure of the law in their state. AIDS Behav **2017**; 21:51–8.

21. National Academy for State Health Policy. Where states stand on Medicaid expansion. Available at: https://www.nashp.org/states-stand-medicaid-expansion-decisions/. Accessed 4 November 2020.
22. National Alliance of State and Territorial AIDS Directors (NASTAD). National Ryan White HIV/AIDS Program (RWHAP) Part B and ADAP Monitoring Project annual reports. Available at: https://www.nastad.org/PartBADAPreport. Accessed 4 November 2020.
23. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. Stat Med **2011**; 30: 377–99.
24. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med **2015**; 162:55–63.
25. Kutner M, Nachtsheim C, Neter J. Applied Linear Statistical Models. 4th ed. New York, NY: McGraw-Hill; **2004**.
26. Dietterich TG. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. Machine Learning **2000**; 40:139–57.
27. Eberhart MG, Yehia BR, Hillier A, et al. Individual and community factors associated with geographic clusters of poor HIV care retention and poor viral suppression. J Acquir Immune Defic Syndr **2015**; 69(Suppl 1):S37–43.
28. Nelson JA, Kinder A, Johnson AS, et al. Differences in selected HIV care continuum outcomes among people residing in rural, urban, and metropolitan areas—28 US jurisdictions. J Rural Health **2018**; 34:63–70.
29. Rebeiro PF, Gange SJ, Horberg MA, et al; North American AIDS Cohort Collaboration on Research and Design (NA-ACCORD). Geographic variations in retention in care among HIV-infected adults in the United States. PLoS One **2016**; 11:e0146119.
30. Kenya S, Chida N, Symes S, Shor-Posner G. Can community health workers improve adherence to highly active antiretroviral therapy in the USA? A review of the literature. HIV Med **2011**; 12:525–34.
31. Freeman HP. The meaning of race in science—considerations for cancer research: concerns of special populations in the National Cancer Program. Cancer **1998**; 82:219–25.
32. Howe CJ, Dulin-Keita A, Cole SR, et al; CFAR Network of Integrated Clinical Systems. Evaluating the population impact on racial/ethnic disparities in HIV in adulthood of intervening on specific targets: a conceptual and methodological framework. Am J Epidemiol **2018**; 187:316–25.
33. United States Health Resources and Services Administration. FY 2020 CARES Act funding for Ryan White HIV/AIDS Program recipients. Available at: https://hab.hrsa.gov/program-grants-management/coronavirus-covid-19-response. Accessed 4 November 2020.
34. Couronné R, Probst P, Boulesteix AL. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics **2018**; 19:270.