2023

Navigating Missed Visits in HIV Primary Care: Exploring Risk Factors Associated with Missed Visits

Jiaying Hao
University Of Alabama At Birmingham

NAVIGATING MISSED VISITS IN HIV PRIMARY CARE: EXPLORING RISK FACTORS ASSOCIATED WITH MISSED VISITS

by

JIAYING HAO

DUSTIN LONG, CO-CHAIR
STACEY COFIELD, CO-CHAIR
HENNA BUDHWANI
MICHAEL MUGAVERO
JEFF SZYCHOWSKI

A DISSERTATION

Submitted to the graduate faculty of The University of Alabama at Birmingham,
in partial fulfillment of the requirements for the degree of
Doctor of Public Health

BIRMINGHAM, ALABAMA

2023

NAVIGATING MISSED VISITS IN HIV PRIMARY CARE: EXPLORING RISK FACTORS ASSOCIATED WITH MISSED VISITS

JIAYING HAO

BIOSTATISTICS

ABSTRACT

The HIV care continuum is a public health model that outlines the steps or stages people with HIV go through from diagnosis to achieving and maintaining viral suppression. Retention in care, a key component of the HIV care continuum, is critical in achieving good health outcomes for people with HIV (PWH) and preventing HIV transmission. Consistent HIV care is essential for the health of PWH as a part of retention in care. Despite no standard measure for retention in care to PWH, the missed (no-show) visits/appointments is one of the most commonly used methods for evaluating retention in HIV care. Reducing the risk of missed visits directly contributes to improving retention in care and overall health for PWH.

This dissertation aims to investigate the risk factors associated with missed primary care visits among PWH. The first paper presents a systematic review summarizing reported risk factors for missed HIV visits in studies conducted between 2000 and 2022. The second paper focuses on youth with HIV (YWH), a vulnerable subpopulation of PWH, exploring the specific risk factors associated with missed visits in this group. The final paper develops a predictive model using machine learning algorithms for new-to-care patients, leveraging advanced statistical techniques to predict missed visits with limited patient information.

Through a combination of methodologies, this dissertation provides a comprehensive understanding of the risk factors associated with missed visits among PWH. The

findings contribute to the existing knowledge base and offer insights into improving retention in HIV care. By identifying and addressing the factors influencing missed visits, healthcare providers and policymakers can develop targeted interventions to enhance retention in care and ultimately improve the health outcomes of PWH.

Keywords: HIV, missed visits, retention in care, risk factors, machine learning

DEDICATION

I dedicate this dissertation to you, my beloved husband and my wonderful parents. Without your ceaseless support, unending love, and unwavering belief in my abilities, this journey would not have been possible. Thank you for being my pillars of strength and my sources of inspiration. This work, this achievement, is a testament to your love and belief in me.

May this dissertation not only be a symbol of my academic achievement but also a testament to the beautiful power of support, love, and belief that you have so generously given me.

ACKNOWLEDGMENTS

The journey of a thousand miles begins with a single step, and in my academic voyage, I have been fortunate to be guided, mentored, and supported by some truly extraordinary individuals.

I express my deepest gratitude to my advisor and co-chair, Dr. Dustin Long. His encouragement in times of doubt, constructive feedback during my research, and unwavering belief in my abilities have enriched my understanding and growth in immeasurable ways.

My sincere thanks go to Dr. Henna Budhwani, who opened doors of opportunity for collaboration and offered her expert guidance as a member of my committee.

I would like to thank my committee members for their guidance and support: Dr. Stacey Cofield, Dr. Jeff Szychowski, and Dr. Michael Mugavero for their wonderful collaboration.

I express my heartfelt gratitude to the Department of Biostatistics. Spending nearly 20% of my life here has been a journey of learning, discovery, and personal growth. The department has provided an environment of intellectual stimulation and a culture of academic rigor that has been indispensable for my research work.

To my fellow students and friends, Rouba Chahine, Jinhong Cui, Li Zhang, Rachel Stuckwisch, Andrew Sims, Lingling Wang, who have been part of this journey, I extend my sincere thanks. Your companionship, shared insights, and mutual encouragement have made the path to this accomplishment less daunting. You have been

an integral part of my academic journey, and your camaraderie and support have been invaluable.

TABLE OF CONTENTS

Page

LIST OF TABLES

Table *Page*

RISK FACTORS FOR MISSED PRIMARY CARE CLINIC VISITS IN PATIENTS WITH HIV: A SYSTEMATIC REVIEW OF US STUDIES 2000-2022

THE IMPACT OF MISSED SCHEDULED VISITS ON HEALTH OUTCOMES AMONG YOUTH WITH HIV IN ALABAMA

PREDICTIVE ANALYTICS FOR MISSED HIV PRIMARY CARE VISITS WITH IENGAGE STUDY USING MACHINE LEARNING METHODS

LIST OF FIGURES

Figures *Page*

RISK FACTORS FOR MISSED PRIMARY CARE CLINIC VISITS IN PATIENTS WITH HIV: A SYSTEMATIC REVIEW OF US STUDIES 2000-2022

PREDICTIVE ANALYTICS FOR MISSED HIV PRIMARY CARE VISITS WITH IENGAGE STUDY USING MACHINE LEARNING METHODS

INTRODUCTION

Background

Approximately 37.7 million people were reported with HIV in 2020, globally. Of these, 36 million were adults and 1.7 million were children aged 0-14 years[1]. More than half (53%) were women and girls. At the end of 2019, an estimated 1,189,700 people aged 13 and older were living with HIV in the United States, including an estimated 158,500 (13%) people whose infections had not been diagnosed[2]. Of diagnosed people living with HIV (PLWH), 81.3% were linked to care within one month of diagnosis, and 65.5% of all diagnosed PLWH are virally suppressed (HIV viral load <200 copies/mL). The annual rate of diagnoses of HIV infection is 11.1 per 100,000 people[2]. According to the HIV surveillance supplemental report 2021, Gay, bisexual and other men who have sex with men (MSM) accounted for 69.0% of new diagnoses; Blacks/African Americans and Hispanics/Latinx represented 44.0% and 30.0% of new diagnoses, respectively; the number of new diagnoses was highest among people aged 25 to 29, followed by the group of aged 13-24 years old[3].

The state of Alabama had 635 newly diagnosed HIV infections, with a total of 14,345 PLWH in 2019. The highest rates of new diagnoses appeared in young adults aged 20-29 years, followed by age 30-39 and 13-19 years. As the age at diagnosis continues to lower, the need for increased prevention efforts prioritizing the younger population increases[4].

Clinicians providing HIV medical care and their clinic administrators face challenges when considering how best to evaluate retention in HIV medical care for their specific clinic population. Retention in care is defined as a patient's regular engagement with medical care at a health care facility after initial entry into the system. More formal definitions and measurements have utilized required follow-up at certain intervals to define retention in care; these definitions have typically conceptualized retention in care based on either appointments missed or medical visits attended at regularly defined intervals[5]. By 2019, only 50.1% of PLWH in the US were retained in care.

Currently, multiple methods exist for estimating retention in care: For reporting purposes, the Centers for Disease Control and Prevention (CDC) defines retention in HIV medical care as documentation of at least two laboratory tests (CD4 cell counts or viral load tests) performed at least 3 months apart during the year of evaluation[6]. The CDC also refers to this as continuous HIV medical care[6]. Health Resources and Services Administration (HRSA) has defined retention in HIV medical care as: persons with diagnosed HIV who had at least two attended medical visit that were at least 90 days apart in the measurement year.[7,8] The most recent publication of 2017 summary data from Ryan White HIV/AIDS Program modified this definition and defined retention in care as persons with diagnosed HIV who attended at least one outpatient ambulatory health services visit by September 1st of the measurement year, with a second visit at least 90 days after.[8] The Institute of Medicine (IOM) has defined retention in HIV medical care as at least two medical visits every 12 months, with a minimum of 90 days between visits.[9] These indicators have largely been used for surveillance and quality reporting purposes

and do not incorporate all scheduled HIV medical care visits, nor incorporate the nuance around the time interval and disposition of visits.

In addition to the methods and criteria listed above for evaluating retention in HIV medical care, missed visits and appointment adherence are typically used as a measure of retention in care for research and clinic-level quality improvement purposes[10]. The purpose of the missed visit measure is to capture the number of missed appointments (no-shows) during an observation period. This quantity is easy to measure, summarize, and analyze, making it a commonly used method for evaluating retention in HIV medical care[11]. Research shows that 61% of new HIV cases are transmitted from PLWH who are not currently engaged in medical care[12]. By finding ways to identify patients at risk for missing future visits and dedicating special efforts to keep them in care, researchers say providers can prevent harmful individual health outcomes for PLWH and lower the likelihood of future virus transmission.[12] Therefore, this study will be focusing on missed HIV primary visits.

Public Health Relevance

The HIV Care Continuum is a public health model that outlines the steps or stages people with HIV go through from diagnosis to achieving and maintaining viral suppression (a very low or undetectable amount of HIV in the blood). The steps include diagnosis of HIV infection, linkage, receipt and retention in HIV medical care, and achievement and maintenance of viral suppression[13]. PLWH who receive regular medical care according to the definition defined by CDC, HRSA or the IOM are considered 'retained in care.'[14] Consistent medical care is essential for the health of

people living with HIV. PLWH who receive regular medical care are more likely to receive antiretroviral therapy, less likely to develop Acquired Immune Deficiency Syndrome (AIDS), and have improved survival rates compared to HIV-positive individuals who do not receive regular medical care[15-17]. According to the CDC's report published in May 2021, in 2019, approximately 1.2 million people in the U.S. were living with HIV[18]. An estimated 87% were diagnosed with HIV with only 81% of those diagnosed linked to care. Approximately 65.9% received some HIV care, 50.1% were retained in care, and 56.8% were virally suppressed[18]. Barriers to retention in care include sociodemographic factors, mental illness, substance use, poverty, stigma, as well as lack of access to healthcare[19-28].

Despite the importance of retention in care to PLWH, there is no recognized gold standard measure with many different indicators linked to health outcomes [29]. A common quality indicator for retention in care is missed clinic visits, which are measured as no-show or missed visits that were not canceled in advance of scheduled appointments[29]. Reducing the risk of missed visits would directly lead to improving retention in care. This study has the potential to provide a more comprehensive understanding of missed HIV visits, provide strategies to measure and predict the HIV retention in care, thereby improving the treatment outcome and overall health status of PLWH. Enhancing the retention in care can lead to increased viral suppression rates, which, in turn, will reduce the transmission of HIV. This not only elevates public health outcomes overall but also lessens the strain on the public health system, both in terms of financial cost and systematic burden.

Research Aims

This dissertation aims to explore and reveal the risk factors associated with HIV missed visits. I conducted a systematic review to have a better understanding of research focused on missed visits, risk factors that have been established, and the relationship between the risk factors and missed visits. Missed visits were determined to be the next study aim. The youth living with HIV (YLWH) were underrepresented in the results of Aim 1, thus I examined factors associated with YLWH missing visits at UAB's family clinic. Finally, a machine learning algorithm will be applied to generate a predictive model using data from newly diagnosed PLWH enrolled in an intervention study focused on engaged HIV care.

Objectives

1: To identify risk factors associated with missed HIV clinic visits in studies published between 2000 to 2022 by a systematic review.

2: To determine the risk factors associated with missed HIV clinic visits among Youth living with HIV in the UAB Family clinic.

3: To develop a model for predicting missed HIV visits among new patients in the Integrating ENGagement and Adherence Goals upon Entry (iENGAGE) study using machine learning algorithms, including the accelerated oblique random survival forest (aorsf) technique.

RISK FACTORS FOR MISSED PRIMARY CARE CLINIC VISITS IN PATIENTS WITH HIV: A SYSTEMATIC REVIEW OF US STUDIES 2000-2022

by

JIAYING HAO, JINHONG CUI, KATHRYN KAISER, DUSTIN M. LONG

In preparation for *BMC Infectious Diseases*

Format adapted for dissertation

Introduction

According to the HIV Surveillance report 2022, an estimated 1.2 million people in the U.S. have HIV, and 30,635 people were newly diagnosis in 2020[1]. Although antiretroviral therapy has substantially reduced HIV-related morbidity and mortality[2], full benefit of treatment depends on consistent access to HIV care[3].

Continuous HIV medical care is essential for improving health outcomes for people living with HIV and reducing the spread of HIV[4, 5]. It is a critical component of the HIV care continuum and an essential goal for public health efforts to end the HIV epidemic.

Missing a scheduled HIV care appointment without prior notification to the healthcare provider is an important measure of HIV retention in care[6] because they provide insight into the level of engagement of patients in their HIV care. Missed visits may indicate inconsistency in engagement in HIV-related medical care[7], which could lead to poor health outcomes, such as disease progression, increased morbidity, and mortality[8-13]. Furthermore, missed HIV visits are a particularly important indicator as they can occur at any point along the HIV care continuum. Patients who miss a scheduled medical appointment may not receive necessary laboratory tests, medication adjustments, prescription refills, or other interventions that are essential for effective HIV treatment[14].

Studying the risk factors of missed HIV visits is crucial for improving HIV care retention and ultimately reducing the transmission of HIV. By identifying the factors that contribute to missed visits, healthcare providers and researchers can develop targeted interventions to address these risk factors and improve retention in care[15]. Additionally,

healthcare providers can use these data to monitor and assess patient engagement in HIV care, identify patients who may be at risk of discontinuing care and provide appropriate interventions to improve retention.

There is a growing body of research focused on understanding the risk factors associated with missed HIV visits and developing interventions to improve HIV care retention. Many studies have identified demographic, clinical, behavioral, and social factors that contribute to missed visits, including younger age, substance use, depression, and lack of social support[16-19]. Moreover, there is increasing recognition of the importance of missed HIV visits as a critical measure of HIV care retention. The use of missed visits as an outcome measure in HIV care research has increased recently, and many studies are now incorporating missed visit data to assess patient engagement and develop targeted interventions to improve retention[20].

However, there is a gap between the current study of risk factors for missed HIV visits and the actual needs of people living with HIV (PLWH). Many studies focus on specific risk factors, such as socio-economic status, substance abuse, or mental health issues. As a result, they may not provide a comprehensive understanding of all factors that contribute to missed HIV visits. Also, PLWH come from various backgrounds, cultures, and regions. Research findings from one population may not necessarily apply to others, making it challenging to generalize results and provide tailored interventions. Additionally, many studies focus on specific time points or short periods, which may not capture the complex and dynamic nature of the risk factors for missed HIV visits. While the topic of retention in care has been extensively studied and reviewed [21], there lacks a synthesis of literature to date to assist researchers and clinicians to summarize and

understand the key risk factors of missed visits. This review's aim is to identify, evaluate studies reporting risk factors of missed HIV visits, and summarize the findings based on selected studies.

Methods

This review is reported using PRISMA reporting guidelines[22] and follows many of the recommendations in the Cochrane Handbook[23]. The protocol has been registered in PROSPERO, the International Prospective Register of Systematic Review (registration number: CRD42022309761). The review question was "what are the risk factors of missed HIV primary care visits among patients in the United States?" To explore the relationship between the risk factors and missed visits, a PEO question framework[24] was used:

Population: All aged HIV patients who initiated HIV care.

Exposure: Risk factors associated with missed visits.

Outcome: Missed HIV primary (routine, clinic) visits. The missed visit is defined as missing (no-show) a given scheduled HIV primary care appointment or being rescheduled for 30 days later. Lost to follow-up is not defined as a missed visit.

Criteria for Inclusion and Exclusion:

This review aimed to explore missed HIV visits in the United States, therefore only longitudinal studies focused on HIV routine visits (for example HIV primary care visits, HIV clinic visits, HIV outpatient visits) conducted in the United States among PLWH of US residency, published between 2000 and 2022 were included. Any study not

focused on the US PLWH; without using missed HIV visits as the study outcome; did not report risk factors associated with missed visits; or reporting cross-sectional analyses or less than a 6-month observational period were excluded. Posters or conference abstracts were not included.

Data Sources and Search Strategies:

To identify studies, four bibliographic search platforms were used: PubMed, Embase, Scopus and the Cochrane library. Example search terms and steps are summarized below:

#1 Search: “HIV” in Title/Abstract,

#2 Search: “missed visit” OR “missed visits” OR “attendance” OR “appointment” OR “appointments” in Title/Abstract,

#3 Search: “risk factors” OR “risk factors” OR “predictive model” OR “prediction” in Full text,

#4 Search: #1 and #2 and #3.

The results were exported to EndNote reference management software.

Screening and selection:

After combining results from four databases, duplicates were removed and two reviewers (JH and JC) reviewed articles by title and abstract independently, determining if the article is eligible for full-text reviewing based on the inclusion and exclusion criteria. Full-text articles were reviewed to confirm eligibility. If disagreement occurred, articles were discussed until consensus was reached.

Data Extraction:

All studies that meet our inclusion criteria were listed in Table 1. A standardized data collection form was be used for data extraction. For each study, data were extracted on first author, year of publication, journal name, study population, study location, sample size, study/observational period, definition of missed visits, statistical method, proportion of missed visits, and definitions of risk factors.

Quality Assessment:

The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomized studies in meta-analyses for observational studies was used for quality assessment [25].

Data Synthesis:

Table 1 summarizes the information on studies included in this review. Table 2 shows the frequency of each reported risk factor and the study to in which it was reported. Missed visits proportion/rate, relative risk, hazard ratios, odds ratios, and coefficients were compared to check the consistency of the effect of each risk factor. Due to the heterogeneity of studies, a meta-analysis was not performed and narrative synthesis of the included papers was conducted. The study selection process is described in the PRISMA flowchart (Figure 1).

Ethics Statement:

This work did not require an ethics review as it is using secondary data from published studies.

Results

A total of 1,473 studies were identified through PubMed, Embase, Scopus database, and the Cochrane Library, of which, 358 duplicates and 86 ineligible by published years were excluded. 862 records were excluded by screening the title and abstract. Nineteen studies were excluded due to full-text being inaccessible. Among the remaining 148 studies, five did not focus on PLWH populations, 55 did not study missed HIV care visits, 10 did not use missed visits as the study outcome, 14 did not report risk factors, 29 were international studies with non-US populations, and three studies had observational period less than 6 months. Consequently, we ended up with 32 studies that satisfied all the eligibility criteria (Figure 1).

Quality and Characteristics of Included Studies:

Table 1 presents a systematic summary of all the selected studies. Two articles were published in 2009 ([5,47]), six were published between 2011-2015([35,48, 3, 11, 49, 50]), 16 were published between 2016-2020 ([51,26,52,37,31,44,53,36,54,27,29,30, 55, 56, 57, 58]), and eight were published in 2021 and 2022 ([59, 60, 61, 62, 63, 28, 43, 64]). Among 32 studies, three included patients under 18 years old ([26,37, 36]), four focused on Men have sex with Men (MSM) ([3,49,52,59]), six focused on female patients ([48, 31, 44,55, 57,60]), and one focused on transgender patients ([37]).

Different measures were found in these studies: 22 studies categorized the outcome as any missed visits using yes versus no (missed vs. not missed). Nine studies ([49,50, 37, 31, 44, 36, 54, 27, 28]) categorized missed visits using self-defined cut-offs or multiple groups. Four studies ([5, 35, 56, 58]) calculated ratios or rates of missed visits or attended visits

as the outcome variable. One study ([56]) used number of missed visits as a continuous outcome. Three studies ([56, 28, 43]) used more than one method to measure the outcome. Multiple statistical methods have been applied among the included studies. The majority of the studies (29, 91%) chose logistic regression or multinomial logistic regression, with odds ratios/risk ratios reported in the results section (Table 1). Twenty-eight studies reported the percentage of missed visits based on the sample of participants, the range variates from 15% to 90%. One study ([53]) reported the person-year of missed visits.

Table 2 presents the quality assessment results using the NOS tool. The tool has eight questions based on three domains: Selection, comparability, and outcome. Based on the Thresholds for converting the Newcastle-Ottawa scales to AHRQ standards[25], 18 studies were rated as good quality, 11 were fair quality and two were poor quality. The main reasons for not being rated as good quality are: 1. Samples are not likely to be representative of the general population of PLWH; 2. Study used a retrospective dataset, which means the outcome was established at the start of the study; 3. Study used self-report records instead of medical records to confirm the outcome.

Risk Factors Associated with Missed Visits:

In eligible studies, 45 different risk factors were identified to have significant associations with missed HIV visits. The risk factors cover aspects from demographic, socioeconomic, behavioral, health-related (HIV-related, clinic, psychiatric and other – see Table 3). The ten most common risk factors reported are age, race/ethnicity, substance use, HIV risk factor (known as HIV transmission mode), gender, income, education, insurance type, depression, and experiences of stigma.

Demographic Risk Factors:

Demographic risk factors, age, race/ethnicity, gender, education, gender identity (heterosexual), clinic site, youth HIV prevalence, partnership, and language have been demonstrated significant associations with missed HIV visits. The effects of the following risk factors are quite consistent: 15 studies found that being white race or non-Hispanic was associated with better visit adherence compared to other races or ethnicities; 8 studies reported that being male was associated with better visit adherence compared to female or patients classified as gender minorities; being homosexual, living in an area with youth HIV prevalence lower than 13%, having partnership/married, and being a Spanish speaker are more likely to have better visit adherence. Twenty studies pointed out that age is a risk factor associated with missed visits, and most of them (18, 90%) have reflect that the younger age groups are more likely to have missed visits. However, Kahana et al.[26], Kay et al.[27] and Lesko et al.[28] presented some different results contradicting the overall trend. In Kahana et al.[26], the likelihood of missed appointments was greater among older youth (OR: 1.85, CI: 1.29-2.66, $p<0.001$). In Kay et al.[27], the bivariate analysis comparing missing 3+ versus 0 visits indicated that older age is more likely to have 3+ missed visits (RR: 1.04, CI: 1.02-1.06, p-value<0.001). This is not consistent with the result in their multivariable analysis (RR: 0.96, CI: 0.94-0.97, $p<0.001$) and results of other studies included in this review. Also, Lesko et al.[28] used different measures defining the study outcome: attended >=75% of scheduled visits, not retained by kept-visit measure, and retained by kept-visit measure, missed >25% of scheduled visits. The directions of the associations between age groups 18-34 years versus 35-49 years and >=50 years versus 35-49 years are incompatible among the two

measures: attended>=75% measurement indicated older age groups are more likely to maintain scheduled visits (OR: 1.20, CI: 1.12-1.30; OR: 0.75, CI: 0.71-0.80, respectively). The missed >25% cutoff showed a trend that older age groups are more likely to miss visits (OR: 1.41. CI:1.31-1.50; OR: 0.65, CI: 0.62-0.69, respectively). Inconsistences were also found in education. Six out of seven studies pointed out that lower education levels have higher risk of missing HIV visits, but Kay et al.[27] again had a differing result in the bivariate model: 3+ versus 0 missed visits when comparing some college and >=college to < high school (RR: 3.15, CI: 1.31-7.57; RR: 4.73, CI: 1.38-16.20, respectively).

Socioeconomic Risk Factors:

Among 32 studies, income, insurance type/status, social support, unemployment or disabled, homeless, health literacy, number of moving, HIV-knowledge, Ryan White HIV/AIDS Program (RWHAP) support service, and need transportation help have been proven with significant associations with missed visits. The insurance type/status didn't show a consistent relationship with the likelihood of missing HIV visits among seven studies: Again, Kay et al.[27], had conflicting results with the trend in bivariate, 3+ versus 0 missed model. The private insurance type were almost 3 times the risk to have 3+ missed visits compared with uninsured patients (RR: 2.74, CI: 1.27-5.91, p=0.01). Other studies all reported that patients with private-paid insurance have higher visit adherence, and patients who have insurance coverage have higher visit adherence. Other than this, patients with lower income/below FPL, unemployment or disabled, not or less receiving social support, with unstable housing status, frequently moved, have poor health literacy

or HIV knowledge, and need help in transportation were reported with higher risk of missing HIV visits.

Behavioral Risk Factors:

Substance use was reported in 10 articles with the result that any type of substance use is associated with higher risk of missed visits. Four articles reported that alcohol use is associated with higher risk of missed visits. Involvement with the legal system, unprotected sex practices, and low appointment expectancy were also found to be associated with higher risk of missed visits.

Health-Related Risk Factors:

Health-related risk factors include four components: HIV-related risk factors, clinical risk factors, psychiatric risk factors and other health-related risk factors.

HIV risk factors/HIV transmission mode have been reported in nine articles. Among these studies, MSM has higher visit adherence compared to heterosexual and patients who were injection drug users (IDU), while the latter has the lowest visit adherence. Year of treatment enrollment is also an important risk factor which was reported in five articles and had a conclusion that longer time of enrollment in HIV care is associated with higher risk of missing visits. Furthermore, PLWH who had missed visits before, larger number of scheduled visits, not on ART, prior AIDS diagnosis, and had HIV disclosure are also associated with lower visit adherence.

Viral load record, baseline CD4 and most recent CD4 are found to be associated with visit adherence. Mannes et al.[29] and Lesko et al.[28] both used the concept of durable viral suppression, defined as all HIV viral suppressed in the study period. Pence et al.[30]

used the most recent VL record, and Safo et al.[31] didn't give a clear description of which record they used. Lesko et al.[28] had differing results with different outcome measures. This study tested both durable viral suppression and most recent viral suppression, in the attended>=75% outcome, viral suppressed patients were less likely to retain in care in both viral suppression outcome (OR:0.79, CI:0.73-0.85; OR: 0.97, CI: 0.96-0.98, respectively). Regardless of this discrepancy, all other studies had the same conclusion that patients with lower VL were more likely to attend scheduled visits. Both baseline CD4 and most recent CD4 records are important: patients with lower CD4 counts are more likely to have higher visit adherence.

Psychiatric factors such as depression, stigma, anxiety, mental health conditions were found to be associated with lower visit adherence. Other health-related factors such as experienced Intimate Partner Violence (IPV), ER visit, cognitive function issue, life chaos, gender affirmation and poor general health are all found to be associated with lower visit adherence[5].

Discussion

The 32 selected studies in this systematic review highlight the complexity of the issues surrounding the topic of HIV retention in care measured by missed visits. This systematic review collated and synthesized findings from these 32 studies, providing a more comprehensive synthesis of the available evidence on factors associated with missed HIV visits in PLWH in the United States by identifying consistent patterns and significant associations across studies.

This review indicates that demographic and socioeconomic risk factors, such as age, gender, race, education level, income level, insurance type, etc., are important risk factors that have significant effects on missed visits. By examining these factors, researchers can identify disparities and inequalities in healthcare access, utilization, and outcomes among different population groups, allowing for prioritization and tailoring of interventions.

Younger PLWH was identified as "high risk" group regarding of HIV retention and ART adherence[32-34] in prior research. The primary predictor of missing HIV visits in this review is age, specifically younger PLWH. Younger HIV patients might miss appointments due to multiple reasons: being less symptomatic[35], having less stable living situations[36], fewer financial resources, irregular schedules[16], and psychological factors like denial, fear, or experiences of stigma. However, in this review, only Kahana et al.[26], Reisner et al.[37] and Tarantino et al.[36] addressed the study on the youth population. Considering the fact that retaining younger HIV-infected persons in care would improve their overall health and reduce HIV transmission[38], more studies are needed on this population, and the need to create youth-friendly environments is urgent[39].

People classified as non-white race, especially African Americans (AA) are less likely to keep their scheduled visits. Although the AA comprise less than 14% in the U.S. population, the proportion of AA in US among PLWH is 42%[40, 41]. Racial/ethnic disparities in HIV care retention are complex and multifaceted. Racial and ethnic minority groups are facing greater challenges related to social determinants of health, such as housing instability, limited access to healthcare facilities, lower healthy literacy[42], and experiences of systemic racism and discrimination, which are associated with missed

HIV visits. Therefore, the risks associated with missed visits cannot be solely explained by racial/ethnic status. Mugavero, et al.[5] raised the point that pathways mediating racial/ethnic healthcare disparities are likely to involve a complex interplay of healthcare, public health, and social factors, and much work needs to be done to explore the intricate processes mediating healthcare disparities. We suggest considering the relationship between racial/ethnic and other socioeconomic factors systemic racism, to design policies and interventions to improve access to quality healthcare service, reduce stigma and mistrust among racial and ethnic minority population, and to reduce the disparities in HIV care retention.

Sohail et al.[43] found that married individuals had better visit constancy in cisgender heterosexual populations. Anderson et al.[44] stated that married PLWH had lower percent visit adherence compared to single or never married PLWH, among persons with less than 100% visit adherence. Waldrop-Valverde et al.[15] similarly found a lower percent visit adherence among PLWH who were married. This disagreement might be attributed to the different sexual orientations in these studies. According to Sohail et al.[43], no other US-based studies focused on heterosexual partnership and retention in care among newly diagnosed PLWH was noticed before, and only a few studies included partnership/marital status as a covariate when assessing associations of other exposures with retention in care. Therefore, sexual orientation may be an important covariance to considerate when estimating the association between missed visits and marital status.

Conflicts between adjusted and unadjusted models were found in multiple studies. Researchers should be cautious with how to report and explain the results of these conflicts. One possible reason for different results is confounding. If there are

confounding factors that significantly influence the relationship between the independent and dependent variables, the unadjusted model may overestimate or underestimate the true association, whereas the adjusted model provides a more accurate estimate by controlling for these factors, or adjusting for incorrect variables, called colliders, may provide biased adjusted estimates. To address the conflict between adjusted and unadjusted models, it is essential to carefully consider which variables to include in the model and to ensure the model is specified correctly. Researchers should also assess the sensitivity of their results to the choice of adjustment variables and report both unadjusted and adjusted estimates to provide a more comprehensive understanding of the relationships between variables.

This review only included studies published between 2000 to 2022. The landscape of HIV treatment and care has undergone significant changes since the advent of the HIV/AIDS pandemic in the early 1980s. The year 2000 marks an important turning point in the history of HIV care and treatment, primarily due to the introduction and widespread use of highly active antiretroviral therapy (HAART) in 1996. The widespread use of HAART led to significant reductions in HIV/AIDS-related morbidity and mortality. HIV became a manageable chronic condition, and life expectancy for people living with HIV increased substantially. The Affordable Care Act (ACA) enacted in 2010, had a significant impact on HIV care by expanding access to health insurance for millions of Americans, including those living with HIV. The ACA allowed states to expand their Medicaid programs to cover more low-income individuals, which has led to increased access to healthcare for people living with HIV.

This systematic review has some limitations. First, there are limitations inherent to the observational and retrospective nature, which means that the investigators have to depend on the availability and accuracy of these records. Some key statistics cannot be measured and significant biases may affect the results of this review. Also, temporal relationships are difficult to assess due to the different observational periods among these studies[45]. Different outcomes and risk factor definitions were used in studies may bias the interpretation across the included studies.

Many of the included studies used self-reported data to evaluate missed visit outcomes, including asking participants how many missed visits in the last year. Self-report bias is a type of measurement error that can occur in any context where random or systematic misreporting is conceivable. The bias is ubiquitous in survey data where cognitive processes, social desirability, and survey conditions can alter interviewee's responses [46]. We also recommend that HIV service providers use electronic medical system to track of missed visits.

Reference

1. Dailey, A., et al., *Diagnoses of HIV infection in the United States and dependent areas, 2020.* 2022.

2. KFF. *The HIV/AIDS Epidemic in the United States: The Basics*. 2021.

3. Traeger, L., et al., *Risk factors for missed HIV primary care visits among men who have sex with men.* J Behav Med, 2012. **35**(5): p. 548-56.

4. Organization, W.H., *The work of WHO on HIV/AIDS progress report*. 2002.

5. Mugavero, M.J., et al., *Racial disparities in HIV virologic failure: do missed visits matter?* J Acquir Immune Defic Syndr, 2009. **50**(1): p. 100-8.

6. Mugavero, M.J., et al., *Measuring retention in HIV care: the elusive gold standard.* J Acquir Immune Defic Syndr, 2012. **61**(5): p. 574-80.

7. Higa, D.H., et al., *Identifying Best Practices for Increasing Linkage to, Retention, and Re-engagement in HIV Medical Care: Findings from a Systematic Review, 1996-2014.* AIDS Behav, 2016. **20**(5): p. 951-66.

8. Majeroni, B.A., et al., *Missed appointments and Medicaid managed care.* Arch Fam Med, 1996. **5**(9): p. 507-11.

9. Sharp, D. and W. Hamilton, *Non-attendance at general practices and outpatient clinics - Local systems are needed to address local problems.* British Medical Journal, 2001. **323**(7321): p. 1081-1082.

10. Waller, J. and P. Hodgkin, *Defaulters in general practice: who are they and what can be done about them?* Family Practice, 2000. **17**(3): p. 252-253.

11. Horberg, M.A., et al., *Missed office visits and risk of mortality among HIV-infected subjects in a large healthcare system in the United States.* AIDS Patient Care STDS, 2013. **27**(8): p. 442-9.

12. Robbins, G.K., et al., *Predictors of antiretroviral treatment failure in an urban HIV clinic.* Jaids-Journal of Acquired Immune Deficiency Syndromes, 2007. **44**(1): p. 30-37.

13. Brennan, A.T., et al., *The importance of clinic attendance in the first six months on antiretroviral treatment: a retrospective analysis at a large public sector HIV clinic in South Africa.* Journal of the International Aids Society, 2010. **13**.

14. Abdulrahman, S.A., et al., *HIV Treatment Adherence - A Shared Burden for Patients, Health-Care Providers, and Other Stakeholders.* AIDS Rev, 2019. **21**(1): p. 28-39.

15. Waldrop-Valverde, D., et al., *Risk and protective factors for retention in HIV care.* AIDS and Behavior, 2014. **18**: p. 1483-1491.

16. Catz, S.L., et al., *Predictors of outpatient medical appointment attendance among persons with HIV.* Aids Care-Psychological and Socio-Medical Aspects of Aids/Hiv, 1999. **11**(3): p. 361-373.

17. Lucas, G.M., R.E. Chaisson, and R.D. Moore, *Highly active antiretroviral therapy in a large urban clinic: Risk factors for virologic failure and adverse drug reactions.* Annals of Internal Medicine, 1999. **131**(2): p. 81-+.

18. Israelski, D., et al., *Sociodemographic characteristics associated with medical appointment adherence among HIV-seropositive patients seeking treatment in a county outpatient facility.* Prev Med, 2001. **33**(5): p. 470-5.

19. Haley, D.F., et al., *Retention strategies and factors associated with missed visits among low income women at increased risk of HIV acquisition in the US (HPTN 064).* AIDS Patient Care STDS, 2014. **28**(4): p. 206-17.

20. Vat, L.E., et al., *Evaluating the "return on patient engagement initiatives" in medicines research and development: A literature review.* Health Expectations, 2020. **23**(1): p. 5-18.

21. Bulsara, S.M., M.L. Wainberg, and T.R. Newton-John, *Predictors of adult retention in HIV care: a systematic review.* AIDS and Behavior, 2018. **22**: p. 752-764.

22. Page, M.J., et al., *The PRISMA 2020 statement: an updated guideline for reporting systematic reviews.* International journal of surgery, 2021. **88**: p. 105906.

23. Higgins, J.P., et al., *Cochrane handbook for systematic reviews of interventions*. 2019: John Wiley & Sons.

24. Moola, S., et al., *Conducting systematic reviews of association (etiology): The Joanna Briggs Institute's approach.* JBI Evidence Implementation, 2015. **13**(3): p. 163-169.

25. Peterson, J., et al., *The Newcastle-Ottawa scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses.* Ottawa: Ottawa Hospital Research Institute, 2011. **2**(1): p. 1-12.

26. Kahana, S.Y., et al., *Structural determinants of antiretroviral therapy use, HIV care attendance, and viral suppression among adolescents and young adults living with HIV.* PloS one, 2016. **11**(4): p. e0151106.

27. Kay, E.S., A. Lacombe-Duncan, and R.M. Pinto, *Predicting Retention in HIV Primary Care: Is There a Missed Visits Continuum Based on Patient Characteristics?* AIDS and Behavior, 2019. **23**: p. 2542-2548.

28. Lesko, C.R., et al., *Exploring definitions of retention in care for people living with HIV in the United States in the modern treatment era.* AIDS, 2022. **36**(8): p. 1181-1189.

29. Mannes, Z.L., et al., *The association between symptoms of generalized anxiety disorder and appointment adherence, overnight hospitalization, and emergency department/urgent care visits among adults living with HIV enrolled in care.* Journal of behavioral medicine, 2019. **42**: p. 330-341.

30. Pence, B.W., et al., *Who will show? Predicting missed visits among patients in routine HIV primary care in the United States.* AIDS and Behavior, 2019. **23**: p. 418-426.

31. Safo, S.A., et al., *Pain is associated with missed clinic visits among HIV-positive women.* AIDS and Behavior, 2017. **21**: p. 1782-1790.

32. Barclay, T.R., et al., *Age-associated predictors of medication adherence in HIV-positive adults: health beliefs, self-efficacy, and neurocognitive status.* Health psychology, 2007. **26**(1): p. 40.

33. Hinkin, C.H., et al., *Medication adherence in HIV-infected adults: effect of patient age, cognitive status, and substance abuse.* AIDS (London, England), 2004. **18**(Suppl 1): p. S19.

34. Silverberg, M.J., et al., *Older age and the response to and tolerability of antiretroviral therapy.* Archives of internal medicine, 2007. **167**(7): p. 684-691.

35. Bofill, L., et al., *Demographic and psychosocial factors associated with appointment attendance among HIV-positive outpatients.* AIDS Care, 2011. **23**(10): p. 1219-25.

36. Tarantino, N., et al., *Correlates of missed clinic visits among youth living with HIV.* AIDS Care, 2018. **30**(8): p. 982-989.

37. Reisner, S.L., et al., *Characterizing the HIV Prevention and Care Continua in a Sample of Transgender Youth in the U.S.* AIDS Behav, 2017. **21**(12): p. 3312-3327.

38. Zanoni, B.C. and K.H. Mayer, *The adolescent and young adult HIV cascade of care in the United States: exaggerated health disparities.* AIDS patient care and STDs, 2014. **28**(3): p. 128-135.

39. Tanner, A.E., et al., *"Youth friendly" clinics: Considerations for linking and engaging HIV-infected adolescents into care.* Aids Care-Psychological and Socio-Medical Aspects of Aids/Hiv, 2014. **26**(2): p. 199-205.

40. Census, U.S., *Population Estimates*. 2022: https://www.census.gov/quickfacts/fact/table/US#.

41. HHS.gov. *HIV/AIDS and African Americans*. 2020 [cited 2023; Available from: https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlid=21.

42. Mallinson, W., *Cyprus: A modern history*. 2005: Bloomsbury Publishing.

43. Sohail, M., et al., *Partnership status and retention in care among cisgender heterosexual newly diagnosed people with HIV: a cohort study.* AIDS care, 2022: p. 1-9.

44. Anderson, J.C., et al., *Impact of intimate partner violence on clinic attendance, viral suppression and CD4 cell count of women living with HIV in an urban clinic setting.* AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV, 2018. **30**(4): p. 399-408.

45. Hyde, J.N., *Retrospective cohort study: Strengths and weaknesses.* Tufts Open Courseware. Tufts University. Archived from the original on, 2017. **12**.

46. Bound, J., C. Brown, and N. Mathiowetz, *Measurement error in survey data*, in *Handbook of econometrics*. 2001, Elsevier. p. 3705-3843.

47. Mugavero, M.J., et al., *Missed visits and mortality among patients establishing initial outpatient HIV treatment.* Clin Infect Dis, 2009. **48**(2): p. 248-56.

48. Sarnquist, C.C., et al., *Rural HIV-infected women's access to medical care: ongoing needs in California.* AIDS Care, 2011. **23**(7): p. 792-6.

49. Horvath, K.J., et al., *Engagement in HIV Medical Care and Technology Use among Stimulant-Using and Nonstimulant-Using Men who have Sex with Men.* AIDS Res Treat, 2013. **2013**: p. 121352.

50. Jones, D., et al., *Personal HIV knowledge, appointment adherence and HIV outcomes.* AIDS Behav, 2013. **17**(1): p. 242-9.

51. Cohen, J.K., et al., *Regular clinic attendance in two large San Francisco HIV primary care settings.* AIDS Care, 2016. **28**(5): p. 579-84.

52. Hightow-Weidman, L., et al., *Exploring the HIV continuum of care among young black MSM.* PLoS ONE, 2017. **12**(6).

53. Pence, B.W., et al., *Association of Increased Chronicity of Depression With HIV Appointment Attendance, Treatment Failure, and Mortality Among HIV-Infected Adults in the United States.* JAMA Psychiatry, 2018. **75**(4): p. 379-385.

54. Christopoulos, K.A., et al., *Internalized HIV Stigma Is Associated with Concurrent Viremia and Poor Retention in a Cohort of US Patients in HIV Care.* Journal of Acquired Immune Deficiency Syndromes, 2019. **82**(2): p. 116-123.

55. Turan, B., et al., *Internal Working Models of Attachment Relationships and HIV Outcomes among Women Living with HIV.* Journal of Acquired Immune Deficiency Syndromes, 2019. **80**(1): p. E1-E8.

56. Batey, D.S., et al., *Are missed- and kept-visit measures capturing different aspects of retention in HIV primary care?* AIDS Care, 2020. **32**(1): p. 98-103.

57. Cressman, A.E., et al., *The Relationship Between Discrimination and Missed HIV Care Appointments Among Women Living with HIV.* AIDS and Behavior, 2020. **24**(1): p. 151-164.

58. Fazeli, P.L., et al., *Critical, and not functional, health literacy is associated with missed HIV clinic visits in adults and older adults living with HIV in the Deep South.* AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV, 2020. **32**(6): p. 694-700.

59. Batchelder, A.W., et al., *Substance Use Stigma, Avoidance Coping, and Missed HIV Appointments Among MSM Who Use Substances.* AIDS and Behavior, 2021. **25**(5): p. 1454-1463.

60. Kay, E.S., et al., *Health insurance and AIDS Drug Assistance Program (ADAP) increases retention in care among women living with HIV in the United States.* AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV, 2021. **33**(8): p. 1044-1051.

61. Menza, T.W., et al., *Social Determinants of Health and Care Outcomes Among People With HIV in the United States.* Open Forum Infect Dis, 2021. **8**(7): p. ofab330.

62. Pearson, C.A., et al., *Internalized HIV Stigma Predicts Suboptimal Retention in Care Among People Living with HIV in the United States.* AIDS Patient Care STDS, 2021. **35**(5): p. 188-193.

63. Judd, R.T., et al., *Association between patient-reported barriers and HIV clinic appointment attendance: A prospective cohort study.* AIDS Care, 2022. **34**(5): p. 545-553.

64. Sohail, M., et al., *Assessing the Impact of COVID-19 on Retention in HIV Primary Care: A Longitudinal Multisite Analysis.* AIDS and Behavior, 2022.

Table 1. Characteristics of included articles (n=32)

Author, article #	Location (Cohort) and criteria	Sample size	Study period	Missed visits measure	Statistical method	% of missed visits
Mugavero et al.[5]	UAB 1917 clinic cohort, >=4 scheduled visits over 6+ months	1221	>=6 mon	ratio: no show/scheduled	GEE with auto-regressive correlation structure	40%
Mugavero et al.[47]	UAB 1917 clinic cohort, initial between 2000-2005, no prior outpatient HIV treatment	543	365 days	dichotomous: missed vs no missed	Chi-square and logistic regression	60%
Bofill et al.[35]	JMH FL, outpatient adult HIV/AIDS clinic, not on ARV for 3 months prior or treatment naïve	178	12 mon	rate: missed/scheduled	hierarchical multiple linear regression	90%
Sarnquist et al.[48]	California (11 facilities), female patients, living in rural areas	64	12 mon	dichotomous: missed vs no missed	Chi-square tests	38%
Traeger et al.[3]	Fenway Health New England, adults, MSM, HIV+ >=3 mon	503	12 mon	dichotomous: missed vs no missed	logistic regression	31%
Horberg et al.[11]	Kaiser Permanente Northern California (KPNC), adults, newly diagnosed between 1997-2007	2811	12 mon	dichotomous: missed vs no missed	logistic regression	66%

Author, article #	Location (Cohort) and criteria	Sample size	Study period	Missed visits measure	Statistical method	% of missed visits
Horvath et al.[49]	Adults, English speaking, MSM, US residence	276	12 mon	Categorical: no missed (ref), missed and not in care.	multinomial regression	31%
Jones et al.[50]	Jackson Memorial Hospital (Miami, FL), adults, in care, English speaking, no psychotic disorder or head trauma	206	7 mon	dichotomous: >=75% adherence vs <75% adherence.	Chi-square and logistic regression	62%
Cohen et al.[51]	San Francisco, adults	469	12 mon	dichotomous: missed vs no missed	logistic regression	45%
Kahana et al.[26]	ATN cohort, 12-26 yrs, aware HIV infected, English or Spanish speaking	1891	12 mon	dichotomous: missed vs no missed	means as intercept model with Bernoulli distribution*	50%
Hightow-Weidman et al.[52]	North Carolina, 18-30 yrs, Biological male, black, reside in North Carolina, have access to mobile device, MSM	193	12 mon	dichotomous: no missed vs missed	logistic regression	48%
Reisner et al.[37]	14 AMTU sites, Transgender, 16-24 years, understand English	56	6 mon	2 dichotomous: missed 2+ vs no missed, missed vs no missed	logistic regression	34%
Safo et al.[31]	9 HRSA demonstration sites, adults, women of color	862	12 mon	dichotomous: >1 vs <=1 missed	Chi-square test, logistic regression	45%

Author, article #	Location (Cohort) and criteria	Sample size	Study period	Missed visits measure	Statistical method	% of missed visits
Anderson et al.[44]	Johns Hopkins, adult, women, English speaking, in clinic 1+ year, with intimate relationship	232	12 mon	dichotomous: missed>33% vs missed <=33%	logistic regression	36%
Pence et al.[53]	CNICS cohort, adults, 2+ PHQ measures completed	5927	12 mon	dichotomous: missed vs no missed	poisson regression	18.8*
Tarantino et al.[36]	ATN cohort, 12-24 yrs, understand English or Spanish	2125	12 mon	dichotomous: >=2 missed vs <=1 missed	logistic regression	64% (36% 2+)
Christopoulos et al.[54]	CNICS cohort, adults	4214	2*6mon	dichotomous: >=2 missed vs <=1 missed	logistic regression	24%
Kay et al.[27]	Southeastern US clinic, adults	1159	12 mon	3 categories: 0 missed, 1-2 missed, 3+ missed	multinomial logistic regression	40%
Mannes et al.[29]	Florida, adults	801	6 mon	dichotomous: missed vs no missed	logistic regression	25%
Pence et al.[30]	CNICS cohort, adults, HIV 2+ visits in 2002-2015	10374	12 mon	dichotomous: missed vs no missed	logistic regression	53%
Turan et al.[55]	4 WIHS sites, adult women	453	24 mon	dichotomous: missed vs no missed	logistic regression	30%
Batey et al.[56]	6 academically-affiliated HIV clinics, HIV, +1 visit, 1+ scheduled visit	10053	12 mon	3 measures: continuous (missed in count), dichotomous (missed vs no missed), ratio (visited/scheduled)	poission, logistic, linear regression	na

Author, article #	Location (Cohort) and criteria	Sample size	Study period	Missed visits measure	Statistical method	% of missed visits
Cressman et al.[57]	WIHS cohort, adult women	1578	6 mon	dichotomous: missed vs no missed	log binomial	15%
Fazeli et al.[58]	Southeastern US urban clinic, HIV+ for 1+ year, 40+ yrs, no comorbidities	95	24 mon	ratio: missed/scheduled	Pearson's correlation/t test, hierarchical linear regression	na
Batchelder et al.[59]	INSIGHT cohort, adults, MSM, substance use, reside in Massachusetts	202	6 mon	dichotomous: missed vs no missed	logistic regression	63%
Kay et al.[60]	WIHS cohort, adult women	1366	6 mon	dichotomous: no missed vs missed	logistic regression	na
Menza et al.[61]	MMP cohort, adults	15964	12 mon	dichotomous: missed vs no missed	logistic regression	24%
Pearson et al.[62]	CNICS cohort, adults, PRO survey completed	5825	12 mon	dichotomous: no missed vs missed	logistic regression	44%
Judd et al.[63]	Urban clinic, AA community, adults, 1+ prior appointment	105	13 mon	dichotomous: no missed vs missed	logistic regression	44%
Lesko et al.[28]	CNICS, adults, 1+ encounter, 1+ visit, 1+ visit in prior year, 1+ visit in study year	21481	18-24mon	2 dichotomous: attended >=75% vs attended <75%, missed>25% vs missed <=25%	logistic regression	42%
Sohail et al.[43]	UAB 1917 clinic cohort, cis-gender, adults, newly diagnosed, heterosexual	152	24 mon	2 dichotomous: attended vs no	logistic regression	33.8%, 35.3%

Author, article #	Location (Cohort) and criteria	Sample size	Study period	Missed visits measure	Statistical method	% of missed visits
				attended, missed vs no missed		
Sohail et al.[64]	Alabama D4C cohort, adults	6410	24 mon	dichotomous: no missed vs missed	logistic regression	28%

Table 2. Quality assessment using Newcastle-Ottowa Scale tool.

Author, article #	**Selection**				**Comparability**	**Outcome**			**Overall rating**
	1. Representative-ness of the Exposed Cohort	2. Selection of the Non-Exposed Cohort	3. Ascertainment of Exposure	4. Demonstration That Outcome of Interest Was Not Present at Start of Study	1. Comparability of Cohorts on the Basis of the Design or Analysis	1. Assessment of Outcome	2. Was Follow-Up Long Enough for Outcomes to Occur	3. Adequacy of Follow Up of Cohorts	Good, fair or poor
Mugavero et al.[5]	1	1	1	0	1	1	1	1	Good
Mugavero et al.[47]	1	1	1	0	1	1	1	1	Good
Bofill et al.[35]	1	1	1	0	1	1	1	1	Good
Sarnquist et al.[48]	0	1	1	0	1	0	1	0	Poor
Traeger et al.[3]	0	1	1	0	1	1	1	1	Fair
Horberg et al.[11]	1	1	1	0	1	1	1	1	Good
Horvath et al.[49]	0	1	1	0	1	0	1	1	Fair
Jones et al.[50]	1	1	1	0	1	1	1	1	Good
Cohen et al.[51]	0	1	1	0	1	0	1	1	Fair
Kahana et al.[26]	0	1	1	0	1	0	1	1	Fair
Hightow-Weidman et al.[52]	0	1	1	0	1	0	1	1	Fair
Reisner#12[37]	0	1	1	0	1	0	1	1	Fair
Safo et al.[31]	0	1	1	1	1	1	1	1	Good
Anderson et al.[44]	0	1	1	0	1	1	1	1	Good
Pence et al.[53]	1	1	1	0	1	1	1	1	Good
Tarantino et al.[36]	0	1	1	0	1	0	1	1	Fair

Author, article #	Selection				Comparability	Outcome			Overall rating
	1. Representative-ness of the Exposed Cohort	2. Selection of the Non-Exposed Cohort	3. Ascertain ment of Exposure	4. Demonstration That Outcome of Interest Was Not Present at Start of Study	1. Comparability of Cohorts on the Basis of the Design or Analysis	1. Assessm ent of Outcome	2. Was Follow-Up Long Enough for Outcomes to Occur	3. Adequac y of Follow Up of Cohorts	Good, fair or poor
Christopoulos et al.[54]	1	1	1	0	1	1	1	1	Good
Kay et al.[27]	1	1	1	0	1	1	1	1	Good
Mannes et al.[29]	1	1	1	0	1	0	1	1	Good
Pence et al.[30]	1	1	1	0	1	1	1	1	Good
Turan et al.[55]	0	1	1	0	1	0	1	1	Fair
Batey et al.[56]	1	1	1	0	1	1	1	1	Good
Cressman et al.[57]	0	1	1	1	0	0	1	1	Poor
Fazeli et al.[58]	0	1	1	0	1	1	1	1	Good
Batchelder et al.[59]	0	1	1	0	1	0	1	1	Fair
Kay et al.[60]	0	1	1	0	1	0	1	1	Fair
Menza et al.[61]	1	1	1	0	1	1	1	1	Good
Pearson et al.[62]	1	1	1	0	1	1	1	1	Good
Judd# et al.[63]	0	1	1	1	1	1	1	1	Good
Lesko et al.[28]	1	1	1	0	1	1	1	1	Good
Sohail et al.[43]	0	1	1	0	0	1	1	1	Fair
Sohail et al.[64]	1	1	1	0	1	1	1	1	Good

Table 3. Summary of predictors of missed visits, and referred articles

Risk factors of missed visits	Number of articles in which risk factor(s) is/are cited	Articles (reference number)	Category	Overall relationship with missed visits	consistency
Age	20	5,47,35,3,11,51,26,31,36,54,27,29,30,55,56,58,61,62,28,64	Demographic	Younger age is associated with missed visits18	N
Race	15	5, 47,3, 11 51,31,36,54,27,30,56,61,62,28,64	Demographic	AA, Hispanic, mixed racial is associated with missed visits (White has better retention)	Y
Substance	10	5,47,51,26,52,44,36,30,58,63	Behavioral	Any type of substance use is associated with missed visits	Y
HIV risk factor/HIV transmission mode	9	47,11,51,36,30,56,61,28,64	HIV related	IDU and Heterosexual are associated with missed visits compared to MSM	PY
Gender	8	5,47,36,54,30,61,62,64	Demographic	Transgender/female is associated with missed visits	Y
Income	8	3,51,52,31,27,61,63	Socioeconomic	Lower-income/below FPL is associated with missed visits	Y
Education	7	3,49,52,36,27,58,61	Demographic	Less education is associated with missed visits, not in 18	N
Insurance	7	5,47,3,27,30,60,61	Socioeconomic	Public or uninsured is associated with missed visits	N
Depression	6	3,50,53,30,58,63	Psychiatric	Depression is associated with missed visits	Y
Stigma	6	52, 37,54,57,59,62	Psychiatric	Stigma/discrimination is associated with missed visits	Y
Year of enrollment	5	11,26,30,62,28	HIV related	Long time of enrollment/treatment/diagnosis is associated with missed visits	Y
Durable VL	4	31,29,30,28	Clinic	Lower VL is associated with missed visits, not in 30.	N

Risk factors of missed visits	Number of articles in which risk factor(s) is/are cited	Articles (reference number)	Category	Overall relationship with missed visits	consistency
Anxiety	4	29,30,55,63	Psychiatric	Anxiety is associated with missed visits	Y
Alcohol	4	5,50,44,30	Behavioral	Alcohol abuse is associated with missed visits	Y
Social support	4	25,52,36,55	Socioeconomic	No/less social support is associated with missed visits	Y
Unemployment or disabled	3	31,36, 63	Socioeconomic	Unemployment or disabled is associated with missed visits	Y
Homeless	3	52,61,64	Socioeconomic	Unstable housing/being homeless is associated with missed visits	Y
Baseline CD4	3	5, 47, 11	Clinic	Lower baseline CD4 counts is associated with missed visits	Y
Previous missed visits	2	30,63	HIV related	Poor previous retention is associated with missed visits	Y
Number scheduled	2	31,62	HIV related	More scheduled visit is associated with missed visits	Y
Heterosexual (gender identity)	2	31,62	Demographic	Heterosexual is associated with missed visits	Y
Site	2	51,64	Demographic	Na	NA
Legal involvement	2	36,61	Behavioral	Legal involvement is associated with missed visits	Y
Health literacy	2	58,61	Socioeconomic	Poor health literacy is associated with missed visits	Y
Art	2	31,30	HIV related	Not on ART is associated with missed visits	Y
Mental health	2	26,36	Psychiatric	Poor mental health or having mental health symptoms is associated with missed visits	Y

Risk factors of missed visits	Number of articles in which risk factor(s) is/are cited	Articles (reference number)	Category	Overall relationship with missed visits	consistency
Ipv	1	61	Health-related	History of IPV is associated with missed visits	NA
Youth HIV prevalence	1	26	Demographic	Youth HIV prevalence >13% is associated with missed visits	NA
Number moved	1	36	Socioeconomic	More number of moving is associated with missed visits miss visits	NA
Partnership	1	43	Demographic	Not married is associated with missed visits.	NA
Language	1	48	Demographic	English speaking is associated with missed visits (compared to Spanish speaking)	NA
Sex behavior	1	36	Behavioral	Unprotected sex more is associated with missed visits	NA
Cd4	1	30	Clinic	Lower CD4 counts is associated with missed visits.	NA
Prior AIDS diagnosis	1	28	HIV related	Prior AIDS diagnosis is associated with missed visits	NA
Pain	1	31	Health-related	Frequent pain is associated with missed visits	NA
ER visit	1	61	Health-related	More ER visits is associated with missed visits	NA
Cognitive function	1	58	Health-related	Poor cognitive function is associated with missed visits	NA
Life chaos	1	49	Health-related	Have life chaos is associated with missed visits	NA
Gender affirmation	1	37	Health-related	Have gender affirmation is associated with missed visits	NA
Disclosure	1	36	HIV related	Had HIV disclosed is associated with missed visits	NA

Risk factors of missed visits	Number of articles in which risk factor(s) is/are cited	Articles (reference number)	Category	Overall relationship with missed visits	consistency
HIV appointment expectancy	1	3	Behavioral	Low appt expectancy is associated with missed visits	NA
HIV-knowledge	1	50	Socioeconomic	Less HIV-knowledge is associated with missed visits	NA
RWHAP support service	1	27	Socioeconomic	Not receiving support is associated with missed visits	NA
Need transportation	1	61	Socioeconomic	Needing transportation is associated with missed visits	NA
General health	1	3	Health-related	Fair/poor health is associated with missed visits (ref: good, very good and excellent health)	NA

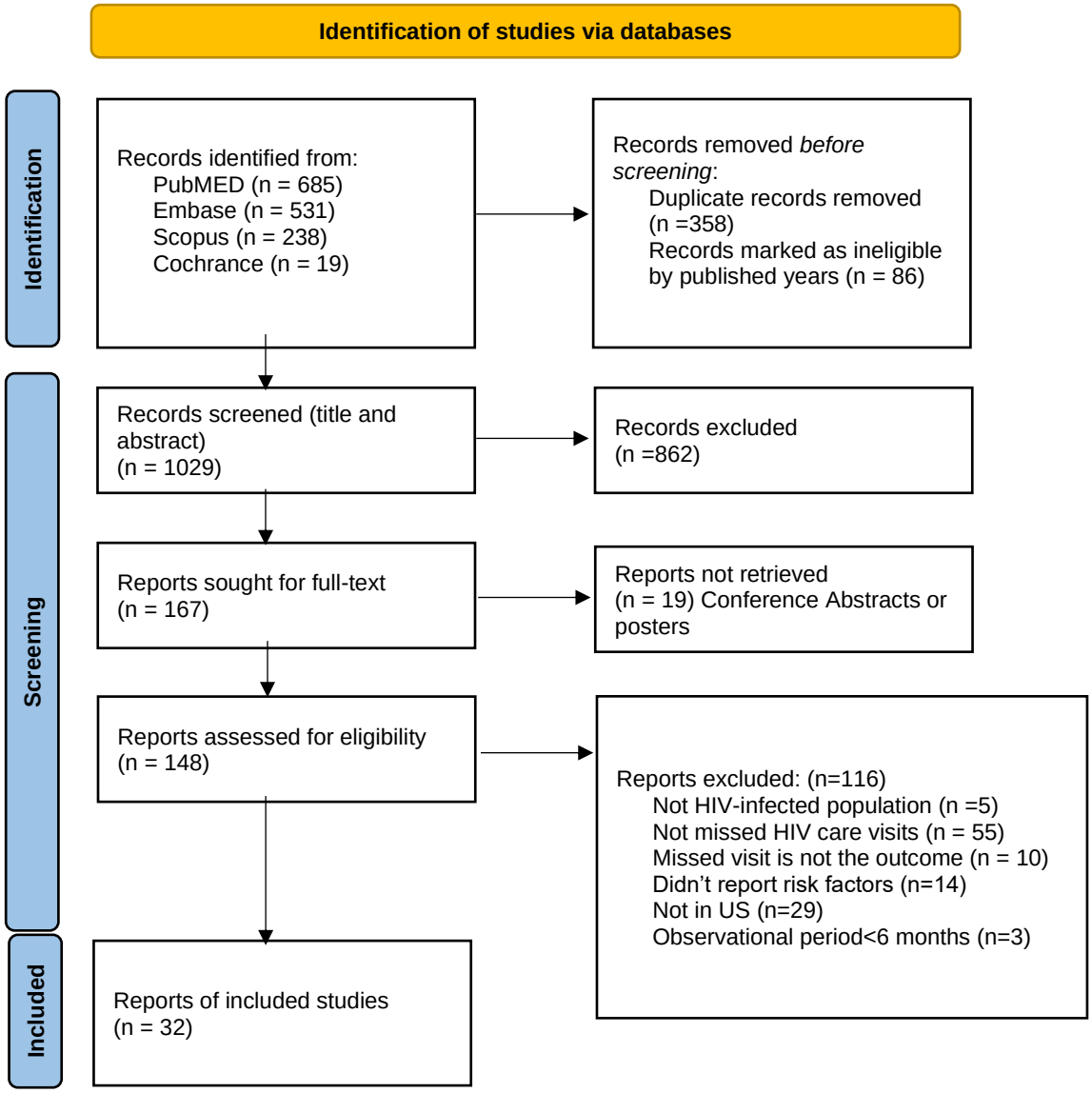


Figure 1. PRISMA-P flowchart of article selection

THE IMPACT OF MISSED SCHEDULED VISITS ON HEALTH OUTCOMES AMONG YOUTH WITH HIV IN ALABAMA

by

JIAYING HAO, DUSTIN M. LONG, HEATHER M. RELYEA-ASHLEY,

HENNA BUDHWANI, TINA Y. SIMPSON, SAMANTHA V. HILL

Submitted to *Population Health Management*

Format adapted for dissertation

ABSTRACT

Introduction: Missed visits are associated with poorer treatment and health outcomes among people with HIV (PWH). Studies have found that demographic factors and viral suppression are associated with missed visits. However, gaps in knowledge remain related to how missed visits may affect youth with HIV (YWH), a vulnerable subpopulation.

Methods: In this exploratory study, the authors examined data from an Alabama academic Ryan White funded HIV clinic to assess correlations of missed visits among clients aged 16 to 24 years old. Ninety-six YWH seen between March 1, 2020 and August 31, 2021 were included. Chi-square test was used to establish the association between missed visits and treatment outcomes. Unadjusted binary logistic regression analysis was performed to estimate the factors associated with missed visits.

Results: Fifty youth clients (46.88%) had at least one missed visit. Missed visits were associated with one negative treatment outcome: YWH with missed visits had more viral load rebound (p=0.04). The study also found that identifying as a cisgender male (OR=3.35, 95% CI: 1.10-10.20) and being virally suppressed at the index visit (OR=0.41, 95% CI: 0.18-0.94) were associated with missed visits among YWH.

Conclusion: Missed visits is an important indicator of retention in HIV care for youth. This study is one of the first to examine missed visits in YWH during the COVID-19 pandemic. More research is needed to elucidate risk factors associated with missed visits in YWH in order to better inform intervention development and improve retention in care for youth.

INTRODUCTION

The southern United States has accounted for the majority of HIV incident cases across the country for decades secondary to both its large and geographically dispersed population and social and political determinates of health that impact health care access and outcomes. More specifically, the southern states accounted for 51% of new HIV diagnoses in the US in 2020[1]. Similarly, 551,600 people with HIV (PWH) were estimated to live in southern states in 2019, which was almost half of the total number of PWH in the US[2]. For PWH, retention in care is a crucial factor for optimal clinical outcomes (e.g. undetectable VL). Studies have demonstrated that clients retained in HIV care have better antiretroviral therapy (ART) adherence and better treatment outcomes[3-5]. However, only 50% of the diagnosed PWH are retained in HIV care[6]. Considering the greater proportion of new HIV diagnoses in the south, more attention is needed on keeping PWH retained in HIV care in the southern states.

Several known measures of HIV retention in care have been used in studies. Among which the missed ("no show") scheduled visits is important for all PWH. Significant associations between missed visits with treatment failure and mortality were found [3,7-10]. Evidence suggests that clients with missed visits are more likely to have interruptions in ART, resulting in negative treatment outcomes, including lower CD4 counts, virologic failure and coinfections with other sexually transmitted infections [8,10,11-19]. Identifying clients at greater risk of "missing HIV visits" could prompt healthcare providers to preemptively intervene and help these clients identify strategies for visit adherence and overall improvement in treatment outcomes.

Adolescents (13-19 years old) and young adults (20-24 years old) account for about 20% of new HIV infections in the United States consistently[20]. In 2019, the rate of new HIV infections among adolescents was 23.3 per 100,000, which was the 3rd highest age group in Alabama. The 2019 Alabama annual report states that new diagnoses are shifting towards younger age groups and calls for increased prevention efforts focusing on younger populations[21]. Tarantino et al. (2018) pointed out that youth with HIV (LWH) have lower rates of testing, diagnosis, treatment engagement, and viral suppression (VS) compared with adults with HIV. Zanoni et al. (2014) estimated only 6% of YWH have suppressed HIV viral loads. Associations between poor retention in care and higher risk of morbidity and mortality have also been identified among YWH[22,23]. Importantly, unique developmental, psychosocial, behavioral, and infrastructural factors affect this vulnerable age group[24]. Although studies have focused on PWH's appointment attendance and risk factors for missed visits, including some that highlighted younger HIV clients being more likely to miss appointment visit than older counterparts; few studies have specifically examined YWH's missed appointment visits[10,25]. Furthermore, the COVID-19 pandemic has significantly altered patients' visit patterns, which in turn has impacted the retention in care for individuals living with HIV. Therefore, this study was designed to explore the relationships between missed visits among YWH and their measurable HIV treatment outcomes (e.g., CD4 count and viral load) and which factors may be associated with missed visits among YWH.

Materials and Methods

Study Setting and Participants:

This cross-sectional study occurred within an Alabama academic center HIV clinic which provides HIV treatment and wrap-around services to mothers, perinatally-infected infants, children, adolescents, young adults up to age 30, and adult women living with HIV[26].

YWH engaged in care at the clinic were included in this study based on the following eligibility criteria:

1. Any clients who had established appointment records within CAREWare.
2. Have complete demographic information available.
3. Aged 16 to 24 at their first service during the study period.
4. Have at least two laboratory records of viral load during the study period.

The study opted to only include YWH 16 to 24 to reduce the effect parents may have on younger YWH who must rely on parents/guardians for transportation to visits, as the legal driving age in Alabama is 16.

Data and Measures of Variables:

Data were extracted from the Human Resources and Service Administration's (HRSA) CAREWare system. CAREWare is an electronic health and social support services information system open to HRSA's Ryan White HIV/AIDS Program recipients and providers[27]. The system began to track appointment status (show, rescheduled, no-

show) on March 1st, 2020, during the early stages of the COVID-19 pandemic when many U.S.-based clinics began transitioning to telehealth appointments. This study included all appointment records regardless of in-person or telehealth. Client records from March 1st, 2020, to August 31st, 2021, were queried from CAREWare.

This study's primary outcomes were missed visits and standard HIV treatment outcomes. A missed visit was defined by 1) any 'no-show' visit without prior notice or rescheduling; or 2) any visit rescheduled for more than 30 days after the original scheduled visit. Standard HIV treatment outcomes include current CD4 count, viral load, and viral suppression. Viral suppression was defined as having less than 200 copies of HIV per milliliter of blood[28].

Missed visits were coded as a dichotomous variable ("missed" as 1 and "not missed" as 0). Treatment outcomes were measured by 1) any decrease of 50 or more in CD4 count; 2) any decrease in CD4 count; 3) any increase in HIV viral load; 4) VS in the last record. The first and last laboratory records among the 18-month study period were extracted from CAREWare. The term "index visit" represents the first record or first visit during the study period to avoid confusion with "initial visit" because in this study, most of these first visits are not initial visits to establish new care. Decreases in CD4 count of 50 or more is generally recognized as a clinically significant decrement and attributed to more than biological variation in the assay [8]. The authors calculated the change in CD4 count from index visit to last visit by subtracting the last CD4 record from the index CD4 record, coded as "1" for "≥50" and "0" for "<50". Also, any decrease in CD4 count is an indication of possible treatment failure[29] and evidence that current treatment may be

unable to control HIV infection[30]. Therefore, a decrease in CD4 count during the study period was considered a negative outcome. The last recorded CD4 count was subtracted from the index and coded as "decrease" (>0) and "no decrease" (≤0). Clients with no change and increased counts were combined due to potential sample size issues (n=1 for no changed group). For viral load, this study looked at the viral load maintenance (subtracting the index viral load recorded from the last viral load recorded, coded as "maintained" (≤, means decreased or no change in viral load) and "rebounded" (>0, means increased viral load)), and VS in last record (coded as "yes" and "no"). Clients with no change in viral load or whose viral load decreased during the study period were considered as viral load "maintained" as some clients' viral loads remained undetectable (Viral Load < 20 copies) throughout the study period.

Statistical Analysis:

Chi-Square tests were performed to identify associations between missed visits and treatment outcomes. To identify the risk factors associated with missed visits, age (continuous, by year), sex (Female and Male), race (African-American (AA), White, and Multi-racial; Asian was excluded due to small sample size), HIV transmission mode (Heterosexual, Perinatal, and Men who have sex with men (MSM); injection drug use was excluded due to small sample size), insurance type (Medicaid, Employer-paid, Private paid, and No insurance), and VS in index visits were used as factors related to missed visits. These factors were selected based on literature and data availability. Age, gender, race, HIV transmission mode, and insurance type were all extracted from demographic information in CAREWare. Bivariate logistic regression was conducted to

identify the risk factors. Odds ratios (OR), 95% confidence intervals (CI), and p-values were reported. All analyses were performed with SAS (Version 9.4, Cary, North Carolina, USA).

Results

A total of 331 clients were found with appointment records from March 1st, 2020, to August 31, 2021, of which, one hundred and one were aged 16-24 at their index visit. Ninety-six were included in this study. The majority were male (69.79%), Black or AA (72.92%), and MSM (58.33%). This study included both in-person clinic visits and telehealth virtual visits. Among 454 visit records, only 15 were telehealth visits, with one missed and other 14 all been attended. Fifty clients (46.88%) have at least one missed visit. (Table 1).

Table 1. Demographic information of clients

Association Between Missed Visits and Treatment Outcomes:

No significant association was found between decreased CD4 count and missed visits (p=0.36) or more than 50 counts of CD4 decrease and missed visits (p=0.89) (Table 2). No association between VS in the last record and missed visits was observed (p=0.75). However, YWH with missed visits had more viral load rebound compared to those without missed visits (p=0.04) (Table 3). Although the other measures did not show any statistical significance, the percentages of clients with missed visits was greater in the groups with adverse outcomes (more than 50 counts of CD4 drop and not virally

suppressed in the last record). Therefore, the evidence to support the hypothesis that missed visits are associated with viral load maintenance, and missed visits have adverse effects on treatment outcomes has been found.

Table 2. CD4 change by missed visits.

Table 3. Viral load change by missed visits.

Risk Factors of Missed Visits:

To understand what factors significantly affected missing HIV visits, continuous age, gender, race, HIV transmission mode, insurance type, and VS from index visits were examined in unadjusted logistic models. As shown in Table 4, only gender and VS in the index visit had significant associations with missed visits (gender: p=0.03, VS in index visit: p=0.04). Male clients were much more likely to have missed visits compared to female clients (OR=3.35, 95% CI: 1.10-10.20). Clients with VS at index visit had lower missed visit rates compared to those without VS (OR=0.41, 95% CI: 0.18-0.94). Age, race, HIV transmission mode, and insurance type were not significantly associated with missed visits.

Table 4. Risk factors associated with missed visits among YLWH (n=101).

Discussion

YWH are a distinct subgroup of PWH as their care is compounded by their continual development cognitively, physically and psychologically.[4] Thus, YWH require more focused studies on their retention in care and health outcomes. This study provides a timely examination of YWH HIV retention and to the knowledge is the only study

examining this during the global pandemic. Within this cohort of YWH aged 16-24, individuals with missed visits were more likely to have greater VL. Men and individuals not virally suppressed at baseline were more likely to have missed visits. Understanding factors influencing HIV visit adherence has the potential to reduce the risk of further HIV transmission, improve long-term health outcomes, and aid in ending the pandemic by achieving the 90-90-90 goals (90% of all people living with HIV will know their HIV status, 90% of all people with diagnosed HIV infection will receive sustained antiretroviral therapy, and 90% of all people receiving antiretroviral therapy will have viral suppression).[31]

The findings are similar to the few other studies that specifically focus on YWH. Other studies found associations between ART adherence and viral suppression with missed clinic visits.[32,33] Additionally, certain demographics were associated with missed visits including young gay and bisexual men who were undergoing development of their identities, youth from economically disadvantaged areas[34], high youth HIV incidence areas[34], and even female gender, suggesting a complexity of factors impacting visit adherence. Early, patient-centered interventions focused on when youth engage in care have the potential to be impactful in retaining YWH in care, and thus should be one area where future research efforts are focused.

This study also demonstrated that missed visits are associated with achieving and maintaining viral suppression. However, there are currently no standard measures for evaluating HIV retention.[12] Therefore, a system tracking each client's appointment status is crucial. Thus, interventions such as Data to Care[35], which focus on identifying and

linking PWH into care and assisting these clients with overcoming barriers to viral suppression may be even more important among YWH subgroups. Additionally, more youth-tailored strategies able to keep YWH engaged in care are needed.

The COVID-19 pandemic has significantly impacted clinic visits for a wide range of health conditions, including HIV. In an effort to minimize potential exposure to COVID-19, many patients and healthcare providers postponed or canceled non-urgent in-person visits. This includes regular check-ups and monitoring visits for chronic conditions like HIV. This can impact the continuity of care and can potentially lead to worsening of the underlying condition. Many healthcare providers rapidly transitioned to telehealth (virtual visits) as a safer alternative to in-person visits. This study is one of the first to examine missed visits in YLWH after the onset of the COVID-19 pandemic. Although this study is not powered to make definitive conclusions, the greater attendance of telehealth visits compared to in-person visits suggests virtual visits could be easier to keep to some patients. Exploring the connection between appointment attendance and virtual visits may help healthcare providers develop interventions to improve the retention in care.

While this is one of the few studies that focuses specifically on YWH and has a regional focus, this study had some limitations. First, although inclusion of a longer time span would have been ideal, neither the clinic nor CAREWare tracked visit status until March 2020, hindering the ability to include any pre-COVID time periods. Additionally, as all visits occurred during the COVID-19 pandemic, it was not able to ascertain how the pandemic affected visit attendance. Furthermore, certain demographic and client

characteristics were unavailable (e.g. year of enrollment/diagnosis, substance abuse, psychiatric disorders, and other factors), which could be potential risk factors for missed visits in YWH. It is also important to note, the age range of YWH is not standardized in literature. This study chose to use the World Health Organization (WHO) definitions where individuals 15 to 24 years of age are considered youth.[36] Furthermore, because this study aimed to identify risk factors associated with missed visits among YWH, it focused on an age range where youth are able to take more responsibility for their behavior. Given the legal driving age in Alabama is 16,[37] using 16 as the lower limit of cut-off was appropriate. However, even with this cutoff, potential bias caused by parental involvement (e.g. transportation, appointment reminders, medication reminders) was likely. Additionally, this study had a small sample size and was focused on YWH within a single state, limiting the generalizability of the findings.

Conclusions

In conclusion, this study supports how missed visits have negative effects on treatment outcomes for YWH, with being a cis-gender male and not being virally suppressed at the index visit serving as risk factors for YWH missed visits. Reducing the risk of missing HIV visits could make a difference in improving retention in care. During the COVID-19 pandemic, it was crucial to adapt these practices, including transitioning to telehealth appointments where appropriate, to ensure continuous care and minimize missed visits. More youth-tailored interventions are needed to identify challenges to keeping YWH engaged in care and virally suppressed.

Abbreviations:

AA: African-American

ART: Antiretroviral therapy

CI: Confidence intervals

HRSA: Human Resources and Service Administration

MSM: Men who have sex with men

OR: Odds ratios

PWH: People with HIV

VS: Viral suppression

WHO: World Health Organization

YWH: Youth with HIV

Declarations:

- Ethics approval and consent to participate: The protocol and project materials were reviewed and approved by the local ethics committee: the Institutional Review Board of the University of Alabama at Birmingham (IRB-981112002). The IRB waived the need to document informed consent for this retrospective study involving reviewing de-identified electronic medical record information.
- Consent for publication: Not applicable.
- Availability of data and material: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.
- Competing interests: Not applicable.
- Funding: Not applicable.
- Authors' contributions: JH: Statistical Conceptualization (lead), Data Curation, Investigation, Formal Analysis, Writing-Original Draft Preparation. DL: Conceptualization (supporting), Supervision (lead), Writing-Review and Editing. HRA: Investigation (supporting), Writing-Review and Editing. HB: Substantive Conceptualization; Supervision (supporting), Writing-Review and Editing. TS: Data Ownership; Writing-Review and Editing. SH: Senior Author, Writing-Review and Editing.
- Acknowledgment: We acknowledge the team at Alabama academic center HIV clinic, specifically Dayna Cooks-Heard, who have provided full support for this study.

- Prior presentation: This work was presented, in part, at ID Week 2022 (19-24 October 2022) virtually, and American Public Health Association 2022 Annual Meeting and Expo (9 July 2022) in Boston.

References

1. Centers for Disease Control and Prevention: HIV diagnoses in 2022. https://www.cdc.gov/hiv/statistics/overview/diagnoses.html (2022). Accessed: 21 November 2022.

2. Centers for Disease Control and Prevention. Estimated HIV incidence and prevalence in the United States, 2015–2019. HIV Surveillance Supplemental Report 2021;26(No. 1). http://www.cdc.gov/ hiv/library/reports/hiv-surveillance.html. Accessed: 21 November 2022.

3. Tripathi A, Youmans E, Gibson JJ, Duffus WA. The impact of retention in early HIV medical care on viro-immunological parameters and survival: a statewide study. AIDS Res Hum Retroviruses. 2011;27(7):751-758. doi:10.1089/AID.2010.0268

4. Yehia BR, French B, Fleishman JA, et al. Retention in care is more strongly associated with viral suppression in HIV-infected patients with lower versus higher CD4 counts. J Acquir Immune Defic Syndr. 2014;65(3):333-339. doi:10.1097/QAI.0000000000000023

5. Shumba C, Atuhaire L, Imakit R, Atukunda R, Memiah P. Missed Doses and Missed Appointments: Adherence to ART among Adult Patients in Uganda. ISRN AIDS. 2013;2013:270914. Published 2013 Jan 14. doi:10.1155/2013/270914

6. HIV.gov: HIV care continuum. https://www.hiv.gov/federal-response/policies-issues/hiv-aids-care-continuum (2022). Accessed: 12 Dec 2022.

7. Mugavero, M. J., Westfall, A. O., Zinski, A., et al. (2012). Measuring retention in HIV care: the elusive gold standard. Journal of acquired immune deficiency syndromes 1999, 61(5), 574–580. https://doi.org/10.1097/QAI.0b013e318273762f

8. Walburn A, Swindells S, Fisher C, et al. Missed visits and decline in CD4 cell count among HIV-infected patients: a mixed method study. Int J Infect Dis. 2012;16(11):e779---e785.

9. Crawford TN, Sanderson WT, Thornton A. Impact of poor retention in HIV medical care on time to viral load suppression. J Int Assoc Provid AIDS Care. 2014;13 (3):242---249.

10. Horberg MA, Hurley LB, Silverberg MJ, et al. Missed office visits and risk of mortality among HIV-infected subjects in a large healthcare system in the United States. AIDS Patient Care STDS. 2013;27(8):442---449. http://doi.org/10.1089/apc.2013.0073

11. Brennan AT, Maskew M, Sanne I, Fox MP. The importance of clinic attendance in the first six months on antiretroviral treatment: a retrospective analysis at a large public sector HIV clinic in South Africa. J Int AIDS Soc. 2010;13(1):49.

12. Nabaggala MS, Parkes-Ratanshi R, Kasirye R, Kiragga A, Castlenuovo B, Ochaka I, et al. Re-engagement in HIV care following a missed visit in rural Uganda. BMC Res Notes. 2018;11(1):9–13. https:// doi. org/ 10. 1186/s13104- 018- 3865-9.

13. Park WB, Choe PG, Kim SH, Jo JH, Bang JH, Kim HB, et al. One-year adherence to clinic visits after highly active antiretroviral therapy: a predictor of clinical progress in HIV patients. J Intern Med. 2007;261(3):268–75.

14. Rastegar DA, Fingerhood MI, Jasinski DR. Highly active antiretroviral therapy outcomes in a primary care clinic. AIDS Care. 2003;15(2):231–7. https:// doi. org/ 10. 1080/ 09540 12031 00006 8371.

15. Zinski A, Westfall AO, Gardner LI, Giordano TP, Wilson TE, Drainoni ML, et al. The contribution of missed clinic visits to disparities in HIV viral load outcomes. Am J Public Health. 2015;105(10):2068–75.

16. Mugavero MJ, Lin H, Willig JH, Westfall AO, Ulett KB, Routman JS, et al. Missed visits and mortality among patients establishing initial outpatient HIV treatment. Clin Infect Dis. 2009;48(2):248–56.

17. Mugavero MJ, Amico KR, Westfall AO, Crane HM, Zinski A, Willig JH, et al. Early retention in HIV care and viral load suppression. J Acquir Immune Defic Syndr. 2012;59(1):86–93.

18. Berg MB, Safren SA, Mimiaga MJ, Grasso C, Boswell S, Mayer KH. 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(7):902–7. https:// doi. org/ 10. 1080/ 09540 12050 01016 58.

19. Lucas GM, Chaisson RE, Moore RD. Highly active antiretroviral therapy in a large urban clinic: risk factors for virologic failure and adverse drug reactions. Ann Intern Med. 1999;131(2):81–7. https:// doi. org/ 10. 7326/0003- 4819- 131-2- 19990 7200-00002.

20. Centers for Disease Control and Prevention: HIV and Youth. https://www.cdc.gov/hiv/group/age/youth/index.html (2020). Accessed: 21 Nov 2022.

21. Alabama department of public health:2019 State of Alabama HIV surveillance annual report. https://www.alabamapublichealth.gov/hiv/assets/hivsurveillanceannualreport_2019.pdf Accessed: November 21, 2022]

22. Brinkhof MW, Pujades-Rodriguez M, Egger M. Mortality of patients lost to follow-up in antiretroviral treatment programmes in resource-limited settings: systematic review and meta-analysis. PLoS One. 2009;4(6):e5790. Published 2009 Jun 4. doi:10.1371/journal.pone.0005790

23. Auld AF, Agolory SG, Shiraishi RW, et al. Antiretroviral therapy enrollment characteristics and outcomes among HIV-infected adolescents and young adults compared with older adults--seven African countries, 2004-2013. MMWR Morb Mortal Wkly Rep. 2014;63(47):1097-1103.

24. Tarantino, N., Brown, L. K., Whiteley, L., et al. Correlates of missed clinic visits among youth living with HIV. AIDS care, 2018; 30(8), 982–989. https://doi.org/10.1080/09540121.2018.1437252

25. Dear N, Esber A, Iroezindu M, et al. Routine HIV clinic visit adherence in the African Cohort Study. AIDS Res Ther. 2022;19(1):1. Published 2022 Jan 7. doi:10.1186/s12981-021-00425-0

26. UAB Family Clinic, Department of Pediatrics. https://www.uab.edu/medicine/peds/patient-care/family-clinic Accessed: 21 Nov 2022.

27. Ryan White HIV/AIDS Program (RWHAP) Website: CareWare 6. https://ryanwhite.hrsa.gov/grants/manage/careware Accessed: 21 Nov 2022.

28. Centers for Disease Control and Prevention: HIV Treatment as Prevention. https://www.cdc.gov/hiv/risk/art/index.html Accessed: 21 Nov 2022.

29. Williams PL, Currier JS, Swindells S. Joint effects of HIV-1 RNA levels and CD4 lymphocyte cells on the risk of specific opportunistic infections. AIDS 1999;13:1035–44.

30. Clinical info HIV.gov: Treatment failure. https://clinicalinfo.hiv.gov/en/glossary/treatment-failure Accessed: 21 Nov 2022.

31. Jamieson D, Kellerman SE. The 90 90 90 strategy to end the HIV Pandemic by 2030: Can the supply chain handle it?. J Int AIDS Soc. 2016;19(1):20917. Published 2016 Jun 30. doi:10.7448/IAS.19.1.20917

32. Ernesto AS, Lemos RM, Huehara MI, et al. Usefulness of pharmacy dispensing records in the evaluation of adherence to antiretroviral therapy in Brazilian children and adolescents. Braz J Infect Dis. 2012 Jul-Aug;16(4):315-20. doi: 10.1016/j.bjid.2012.06.006.

33. Kahana SY, Fernandez MI, Wilson PA, et al. Rates and correlates of antiretroviral therapy use and virologic suppression among perinatally and behaviorally HIV-infected youth linked to care in the United States. J Acquir Immune Defic Syndr. 2015 Feb 1;68(2):169-77. doi: 10.1097/QAI.0000000000000408.

34. Kahana SY, Jenkins RA, Bruce D, et al. Adolescent Medicine Trials Network for HIV/AIDS Interventions. Structural Determinants of Antiretroviral Therapy Use, HIV Care Attendance, and Viral Suppression among Adolescents and Young Adults Living with HIV. PLoS One. 2016 Apr 1;11(4):e0151106. doi: 10.1371/journal.pone.0151106.

35. Centers for Disease Control and Prevention: Data to care. https://www.cdc.gov/hiv/effective-interventions/treat/data-to-care/index.html. Accessed: 21 Nov 2022.

36. World Health Organization webpage: Adolescent health. https://www.who.int/southeastasia/health-topics/adolescent-health Accessed: 21 Nov 2022.

37. Alabama law enforcement agency webpage. https://www.alea.gov/dps/driver-license/driver-license-information Accessed: 21 Nov 2022.

Table 1. Demographic information of clients (N=96)

Between March 2020 and August 2021, 96 youth with HIV were included in this study.

Variable		Clients
Age	Mean (SD)	20.90 (2.40)
Gender	Female	24 (25.00%)
	Male	67 (69.79%)
	Transgender	1 (1.04%)
	Unknown	4 (4.17%)
Race	AA	70 (72.92%)
	White	17 (17.71%)
	Asian	2 (2.08%)
	More than one race	4 (4.17%)
	Unknown	3 (3.13%)
HIV transmission mode	Heterosexual	18 (18.75%)
	MSM	56 (58.33%)
	Perinatal	17 (17.71%)
	MSM and IDU[a]	1 (1.04%)
	Unknown	4 (4.17%)
Insurance type	Medicaid	19 (19.79%)
	Private Individual	41 (42.71%)
	Private Employer	10(10.42%)
	VA, Tricare and other military health care	2 (2.08%)
	Other	3 (3.12%)
	No insurance	8 (8.33%)
	Unknown	13 (13.54%)
Any missed visits	Yes	51 (53.13%)
	No	45 (46.88%)
Index viral suppression	Yes	53 (55.21%)
	No	43 (44.79%)

Characteristics are summarized from their first visit record during the study period. Data are presented as n (%).

[a]The client under the MSM and IDU group was assigned to MSM group in data analysis to avoid sample size issue.

Table 2. CD4 change by missed visits.

Visit	Change (p=0.36)		More than 50 counts lose (p=0.89)	
	Fall	Not fall	Yes (>50)	No (<=50)
Miss	11 (25.00%)	33 (75.00%)	8 (18.18%)	36 (81.82%)
No miss	14 (34.15%)	27 (65.85%)	7 (17.07%)	34 (82.93%)

Table 3. Viral load change by missed visits.

Visit	Viral suppression (p=0.75)		Viral load Maintenance (**p=0.04**)	
	No(>=200)	Yes (<200)	Increase	Decrease or no change
Miss	6 (11.76%)	45 (83.33%)	31 (60.78%)	20 (39.22%)
No miss	4 (8.89%)	41 (91.11%)	18 (40.00%)	27 (60.00%)

Table 4. Risk factors associated with missed visits among YWH (n=101).

Effect		**OR**	**95% CI**	**p-value**
Age				0.32
		1.09	(0.92, 1.28)	
Sex				**0.03**
	Male vs Female	3.35	(1.10, 10.20)	
Race				0.71
	African-American vs White	1.42	(0.46, 4.40)	
	More than one race vs White	2.25	(0.29, 17.76)	
HIV risk factor				0.20
	Heterosexual vs Perinatal	0.25	(0.04, 1.45)	
	MSM vs Perinatal	1.10	(0.42, 2.93)	
Insurance type				0.64
	Medicaid vs Private - Individual	0.66	(0.20, 2.17)	
	Private - Employer vs Private - Individual	0.38	(0.10, 1.46)	
	Other vs Private - Individual	0.94	(0.12, 7.52)	
	No Insurance vs Private - Individual	1.09	(0.30, 4.01)	
Viral Suppression in first record				**0.04**
	Yes vs No	0.41	(0.18, 0.94)	

Bivariate logistic regression was conducted. Odds ratios, 95% confidence intervals, and p-values were reported. Bold p-value indicates significance at p<0.05.

PREDICTIVE ANALYTICS FOR MISSED HIV PRIMARY CARE VISITS WITH IENGAGE STUDY USING MACHINE LEARNING METHODS

by

JIAYING HAO, BYRON C. JAEGER, MEI LI, DUSTIN M. LONG

In preparation for *Journal of Acquired Immune Deficiency Syndromes*

Format adapted for dissertation

Introduction

Retention in health care is critical in achieving optimal health outcomes particularly for people living with HIV (PLWH) and also play a role in preventing HIV transmission[1]. Retention measures are numerous and include missed (no-show) clinic visits[2]. Missed visits are uniquely captured in real time by clinics, amenable to immediate intervention, and associated with deleterious HIV outcomes[3]. It has been identified that HIV patients missing three or more HIV clinic appointments have more than triple the risk of death from any cause[2]. Also, patients who missed visits within the first year after initiating HIV care had more than twice the rate of long-term mortality[4]. Paying attention to newly-initiated HIV care patients can help ensure proper and timely initiation of antiretroviral therapy (ART), monitor the effectiveness of ART, manage side effects, prevent drug resistance, and maintain low viral loads. Tracking first-year visit attendance can help identify trends and patterns in patient engagement and care and be used to inform future interventions, improve healthcare services, and contribute to HIV research.

Many studies have developed predictive models to help identify high risk in patients' HIV appointment visit attendance[5-8]. By accurately predicting missed visits, healthcare providers can proactively follow up with patients who are at higher risk of not showing up. Healthcare facilities can allocate resources more effectively to patients more likely to miss appointments. With the development of machine learning, researchers are using new techniques to identify the best-performing predictive model for the missed visits[9-11]. Machine learning algorithms can automatically identify complex patterns and relationships within large datasets and can be particularly useful in healthcare settings where patient records, demographics, and other relevant factors can be extensive[12,13]. As

more data is collected, models can be trained and updated to improve their predictive accuracy. Machine learning models can be integrated with existing electronic health record (EHR) systems and other healthcare IT infrastructure, allowing for seamless access to the relevant data and predictions, streamlining the process for healthcare providers. These techniques can ultimately contribute to improving patient care and better health outcomes for PLWH.

However, seldom have researchers used machine learning techniques in the newly-initiated HIV care patients. This study is aimed at developing and validating a predictive model for missing scheduled HIV care visits with data from a study conducted among new-to-care patients. To achieve this, we explore the innovative technique of oblique survival random forests, which offers a unique approach for analyzing time-to-event data while considering interactions among covariates[14]. The oblique survival random forests method has gained attention in recent years due to its ability to capture non-linear and interactive effects between predictors, which are often present in complex healthcare datasets. To comprehensively evaluate the performance of the oblique survival random forests model, we will compare it with other commonly used models, including axis-based random survival forests, conditional inference random survival forests, the proportional hazards Cox model, penalized Cox proportional hazards, and xgboost, and we will identify the key variables that emerge as influential predictors within the oblique survival random forests framework, shedding light on the factors that significantly impact the likelihood of missed primary care visits among HIV patients.

Methods

Data Sources and Study Population:

We used data records of Integrating ENGagement and Adherence Goals Upon Entry (iENGAGE) to Control HIV intervention trial to generate the predictive model. iENGAGE is a 4 session, in-clinic behavioral intervention that is delivered to new clinic patients during the first year of HIV care on a flexible delivery schedule, with intervention visits scheduled to coincide with HIV medical care visits[15]. iENGAGE was funded by the National Institutes of Health (NIH)/NIAID-funded (R01 AI 103661), registered in ClinicalTrials.gov (NCT01900236). The participants were assigned to intervention arm or control arm (1:1) randomly. The eligible participants were 18 years and older English-speaking adults with documented HIV infection who newly establishing HIV care at study site. The study implementation sites were the University of Alabama at Birmingham (UAB), the University of North Carolina at Chapel Hill (UNC), Johns Hopkins University (JHU) in Baltimore, and the University of Washington in Seattle (UW)[16]. Three hundred seventy-one participants were enrolled across the study sites, 369 were used for analysis, as one participant was withdrawn due to eligibility, two participants didn't show in any of the visits based on the data records.

Predictor and Outcome Variables:

The iEngage study collected data on multiple aspects: Demographic information (gender, race/ethnicity, age at first primary care visit, HIV risk factor (or HIV transmission mode), baseline insurance), Study-related information (whether started ART by a week after baseline, study site, study arm, baseline CD4 count, baseline Viral load)

and computer-administered self-interview (CASI) data. Considering of the validation of data, only the baseline variables of following instruments were included in the predictive models: Patient Health Questionnaire depression scale-PHQ-8[17], PHQ-Anxiety scale[18-20], The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST)[21], The AUDIT Alcohol Consumption Questions (AUDIT-C)[22], Coping styles assessed with Brief COPE scale[23], Medical Outcomes Study social support score (MOS-4)[24], HIV Stigma using Bunn, Solomon, Miller, and Forehand's (2007) version of the HIV stigma scale [25,26] (includes enacted, disclosure concerns, negative self-image, and public stigma) and Chronic Illness Anticipated Stigma Scale (CIASS) by Earnshaw[27] (includes family, friends, healthcare stigma), and Self-efficacy scales[28]. Table 1 lists the included instruments and names and explanations.

The outcome variable Is defined as not showing up for a scheduled HIV primary care visit based on the study records. Cancelled, or bumped visits were not considered no-show visits. Time to event was computed as the time elapsed between the initial HIV care visit and the first missed visit in days.

Time Frames:

This study used two time frames to evaluate the models: the first one is based on the iEngage study time frame, which set 96 week (672 days) observational periods as the study time frame. Event times included were between 1 to 672 days, and right-censored observations. Those with event times or last date of contact more than 672 days were censored at 672 days. However, iEngage study kept the visit record after the 96 weeks study period, up to 3 years for some participants. For this research, we considered a

longer time frame based on the data records we have collected, and performed models on both time frames. This methodology allows us to compare the models on both fixed time frame and continuous time frame.

Models:

As we have a time-to-event outcome, models included were related to either Cox proportional hazards (CPH) model or random survival forests (RSF) model. CPH models are commonly used in medical research[29] and we included 3 types: traditional CPH, penalized CPH, and extreme gradient-boosted CPH. The traditional CPH estimates the hazard ratio associated with each predictor and provides valuable insights into the relative impact of each predictor on the risk of the event[30]. Penalized CPH achieves more stable coefficient estimates, mitigating the potential effects of overfitting by applying penalties to the likelihood function[31]. Extreme gradient boosted CPH, specifically, XGBoost-COX, provides enhanced predictive performance by sequentially boosting the individual weak learners, allowing for the identification of complex interactions and non-linear relationships[32-34]. RSF models included were oblique random survival forest (ORSF), axis-based RSF, and conditional inference (CIF) RSF model. Axis-based RSF focuses on identifying meaningful splits based on survival times and censoring status[35]. ORSF model captures non-linear relationships and interactive effects among predictors, allowing for the identification of complex relationship in the data[36]. CIF RSF provides robust estimates of the associations with event by considering the conditional relationships between predictors and the survival outcome[37-39]. The details of each model and implemented packages are listed in Table 2.

Model Tuning and Evaluation:

Monte-Carlo cross validation was used with 25 runs for the data preparation. Each run was separated into 75% and 25% for training and testing respectively. For each learner, we computed Concordance (C)-statistic, and index of prediction accuracy (IPA)[40]. The C-statistic is the most frequently used evaluation metric of survival models[41]. The C-statistic measures goodness of fit for binary outcomes models and yields the probability a randomly selected patient who experienced an event had a higher risk score than a patient who had not experienced the event[42]. IPA quantifies the performance of a risk prediction model by reflecting calibration and not just discrimination[40]. It assesses how well the model's predicted probabilities align with the actual outcomes in a probabilistic sense. This means that a high index of prediction indicates that the model not only accurately discriminates between high and low-risk individuals but also assigns probabilities that are well-calibrated, increasing the confidence in the model's predictions. Both C-statistic and IPA were scaled by 100 to avoid any excessive amount of leading zeros.

Implementation:

All analyses were performed in R studio based on R version 4.2.1[43]. Missing values in predictors were imputed using nearest neighbors with the recipe function under the tidymodels package[44]. The Monte Carlo cross-validation was conducted with the package purrr [45].

Results

Study Population:

In 371 iEngage participants, 369 participants with complete records were included in this study. The participants are mostly male (79%), African American (63%), non-Hispanic ethnicity (95%), with the risk factor of men having sex with men (60%), with public paid insurance (47%), not having ART initialed at the time of study enrollment (58%). Two hundreds and sixty-seven (72.4%) participants had baseline viral load more than 10,000 copies/mL, and 234 (63.4%) participants had CD4 counts higher than 250 at baseline. Eighty-six patients either have the first missed visits after the 96-week time range or did not miss any visits even during the 96 weeks.

Table 3 shows the details of characteristics by visit status based on 96 weeks observational period. Among 369 participants, 216 (58.5%) have missed at least one scheduled visit, and 153 (41.5%) attended all visits during the 96 weeks study period.

Table 4 is based on the continuous time frame. One hundred and thirty-four participants (36%) didn't miss any scheduled HIV primary visit, while 235 participants (64%) had at least one missed HIV primary visit. The survival time of the missed visits participants ranges from 7 days to 1098 days.

Model Performance:

The data was divided into two sets: a training set consisting of 6,900 observations (randomly selected from 276 patients, repeated 25 times) and a test set consisting of 2,325 observations (remaining patients from the random selection, repeated 25 times). The observational period spanned 96 weeks. The overall C-statistics for the RSF and

CPH models were tied, with both models achieving a mean C-statistic of 72.7. Among all six models, the CIF-RSF and penalized CPH models demonstrated the highest C-statistics. Regarding the IPA score, RSF models had a slightly higher mean score compared to CPH models, with values of 10.4 and 9.7, respectively. While the ORSF model did not stand out significantly in terms of either C-statistics or IPA, its performance was still good with a C-statistic of 72 and an IPA score of 10.

With the continuous time frame, among the six learners, the CIF-CPH and Axis-based RSF models demonstrated the highest IPAs, with values of 10 and 9.4, respectively. These models excelled in capturing important survival patterns and relationships. ORSF, along with other random survival forest models, achieved an overall mean C-statistic of 71, indicating a reasonably good discriminative ability. Additionally, the ORSF model exhibited an overall mean IPA of 9.1, indicating its capability to capture relevant information within the survival curves.

With the assessments of both datasets, overall, the ORSF model demonstrated respectable performance, although it did not achieve the highest C-statistics or IPA scores compared to other models. While other models may have slightly higher C-statistics or lower IPA scores, the ORSF model demonstrates a balanced performance, making it a valuable approach within the context of this study. Nevertheless, with C-statistics of more than 70 and IPA scores of 10 and 9.4, the ORSF model exhibited good predictive capabilities within the context of this study. Table 5 and Table 6 show the model summary statistics of 96 week time frame and continuous time frame.

Predictive Model and Variable of Importance:

The most important 10 variables based on the variable of importance (VI) of both data frames were listed (Figure 1). Both plots indicate that age is the most important variable in the predictive model, followed by insurance. Race, Bunn disclosure concerns score, anticipate stigma of healthcare and friends, coping style of use of emotional support, self efficacy score, and using study intervention are all in the top 10 in both plots.

Age and insurance type emerged as the most influential variables with significantly higher Variable Importance (VI) scores. Figure 2 illustrates the relationship between insurance type, age, and the risk of missing visits based on the continuous time frame. Specifically, the analysis focused on patients aged between 20 and 50, representing the majority of the patient population. Patients without insurance coverage exhibited the lowest risk of missing visits throughout the age range. The probability of missing a visit started at 43% and gradually decreased to 35% by the age of 50. In contrast, patients with privately paid insurance began with a 48% probability and ended with a 40% probability of missing visits within the same age range. In comparison to these insurance types, patients with publicly funded insurance (such as Medicaid or Medicare) had a higher likelihood of missing visits. At the age of 20, their probability of missing a visit stood at 58%, gradually decreasing to 50% by the age of 50. In other words, individuals under publicly funded insurance coverage had a probability exceeding 50% of missing their scheduled appointments until the age of 50. Conversely, for those with private insurance or no insurance at all, the likelihood of missing appointments was less than 50% from the age of 20 onwards. Additionally, irrespective of insurance type, older individuals demonstrated a higher likelihood of adhering to their scheduled visits.

The 96 weeks dataset generated plot with high similarity, the corresponding plot has been included in the figure 3.

Discussion

The latest prevalence-based HIV continuum data reported an estimated 50% PLWH were retained in care[46]. In the National HIV/AIDS strategy for the United States 2022-2025, increasing retention in care and adherence to HIV treatment is defined as a key objective, with the goal of improving HIV-related health outcomes of people with HIV[47]. Predictive analytics can be a powerful tool to achieve this objective. Some studies have used machine learning algorithms to predict missed appointments or engagement in HIV care[48,11]. They all pointed out that retention history or previous attendance status are important variables in their models. This study applied machine learning to patients' HIV care appointment attendance data, and built a predictive model able to predict the visit status of a new to HIV care patient with the information can be collected during the initial visit. This approach allows care providers to separate high-risk and low-risk patients effectively using limited medical records.

Consistent with prior studies, our study found that specific health insurance type can make contributions to poor retention in HIV care[49]. However, there are research pointing out that patients with no insurance are more likely to miss their visits[50], which conflicts with our conclusion. Patients without health insurance may have limited access to healthcare services. Consequently, when they do have an opportunity to receive primary care, they may be more motivated to attend their appointments and make the most of the available resources. Once the patients entered the iENGAGE study, those

uninsured patients were linked to Ryan White HIV/AIDS Program to get insurance coverage. Recognizing the importance of managing their health and being linked to insurance coverage, these individuals may prioritize attending primary care appointments to address their health needs and proactively manage their conditions. Linkage to care and insurance are important to newly diagnosed HIV patients.

This study aimed to use machine learning algorithms, seek the risk factors that compose an effective predictive model on missed visits. Table 3 and Table 4 also provide insight into the possible risk factors with a simple contingency table with p-values. With different time frames, the significant risk factors from the two tables are consistent, but when we compared the results here to the results of machine learning models, it's hard to say we have found the same associations with different methods. Traditional predictive analytics can be simple and more audience-friendly since they are easier to interpret, while the machine learning algorithms can handle more model complex, non-linear relationships and interactions between variables. With machine learning techniques, we can handle large datasets. This is increasingly important when processing massive electronic medical records.

We compared the overall performances of ORSF model and other machine learning models, which are all powerful tools for survival analysis. ORSF exhibited competitive and balanced performance with the C-statistic and the IPA. Unlike traditional survival models that assume linear relationships, ORSF allows for non-linear and interactive effects, making it well-suited for identifying intricate patterns in survival data. This flexibility enables the model to better handle complex real-world scenarios where predictors may have nonlinear associations with the survival outcome. The aorsf

package incorporates various functions that streamline and expedite the model development process, making it both easier and faster compared to previous approaches. These functions facilitated the generation of VI and figures with significantly reduced effort. Users can now obtain informative and visually appealing outputs without having to invest extensive time and resources.

This study has some limitations: firstly, the sample size is relatively small. Although cross-validation has been conducted, there are chances that the complexity of the machine learning was limited. Also, small sample size may not be representative of the population we're interested in. This means the predictive model might not generalized well to new data from the same population. Secondly, although we did imputation to deal with the missingness, we are essentially making educated guesses about the missing values. Imputations can either add random noise or oversmooth the data, either may lead to misleading conclusions. Also, the iENGAGE study data doesn't include some other variables which have been identified as important risk factors of missed visits, for example, education level[7,52-54], income level[7,51,54], legal involvement[52,54], or health literacy[53,54].

Conclusions

Retention in care is crucial for HIV patients who initiated care. It is important to keep all visits to receive the optimal care and to improve overall health. This study developed a predictive model of missed visits among new to care patients using iENGAGE data by employing the oblique survival random forests technique. We compared various models utilizing different machine learning methods, and found that

ORSF model has good performance. Such a model will allow health care providers predict who is at higher risk of missing visits with limited information. And targeted intervention should be developed to patients likely to benefit most.

References

1. Clay Roscoe, MD, David M. Hachey, PharmD, "Retention in HIV Care," 2020. [Online]. Available: https://www.hiv.uw.edu/go/basic-primary-care/retention-care/core-concept/all#figures.

2. Mugavero, M. J., Westfall, A. O., Cole, S. R., Geng, E. H., Crane, H. M., Kitahata, M. M., Mathews, W. C., Napravnik, S., Eron, J. J., Moore, R. D., Keruly, J. C., Mayer, K. H., Giordano, T. P., Raper, J. L., & Centers for AIDS Research Network of Integrated Clinical Systems (CNICS) (2014). Beyond core indicators of retention in HIV care: missed clinic visits are independently associated with all-cause mortality. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 59(10), 1471–1479. https://doi.org/10.1093/cid/ciu603

3. Mugavero, M. J., Lin, H. Y., Willig, J. H., Westfall, A. O., Ulett, K. B., Routman, J. S., Abroms, S., Raper, J. L., Saag, M. S., & Allison, J. J. (2009). Missed visits and mortality among patients establishing initial outpatient HIV treatment. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 48(2), 248–256. https://doi.org/10.1086/595705

4. stfall, A. O., Ulett, K. B., Routman, J. S., Abroms, S., Raper, J. L., Saag, M. S., & Allison, J. J. (2009). Missed visits and mortality among patients establishing initial outpatient HIV treatment. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America, 48(2), 248–256. https://doi.org/10.1086/595705

5. Pence, B. W., Bengtson, A. M., Boswell, S., Christopoulos, K. A., Crane, H. M., Geng, E., Keruly, J. C., Mathews, W. C., & Mugavero, M. J. (2019). Who Will Show? Predicting Missed Visits Among Patients in Routine HIV Primary Care in the United States. AIDS and behavior, 23(2), 418–426. https://doi.org/10.1007/s10461-018-2215-1

6. Pettit, A. C., Bian, A., Schember, C. O., Rebeiro, P. F., Keruly, J. C., Mayer, K. H., Mathews, W. C., Moore, R. D., Crane, H. M., Geng, E., Napravnik, S., Shepherd, B. E., & Mugavero, M. J. (2021). Development and Validation of a Multivariable Prediction Model for Missed HIV Health Care Provider Visits in a Large US Clinical Cohort. Open forum infectious diseases, 8(7), ofab130. https://doi.org/10.1093/ofid/ofab130

7. Traeger, L., et al., Risk factors for missed HIV primary care visits among men who have sex with men. J Behav Med, 2012. 35(5): p. 548-56.

8. Kay, E.S., A. Lacombe-Duncan, and R.M. Pinto, Predicting Retention in HIV Primary Care: Is There a Missed Visits Continuum Based on Patient Characteristics? AIDS and Behavior, 2019. 23: p. 2542-2548.

9. Weissman, S., Yang, X., Zhang, J., Chen, S., Olatosi, B., & Li, X. (2021). Using a machine learning approach to explore predictors of healthcare visits as missed opportunities for HIV diagnosis. *AIDS (London, England)*, *35*(Suppl 1), S7–S18. https://doi.org/10.1097/QAD.0000000000002735 ,

10. Mason JA, Friedman EE, Rojas JC, Ridgway JP
No-show Prediction Model Performance Among People With HIV: External Validation Study J Med Internet Res 2023;25:e43277
doi: 10.2196/43277PMID: 36989038PMCID: 10131690 ,

11. Ramachandran, A., Kumar, A., Koenig, H. *et al.* Predictive Analytics for Retention in Care in an Urban HIV Clinic. *Sci Rep* **10**, 6421 (2020). https://doi.org/10.1038/s41598-020-62729-x

12. Habehh, H., & Gohel, S. (2021). Machine Learning in Healthcare. Current genomics, 22(4), 291–300. https://doi.org/10.2174/1389202922666210705124359

13. Javaid, M.; Haleem, A.; Singh, R.P.; Suman, R.; Rab, S. Significance of machine learning in healthcare: Features, pillars and applications. Int. J. Intell. Networks 2022, 3, 58–73. [Google Scholar] [CrossRef]

14. Byron C. Jaeger, D. Leann Long, Dustin M. Long, Mario Sims, Jeff M. Szychowski, Yuan-I Min, Leslie A. Mcclure, George Howard, Noah Simon. "Oblique random survival forests." The Annals of Applied Statistics, 13(3) 1847-1883 September 2019. https://doi.org/10.1214/19-AOAS1261

15. Integrating ENGagement and Adherence Goals Upon Entry iENGAGE to Control HIV (iENGAGE). ClinicalTrails.gov website. Available:https://clinicaltrials.gov/ct2/show/NCT01900236

16. Modi, R. A., Mugavero, M. J., Amico, R. K., Keruly, J., Quinlivan, E. B., Crane, H. M., Guzman, A., Zinski, A., Montue, S., Roytburd, K., Church, A., & Willig, J. H. (2017). A Web-Based Data Collection Platform for Multisite Randomized Behavioral Intervention Trials: Development, Key Software Features, and Results of a User Survey. JMIR research protocols, 6(6), e115. https://doi.org/10.2196/resprot.6768

17. Kroenke K, Spitzer RL. The PHQ-9: a new depression diagnostic and severity measure. Psychiatric annals 2002; 32(9):509–515.

18. Kroenke K, Spitzer RL, Williams JB, Lowe B. The Patient Health Questionnaire Somatic, Anxiety, and Depressive Symptom Scales: a systematic review. Gen Hosp Psychiatry. 2010;32(4):345–59. [PubMed] [Google Scholar]

19. Gilbody S, Richards D, Brealey S, Hewitt C. Screening for depression in medical settings with the Patient Health Questionnaire (PHQ): a diagnostic meta-analysis. J Gen Intern Med. 2007;22:1596–602. [PMC free article] [PubMed] [Google Scholar]

20. Kroenke K, Spitzer RL, Williams JBW, Monahan PO, Lowe B. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. Ann Intern Med. 2007;146(5):317–25. [PubMed] [Google Scholar]

21. Group, W.A.W. (2002), The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST): development, reliability and feasibility. Addiction, 97: 1183-1194. https://doi.org/10.1046/j.1360-0443.2002.00185.x

22. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA, for the Ambulatory Care Quality Improvement Project (ACQUIP). The AUDIT Alcohol Consumption Questions (AUDIT-C): An Effective Brief Screening Test for Problem Drinking. Arch Intern Med. 1998;158(16):1789–1795. doi:10.1001/archinte.158.16.1789

23. Carver CS. You want to measure coping but your protocol's too long: consider the brief COPE. Int J Behav Med 1997; 4(1):92–100. [PubMed] [Google Scholar] [Ref list]

24. Sherbourne CD and Stewart AL. The MOS Social Support Survey. Soc Sci Med, 1991;32(6):705-714. PMID:2035047

25. Bunn JY, Solomon SE, Miller C, Forehand R. Measurement of stigma in people with HIV: a reexamination of the HIV Stigma Scale. AIDS Educ Prev 2007; 19(3):198–208. [PubMed] [Google Scholar] [Ref list]

26. Berger B., Ferrans C.E., Lashley F.R. (2001). Measuring stigma in people with HIV: Psychometric assessment of the HIV Stigma Scale. Research in Nursing & Health, 24, 518–529.

27. Earnshaw, V. A., Quinn, D. M., & Park, C. L. (2012). Anticipated stigma and quality of life among people living with chronic illnesses. Chronic illness, 8(2), 79-88.

28. Bandura A. Guide for constructing self-efficacy scales. In: Pajares F, Urdan TC, editors. Self-efficacy beliefs of adolescents. Greenwich: Information Age Publishing; 2006. p. 307–37.

29. Cox DR (1972). Regression models and life tables (with discussion). J R Statist Soc B 34: 187–220

30. Deo, S. V., Deo, V., & Sundaram, V. (2021). Survival analysis-part 2: Cox proportional hazards model. Indian journal of thoracic and cardiovascular surgery, 37(2), 229–233. https://doi.org/10.1007/s12055-020-01108-7

31. Goeman J. J. (2010). L1 penalized estimation in the Cox proportional hazards model. Biometrical journal. Biometrische Zeitschrift, 52(1), 70–84. https://doi.org/10.1002/bimj.200900028

32. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system", in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, pp. 785-794, 2016.

33. L. Yang and K. Pelckmans, "Machine learning approaches to survival analysis: Case studies in microarray for breast cancer", Int. J. Mach. Learn.Comput., vol. 4, no. 6, Dec. 2014.

34. P. Liu, B. Fu, S. X. Yang, L. Deng, X. Zhong and H. Zheng, "Optimizing Survival Analysis of XGBoost for Ties to Predict Disease Progression of Breast Cancer," in IEEE Transactions on Biomedical Engineering, vol. 68, no. 1, pp. 148-160, Jan. 2021, doi: 10.1109/TBME.2020.2993278.

35. Stefan Dietrich and others, Random Survival Forest in practice: a method for modelling complex metabolomics data in time to event analysis, International Journal of Epidemiology, Volume 45, Issue 5, October 2016, Pages 1406–1420, https://doi.org/10.1093/ije/dyw145

36. Jaeger, B. C., Long, D. L., Long, D. M., Sims, M., Szychowski, J. M., Min, Y.-I., Mcclure, L. A., Howard, G., & Simon, N. (2019). OBLIQUE RANDOM SURVIVAL FORESTS. The Annals of Applied Statistics, 13(3), 1847–1883. https://www.jstor.org/stable/26866706

37. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. BMC Bioinforma. 2007; 8(1):1.

38. Wright MN, Dankowski T, Ziegler A. Unbiased split variable selection for random survival forests using maximally selected rank statistics. Stat Med. 2017; 36(8):1272–84. doi:10.1002/sim.7212.sim.7212.

39. Das A, Abdel-Aty M, Pande A. Using conditional inference forests to identify the factors affecting crash severity on arterial corridors. J Saf Res. 2009; 40(4):317–27.

40. Kattan, M.W., Gerds, T.A. The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. Diagn Progn Res 2, 7 (2018). https://doi.org/10.1186/s41512-018-0029-2

41. S. Pölsterl, “scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn,” Journal of Machine Learning Research, vol. 21, no. 212, pp. 1–6, 2020.

42. Stephanie Glen. "C-Statistic: Definition, Examples, Weighting and Significance" From StatisticsHowTo.com: Elementary Statistics for the rest of us! https://www.statisticshowto.com/c-statistic/

43. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

44. Kuhn et al., (2020). Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. https://www.tidymodels.org

45. Wickham H, Henry L (2023). _purrr: Functional Programming Tools_. R package version

46. HIV Care Continuum. HIV.gov website. 2022. https://www.hiv.gov/federal-response/policies-issues/hiv-aids-care-continuum/

47. National HIV/AIDS strategy for the United States 2022-2025. Website: Whitehouse.gov, 2021. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.whitehouse.gov/wp-content/uploads/2021/11/National-HIV-AIDS-Strategy.pdf

48. Ridgway, J. P., Ajith, A., Friedman, E. E., Mugavero, M. J., Kitahata, M. M., Crane, H. M., Moore, R. D., Webel, A., Cachay, E. R., Christopoulos, K. A., Mayer, K. H., Napravnik, S., & Mayampurath, A. (2022). Multicenter Development and Validation of a Model for Predicting Retention in Care Among People with HIV. AIDS and behavior, 26(10), 3279–3288. https://doi.org/10.1007/s10461-022-03672-y

49. Hessol NA, Weber KM, Holman S, et al. Retention and attendance of women enrolled in a large prospective study of HIV-1 in the United States. J Womens Health (Larchmt) 2009; 18: 1627–1637.

50. Kay, E.S., et al., Health insurance and AIDS Drug Assistance Program (ADAP) increases retention in care among women living with HIV in the United States. AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV, 2021. 33(8): p. 1044-1051.

51. Hightow-Weidman, L., et al., Exploring the HIV continuum of care among young black MSM. PLoS ONE, 2017. 12(6).

52. Tarantino, N., et al., Correlates of missed clinic visits among youth living with HIV. AIDS Care, 2018. 30(8): p. 982-989.

53. Fazeli, P.L., et al., Critical, and not functional, health literacy is associated with missed HIV clinic visits in adults and older adults living with HIV in the Deep South. AIDS Care - Psychological and Socio-Medical Aspects of AIDS/HIV, 2020. 32(6): p. 694-700.

54. Menza, T.W., et al., Social Determinants of Health and Care Outcomes Among People With HIV in the United States. Open Forum Infect Dis, 2021. 8(7): p. ofab330.

55. Jaeger et al. (2022). aorsf: An R package for supervised learning using the oblique random survival forest. Journal of Open Source Software, 7(77), 4705. https://doi.org/10.21105/joss.04705.

56. Marvin N. Wright, Andreas Ziegler (2017). ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software, 77(1), 1-17. doi:10.18637/jss.v077.i01

57. Hothorn T, Hornik K, Zeileis A (2006). “Unbiased Recursive Partitioning: A Conditional Inference Framework.” Journal of Computational and Graphical Statistics, *15*(3), 651-674. doi:10.1198/106186006X133933 https://doi.org/10.1198/106186006X133933.

58. Zeileis A, Hothorn T, Hornik K (2008). “Model-Based Recursive Partitioning.” Journal of Computational and Graphical Statistics, 17(2), 492-514. doi:10.1198/106186008X319331 https://doi.org/10.1198/106186008X319331.

59. Hothorn T, Buehlmann P, Dudoit S, Molinaro A, Van Der Laan M (2006). “Survival Ensembles.” Biostatistics, 7(3), 355-373.

60. Strobl C, Boulesteix A, Zeileis A, Hothorn T (2007). “Bias in Random Forest Variable Importance Measures: Illustrations, Sources and a Solution.” BMC Bioinformatics, 8(25). doi:10.1186/1471-2105-8-25. https://doi.org/10.1186/1471-2105-8-25.

61. Strobl C, Boulesteix A, Kneib T, Augustin T, Zeileis A (2008). “Conditional Variable Importance for Random Forests.” BMC Bioinformatics, 9(307). doi:10.1186/1471-2105-9-307. https://doi.org/10.1186/1471-2105-9-307.

62. Therneau T (2023). A Package for Survival Analysis in R. R package version 3.5-0, https://CRAN.R-project.org/package=survival.

63. Terry M. Therneau, Patricia M. Grambsch (2000). Modeling Survival Data: Extending the Cox Model. Springer, New York. ISBN 0-387-98784-3.

64. Jerome Friedman, Trevor Hastie, Robert Tibshirani (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. Journal of Statistical Software, 33(1), 1-22. URL https://www.jstatsoft.org/v33/i01/.

65. Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, Chen K, Mitchell R, Cano I, Zhou T, Li M, Xie J, Lin M, Geng Y, Li Y, Yuan J (2023). xgboost: Extreme Gradient Boosting. R package version 1.7.3.1, <https://CRAN.R-project.org/package=xgboost>.

Table 1. Instruments implemented in iENGAGE study.

Instrument	Variable Name	Explanation	Measure
Patient Health Questionnaire depression scale	PHQ8	Assessed based on 8 items	3 categories: a. no depressive disorder (<10); b.Major depression (10-19); c. Severe depression (>=20)
PHQ-Anxiety scale	PHQAnx	Assessed based on 5 items	3 categories: a. None; b. Panic symtoms; c.Panic disorder
The Alcohol, Smoking and Substance Involvement Screening Test (ASSIST)	ASSIST_Crack	ASSIST: Cocaine or crack	3 categories: a. Never; b.Prior; c.Current
	ASSIST_Amp	ASSIST: Amphetamines	3 categories: a. Never; b.Prior; c.Current
	ASSIST_Opi	ASSIST: Opiates	3 categories: a. Never; b.Prior; c.Current
	ASSIST_Mari	ASSIST: Marijuana	3 categories: a. Never; b.Prior; c.Current
	ASSIST_IVDU	ASSIST: IVDU	3 categories: a. Never; b.Prior; c.Current
	ASSIST_Substance	ASSIST: Substance use (not including marijuana)	3 categories: a. Never; b.Prior; c.Current
The AUDIT Alcohol Consumption Questions	AUDIT-C	Assessed based on 3 items	3 categories: a. No risk; b. Low risk; c. High risk
Coping styles assessed with Brief COPE scale	BriefCOPE_ActiveCoping	BriefCOPE: Active coping, items 2 and 7	Continuous: 2-8
	BriefCOPE_Denial	BriefCOPE: Denial, items 3 and 8	Continuous: 2-8
	BriefCOPE_SubstanceUse	BriefCOPE: Substance use, items 4 and 11	Continuous: 2-8
	BriefCOPE_UseOfEmotionalSupport	BriefCOPE: Use of emotional support, items 5 and 15	Continuous: 2-8

Instrument	Variable Name	Explanation	Measure
	BriefCOPE_Behavioral Disengage	BriefCOPE: Behavioral disengagement, items 6 and 16	Continuous: 2-8
	BriefCOPE_PositiveReframing	BriefCOPE: Positive reframing, items 12 and 17	Continuous: 2-8
	BriefCOPE_Acceptance	BriefCOPE: Acceptance, items 20 and 24	Continuous: 2-8
	BriefCOPE_Religion	BriefCOPE: Religion, items 22 and 27	Continuous: 2-8
	BriefCOPE_SelfBlame	BriefCOPE: Self-blame, items 13 and 26	Continuous: 2-8
Medical Outcomes Study social support score	MOS4Score	Assessed on 4 items.	Continuous: 0-100
HIV Stigma using Bunn, Solomon, Miller, and Forehand's version of the HIV stigma scale	BunnEnactedStigma	Bunn ENACTED STIGMA score (29,21,24,28,25,32,27,16,31,26,23)	Continuous: 1-4
	BunnDisclosureConcerns	Bunn DISCLOSURE CONCERNS score (18R,22,15,6,30,1,4,19)	Continuous: 1-4
	BunnNegativeSelfImage	Bunn NEGATIVE SELF IMAGE score (13,7,11,20,3,2,8R)	Continuous: 1-4
	BunnPublicStigma	Bunn PUBLIC STIGMA score (12,14,5,10,17,9)	Continuous: 1-4
Chronic Illness Anticipated Stigma Scale (CIASS) by Earnshaw	AnticipatedStigma_Family	Earnshaw anticipated FAMILY stigma	Continuous: 1-5
	AnticipatedStigma_Friends	Earnshaw anticipated FRIENDS stigma	Continuous: 1-5
	AnticipatedStigma_Healthcare	Earnshaw anticipated HEALTHCARE stigma	Continuous: 1-5
Self-efficacy scales	SelfEfficacyScore2	Self efficacy score 2 (average of non-missing responses)	Continuous: 0-10

Table 2. Machine learning algorithms assessed in this study.

Model	Package	Description
Random Survival Forests		The ORSF model combines elements of survival analysis and random forests. The Axis-based RSF is a variant of the RSF approach that utilizes axis splits to divide the predictor space. The CIF-RSF is an extension of the RSF approach that incorporates conditional inference trees.
Oblique RSF (ORSF)	Aorsf[55]	
Axis based RSF	Ranger[56]	
Conditional Inference RSF (CIF-RSF)	Party[57-61]	
Cox Proportional Hazards		The CPH assumes that the hazard function is proportional across different levels of the predictors. The Penalized CPH extends the original CPH model by incorporating a penalized or restricted likelihood approach. The XGBoost-Cox implemented using the XGBoost framework, combines gradient boosting techniques with the CPH model.
Original CPH	Survival[62,63]	
Penalized CPH	Glmnet[64]	
Gradient Boosted CPH(XGBoost-Cox)	Xgboost[65]	

Table 3. Characteristics of participants by visit status in 96 weeks

		No missed	%	Missed	%	Total	p-value
Race							
	Black or African American	83	35.93	148	64.07	231	0.02
	White	56	51.85	52	48.15	108	
	Asian	6	66.67	3	33.33	9	
	Native American	1	50	1	50	2	
	Other	7	36.84	12	63.16	19	
Ethnicity							
	Hispanic	9	47.37	10	52.63	19	0.59
	Non-Hispanic	144	41.14	206	58.86	350	
Gender							
	Female	27	38.03	44	61.97	71	0.75
	Male	104	35.62	188	64.38	292	
	Transgender	3	50	3	50	6	
Risk Factor							
	MSM	97	44.5	121	55.5	218	0.05
	Heterosexual	48	41.03	69	58.97	117	
	IVDU	6	20.69	23	79.31	29	
Insurance							
	Public	49	28.32	124	71.68	173	<0.001
	Private	56	52.34	51	47.66	107	
	None	46	52.87	41	47.13	87	
Started ART by a week after baseline							
	Yes	69	44.52	86	55.48	155	0.31
Study Site							
	UAB	71	46.41	82	53.59	153	0.01
	UNC	37	49.33	38	50.67	75	
	JHU	21	26.92	57	73.08	78	
	UW	24	38.1	39	61.9	63	
Study Arm							
	Control	71	38.38	114	61.62	185	0.23
	Intervention	82	44.57	102	55.43	184	
Age							
		38.44	12.58	33.77	11.09	35.71	<0.001
Baseline Viral Load							
	<=20	1	33.33	2	66.67	3	0.61
	21-200	5	41.67	7	58.33	12	
	201-1000	4	25	12	75	16	
	1001-10000	25	39.06	39	60.94	64	
	>10000	118	44.19	149	55.81	267	
Baseline CD4							
	<=250	44	41.12	63	58.88	107	0.78
	>250	100	42.74	134	57.26	234	

Table 4. Characteristics of participants by visit status in continuous time frame

	No missed	%	Missed	%	Total	p-value
Race						
Black or African American	70	30.3	161	69.7	231	0.01
White	50	46.3	58	53.7	108	
Asian	6	66.67	3	33.33	9	
Native American	1	50	1	50	2	
Other	7	36.84	12	63.16	19	
Ethnicity						
Hispanic	9	47.37	10	52.63	19	0.3
Non-Hispanic	125	35.71	225	64.29	350	
Gender						
Female	27	38.03	44	61.97	71	0.75
Male	104	35.62	188	64.38	292	
Transgender	3	50	3	50	6	
Risk Factor						
MSM	87	39.91	131	60.09	218	0.06
Heterosexual	42	35.9	75	64.1	117	
IVDU	5	17.24	24	82.76	29	
Insurance						
Public	43	24.86	130	75.14	173	<0.001
Private	54	50.47	53	49.53	107	
None	35	40.23	52	59.77	87	
Started ART by a week after baseline						
Yes	61	39.35	94	60.65	155	0.3
Study Site						
UAB	64	41.83	89	58.17	153	0.17
UNC	26	34.67	49	65.33	75	
JHU	21	26.92	57	73.08	78	
UW	23	36.51	40	63.49	63	
Study Arm						
Control	62	33.51	123	66.49	185	0.26
Intervention	72	39.13	112	60.87	184	
Age						
	38.57	12.9	32.08	11.05	35.71	0.02
Baseline Viral Load						
<=20	1	33.33	2	66.67	3	0.07
21-200	3	25	9	75	12	
201-1000	2	12.5	14	87.5	16	
1001-10000	19	29.69	45	70.31	64	
>10000	109	40.82	158	59.18	267	
Baseline CD4						
<=250	40	37.38	67	62.62	107	0.9
>250	86	36.75	148	63.25	234	

Table 5. Model performance statistics: 96 weeks

	Random Survival Forest (RSF)			Cox Proportional Hazards (CPH)		
	Oblique	Axis-based	Conditional Inference	Original	Penalized	XGBoost
C-Statistic	72	72	74	73	74	71
Index of Prediction Accuracy	10	8.1	13	11	9.1	8.9

Table 6. Model performance statistics: Continuous time frame

	Random Survival Forest (RSF)			Cox Proportional Hazards (CPH)		
	Oblique	Axis-based	Conditional Inference	Original	Penalized	XGBoost
C-Statistic	71	71	71	69	72	68
Index of Prediction Accuracy	9.4	7.8	10	7.3	8.1	5.9

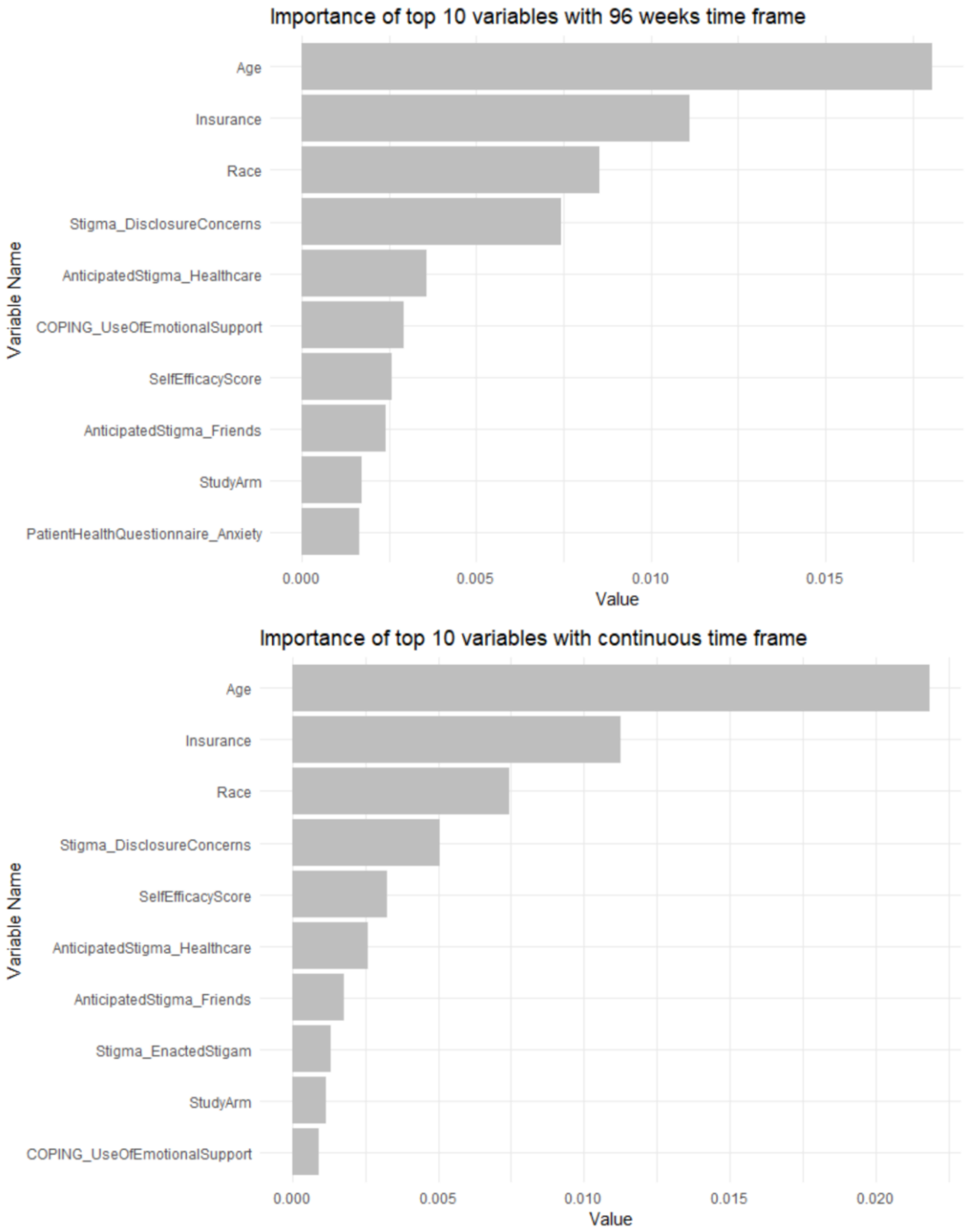


Figure 1. Variable of Importance (VI) of top 10 variables.

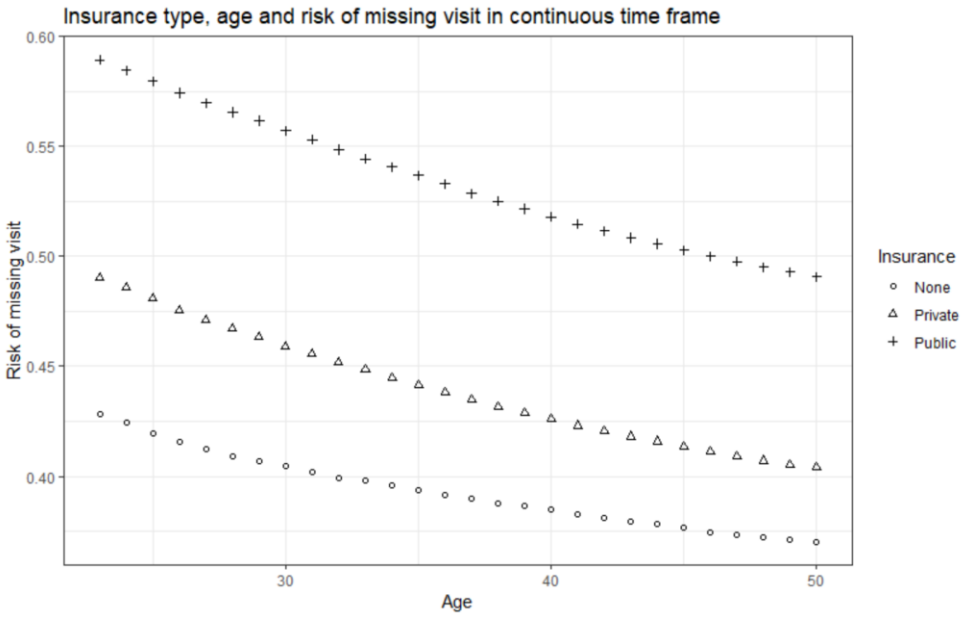


Figure 2. Survival of different insurance types of 20-50 years old patients by ORSF model based on continuous time frame.

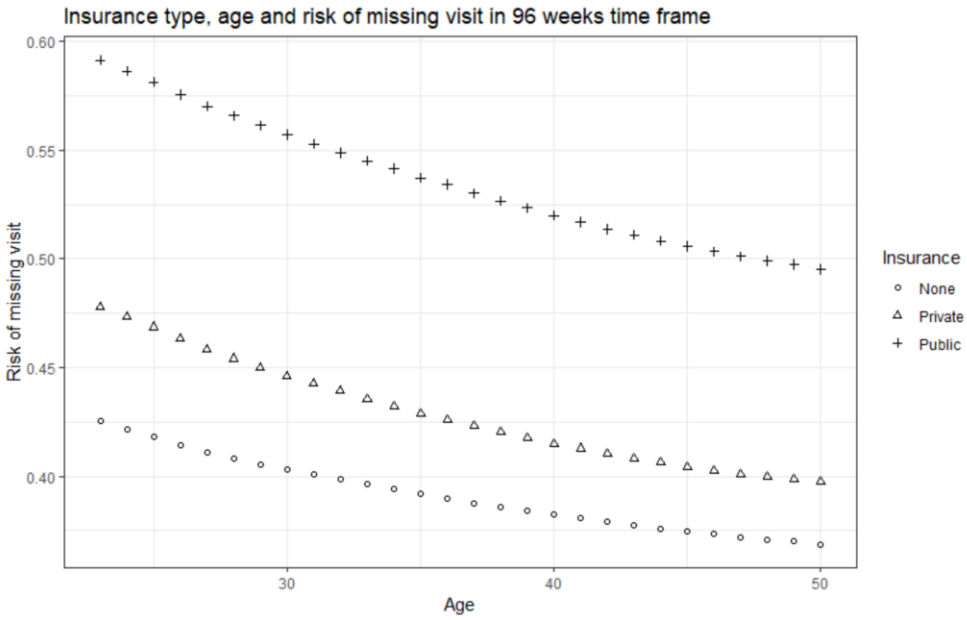


Figure 3. Survival of different insurance types of 20-50 years old patients by ORSF model based on 96 weeks data

CONCLUSION AND FUTURE DIRECTIONS

Keeping HIV care appointments is important for individuals with HIV[30]. HIV primary visit is necessary for monitoring the progress of the disease, adjusting treatment plans, providing emotional support, and educating patients[31]. Missed visits could potentially lead to poorer health outcomes, such as lack of viral suppression, increased chance for transmitting the virus, and higher mortality rates[32-38].

Research was underway to understand the factors that contribute to missed visits. Some of these included socioeconomic factors, mental health issues, stigma associated with HIV, and logistical barriers such as transportation or scheduling issues[39-42].

The goal of this dissertation was to provide a comprehensive understanding of HIV primary visits by exploring the risk factors associated with missed visits. The systematic review highlighted the complexity of the issues of HIV missed visits. The review indicates that demographic and socioeconomic risk factors, such as age, gender, race, education level, income level, insurance type, etc., are important risk factors that have significant effects on missed visits. By examining these factors, researchers can identify disparities and inequalities in healthcare access, utilization, and outcomes among different population groups. It is evident that results vary across studies due to differences in population characteristics, methodological approaches, and the use of different

statistical methods. It is essential to acknowledge this heterogeneity when interpreting findings and formulating implications for future research and clinical practice.

In the systematic review, only three included patients under 18 years old[43-45]. The lack of focused studies on missed HIV visits among adolescents and youth is a concerning issue considering their elevated risk of HIV transmission and potential challenges in retaining them in care. This dissertation explores the relationships between missed visits among youth with HIV (YWH) in Alabama who were 16 to 24 years old and their measurable HIV treatment outcomes and risk factors associated with missed visits among this population. We found evidence to support the idea that missed visits are associated with treatment outcomes among YWH. Specifically, being male and being virally unsuppressed at their index visit were associated with missed visits in our sample. Various demographic factors, such as young gay and bisexual men navigating identity development, economically disadvantaged youth, those residing in high HIV incidence areas, and even females, have been linked to missed visits, indicating a multifaceted range of influences on adherence to healthcare appointments. The focused investigation on the impact of missed visits on health outcomes among youth with HIV in Alabama unveiled significant findings that extend our understanding within this specific demographic. It reinforces the need for interventions to ensure better engagement with care in this particularly vulnerable population.

Finally, we ventured into the realm of predictive analytics using machine learning methods among new-to-care patients. This study demonstrated how these cutting-edge techniques can leverage complex data to identify patterns that might be less apparent with

traditional statistical methods. This approach offers exciting possibilities for developing predictive models that could be used to flag patients at risk of missing appointments with limited information, enabling early interventions.

Taken together, these three studies illustrate the multifaceted nature of missed HIV visits, both in terms of the factors that contribute to it and the potential strategies to address it. The findings underscore the need for a comprehensive approach to HIV care that not only provides medical treatment but also addresses the broader social, economic, and psychological factors that impact care engagement. The potential of predictive analytics to inform such an approach is promising, suggesting a direction for future research and practice in this field.

Tracking missed visits provides a straightforward and practical way to assess retention in HIV care, it is easily measured and can be monitored over time[11]. It can be identified and addressed in real-time, allowing for timely interventions to prevent further disengagement from care. Using missed HIV visits as a measure of retention in HIV care provides a practical and valuable approach to assessing the continuity of care and identifying individuals who may require additional support or interventions to remain engaged in HIV treatment and prevention services[46]. Future research should continue to explore the integration of predictive analytics in clinical practice, including testing the efficacy of interventions informed by these methods. Studies should also seek to investigate the impact of specific interventions targeted at the risk factors identified in this research, particularly for vulnerable groups such as youth. Ultimately, the goal is to provide the most effective and personalized care to people living with HIV, supporting

them in maintaining consistent engagement with care to achieve the best possible health outcomes.

GENERAL LIST OF REFERENCES

1. Global Statistics, HIV.gov: https://www.hiv.gov/hiv-basics/overview/data-and-trends/global-statistics

2. CDC. Estimated HIV incidence and prevalence in the United States 2015–2019pdf icon. HIV Surveillance Supplemental Report 2021;26(1).

3. U.S. Statistics, HIV.gov: https://www.hiv.gov/hiv-basics/overview/data-and-trends/statistics

4. 2019 State of Alabama HIV Surveillance annual report, Alabama department of Public Health: https://www.alabamapublichealth.gov/hiv/assets/hivsurveillanceannualreport_2019.pdf

5. Horstmann E, Brown J, Islam F, Buck J, Agins BD. Retaining HIV-infected patients in care: Where are we? Where do we go from here? Clin Infect Dis. 2010;50:752-61.

6. Centers for Disease Control and Prevention. Monitoring Selected National HIV Prevention and Care Objectives by Using HIV Surveillance Data United States and 6 Dependent Areas, 2018 HIV Surveillance Supplemental Report. 2020;25(No. 2):1-104. Published May 2020.

7. Doshi RK, Milberg J, Isenberg D, et al. High rates of retention and viral suppression in the US HIV safety net system: HIV care continuum in the Ryan White HIV/AIDS Program, 2011. Clin Infect Dis. 2015;60:117-25.

8. Health Resources and Services Administration. Ryan White HIV/AIDS Program Annual Client-Level Data Report 2018. December 2019:1-139.

9. Rebeiro PF, Horberg MA, Gange SJ, et al. Strong agreement of nationally recommended retention measures from the Institute of Medicine and Department of Health and Human Services. PLoS One. 2014;9:e111772.

10. Mugavero MJ, Davila JA, Nevin CR, Giordano TP. From access to engagement: measuring retention in outpatient HIV clinical care. AIDS Patient Care STDS. 2010;24:607-13.

11. Retention in HIV Care: Core concept. https://www.hiv.uw.edu/go/basic-primary-care/retention-care/core-concept/all#citations/

12. History of missed medical visits predicts who is most likely to fall out of HIV care. Gillings school news. https://sph.unc.edu/sph-news/history-of-missed-medical-visits-predicts-who-is-most-likely-to-fall-out-of-hiv-care/

13. HIV Care Continuum. HIV.gov: https://www.hiv.gov/federal-response/policies-issues/hiv-aids-care-continuum

14. Ramachandran, A., Kumar, A., Koenig, H. et al. Predictive Analytics for Retention in Care in an Urban HIV Clinic. Sci Rep 10, 6421 (2020). https://doi.org/10.1038/s41598-020-62729-x

15. Ulett, K. B. et al. The therapeutic implications of timely linkage and early retention in HIV care. AIDS Patient Care STDS 23, 41–49, https://doi.org/10.1089/apc.2008.0132 (2009).

16. Gardner, E. M., McLees, M. P., Steiner, J. F., Del Rio, C. & Burman, W. J. The spectrum of engagement in HIV care and its relevance to test-and-treat strategies for prevention of HIV infection. Clin. Infect. Dis. 52, 793–800, https://doi.org/10.1093/cid/ciq243 (2011).

17. The Lancet. H. I. V. U=U taking off in 2017. Lancet HIV. 4, e475, https://doi.org/10.1016/S2352-3018(17)30183-2 (2017).

18. Centers for Disease Control and Prevention. Monitoring selected national HIV prevention and care objectives by using HIV surveillance data—United States and 6 dependent areas, 2019. HIV Surveillance Supplemental Report 2021;26(No.2). http://www.cdc.gov/hiv/library/reports/hiv-surveillance.html. Published May 2021.

19. Wohl, D. A. et al. Financial Barriers and Lapses in Treatment and Care of HIV-Infected Adults in a Southern State in the United States. AIDS Patient Care STDS 31, 463–469, https://doi.org/10.1089/apc.2017.0125 (2017).

20. Kim, M. M. et al. Healthcare barriers among severely mentally ill homeless adults: evidence from the five-site health and risk study. Adm. Policy Ment. Health 34, 363–375, https://doi.org/10.1007/s10488-007-0115-1 (2007).

21. Masson, C. L., Sorensen, J. L., Phibbs, C. S. & Okin, R. L. Predictors of medical service utilization among individuals with co-occurring HIV infection and substance abuse disorders. AIDS Care 16, 744–755, https://doi.org/10.1080/09540120412331269585 (2004).

22. Aidala, A. Inequality and HIV: The role of housing. Psychology and AIDS 34 (2006).

23. Pecoraro, A. et al. Factors contributing to dropping out from and returning to HIV treatment in an inner city primary care HIV clinic in the United States. AIDS Care 25, 1399–1406, https://doi.org/10.1080/09540121.2013.772273 (2013).

24. Cunningham, C. O. et al. Factors associated with returning to HIV care after a gap in care in New York State. J. Acquir. Immune Defic. Syndr. 66, 419–427, https://doi.org/10.1097/qai.0000000000000171 (2014).

25. Giordano, T. P., Hartman, C., Gifford, A. L., Backus, L. I. & Morgan, R. O. Predictors of retention in HIV care among a national cohort of US veterans. HIV. Clin. Trials 10, 299–305, https://doi.org/10.1310/hct1005-299 (2009).

26. Giordano, T. et al. Patients referred to an urban HIV clinic frequently fail to establish care: Factors predicting failure. AIDS care 17, 773–783, https://doi.org/10.1080/09540120412331336652 (2005).

27. Cook, J. A. et al. Illicit drug use, depression and their association with highly active antiretroviral therapy in HIV-positive women. Drug. Alcohol. Dependence 89, 74–81, https://doi.org/10.1016/j.drugalcdep.2006.12.002 (2007).

28. Zuniga, J. A., Yoo-Jeong, M., Dai, T., Guo, Y. & Waldrop-Valverde, D. The Role of Depression in Retention in Care for Persons Living with HIV. AIDS Patient Care STDS 30, 34–38, https://doi.org/10.1089/apc.2015.0214 (2016).

29. Mugavero, M. J., Westfall, A. O., Zinski, A., Davila, J., Drainoni, M. L., Gardner, L. I., Keruly, J. C., Malitz, F., Marks, G., Metsch, L., Wilson, T. E., Giordano, T. P., & Retention in Care (RIC) Study Group (2012). Measuring retention in HIV care: the elusive gold standard. Journal of acquired immune deficiency syndromes (1999), 61(5), 574–580. https://doi.org/10.1097/QAI.0b013e318273762f

30. Dietz, E., Clum, G. A., Chung, S. E., Leonard, L., Murphy, D. A., Perez, L. V., Harper, G. W., & Ellen, J. M. (2010). Adherence to scheduled appointments among HIV-infected female youth in five U.S. cities. The Journal of adolescent health : official publication of the Society for Adolescent Medicine, 46(3), 278–283. https://doi.org/10.1016/j.jadohealth.2009.06.013

31. Abdulrahman, S.A., et al., HIV Treatment Adherence - A Shared Burden for Patients, Health-Care Providers, and Other Stakeholders. AIDS Rev, 2019. 21(1): p. 28-39.

32. Higa, D.H., et al., Identifying Best Practices for Increasing Linkage to, Retention, and Re-engagement in HIV Medical Care: Findings from a Systematic Review, 1996-2014. AIDS Behav, 2016. 20(5): p. 951-66.

33. Majeroni, B.A., et al., Missed appointments and Medicaid managed care. Arch Fam Med, 1996. 5(9): p. 507-11.

34. Sharp, D. and W. Hamilton, Non-attendance at general practices and outpatient clinics - Local systems are needed to address local problems. British Medical Journal, 2001. 323(7321): p. 1081-1082.

35. Waller, J. and P. Hodgkin, Defaulters in general practice: who are they and what can be done about them? Family Practice, 2000. 17(3): p. 252-253.

36. Horberg, M.A., et al., Missed office visits and risk of mortality among HIV-infected subjects in a large healthcare system in the United States. AIDS Patient Care STDS, 2013. 27(8): p. 442-9.

37. Robbins, G.K., et al., Predictors of antiretroviral treatment failure in an urban HIV clinic. Jaids-Journal of Acquired Immune Deficiency Syndromes, 2007. 44(1): p. 30-37.

38. Brennan, A.T., et al., The importance of clinic attendance in the first six months on antiretroviral treatment: a retrospective analysis at a large public sector HIV clinic in South Africa. Journal of the International Aids Society, 2010. 13.

39. Catz, S.L., et al., Predictors of outpatient medical appointment attendance among persons with HIV. Aids Care-Psychological and Socio-Medical Aspects of Aids/Hiv, 1999. 11(3): p. 361-373.

40. Lucas, G.M., R.E. Chaisson, and R.D. Moore, Highly active antiretroviral therapy in a large urban clinic: Risk factors for virologic failure and adverse drug reactions. Annals of Internal Medicine, 1999. 131(2): p. 81-+.

41. Israelski, D., et al., Sociodemographic characteristics associated with medical appointment adherence among HIV-seropositive patients seeking treatment in a county outpatient facility. Prev Med, 2001. 33(5): p. 470-5.

42. Haley, D.F., et al., Retention strategies and factors associated with missed visits among low income women at increased risk of HIV acquisition in the US (HPTN 064). AIDS Patient Care STDS, 2014. 28(4): p. 206-17.

43. Kahana, S.Y., et al., Structural determinants of antiretroviral therapy use, HIV care attendance, and viral suppression among adolescents and young adults living with HIV. PloS one, 2016. 11(4): p. e0151106.

44. Tarantino, N., et al., Correlates of missed clinic visits among youth living with HIV. AIDS Care, 2018. 30(8): p. 982-989.

45. Reisner, S.L., et al., Characterizing the HIV Prevention and Care Continua in a Sample of Transgender Youth in the U.S. AIDS Behav, 2017. 21(12): p. 3312-3327.

46. Waldrop-Valverde, D., et al., Risk and protective factors for retention in HIV care. AIDS and Behavior, 2014. 18: p. 1483-1491.

APPENDIX A

UAB IRB APPROVAL

THE UNIVERSITY OF ALABAMA AT BIRMINGHAM
Office of the Institutional Review Board for Human Use

470 Administration Building
701 20th Street South
Birmingham, AL 35294-0104
205.934.3789 | Fax 205.934.1301 |
irb@uab.edu

APPROVAL LETTER

TO: Simpson, Tina Y

FROM: University of Alabama at Birmingham Institutional Review Board
Federalwide Assurance # FWA00005960
IORG Registration # IRB00000196 (IRB 01)
IORG Registration # IRB00000726 (IRB 02)
IORG Registration # IRB00012550 (IRB 03)

DATE: 29-Jul-2022

RE: IRB-981112002
Alabama Ryan White Part D Program (Alabama Ryan White Part D Program - The Family Clinic)

The IRB reviewed and approved the Continuing Review submitted on 22-Jul-2022 for the above referenced project. The review was conducted in accordance with UAB's Assurance of Compliance approved by the Department of Health and Human Services.

Type of Review:	Expedited
Expedited Categories:	5,
Determination:	Approved
Approval Date:	28-Jul-2022
Approval Period:	One Year
Expiration Date:	27-Jul-2023

Linked Records:

- 000436508
- 000357263
- 000516217
- 000515733

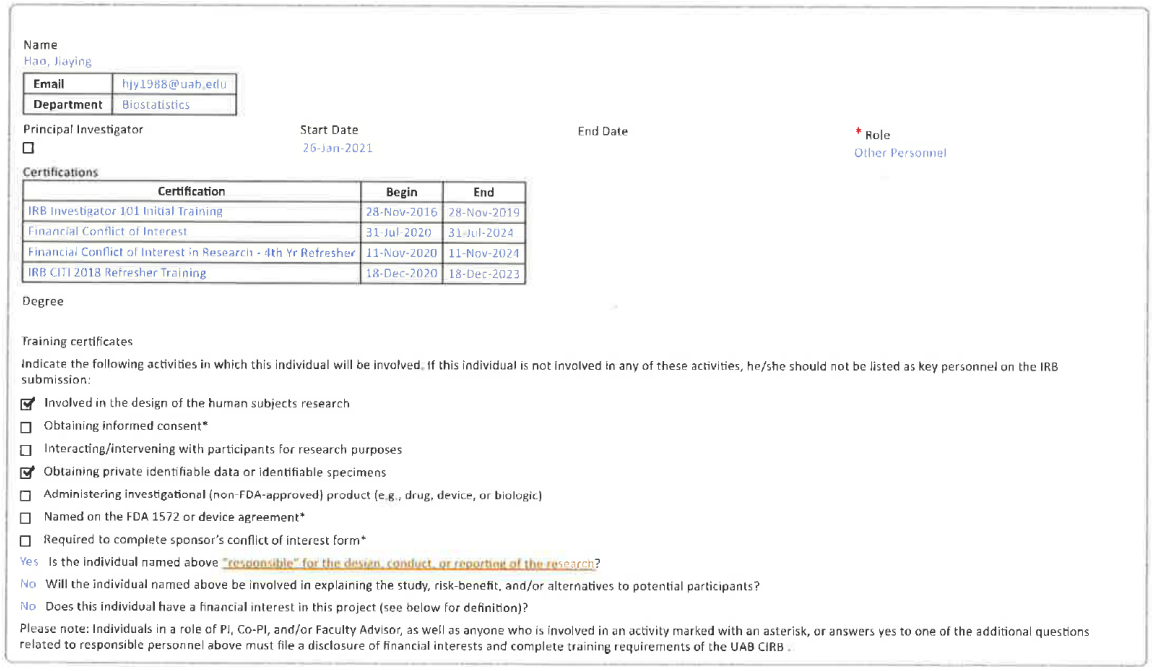
The following populations are approved for inclusion in this project:

- Children – CRL 1

Documents Included in Review:

- CONTINUING REVIEW EFORM

To access stamped consent/assent forms (full and expedited protocols only) and/or other approved documents:
1. Open your protocol in IRAP.
2. On the Submissions page, open the submission corresponding to this approval letter. NOTE: The Determination for the submission will be "Approved."
3. In the list of documents, select and download the desired approved documents. The stamped consent/assent form(s) will be listed with a category of Consent/Assent Document (CF, AF, Info Sheet, Phone Script, etc.)



Name
Hao, Jiaying

Email	hjy1988@uab.edu
Department	Biostatistics

Principal Investigator	Start Date	End Date	* Role
☐	26-Jan-2021		Other Personnel

Certifications

Certification	Begin	End
IRB Investigator 101 Initial Training	28-Nov-2016	28-Nov-2019
Financial Conflict of Interest	31-Jul-2020	31-Jul-2024
Financial Conflict of Interest in Research - 4th Yr Refresher	11-Nov-2020	11-Nov-2024
IRB CITI 2018 Refresher Training	18-Dec-2020	18-Dec-2023

Degree

Training certificates

Indicate the following activities in which this individual will be involved. If this individual is not involved in any of these activities, he/she should not be listed as key personnel on the IRB submission:

- ☑ Involved in the design of the human subjects research
- ☐ Obtaining informed consent*
- ☐ Interacting/intervening with participants for research purposes
- ☑ Obtaining private identifiable data or identifiable specimens
- ☐ Administering investigational (non-FDA-approved) product (e.g., drug, device, or biologic)
- ☐ Named on the FDA 1572 or device agreement*
- ☐ Required to complete sponsor's conflict of interest form*

Yes Is the individual named above "responsible" for the design, conduct, or reporting of the research?
No Will the individual named above be involved in explaining the study, risk-benefit, and/or alternatives to potential participants?
No Does this individual have a financial interest in this project (see below for definition)?

Please note: Individuals in a role of PI, Co-PI, and/or Faculty Advisor, as well as anyone who is involved in an activity marked with an asterisk, or answers yes to one of the additional questions related to responsible personnel above must file a disclosure of financial interests and complete training requirements of the UAB CIRB .

APPENDIX B

R CODE FOR AIM 3

```
library(aorsf)
library(tidyverse)
library(tidymodels)
library(survival)
library(riskRegression)
library(data.table)
library(ranger)
library(glmnet)
library(xgboost)
library(party)
library(glue)
library(table.glue)
library(gt)
library(haven)
library(gdata)#for drop.level function
#library(purrr)
tidymodels_prefer()

R.utils::sourceDirectory('forLing/functions/')

# my code assumes variables called time and status are the outcomes.
# here you should make sure your data match this expectation.

dffull = read_sas("updatemodel1.sas7bdat") %>%
   filter(eventtime>0)%>%drop_na(eventtime) %>%
 select(-contains("SelfEff_")) %>%
 select(-contains("SelfEfficacyScore1")) %>%
```

```
 select(-contains("BriefCOPE2")) %>%
 rename(time = eventtime, status = VStatus)  %>%
 mutate_if(is.character, as.factor)
df1<-replace(df,df=="",NA)

data <- df1 #%>%
 # remove variables that are not time, status, or predictors
 # e.g., drop the id column

# reproducibility

set.seed(8675309)

# make the monte carlo cross validation object
splits <- mc_cv(data, times = 25)
#splits <- vfold_cv(data, v = 5)

# make a preprocessing recipe (no computing yet)
preproc_recipe <- recipe(time + status ~ ., data = data) %>%
 # drop constants
 step_nzv(all_predictors()) %>%
 # impute missing values using nearest neighbors
 step_impute_knn(all_predictors()) %>%
 # if corr > 0.9, drop one at random
 step_corr(all_numeric_predictors(), threshold = 0.9)

# some learners need data with categorical variables one-hot encoded.
```

```
# modify default dummy naming function
ref_code_namer <- function(var, lvl, ordinal = FALSE, sep = '..'){
 dummy_names(var, lvl, ordinal, sep)
}

preproc_recipe_ref_coded <- preproc_recipe %>%
 step_dummy(all_nominal_predictors(), naming = ref_code_namer)

mccv_prep <- function(splits, recipe){
 # run preprocessor on all splits of the data
 # dont worry, tidymodels has made this memory efficient.
 # although it looks like you're copying your data over and over,
 # that is not what is happening under the hood.
 splits %>%
  # just grabbing the unprocessed training and testing sets
  mutate(train = map(splits, training),
         test = map(splits, testing)) %>%
  # now modify them by executing the pre-process steps
  mutate(
   # convert train into a prepped recipe
   # (prepped recipe contains both the preprocessed data
   #  AND the routines to preprocess new data)
   train = map(train, ~ prep(recipe, training = .x)),
   # calling bake(train, new_data = test) to preprocess the
   # testing data using just the training data
   test = map2(train, test, ~ bake(.x, new_data = .y)),
   # now that testing data are preprocessed, we don't need to
   # keep the recipe around, so we can just juice it to get
```

```
  # only the preprocessed data
  train = map(train, juice)
 )
}

mccv_standard <- mccv_prep(splits, recipe = preproc_recipe)
mccv_ref_coded <- mccv_prep(splits, recipe = preproc_recipe_ref_coded)

# do these one by one in case of errors or just taking a while to run

mccv_results_orsf <- mccv_standard %>%
 mutate(results = map2(train, test, mccv_iterate, model_type = 'orsf'))
Sys.time()

mccv_results_rsf <- mccv_standard %>%
 mutate(results = map2(train, test, mccv_iterate, model_type = 'rsf'))

mccv_results_cif <- mccv_standard %>%
 mutate(results = map2(train, test, mccv_iterate, model_type = 'cif'))

mccv_results_cox_ph <- mccv_standard %>%
 mutate(results = map2(train, test, mccv_iterate, model_type = 'cox_ph'))

mccv_results_cox_net <- mccv_ref_coded %>%
 mutate(results = map2(train, test, mccv_iterate, model_type = 'cox_net'))

mccv_results_xgb_cox <- mccv_ref_coded %>%
mutate(results = map2(train, test, mccv_iterate, model_type = 'xgb_cox'))
```

```
Sys.time()



# combine results

mccv_results_all <- list(mccv_results_orsf,
                         mccv_results_rsf,
                         mccv_results_cif,
                         mccv_results_cox_ph,
                         mccv_results_cox_net,
                         mccv_results_xgb_cox) %>%
 map(select, id_resample = id, results) %>%
 map(unnest_wider, results) %>%
 map(rename, id_model = model) %>%
 bind_rows()

# analyze mean C-stat and IPA
mccv_results_smry_pred <- mccv_results_all %>%
 dplyr::select(id_model, cstat, ibs_scaled) %>%
 group_by(id_model) %>%
 summarize(across(gt::everything(), .fns = mean)) %>%
 # the pivoting below makes this easy to present in tables
 pivot_longer(cols = -id_model) %>%
 pivot_wider(names_from = id_model, values_from = value)

# analyze which variables were selected and by which model
mccv_results_smry_vars <- mccv_results_all %>%
 select(id_model, vars_selected) %>%
```

```
# you have to unnest twice because vars_selected is double listed
# (cound probably do this without double listing but why bother)
unnest(vars_selected) %>%
unnest_longer(vars_selected) %>%
group_by(id_model) %>%
# for each model, count how many times a variable was picked
count(vars_selected) %>%
# the proportion of times picked depends on how many mccv splits u did
mutate(n = n / nrow(splits)) %>%
pivot_wider(names_from = id_model,
          values_from = n,
          values_fill = 0) %>%
# if a variable was picked >20% of the time by any of the models,
# keep it in our summary. Nothing magic about using 20% as a cut-point,
# you can modify it to be whatever, including 0
filter( xgb_cox > 0.20 |
       cox_net > 0.20 |
       cox_ph  > 0.20 |
       rsf     > 0.20 |
       cif     > 0.20 |
       orsf    > 0.20 ) %>%
rowwise() %>%
mutate(total = sum(across(-vars_selected))) %>%
arrange(desc(total)) %>%
select(-total) %>%
rename(name = vars_selected)

rspec <- round_spec() %>%
```

```
 round_using_magnitude(digits = c(2,1,0),
                       breaks = c(1,10,100))


bind_rows(pred = mccv_results_smry_pred,
          vars = mccv_results_smry_vars,
          .id = 'type') %>%
 mutate(
  across(where(is.numeric), ~ table_value(.x * 100, rspec = rspec)),
  across(where(is.character), ~str_replace_all(.x, '0.00', '0')),
  type = recode(type,
                'pred' = 'Model summary statistics',
                'vars' = 'Variables selected, %'),
  name = recode(name,
                cstat = 'C-statistic',
                ibs_scaled = 'Index of Prediction Accuracy')
) %>%
gt(rowname_col = 'name',
   groupname_col = 'type') %>%
cols_label(cox_ph = 'Proportional Hazards',
           xgb_cox=  'XGBoost',
           cox_net = 'Penalized',
                   orsf = 'Oblique',
           rsf = 'Axis based',
           cif = 'Conditional inference') %>%
tab_spanner(columns = c('cox_ph', 'cox_net', 'xgb_cox'),
            label = 'Cox proportional hazards') %>%
tab_spanner(columns = c('orsf', 'rsf', 'cif'),
            label = 'Random survival forest') %>%
```

```
cols_align('center', columns =  c('orsf', 'rsf', 'cif',
                                  'cox_net'))
```