THE ROLE OF VIRAL LOAD IN THE PROGRESSION OF HIV DIAGNOSIS TO DEVELOPMENT OF AIDS IN MICHIGAN PATIENTS POPULATION

By

Souliya Channavong

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Epidemiology—Master of Science

2018

ABSTRACT

THE ROLE OF VIRAL LOAD IN THE PROGRESSION OF HIV DIAGNOSIS TO DEVELOPMENT OF AIDS IN MICHIGAN PATIENTS POPULATION

By

Souliya Channavong

Objective: The primary objective is to emphasize the prognostic value of viral load in the progression of HIV infection diagnosis to development of AIDS in the state of Michigan.

Methods: This study included HIV positive cases whose diagnosis status and laboratory tests information were reported to Michigan State health department's HIV surveillance system between Jan 2001 to Dec 2008 and information were available till Dec 2016. Kaplan Meier analyses were used to describe AIDS-free survival probability. Additionally, Cox regression models were used to evaluate baseline and time-varying viral load on risks of progression from HIV diagnosis to AIDS diagnosis. The models were adjusted for baseline CD4 cells count, race, gender, and age at HIV diagnosis.

Results: 2,292 cases were eligible for analysis. Kaplan Meier curve shows that 56% of the patients were AIDS-free at 16 years after HIV diagnosis. For every unit increase of log copies/mL of baseline viral load, the hazard ratio of AIDS progression is 1.09 (95% CI: 1.05-1.13). The time-varying model shows the hazard ratio of AIDS progression of 1.05 (95% CI: 1.03-1.07) for every unit increase in log viral load.

Conclusion: The higher the viral load, the shorter the progression window for HIV patients to develop AIDS. Median progression time from HIV diagnosis to AIDS development in Michigan population is estimated to be longer than 16 years.

ACKNOWLEDGMENTS

First of all, I wish to acknowledge Michigan State University (MSU), College of Human Medicine, and Department of Epidemiology and Biostatistics in providing data analysis facilities and access to previous research. Sincere gratitude to the Michigan Department of Health and Human Services (MDHHS) for providing necessary datasets.

Furthermore, I would like to acknowledge Dr. Joseph Gardiner who had advised me on the statistical analysis of my project. I am also grateful for the support and guidance of Dr. Mary-Grace Brandt who supported this project by allowing access to the datasets and giving insightful comments on how to manage the data.

Finally, I would like to acknowledge the guidance of my academic advisor Dr. A. Mahdi Saeed who had helped me through every process of the project. Not only did Dr. Saeed suggest the topic to me, but he also connected me with MDHHS to obtain the dataset. He always gave valuable advice throughout my academic years as a master's student. I am honored to have him as my advisor.

TABLE OF CONTENTS

LIST OF TABLES vii

LIST OF FIGURES viii

KEY TO ABBREVIATIONS ix

INTRODUCTION 1
- Human Immunodeficiency Virus (HIV) 1
- Stages of HIV Infection 1
- Epidemiology of HIV/AIDS 2
- HIV/AIDS in the US 3
- HIV/AIDS in the State of Michigan 3
- Progression Time from HIV to AIDS 4
- Factors Affecting Disease Progression 5
  - *Human immune response* 5
  - *Antiretroviral Therapy (ART)* 6
  - *Coinfection* 7
  - *Other factors* 7
- HIV Surveillance in the US 8
- Viral Load (Plasma HIV-1 RNA) 9
- Rationale 9
- Objectives 10

METHODS 11
- Study Population and Design 11
- Measurements 14
  - *HIV diagnosis* 14
  - *AIDS diagnosis* 14
  - *Follow-up Status* 14
  - *Follow-up time* 14
    - Follow-up time for baseline model 14
    - Follow-up time for time-varying model 15
  - *Viral load (Plasma HIV-1 RNA)* 16
    - Baseline viral load 16
    - Baseline viral load category 16
    - Time-varying viral load 16
  - *Absolute CD4 T-lymphocyte count (CD4 cells count)* 17
    - Baseline CD4 cells count 17
    - Baseline CD4 cells count category 17
    - Time-varying CD4 cells count 17
  - *Serological Testing Algorithm for Recent HIV Seroconversion (STARHS)* 17
  - *Race/Ethnicity* 18
  - *Gender* 18
  - *Age at HIV diagnosis* 18

Statistical Analysis 18
Kaplan Meier Analysis 19
Cox Proportional Hazard Models 19

RESULTS 21

DISCUSSION 37

CONCLUSION 42

REFERENCES 43

LIST OF TABLES

Table 1. Characteristics of 2,290 HIV patients diagnosed with HIV between 2001-2008 22

Table 2. Baseline log viral load and its relationship with other covariates 29

Table 3. AIDS diagnosis status and its relationship with other covariates 30

Table 4. Cox proportional hazards model with baseline variables 32

Table 5. Cox proportional hazards model with time-varying variables 35

Table 6. Unadjusted and adjusted hazard ratios 36

LIST OF FIGURES

Figure 1. Inclusion and exclusion criteria flowchart 13

Figure 2. Kaplan-Meier curve of AIDS-free probability after HIV diagnosis with the number of subjects at risk 23

Figure 3. Kaplan-Meier curves of AIDS-free probability after HIV diagnosis with the number of subjects at risk by baseline viral load categories 24

Figure 4. Kaplan-Meier curves of AIDS-free probability after HIV diagnosis with the number of subjects at risk by baseline CD4 cells count categories 26

Figure 5. Kaplan-Meier curves of AIDS-free probability after HIV diagnosis with the number of subjects at risk by race 27

Figure 6. Kaplan-Meier curves of AIDS-free probability after HIV diagnosis with the number of subjects at risk by gender 28

Figure 7. Percentage of patients' contribution to the number of viral load tests in time-varying model 33

Figure 8. Viral load test results of five patients follow HIV diagnosis date 34

KEY TO ABBREVIATIONS

Ab	Antibody
Ag	Antigen
AIDS	Acquired Immune Deficiency Syndrome
ART	Antiretroviral Therapy
AZT	Azidothymidine
CD3	Cluster of Differentiation 3
CD4	Cluster of Differentiation 4
CD8	Cluster of Differentiation 8
CDC	Centers for Disease Control and Prevention
CI	Confidence Interval
CMV	Cytomegalovirus
DNA	Deoxyribonucleic Acid
FDA	Food and Drug Administration
HAART	Highly Active Antiretroviral Therapy
HCV	Hepatitis C Virus
HSV-2	Herpes Simplex Virus type 2
HIV	Human-Immunodeficiency Virus
IRB	Institutional Review Board
MDHHS	Michigan Department of Health and Human Services
MDSS	Michigan Disease Surveillance System
mL	Milliliter
MSU	Michigan State University
NAATs	Nucleic Acid Amplification-Based Tests
NRTI	Nucleoside Reverse Transcriptase Inhibitor
OIs	Opportunistic Illnesses
PCP	Pneumocystis pneumonia
PLWH	Persons Living With HIV
RNA	Ribonucleic Acid

SIV	Simian Immunodeficiency Virus
U=U	Undetectable = Untransmittable
WB	Western Blot
WHO	World Health Organization
μL	Microliter

INTRODUCTION

Human Immunodeficiency Virus (HIV)

Human Immunodeficiency Virus (HIV) is a double-stranded retrovirus of lentivirus family (1). The virus was first isolated in 1983 (2). Two serologically distinct species (HIV-1 and HIV-2) were identified (2). HIV in this paper refers to HIV-1 unless specified otherwise. As the name implies, the human is a natural reservoir of HIV. It is believed that HIV evolved from chimpanzee's Simian Immunodeficiency Virus (SIV) which crossed over to human from the butchering of bush meat (2). Since then, the virus has spread around the world and become one of the world's largest pandemic diseases in human history (1).

HIV transmits from person to person via body fluid, semen, and breast milk (2). High-risk activities include unprotected sexual intercourse, the use of HIV contaminated needles and skin piercing products. For instance, sharing needles between injection drugs users, transfusion of infected blood, breastfeeding by an infected mother leads to vertical transmission from mother to child (2). Among these transmission modes, heterosexual transmission is the predominant mode of transmission with more than 75% of HIV transmission, and 5-10% of HIV transmission is from mother to child (3).

Stages of HIV Infection

There are three stages of HIV infection. An HIV infected patient will first experience acute HIV infection, followed by chronic HIV infection and later, the final stage: Acquired Immunodeficiency Syndrome (AIDS) (4). In the acute HIV infection stage, some people will develop flu-like symptoms such as fever, headache, and rash (4). This stage develops typically within 2-4 weeks after initial HIV infection. During the acute stage, the virus' Ribonucleic Acid (RNA) genome transcribes into

Deoxyribonucleic Acid (DNA) by integrating into the host immune cell genome and duplicates itself (4). The virus multiplies and replicates approximately 10 billion copies each day (4). The viruses also disseminate throughout the human body (4). The virus mainly targets a type of white blood cell called T-lymphocyte, especially CD4 T-lymphocyte. Due to the high level of replication and dissemination, the number of HIV in the blood is very high (4). At the same time, the body's immune cells can drop very low. After the acute stage, HIV infection progresses into chronic HIV infection also known as clinical latency (4). In chronic HIV infection stage, an HIV infected person might not show any HIV-related symptom for many years (4). HIV continues to replicate during this period but at a lower rate. Generally, CD4 cells level return to normal range (4). The final stage of HIV infection, or commonly known as AIDS, is developed after HIV destroys the host body's immune system and decreases the body's ability to defend itself from other infections. This puts the host at risk of opportunistic infections (OIs) (4). The most common conditions that alert the presence of AIDS are Kaposi's Sarcoma, Pneumocystis Pneumonia (PCP) cachexia, and esophageal candidiasis. (Detailed list of OIs are described elsewhere,(5,6)).

Epidemiology of HIV/AIDS

HIV/AIDS is a significant public health concern around the globe. The AIDS pandemic has affected more than 60 million people and caused more than 25 million deaths over the past several decades (1). In the year 2016, the World Health Organization (WHO) estimated 36.7 million people infected with HIV and over one million deaths directly associated with HIV (7). According to WHO 2016 care continuum, out of all the people with HIV infection, an estimation of only 70% of them were aware of their HIV status (7). The remaining 30% (over 11 million people) still

need access to HIV testing services (7). Approximately 19.5 million people are receiving treatment, and 16 million reached viral suppression (7).

HIV/AIDS in the US

The first official report of what would have been known as AIDS was a report describing five cases of Pneumocystis pneumonia (PCP) in previously healthy homosexual men on June 5th, 1981 in Los Angeles (8). The Center for Disease Control and Prevention (CDC) estimated 42,000 people were living with HIV by the end of 1981 (8). Since the first AIDS case was reported, the number of Persons Living With HIV (PLWH) has increased remarkably (9). According to CDC by the end of 2015, there was an estimation of 1.1 million PLWH (9). Of those, 162,500 (15%) did not receive diagnosis (9). Approximately 40,000 people were newly diagnosed in 2015 (9). There was a constant number of new HIV cases throughout the past 30 years.

HIV/AIDS in the State of Michigan

The estimated number of people living with HIV infection in the state of Michigan was 15,180 by the end of 2015 (10). Of those, approximately 86% were diagnosed with HIV and were aware of their status. Around 67% were linked to medical care, and 55% were virally suppressed (11). In the same year, there were 726 newly diagnosed cases, and 238 of PLWH died (10). The number of PLWH has increased gradually (12). The increasing number of PLWH is due to the fact that HIV incidence rates were stable while HIV mortality rates decrease (12,13). The incidence rates were stable at around 8 cases per 100,000 individuals (12) while death rates among PLWH steadily declined from 44 to 17 cases per 1000 PLWH (14).

HIV primarily spreads through unprotected sexual intercourse and needles sharing (15). Despite HIV prevention efforts such as the promotion of safe sex practices and disposable needles use, incidence rates of HIV infection in Michigan has remained at the same level over the past decade (13). The incidence rate reported in 2014 was 7.8 per 100,000 cases, dropping only 0.2 per 100,000 cases since 2006 (13). In 2015, among the 15,180 PLWH in Michigan, 8,033 of them have developed AIDS (10). AIDS-defining illnesses are the leading causes of fatality in PLWH, with 50% of deaths among PLWH related to AIDS (16).

Progression Time from HIV to AIDS

Preventing HIV infected patients from developing AIDS is essential in reducing AIDS-related mortality because AIDS-defining illnesses are the primary cause of mortality among HIV infected patients (17). The estimated progression time from HIV to AIDS is an important public health statistic that is used to assess the impact of the HIV epidemic. It is an important parameter to measure the effectiveness of HIV treatments and controls at the population level (18). Even though the time from HIV diagnosis to onset of AIDS has constantly been updated, the median AIDS progression time and AIDS-free probability reported in previous articles have varied (19–22). The median progression time from HIV diagnosis to onset of AIDS ranges from 2-9 years (21–23). Additionally, the AIDS-free probability at ten years after HIV seroconversion ranges from 10-40% (21,22). A study from Tehran, Iran shows that 90.4% of HIV infected patients developed AIDS within ten years (22). Nevertheless, a study from the UK study shows only 60.4% of HIV infected patients developing AIDS within ten years (21). These differences might be due to the differences in study population, study period and study duration. In the

state of Michigan, the progression time from HIV to AIDS development among HIV patients has not been previously studied or reported.

Factors Affecting Disease Progression

Several important factors are used to predict the progression time from HIV seroconversion to the development of AIDS. The main factors that profoundly influence AIDS progression time are patient's immune response to Antiretroviral Therapy (ART). Several additional factors affect HIV to AIDS progression time. Examples of those factors are coinfection with other viruses, race, gender, and age.

Human immune response

The development of AIDS appears to be strongly linked to human immune response since AIDS is principally defined by the deficiency of human acquired immune system (3). The human acquired immune system operates by creating specific antibodies to mark specific invaders to eliminate these foreign organisms. The primary type of cells that plays an essential role in creating antibodies is T-lymphocyte cell. When HIV infected a host, they mainly target the host's T-lymphocyte cells and demolishes the number of T-cells (3). As a result, the host's immune system weakens, and the body is vulnerable to other infections. Thus, a group of symptoms and infections caused by HIV are named AIDS (3).

Researches are exploring several markers to measure the activation of the human immune system against HIV. The markers include CD4 cells count, CD3 cells count, CD8 cells count, β_2-microglobulin and HIV Plasma RNA/DNA (viral load) (24). Prior to 1996, it was commonly accepted that CD4 cells count was the most suitable parameter for HIV treatment guideline and as a basis evaluator for the disease prognosis (24). However, later studies showed that the viral load provided more reliable information about the disease prognosis. Mellors et al. (1997) concluded that

among HIV/AIDS laboratory tests, viral load is the best predictor of progression time from HIV to AIDS (24). Moore et al. (2009) showed that in viral load suppressed patients, change in CD4 cells count had no significant association with progression time to AIDS (25).

In recent practice, both viral load and CD4 cells count are the most routinely prescribed by physicians for evaluating disease progression in HIV patients, and they are used as parameters for initiating and modifying ART (24,26). Prior studies have shown that baseline viral load, and CD4 cells count together are valid predictors of the disease prognosis (24,27,28). Likewise, low viral load and high CD4 cells count at follow-up predicts AIDS progression more accurately (29).

Antiretroviral Therapy (ART)

The first commonly used antiretroviral drug (HIV viral suppressor) was a Nucleoside Reverse Transcriptase Inhibitor (NRTI) or commonly known as Azidothymidine (AZT). The medicine was approved by the US Food and Drug Administration (FDA) in 1987 (30). However, there was little evidence that AZT used alone, as a monotherapy, increased the survival of HIV/AIDS patients. Kaufmann et al. found that monotherapy with this and other early antiretroviral drugs increased patients' CD4 cells count but did not reduce or suppress the viral load (31). The drugs were not an effective suppressor of the viruses and patients still developed AIDS after a period of HIV infection (31). This raised the hypothesis that viral load is a better proxy for HIV/AIDS prognosis and a better treatment indicator.

The introduction of Highly Active Antiretroviral Therapy (HAART) in 1997 elevated the importance of viral load's role in HIV/AIDS prognosis. HAART is the most effective treatment for HIV/AIDS; it combines three types of antiretroviral drugs (30). Studies found that HAART effectively reduced viral load in HIV positive patients

(32). Moreover, a study of HIV to AIDS progression and times to death after AIDS diagnosis between the pre- and post- HAART era showed that the hazard ratio of progression from HIV to AIDS decreased by 80% after the introduction of HAART (33).

Coinfection

Coinfection of HIV with viruses such as herpes, and Hepatitis C Virus (HCV) can affect HIV to AIDS progression. Studies have shown that coinfection of HIV and various members of herpes virus family may increase the risk of HIV to AIDS progression (34). The proposed mechanism was a protein produced by Herpes viruses, especially Human Simplex Virus type 2 (HSV-2), increasing HIV replication in CD4 T-lymphocyte (34). Thus, Herpes viruses accelerate the disease progression by increasing viral load (34). Patients who are infected with both HIV and Cytomegalovirus (CMV) are found to develop AIDS-related symptoms faster than those who are infected with HIV alone (35). Another virus that interacts with HIV and is associated with faster disease progression is HCV (36). Recent research suggests that HIV coinfection with multiple HCV genotypes increases the HIV progression time to AIDS and death (36).

Other factors

Factors such as race/ethnicity, gender, and age at HIV diagnosis are considered to affect progression time from HIV to AIDS. The differences in AIDS progression among different races were observed in many previous studies. A study from CDC surveillance data showed that Non-Hispanic Blacks had the worst HIV/AIDS prognosis (37). Whereas, a study by Grigoryan et al. in 2009 showed that Asian/Pacific Islanders had the worst AIDS-free survival probability (38).

The effect of gender on AIDS progression is controversial. A study of 4,643 HIV patients in Spain suggested that although there were significant differences in the viral load between gender, the disease progression times were not different between male and female (19). Similarly, a US study by Sterling et al. (2001) showed that the rates of AIDS progression in male and female patients were similar (39). However, a recent Kenyan study showed that a hazard ratio of AIDS progression was 1.98 (CI: 1.69-2.33) comparing male to female (40). A meta-analysis of 38 studies from the pre-HAART period showed that the median survival of HIV/AIDS varied. The median time of progression to AIDS among the 15-24 year-old and 45-54 year-old were 11.0 years and 7.7 years, respectively (41).

HIV Surveillance in the US

CDC's Division of HIV/AIDS Prevention (DHAP) has developed a program of HIV surveillance in the US (42). This National HIV Surveillance System uses a uniform surveillance report form and case definition (42). AIDS reporting has started since 1981. As of 2018, all 50 states, the District of Columbia and six US dependent areas report confirmed HIV infection and AIDS cases to CDC (42). CDC provides funding and supports state and local health departments to collect the data (42). Information such as demographics (gender, age, race and place of diagnosis), transmission category (homosexual transmission, heterosexual transmission or mother to child transmission), opportunistic infections, antiretroviral drug use and vital status are collected (42). The CDC uses this information to understand the burden of HIV/AIDS and guides public health intervention at the federal, state and local level (42).

Viral Load (Plasma HIV-1 RNA)

Viral load is a measure of the quantitative volume of HIV-1 RNA in blood plasma. Examples of viral load tests include reverse transcriptase-polymerase chain reaction (RT-PCR), and qualitative transcription-mediated amplification. The test results are expressed in RNA copies per milliliter (copies/mL) (43).

Viral load has been used to monitor HIV/AIDS prognosis. It has been used to monitor the effectiveness of treatment in HIV infected individuals. Apart from the important use of viral load to modify treatment, it is also used to determine whether an HIV infected patient will transmit the virus to another person. An undetectable viral load is defined as a viral load below 40 to 50 copies/ml. In some very sensitive tests, the cut point is set at 20 copies/ml. Undetectable equal Untransmittable (U = U) is a campaign advocating that HIV patients with undetectable viral load are a non-transmittable source of HIV infection. These patients have no risk of transmitting HIV to their partner when engaging in unprotected sexual intercourse.

Another important application of viral load is the concept of community viral suppression. Community viral suppression means the proportion of virally suppressed cases out of all people living with HIV. The CDC suggested that higher levels of community viral suppression reduce the HIV transmission rate in the community.

Rationale

In previous literature, viral load is an effective surrogate that is found to be associated with HIV to AIDS progression time. It is also widely used by physicians to initiate and modify treatment among HIV infected patients. Despite this critical application of viral load, some studies found that the association between viral load and HIV to AIDS progression time is not strong. In order to emphasize the

importance of viral load as an HIV/AIDS prognostic parameter, it is essential to test the association on a different population. This study will confirm the direction of the effect of viral load on HIV to AIDS progression. Also, this evidence will support the concept of two important public health indicators which are U=U and community viral suppression.

The state of Michigan has the 10th largest population in the US (44). Although MDHHS continually reports prevalence, trend, mortality and other public health indicators, HIV to AIDS progression time and probabilities of HIV to AIDS development have not been reported. Our study is designed to describe the progression time from HIV diagnosis to AIDS development. Describing the progression time from HIV to AIDS in the state of Michigan will provide valuable information for MDHHS policymakers to classify HIV/AIDS as a chronic disease.

Objectives

The main objectives of the present study are 1.) to estimate AIDS-free probability from HIV diagnosis to development of AIDS in HIV positive cases over a period of 16 years in Michigan. 2.) to examine the role of viral load at the baseline level and its subsequent updates on progression time from HIV diagnosis to onset of AIDS.

METHODS

Study Population and Design

The study is designed as a retrospective cohort. The study population is sampled from the Michigan Department of Health and Human Services (MDHHS) HIV surveillance system.

Since 1981, all HIV cases in the state of Michigan have been reported to MDHHS either through Confidential Case Report Forms or the Michigan Disease Surveillance System (MDSS-computer-based system). HIV/AIDS diagnoses follow the CDC protocols and guidelines (6,43). Patients' personal information is de-identified and stored in the HIV/AIDS reporting system (43).

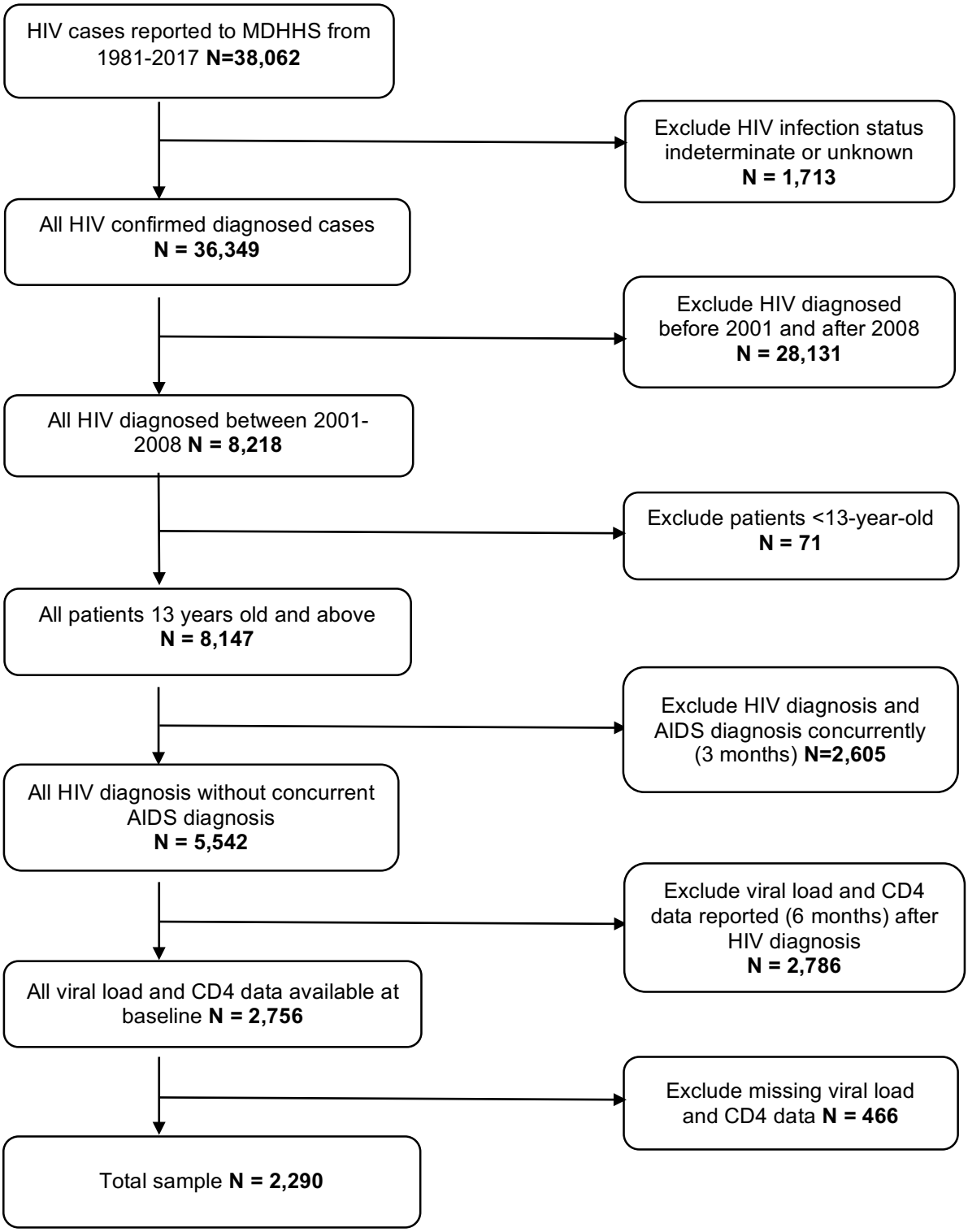
The de-identified data were obtained in collaboration with MDHHS. We built a retrospective cohort from the surveillance data. The study cohort includes all confirmed HIV positive cases between January 1, 2001, to December 31, 2008 (N= 8,218). We examined the cohort's HIV/AIDS laboratory tests and diagnosis status up to December 31, 2016. The data were received from MDHHS in August 2017. To ensure timeliness and completeness of the data, we decided to set the study end date on December 31, 2016. The median progression time to AIDS development from previous Western studies was approximately eight years (20),(21). Therefore, we will have at least an eight-year follow-up period.

The analysis focuses on adults and adolescents; therefore, patients who were younger than 13-year-old were excluded (N = 8,147). Cases with an AIDS diagnosis within three months of an HIV diagnosis were excluded to eliminate the misclassification of acute HIV infection cases as AIDS cases (N = 5,542). In this study, we focused on baseline laboratory tests and its association with progression

time to AIDS. We define baseline laboratory tests as an initial test within six months after HIV diagnosis. Therefore, patients with initial viral load and CD4 cells count after six months of HIV diagnosis were also excluded to align with the baseline laboratory tests definition (N = 2,756). Lastly, we excluded those cases with missing baseline viral load, and CD4 cells count information (N=2,290). Below, a flowchart of inclusion and exclusion criteria is presented in Figure 1.

After the application of all the inclusion and exclusion criteria, the sample consists of 2,290 subjects. The study protocol was reviewed and approved by MDHHS Institutional Review Board (IRB). The data used in the study were collected for surveillance purposes and were de-identified by MDHHS staff before they were sent to the investigators. Therefore, the study was determined as not using "Human Subjects" and was exempt from further review by MSU and IRB.

Figure 1. Inclusion and exclusion criteria flowchart



Measurements

HIV diagnosis

HIV diagnosis included those individuals who were classified as HIV positive either from laboratory tests or physician's documentation according to CDC 1993 guidelines (5). HIV laboratory positive tests were confirmed by positive results from HIV-antibody screening tests (HIV-1 IA, HIV-1 Ag/Ab, etc.) and followed by positive results on confirmatory tests, e.g., Western Blot (WB) test (43,45).

AIDS diagnosis

AIDS diagnosis was defined by patients who have the test result of CD4 cells count < 200 cells/μL or CD4 cells percentage <14%. Alternatively, patients with an AIDS-defining opportunistic illnesses (OIs) were also considered as having AIDS regardless of their CD4 status. The AIDS stage can change in only one direction over time. For example, AIDS diagnosis status will not return to non-AIDS status when CD4 cells counts return to the level above 200 cells/μL (6).

Follow-up Status

Follow-up status was either the diagnosis or vital status of a patient at the end of the follow-up time. Patients who were diagnosed as having AIDS were defined as AIDS cases (Follow-up status = 0). Patients who did not develop AIDS by the end of the study were defined as non-AIDS cases (Follow-up status = 1). Patients who die of other causes before the study end date were defined as death from other causes (Follow-up status = 2).

Follow-up time

Follow-up time for baseline model

The follow-up time for AIDS cases was defined as the duration between HIV diagnosis date and AIDS diagnosis date. The date of HIV diagnosis was defined as

the earliest date that the specimen was collected and had a positive HIV test result if the diagnosis was based on laboratory evidence. If the diagnosis was based on physician documented evidence, the diagnosis date was defined as the on physician documented date on the medical record (6).

The date of AIDS diagnosis was defined as the date of the earliest condition classifying the case as stage 3 HIV infection, either by CD4 laboratory test date or OIs diagnosed date (43). For patients who did not develop AIDS, the follow-up time was defined as the duration between HIV diagnosis date and the study end date. In case the patient died of other causes before the study end date, their follow-up time was defined as the duration from HIV diagnosis to their time of death.

Follow-up time for time-varying model

Time-varying model had multiple instances of viral load measurements taken at multiple time points. The first follow-up time for every patient was the duration between the HIV diagnosis date and the first laboratory test sample date (either viral load or CD4 cells count). From this point on, the length between every laboratory sample date counted as a distinct follow-up time. Because most patients had multiple viral load and CD4 cells count tests, their second follow-up time was defined as the time of the first laboratory test date to the second laboratory test date. Their third follow-up time was defined as the time between the second laboratory test date and the third laboratory test date, and so on. The final follow-up time is the time between their final laboratory test and AIDS diagnosis date. Alternately, if a patient dies before AIDS diagnosis, the time between the final laboratory test and death was considered the final follow-up time. In case the patient did not die or develop AIDS until the end of the study, the time between the final laboratory test date and the end of the study was considered to be the final follow-up time.

The follow-up time was calculated by SAS using “INTCK” function. This function calculated the interval between two given dates and expressed the follow-up time in years.

Viral load (Plasma HIV-1 RNA)

Baseline viral load

Baseline viral load was defined as the first viral load test reported within six months after HIV diagnosis.

Baseline viral load category

Given that viral suppression is an important parameter used to measure HIV/AIDS progression, we categorize all viral loads into three magnitudes to observe each category’s AIDS-free probability. The viral load was categorized and coded as: 1 = <50 copies/mL, 2 = 50-200 copies/mL*,* and 3 = >200 copies/mL. Clinically, optimal viral suppression was defined as <50copies/mL (26). Viral suppression was defined as a confirmed detectable viral load ≤ 200 copies/mL (11). Viral load >200copies/mL was considered incomplete virologic response or virologic failure (26).

Time-varying viral load

The time-varying viral load was a collection of various viral load tests that started from the initial baseline test and continued until AIDS diagnosis date for those who were diagnosed with AIDS. For those who died within the study period, the test conducted until their death will be included in the study. Finally, for those who did not develop AIDS or die, the viral load tests from HIV diagnosis until the end of the study date was considered.

Absolute CD4 T-lymphocyte count (CD4 cells count)

Baseline CD4 cells count

CD4 cells count is a laboratory test of a blood specimen. Commonly, the technique of flow cytometry is used to measure CD4 in a blood sample. Unit of measurement is cells/μL. Baseline CD4 cells count was defined as the first result reported within six months after HIV diagnosis.

Baseline CD4 cells count category

We categorized baseline CD4 cells count into four groups according to ART guidelines. Baseline CD4 cells count categories were coded as: 1 = <200 cells/μL, 2 = 200-350 cells/μL, 3 = 350-500 cells/μL and 4 = >500 cells/μL. CD4 cells count were found to be more effective if starting ART when CD4 cells count >500 cells/μL(26). However, the international AIDS society suggested starting the treatment when CD4 cells count drop below 350 cells/μL(46).

Time-varying CD4 cells count

Time-varying CD4 cells count is a collection of various CD4 tests that started from the initial baseline test and continued until AIDS diagnosis date for those who were diagnosed with AIDS. For those who died within the study period, the test conducted until their death was included in the data. Finally, for those who did not develop AIDS or die, the CD4 tests from HIV diagnosis until the end of the study date was considered.

Serological Testing Algorithm for Recent HIV Seroconversion (STARHS)

This serological test indicates whether a patient is likely to have been infected within the last six months by testing HIV specific IgG Capture BED Enzyme Immunoassay (BED EIA) (43). The BED EIA quantifies the proportion of HIV-antibodies IgG antibodies out of all IgG antibodies in the collected sample (43). In an

individual whose infection is long-term, there is a high level of HIV-specific antibodies in the sample. On the other hand, the level of HIV-specific IgG antibody in the sample is low in an individual with recent infection (43). Long-term infection indicates that the patients were infected with HIV more than six months before HIV diagnosis (43). Recent infection indicates that infection of HIV was within six months of HIV diagnosis (43).

Race/Ethnicity

Race was categorized into four categories. The categories were defined and coded as 1 = Non-Hispanic Black, 2 = Non-Hispanic White, 3 = Hispanic, All races, and 4 = Others. The Hispanic, Non-Hispanic Black and White are original categories from the HIV surveillance system (43). Due to the small number of sample, we grouped American Indian, Pacific Islander, Not Hispanic American Indian, Native Alaska, Native Hawaiian, Pacific Islander, Asian and Multi-Race into Others.

Gender

Gender was defined by a person's sex at birth which includes male and female. We did not consider male to female or female to male transgender. The variable was coded as Male = M and Female = F.

Age at HIV diagnosis

Age at HIV diagnosis was defined as the patient's age at the time of HIV diagnosis that was reported in the surveillance data and calculated in years.

Statistical Analysis

The means of baseline viral load were compared by gender, race/ethnicity using t-tests or ANOVA F-tests, as appropriate. As continuous variables, correlations were obtained between baseline viral load, age at HIV diagnosis, and CD4 cell count. Fisher Z test was used to determine the correlation between baseline CD4

cells count and baseline viral load. Also, the Z test was used to examine the relationship between age at HIV diagnosis and baseline viral load. As a preliminary analysis compared AIDS cases and non-AIDS cases, by gender, race/ethnicity, CD4 cell count and age at HIV diagnosis using Chi-square tests or t-tests.

Kaplan Meier Analysis

We carried out a Kaplan Meier analysis of duration from HIV diagnosis to development of AIDS. Duration is measured from HIV diagnosis date to the AIDS diagnosis date if observed. AIDS patients were not followed beyond their AIDS diagnosis date. If the AIDS event was not observed, the last follow-up date was used. Censoring events are: death from any cause (prior to AIDS diagnosis), or alive without AIDS, at last, follow up. We estimated Kaplan-Meier survival curves—that is, the probability of being AIDS-free, in strata defined by race, gender, baseline CD4 cells count categories and baseline viral load categories. The log-rank test was used to compare survival curves.

Cox Proportional Hazard Models

Cox proportional hazard models were built to analyze the associations between baseline viral load or time-varying viral load and their associations with the progression time from HIV diagnosis to onset of AIDS. The event of interest was AIDS diagnosis. Time variables used in these models were the follow-up times described above. The independent variable was log baseline or time-varying viral load. Previous research suggest that the baseline CD4 cells count, race, gender, and age at HIV diagnosis might affect AIDS progression time (37,47,48). Therefore, we used these covariates in our multivariable proportional hazard model.

In the time-dependent Cox proportional hazard model, the viral load and CD4 cells count for each patient were updated at different time points throughout the

course of treatment. Time points at a determination of viral load and CD4 cells count were not necessarily the same. Therefore, if an observation time point had the viral load, but not the CD4 value, we replaced the 'missing' value by the most recent non-missing value. The same method was applied to fill a missing viral load at a time point that had a CD4 value.

Due to the wide range of viral load (from 0 - 29,000,000 copies/mL), and positive skewness of its distribution, we used the log-transformed viral load in our models. However, because a small proportion of our sample had zero values, we set up two variables: [1] a binary indicator X_1 : $X_1 = 0$ if viral load =0, and $X_1=1$ if viral load > 0.

[2] X_2 = log (viral load/c) if $X_1=1$. The value c is an arbitrary. If $X_1 = 0$ the value of X_2 is arbitrary (not missing). For proper interpretation in the proportional hazard model, both X_1 and X_2 are used: for example, suppose the log relative hazard:

$\log\left(h(t \mid X_1, X_2) / h_0(t)\right) = \beta_1 X_1 + \beta_2 X_1 X_2$. Then

$\log\left(h(t \mid X_1 = 1, X_2) / h_0(t)\right) - \log\left(h(t \mid X_1 = 0, X_2) / h_0(t)\right) = \beta_1 + \beta_2 X_2$, and β_1 is the difference in log relative hazard comparing a patient is a positive viral load (=c) versus a patient who has zero viral load. Similarly compare a subject with viral load =*k*x, to a subject with viral load =x,

$\log\left(h(t \mid X_1 = 1, X_2 = \log(kx / c)) / h_0(t)\right) - \log\left(h(t \mid X_1 = 1, X_2 = \log(x / c)) / h_0(t)\right) = \beta_2 \log k$,

which gives an interpretation of β_2.

In this study, we stressed precision of the study estimates with a focus on 95% confidence intervals; p-values were presented as an aid to interpretation. All data analysis for this study was generated using SAS software, Version 9.4.

RESULTS

Table 1 presents a description of the study sample. Among the 2,290 subjects diagnosed with HIV, by the end of the study, 37% were diagnosed with AIDS whereas 57% were without AIDS, and 6% died from other causes. The majority of the sample (57%) are Non-Hispanic Black followed by 35% Non-Hispanic Whites. In contrast, Hispanic and Others races only made up 8% of the total sample. The study sample consisted of 75% male and 25% female. STARHS shows that 10% of the study sample had a recent HIV infection. Mean age at HIV diagnosis was 34.06 years (SD: 11.18). The mean baseline CD4 cells count was 500.25 cells/μL (SD: 255.17) and mean log viral load at baseline was 9.65 (Log copies/mL) with the standard deviation of 2.05.

The comparison between the analytic sample and the subjects excluded due to missing viral load and CD4 value is also presented in Table 1. The number of subjects excluded due to missing value (N = 466) made up of 17% of the dataset (N = 2756). Comparisons between the two samples show that there is a significant difference in the distribution of race and follow-up status. Also, the mean age at HIV diagnosis between the two samples is significantly different. The other variables such as gender and follow-up time are not significantly different between the two samples. Although some of the demographic variables between the two sample are significantly different, it is considered acceptable to exclude missing observations from the dataset if the number of missing observations are below 20% of the dataset (49). Additionally, the analytic sample is large with enough power to detect the effect of viral load between AIDS cases and non-AIDS cases.

Table 1. Characteristics of 2,290 HIV patients diagnosed with HIV between 2001-2008

Variables	Analytic Sample N = 2290	Excluded Sample N = 466	P value
Gender, n (%)			
Male	1728 (75.5)	333 (71.5)	0.07**
Female	562 (24.5)	133 (28.5)	
Race, n (%)			
Non-Hispanic Black	1304 (56.9)	300 (64.4)	<.0001**
Non-Hispanic White	810 (35.4)	119 (25.5)	
Hispanic	96 (4.2)	30 (6.4)	
Others	80 (3.5)	17 (3.7)	
Follow-up status, n (%)			
Non-AIDS cases	1296 (56.6)	285 (61.2)	<.0001**
AIDS cases	849 (37.1)	51 (10.9)	
Death with other causes	145 (6.3)	130 (27.9)	
STARHS Interpretation*, n (%)			
Long-term infection	444 (19.4)	38 (8.2)	0.94**
Recent infection	229 (10.0)	20 (4.3)	
Age at HIV diagnosis (Years)			
Mean (SD)	34.06 (11.18)	37.16 (12.01)	<.0001***
Range	13 - 79	16 - 78	
Follow-up time (Years)			
Mean (SD)	8.14 (4.49)	8.36 (5.38)	.27***
Range	4.49 - 15.83	1.83 - 15.25	
Baseline CD4 Cells count (cells/μL)			
Mean (SD)	500.3 (255.2)	N/A	
Range	0 - 2610		
Baseline viral load (copies/mL)			
Mean (SD)	0.1×10^6 (0.8×10^6)	N/A	
Range	0 - 29×10^6		
Baseline Log viral load* (Log of copies/mL)			
Mean (SD)	9.65 (2.05)	N/A	
Range	0.96 - 17.19		

Missing data analytic sample: STARHS interpretation (N = 1617)

Missing data excluded sample: STARHS interpretation (N = 408)

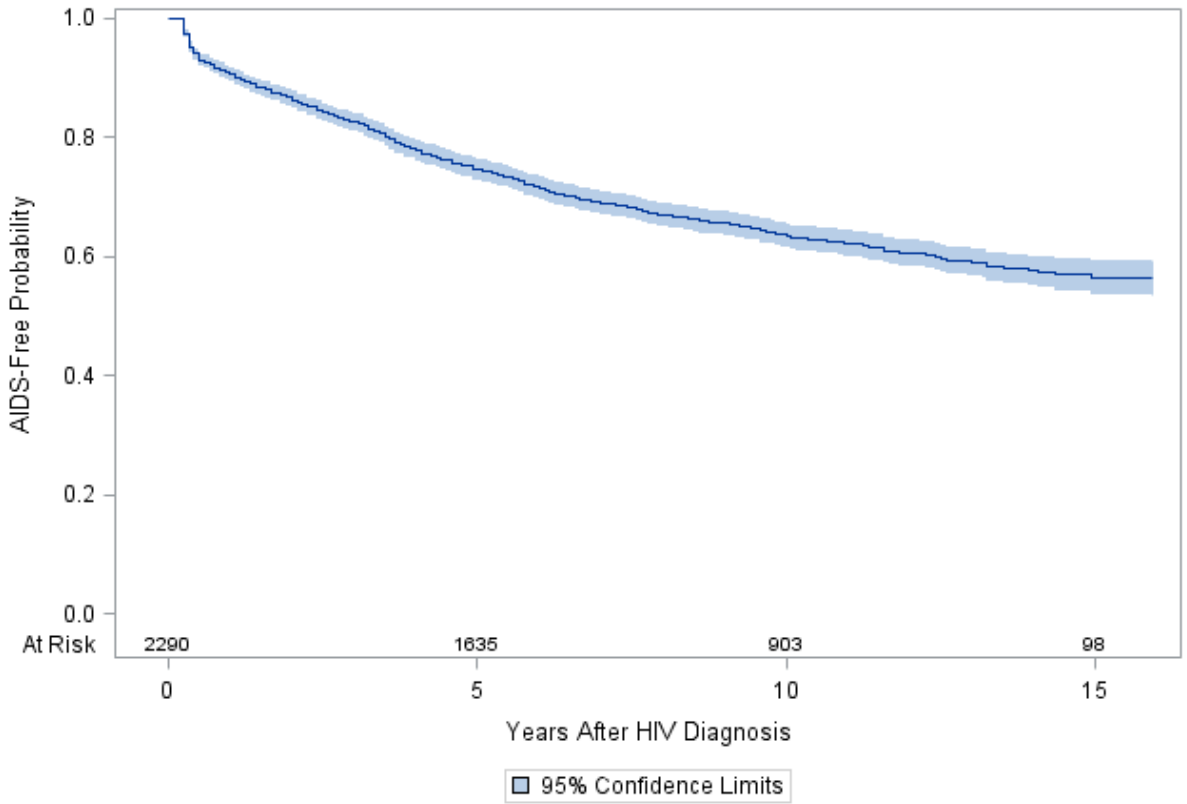
*Baseline log viral load did not include zero value observations (N = 13)

** Chi-square test of two-sample proportion

*** Pooled variance two-sample t-test

Kaplan Meier analysis in Figure 2. shows that among 2,290 people with HIV diagnosis, 25% of the patients developed AIDS by 4.92 years after HIV diagnosis. At 5 years after HIV diagnosis 74.6% did not develop AIDS (95% Confidence Interval [CI]: 72.7-76.3%). The AIDS-free probability dropped drastically to 63.5% (CI: 61.4-65.5%) by the 10th year after HIV diagnosis. After that, the AIDS-free probability gradually declined. At the 15th year, 56.4% (CI: 53.5-59.2%) of the patients were without AIDS. The median progression time from HIV to AIDS development was expected to be longer than 15 years.

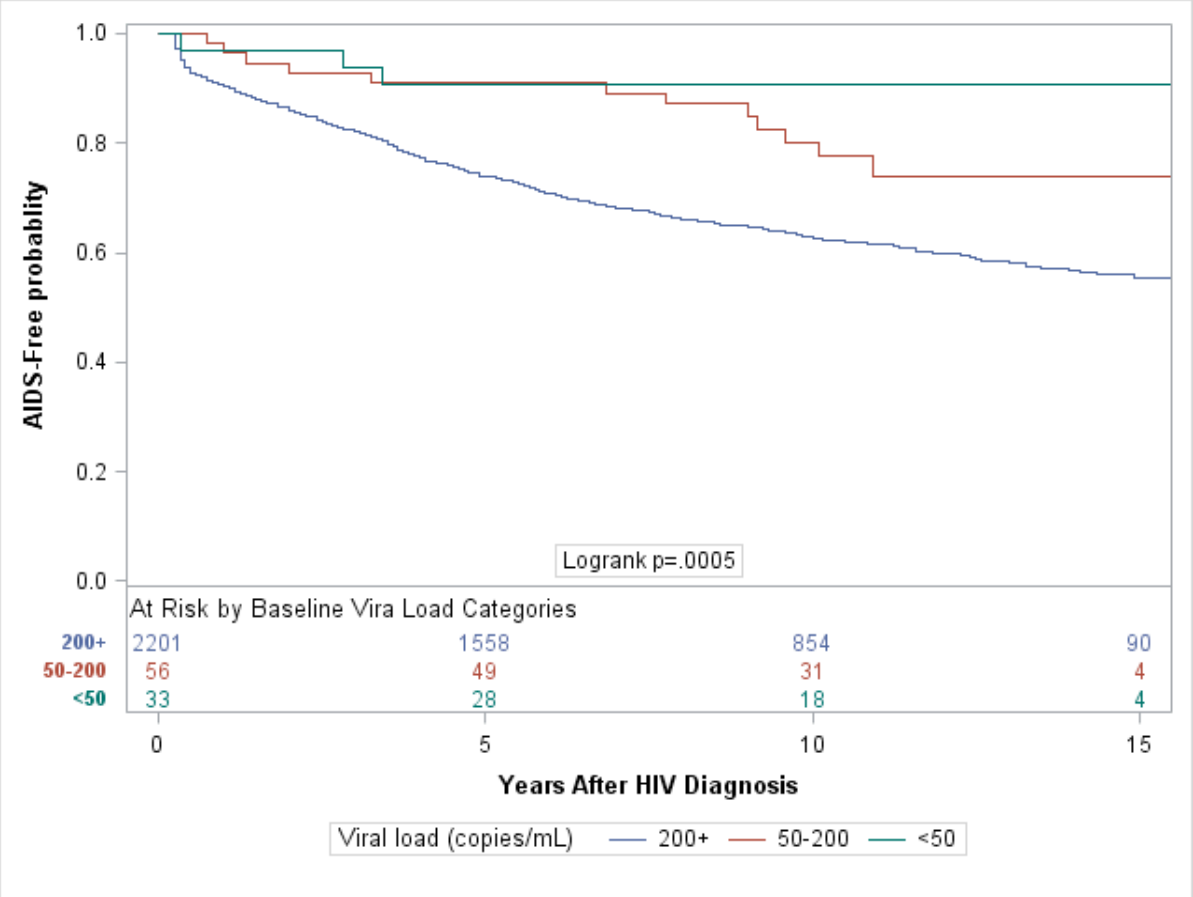
Figure 2. Kaplan-Meier curve of AIDS-free probability after HIV diagnosis with the number of subjects at risk



We obtained Kaplan Meier curves stratified by baseline viral load categories; baseline CD4 cells count categories, race, and gender.

In Figure 3, we present the AIDS-free probability by baseline viral load categories. Overall, there is a significant difference in the AIDS-free probabilities by baseline viral load (log-rank test p<.0005). The group with the lowest viral load had the highest AIDS-free probabilities. At 10 years from HIV diagnosis the AIDS-free probability in patients with baseline viral load below 50 copies/mL was 90.6% (CI: 73.7%-96.9%); in patients with baseline viral load between 50 to 200 copies/mL, the estimate was 77.6%(CI: 62.8-87%), and in patients whose baseline viral loads were above 200 copies/mL, the estimate was 62.4% (CI: 60.3-64.5%).

Figure 3. Kaplan-Meier curves of AIDS-free probability after HIV diagnosis with the number of subjects at risk by baseline viral load categories



Demonstrated in Figure 4, are the Kaplan Meier curves stratified by three categories of CD4 cells count. Overall, there is a significant difference in AIDS-free probabilities ($p < .0001$). However, comparing patients with baseline CD4 cells count of 200-350 and 350-500 cells/μl, the AIDS-free probability is not significantly different ($p = 0.98$). On the other hand, the probability of developing AIDS for patients with CD4 cells count above 500 cells/μl was significantly different from the other groups. By the end of the study, the AIDS-free probability of patients with baseline CD4 cells count above 500 cells/μl was 62.9% (CI: 60-67.6%). For patients with baseline CD4 cells count of 200-350 and 350-500 cells/μl, the AIDS-free probabilities were 55.6% (CI: 49.6-61.2%) and 56.2% (CI: 51.3-60.8%), respectively.

AIDS diagnosis is heavily influenced by the level of CD4 cells count due to the nature of AIDS definition. In our sample, 95% of patients with baseline CD4 cells count <200 cells/μl were diagnosed as having AIDS within one year of HIV diagnosis. Therefore, in Figure 4 we did not show patients with CD4 cells count below 200 cells/μl.

Figure 4. Kaplan-Meier curves of AIDS-free probability after HIV diagnosis with the number of subjects at risk by baseline CD4 cells count categories

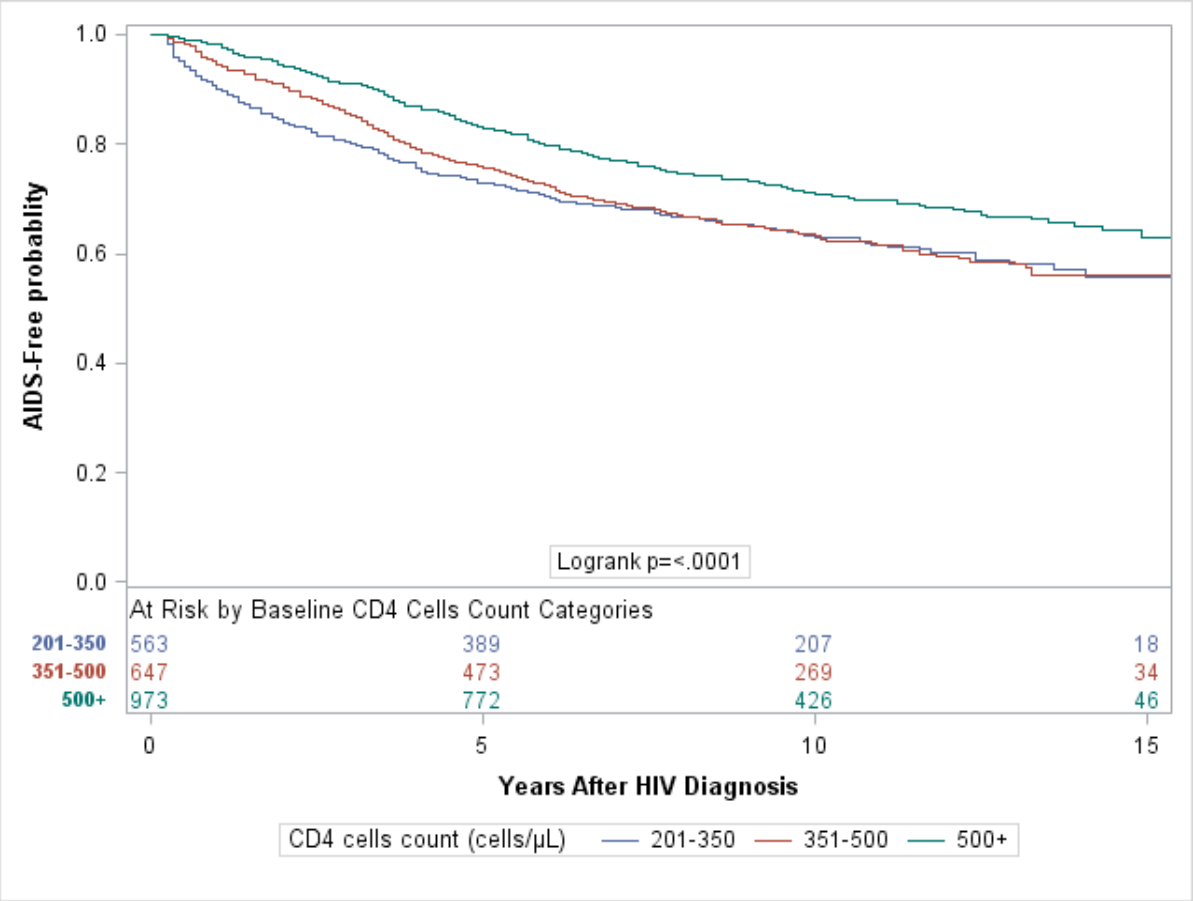
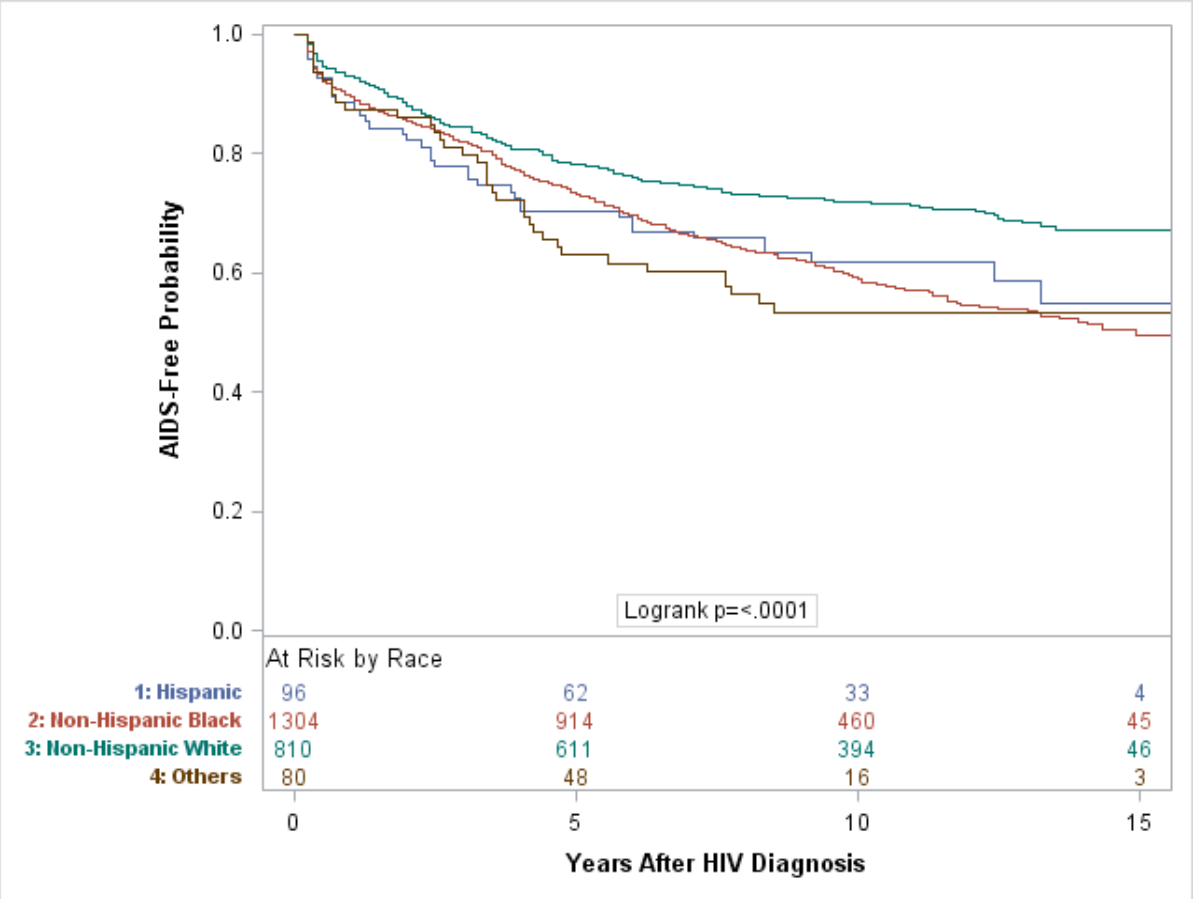


Figure 5. shows that there was a significant difference in the AIDS-free probability by race (p <.0001). Non-Hispanic White had the highest AIDS-free survival probability 15 years after HIV diagnosis at 67.4% (CI: 63.3-71.1%). They were followed by Hispanic and Others races with 55% (CI: 41.5-66.6%) and 53.5% (CI: 41.7-63.9%) of AIDS-free probability, respectively. On the contrary, the worst AIDS-free probability was among Non-Hispanic Black at 49.6% (CI: 45.3-53.8%) after 15 years.

Figure 5. Kaplan-Meier curves of AIDS-free probability after HIV diagnosis with the number of subjects at risk by race



There was no significant difference in AIDS-free survival probability between gender. AIDS-free survival probability at 15 years after HIV diagnosis was 57.2% (CI: 53.7-60.5%) for males and 53.7% (CI: 48.4-58.8%) for females (Figure 6).

Figure 6. Kaplan-Meier curves of AIDS-free probability after HIV diagnosis with the number of subjects at risk by gender

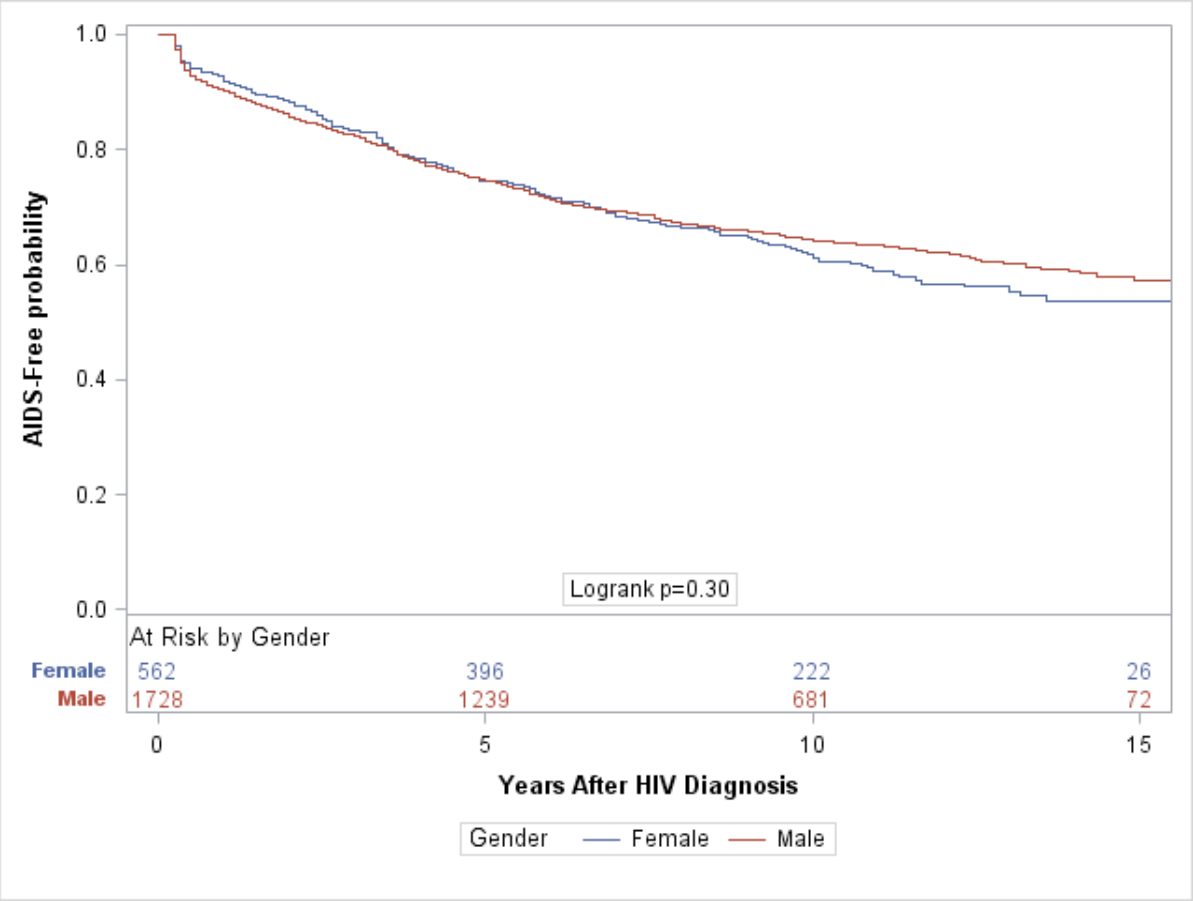


Table 2 shows the univariate analysis between the log of baseline viral load and other covariates. We present the mean of baseline log viral load of each stratum of gender and race. Males had a higher mean baseline log viral load than females. The statistical test shows that between male and female, the means of baseline log viral load were significantly different ($p < .0001$). Non-Hispanic White had the highest mean baseline viral load, followed by Hispanic, Others, and Non-Hispanic Black. ANOVA test shows a significant difference of viral load among races ($p < .0005$). Baseline CD4 cells count was negatively correlated with baseline log viral load, while the age at HIV diagnosis was not associated with baseline log viral load.

Table 2. Baseline log viral load and its relationship with other covariates

	Baseline log viral load[#] (Log of copies/mL), Mean (SD)	P-Value
Gender		
Male	9.19 (2.03)	< .0001[*]
Female	9.80 (2.04)	
Race		
Non-Hispanic Black	9.51 (1.97)	< .0001[**]
Non-Hispanic White	9.88 (2.17)	
Hispanic	9.67 (2.03)	
Others	9.48 (1.99)	
	Baseline log viral load (Log of copies/mL)	**P-Value**
Baseline CD4 cells count (cells/μL)	ρ = - 0.26	< .0001[***]
Age at HIV diagnosis (Years)	ρ = 0.04	.04[***]

[#]Baseline log viral load did not include zero value observations (N = 13)
[*]Pooled t-test
[**]ANOVA F test
[***]Fisher Z test

In Table 3, we compare the distribution of covariates between AIDS and non-AIDS cases. Non-AIDS cases here include patients who did not develop AIDS and patients who died of other causes before the study end date. Means of baseline CD4 cells count and age at HIV diagnosis among AIDS and non-AIDS cases are also reported. The Chi-square and t-test were performed, and the results show that there are significant differences by race and baseline CD4 between the AIDS cases and non-AIDS cases. However, there is no significant difference by gender and age at HIV diagnosis.

Table 3. AIDS diagnosis status and its relationship with other covariates

	Non-AIDS Cases[#] (n = 1441)	AIDS Cases (n = 849)	P-Value
Gender, n (%)			
Male	1101 (76)	627 (74)	.17*
Female	340 (24)	222 (26)	
Race, n (%)			
Non-Hispanic Black	762 (53)	542 (64)	< .0001*
Non-Hispanic White	576 (40)	234 (28)	
Hispanic	59 (4)	37 (4)	
Others	44 (3)	36 (4)	
Baseline CD4 cells count (cells/μL), Mean (SD)	538.93 (261.01)	434.60 (230.49)	< .0001**
Age at HIV diagnosis (Years), Mean (SD)	34.06 (11.08)	34.06 (11.35)	.99**

[#]Non-AIDS cases included HIV negative cases and death from other causes prior to AIDS development

**Pooled t-test

We built Cox regression models for both baseline and time-varying viral load. We included gender and age at HIV diagnosis in the model. The univariate analysis does not show a significant association between these variables and our primary variables of interest, but these two variables were found to be correlated with the progression time to development of AIDS according to previous literature (40). Therefore, we included them in our models.

Multivariable Cox regression models were used. Table 4 shows the baseline model with only baseline information on viral load and CD4 cells count. The baseline model shows that for every increase in 1 unit of Log copies/mL baseline viral load, the hazard ratio of AIDS diagnosis was 1.09 (CI: 1.05-1.13). High baseline CD4 cells count was found to prolong the progression time to AIDS with a hazard ratio of 0.91 (CI: 0.90-0.93) for every increment of CD4 cells count of 50 cells/μL. The model shows that the hazard ratio of developing AIDS among Non-Hispanic White was 0.64 (CI: 0.55-0.75) compared to Non-Hispanic Black. There was no significant difference in the hazard ratio between Hispanic and Others compared to Non-Hispanic Black. In both baseline and time-varying models, gender and age at HIV diagnosis were not significantly associated with progression time to AIDS.

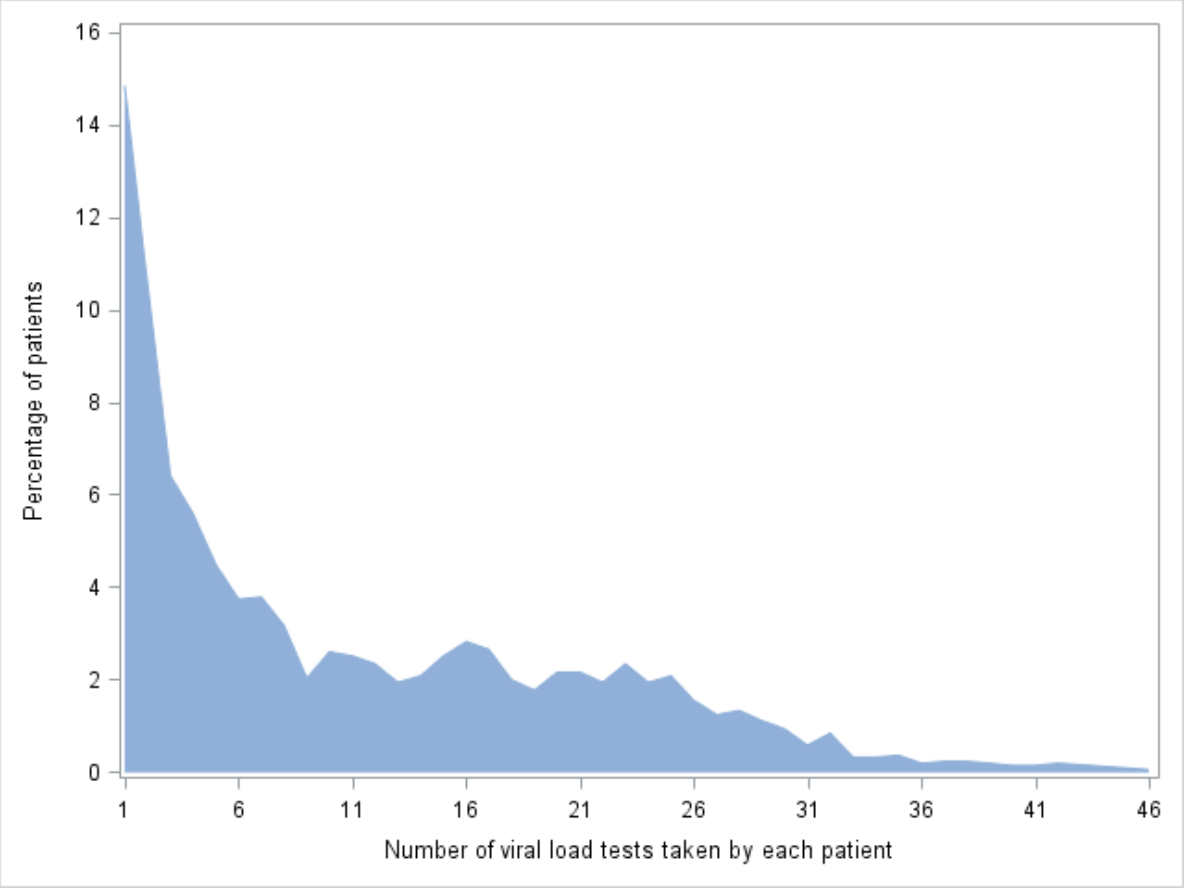
Table 4. Cox proportional multivariate model with baseline variables

	Baseline model Hazard ratio (95% CI)	**P-Value**
Log baseline viral load*	1.09 (1.05-1.13)	< .0001
Baseline CD4 cells count**	0.91 (0.90-0.93)	<.0001
Age at HIV diagnosis***	1.06 (0.99-1.13)	.06
Gender		
Male	Ref.	
Female	1.10 (0.94-1.29)	.22
Race		
Non-Hispanic Black	Ref.	<.0001
Non-Hispanic White	0.64 (0.55-0.75)	
Hispanic	0.95 (0.68-1.33)	
Others	1.18 (0.84-1.65)	

* For every 1-unit increase of log copies/mL
** For every increase of 50 cells/μL
*** For every increase of 10 years

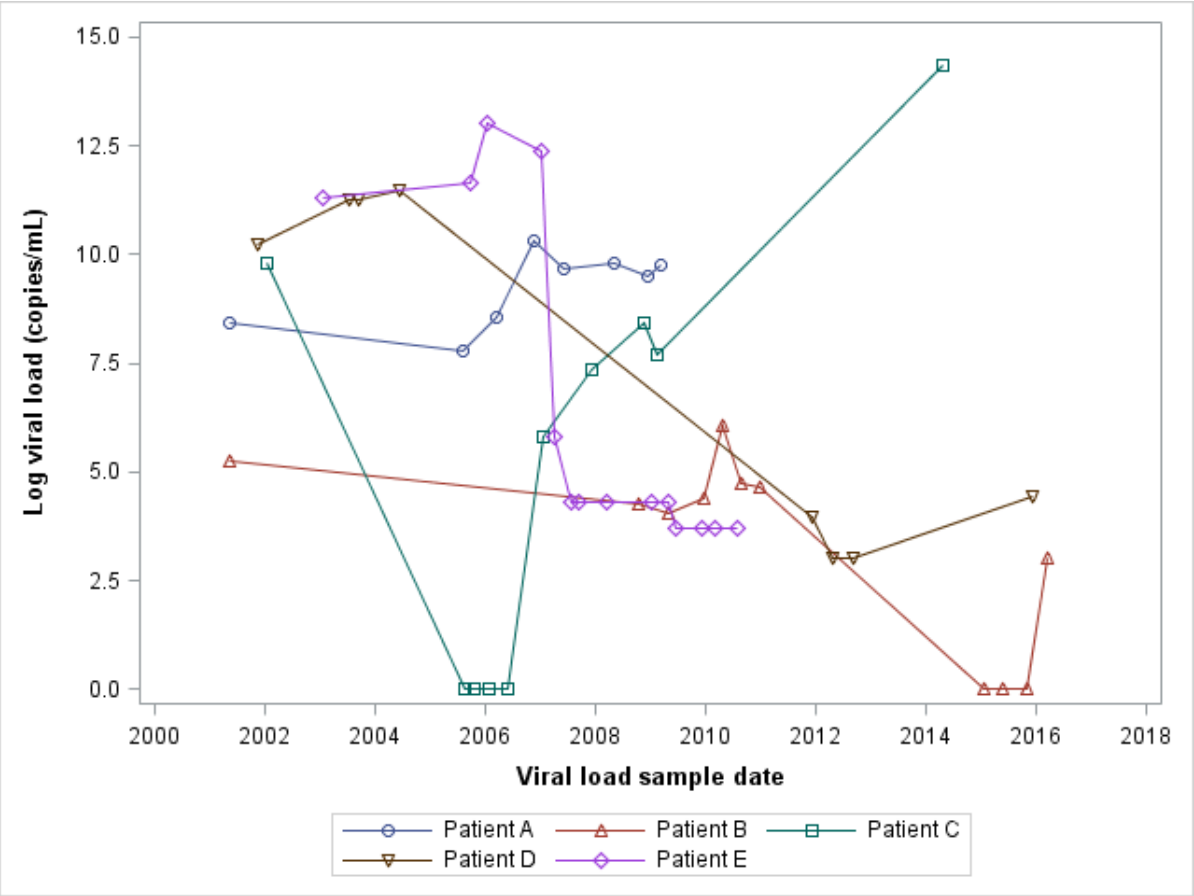
The time-varying variables of viral load and CD4 cells count were added to a new model. This model included not only the baseline laboratory tests but also their subsequent updates until the study endpoint. Figure 7 shows the percentage of the number of viral load tests taken by each patient. Some patients had only one record of the viral load while some patients had up to 46 records. The majority of patients in our sample had taken one to five viral load tests during the study period.

Figure 7. Percentage of patients' contribution to the number of viral load tests in time-varying model



Five patients were randomly selected from the sample to illustrate the level of time-varying viral load in Figure 8. Patient A's viral load never reached an undetectable level during the study. Patient B's viral load were maintained at the same level for an extended period and dropped to zero towards the end of the study. Meanwhile, Patient C's viral load dropped to zero for several months but then increased again, and the level was higher than the initial level. In contrast, Patient E could maintain a non-detectable viral load. This figure shows that time-varying viral load of each patient in our dataset fluctuated and varied.

Figure 8. Viral load test results of five patients follow HIV diagnosis date



The hazard ratio of AIDS diagnosis for every unit increase in log viral load was 1.05 (CI: 1.03-1.07) for the time-varying model. Similar to the baseline model, the higher viral load suggested a shorter progression time to AIDS. The association between the viral load and the disease progression time was significant ($p < .0001$). Low CD cells count was also found to be significantly associated with HIV to AIDS progression. For every increase in 50 cells/μL of CD4 cells count, the hazard ratio of AIDS diagnosis was 0.55 (CI: 0.54-0.56). The associations of other covariates were similar to the baseline model. Age at HIV diagnosis, gender, and race were found to be non-significant in this model.

Table 5. Cox proportional multivariate model with time-varying variables

	Longitudinal model Hazard ratio (95% CI)	**P-Value**
Log viral load*	1.05 (1.03-1.07)	<.0001
CD4 cells count**	0.55 (0.54-0.56)	<.0001
Age at HIV diagnosis***	1.03 (0.97-1.10)	.30
Gender		
Male	Ref.	
Female	1.06 (0.91-1.23)	.46
Race		
Non-Hispanic Black	Ref.	.28
Non-Hispanic White	0.90 (0.77-1.06)	
Hispanic	1.19 (0.85-1.67)	
Others	1.12 (0.80-1.56)	

* For every 1-unit increase of log (1+ viral load copies/mL)
** For every increase of 50 cells/μL
*** For every increase of 10 years

Table 6 presents the association of viral load and HIV diagnosis to AIDS development in both adjusted and unadjusted models. In the unadjusted model, the hazard ratio of AIDS development was 1.27 (CI:1.09-1.17) for every unit increase of baseline log viral load. After adjusting for CD4 cells count, age at HIV diagnosis, gender, and race, the hazard ratio dropped to 1.09 (CI:1.05-1.13). The same is true for the time-varying viral load; the hazard ratio for every unit increase in time-varying log viral load was 1.35 (CI:1.31-1.38). When adjusted for the covariates mentioned above, the hazard ratio fell to 1.05 (CI:1.03-1.07).

Table 6. Unadjusted and adjusted hazard ratios

	Unadjusted Hazard ratio (95% CI)	**P-Value**	**Adjusted Hazard ratio*** **(95% CI)**	**P-Value**
Log viral load (baseline model)	1.27 (1.09-1.17)	< .0001	1.09 (1.05-1.13)	< .0001
Log viral load (time-varying model)	1.35 (1.31-1.38)	< .0001	1.05 (1.03-1.07)	< .0001

* The models were adjusted for CD4 cells count, race, gender, and age at HIV diagnosis

DISCUSSION

Our study was conducted to estimate the progression time from HIV diagnosis to the development of AIDS. Furthermore, the study confirmed that viral load is a valuable predictor of AIDS progression. Before a detailed discussion of these results, several important limitations of the study need to be addressed.

First of all, our data are secondary data extracted from HIV surveillance data. Our analyses were restricted to data that were available to us. In our dataset, we could not get information on patients that were not in medical care. Moreover, we assume that by the end of the study patients who did not develop AIDS were still in the surveillance system and their diagnosis status was updated and concluded.

Another limitation is with respect to the study assessment of HIV diagnosis date. The dates recorded in the surveillance system were dates that HIV infected patients came to seek medical care. There might be an underlying period between the HIV infection date to HIV diagnosis date. Therefore, HIV diagnosis dates used in this study cannot indicate the actual HIV infection dates. The time to event (AIDS) is assessed from the HIV diagnosis date to AIDS diagnosis date.

Another limitation of the study is the precision of AIDS diagnosis. In this study, AIDS diagnosis is defined by surveillance definition which considered every case that has CD4 cells count below 200 cells/μL as an AIDS case. However, as mentioned above, the CD4 count can drop below 200 cells/μL in the acute HIV infection phase. The acute HIV infection might be misclassified as AIDS. In the analysis, we excluded those with concurrent HIV and AIDS diagnosis to account for the misclassification. Nonetheless, the misclassification was not guaranteed to be all excluded.

With respect to the availability of the data, ART which is considered a significant factor that affects progression time to AIDS was not included in the model due to more than 70% of missing information on ART used in our dataset. Similarly, 38% of patients diagnosed between 2001 to 2008 were excluded because of missing baseline viral load and CD4 cells counts data.

Despite the study limitations, our results show the important role of baseline and its subsequent updates of viral load and CD4 cells count in predicting HIV/AIDS disease progression.

Our study reported the progression time from HIV diagnosis to AIDS diagnosis. Among HIV patients in Michigan, 56% of HIV diagnosed patients did not develop AIDS at the 16^{th} year after HIV diagnosis. The median time to AIDS development is expected to be longer than 16 years. Compared to previous data where the median time to AIDS diagnosis was 2-9 years (21,22), our estimation is relatively long. In our Kaplan Meier analyses, precise statistical estimation of the median time was not feasible. We suspect that the median time of AIDS development might be overestimated because we excluded from study those subjects who were concurrently diagnosed with HIV and AIDS. Although we have the precision of AIDS cases by excluding the misclassification between acute HIV infection and AIDS, we still lost patients who developed AIDS at an early stage following their HIV infection. Another reason for the difference in progression time that was observed in our study might be due to the extensive use of ART and other medical intervention in the state of Michigan. Future study should focus on the role of ART and disease progression time. Furthermore, different types of ART and the variability of initiation time of ART are essential areas to explore.

Kaplan Meier curves show that patients with the initial viral load below 50 copies/mL had a 90% AIDS-free probability after 15 years. Furthermore, higher categories of viral load had lower AIDS-free probabilities. In other words, virally suppressed patients are not likely to develop AIDS within 15 years. This indicates that viral load is associated with disease progression.

The study shows that viral load is an effective predictor of AIDS progression. The baseline viral load data is a strong predictor with every one-unit increment predicting a 27% increase in the hazard of AIDS development. After adjusting for covariates, the baseline viral load predicts 9% of AIDS development hazard. This result reassures the importance of using viral load as HIV/AIDS prognosis. In other words, higher viral load at initial test within six months after HIV diagnosis is associated with shorter time to AIDS development. The time-varying Cox regression model also shows that time-varying viral load is a reliable predictor of disease progression, with a 5% increase in the hazard of AIDS diagnosis after adjustment. The results suggest that constantly measuring the viral load is crucial to understanding the disease projection in an HIV positive patient. After adjusting for covariates such as CD4 cells count, race, gender, and age at HIV diagnosis, the hazard ratios dropped from 1.27 to 1.09 in the baseline model and from 1.35 to 1.05 in the time-varying model. This might be due to the confounding effect of CD4 cells count. Nonetheless, the results reported in this study are consistent with previous studies that viral load is an important predictor of AIDS development (50). Furthermore, it is important to measure the viral load longitudinally as it provides substantial evidence for disease progression.

Our models show that CD4 cells count can also reliably predicts disease progression. Low baseline CD4 cells count is associated with shorter progression

time to AIDS. For an increment of baseline CD4 by every 50 cells/μL, it is associated with a 9% decrease of the expected hazard of AIDS development. The time-varying model shows the stronger effect of CD4 cells count, the hazard ratios for every increment of 50 cells/μL of the CD4 count is 0.55 (CI: 0.54-0.56).

The results also reinforced the importance of CD4 cells counts on AIDS progression. Our results are consistent with previous studies exploring the predictive value of CD4 in AIDS progression (24,27,51). However, Moore et al. (2009) suggested that among virally suppressed patients, CD4 cells count are not significantly associated with progression time to AIDS (25). The effect of CD4 among virally suppressed patients and its interaction with viral load is another interesting area to explore in future research.

Our study confirms that HIV to AIDS progression differs by race. This result is consistent with previous studies where Non-Hispanic Black had the worst prognosis after HIV diagnosis (37). Poorer survival of Non-Hispanic Black compared to White may be due to inadequate access to medical care, inconsistent treatment, and diagnosis at the later stage of the disease. Nevertheless, a study by Grigoryan et al. (2009) shows that Asian/Pacific Islander had the worst AIDS-free survival probability (38). We consider that the differences observed are due to the small number of Asian/Pacific Islanders in our sample. Moreover, we collapsed this group into Others races which makes it impossible to analyze Asian/Pacific Islander separately.

We did not observe a difference in AIDS-free probability by gender. Some previous studies regarding progression time to AIDS have shown differences between gender. Our result is consistent with past studies conducted in the US and Spain in the 1990s (19,39). However, it contradicts a recent study from Kenya which

suggested that females have higher AIDS-free probability than males (40). The contradiction might be due to differences in the study period and study population.

Hall et al. (2006) examined demographic factors and their association with AIDS progression using US national HIV surveillance data. Their results show that age at HIV diagnosis is an important indicator for disease prognosis because the immune function of older subjects is weaker than in young people. Thus, the progression time from HIV diagnosis to AIDS when diagnosed at an older age is shorter (37). Generally, the human immune system diminishes significantly after the age of 60 (52). However, our study did not show an association between age at HIV diagnosis and progression time to AIDS. This might be due to an inadequate number of older ages in our sample size. Our sample only consisted of 2% of HIV cases that were diagnosed when they were above the age of 60.

CONCLUSION

Our study observes that the viral load at baseline and its subsequent updates over time are essential in clinical use for HIV/AIDS prognosis. This is especially true for the value of viral load in predicting progression time to AIDS development, following an HIV diagnosis. Confirming the effect of viral load on the progression of HIV infection to AIDS will guide future studies to evaluate different combinations of anti-retroviral drugs to slow such progression of HIV infection to AIDS. By monitoring the viral load over time, researchers can predict the progression time and evaluate the effectiveness of ART. In addition, the study emphasizes that HIV-infected individuals with high viral load should receive better preventive care and reduction of their exposure to infectious diseases associated with fatal outcomes among AIDS patients.

REFERENCES

REFERENCES

1. Sharp PM, Hahn BH. Origins of HIV and the AIDS Pandemic. Cold Spring Harb Perspect Med. 2011;1(1):1–22.
2. Heymann DL. Control of communicable disease manual. 20th ed. Washington DC; 2015. 287-294 p.
3. Lewthwaite P, Wilkins E. Natural history of HIV/AIDS. Medicine (Baltimore) [Internet]. 2009;37(7):333–7. Available from: http://dx.doi.org/10.1016/j.mpmed.2009.04.015
4. U.S. Department of Health and Human Services. The Stages of HIV Infection [Internet]. 2017. Available from: https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/19/46/the-stages-of-hiv-infection
5. Centers for Disease Control and Prevention (CDC). 1993 Revised Classification System for HIV infection and Expanded Surveillance Case Definition for AIDS Among Adolescents and Adults [Internet]. 1993. Available from: https://www.cdc.gov/mmwr/preview/mmwrhtml/00018871.htm
6. Centers for Disease Control and Prevention (CDC). Revised Surveillance Case Definition for HIV Infection — United States , 2014. MMWR Morb Mortal Wkly Rep. 2014;63(3).
7. WHO. 2016 HIV estimates and WHO policy uptake [Internet]. 2016. Available from: http://www.who.int/hiv/data/en/
8. Centers for Disease Control and Prevention (CDC). HIV and AIDS timeline [Internet]. 2017. Available from: https://npin.cdc.gov/pages/hiv-and-aids-timeline#1980
9. Centers for Disease Control and Prevention (CDC). HIV/AIDS [Internet]. 2017. Available from: https://www.cdc.gov/hiv/basics/statistics.html
10. Michigan Department of Health and Human Services. Michigan Statewide HIV Surveillance Report New Diagnoses and Prevalence Tables [Internet]. 2017. Available from: https://www.cdc.gov/hiv/basics/statistics.html
11. Michigan Department of Health & Human Services. The 2020 National HIV/AIDS Strategy (NHAS) Indicators Assessment of Michigan’s Progress 2010-2015. 2017.
12. Michigan Department of Community Health. Estimates of HIV incidence rates in Michigan, 2006-2010. 2013.
13. Michigan Department of Health & Human Services. Estimates of HIV Incidence Rates in Michigan, 2010-2014. 2014.
14. Michigan Department of Health & Human Services. HIV Mortality in Michigan. 2016;(October):1–2.
15. Centers for Disease Control and Prevention (CDC). HIV Transmission [Internet]. 2018. Available from: https://www.cdc.gov/hiv/basics/transmission.html

16. Gill J, May M, Lewden C, Saag M, Reiss P, Ledergerber B. Causes of Death in HIV-1–Infected Patients Treated with Antiretroviral Therapy, 1996–2006: Collaborative Analysis of 13 HIV Cohort Studies. HIV/AIDS [Internet]. 2010;50:1387–96. Available from: https://academic.oup.com/cid/article-lookup/doi/10.1086/652283

17. Lewden C, Salmon D, Morlat P, Bévilacqua S, Jougla E, Bonnet F, et al. Causes of death among human immunodeficiency virus (HIV)-infected adults in the era of potent antiretroviral therapy: Emerging role of hepatitis and cancers, persistent role of AIDS. Int J Epidemiol. 2005;34(1):121–30.

18. Munoz A, Gange SJ, Jacobson LP. Distinguishing efficacy, individual effectiveness and population effectiveness of therapies. AIDS. 2000;14.(6):754–6.

19. Perez-Hoyoz S, García M, Hera DELA, Iribarren JA, Moreno S, Saumoy M, et al. Progression to AIDS and Death and Response to Hospital-Based Cohort. J Women's Heal. 2007;16(7):1052–60.

20. Tassie JM, Grabar S, Lancar R, Deloumeaux J, Bentata M, Costagliola D. Time to AIDS from 1992 to 1999 in HIV-1-Infected Subjects with Known Date of Infection. J Acquir Immune Defic Syndr. 2002;30(1):81–7.

21. Beral V, Babiker A, Brettle R, Carne C. The AIDS Incubation Period in the UK Estimated from a National Register of HIV Seroconverters. Aids. 1998;12(6):659–67.

22. Poorolajal J, Molaeipoor L, Mohraz M, Mahjub H, Ardekani MT, Mirzapour P, et al. Predictors of Progression to AIDS and Mortality Post-HIV Infection: A Long-Term Retrospective Cohort Study. AIDS Care. 2015;27(10):1205–12.

23. Kilmarx PH, Limpakarnjanarat K, Kaewkungwal J, Srismith R, Saisorn S, Uthaivoravit W, et al. Disease Progression and Survival with Human Immunodeficiency Virus Type 1 Subtype E Infection among Female Sex Workers in Thailand. J Infect Dis. 2000;181(5):1598–606.

24. Mellors JW, Munoz A, Giorgi JV, Margolick JB, Tassoni CJ, Gupta P et al. Plasma Viral Load and CD4+ Lymphocytes as Prognostic Markers of HIV-1 Infection. Ann Intern Med. 1997;126:946–54.

25. Moore DM, Harris R, Lima V, Hogg B, May M, Yip B, et al. Effect of Baseline CD4 Cell Counts on the Clinical Significance of Short-Term Immunologic Response to Antiretroviral Therapy in Individuals With Virologic Suppression. 2009;52(3):357–63.

26. A Working Group of the Office of AIDS Research Advisory Council (OARAC). AIDS info: Guidelines for the Use of Antiretroviral Agents in Adults and Adolescents Living with HIV. 2017;

27. O'brien WA, Hartigan PM, Martin D, Esinhart J, Hill A, Benoit S, et al. Changes in Plasma HIV-1 RNA and CD4+ Lymphocyte Counts and the Risk of Progression to AIDS. N Engl J Med. 1996;(Feb. 15 1991):426–31.

28. Macarthur RD, Perez G, Walmsley S, Baxter JD, Mullin CM, Neaton JD, et al. Comparison of Prognostic Importance of Latest CD4+ Cell Count and HIV

RNA Levels in Patients with Advanced HIV Infection on Highly Active Antiretroviral Therapy. HIV Clin Trials. 2005;6:3(May-June):127–35.

29. Ferrer E, Curto J, Esteve A, Miro JM, Tural C, Murillas J, et al. Clinical Progression of Severely Immunosuppressed HIV-infected Patients Depends on Virological and Immunological Improvement Irrespective of Baseline Status. J Antimicrob Chemother. 2015;70(12):3332–8.

30. Lange JMA, Ananworanich J. The Discovery and Development of Antiretroviral Agents. Antivir Ther. 2014;19:5–14.

31. Kaufmann GR, Bloch M, Finlayson R, Zaunders J, Smith D, Cooper DA. The Extent of HIV-1-related Immunodeficiency and Age Predict the Long-term CD4 T Lymphocyte Response to Potent Antiretroviral Therapy. AIDS. 2002;16:359–67.

32. Ledergerber B, Egger M, Opravil M, Telenti A, Hirschel B, Battegay M. Clinical Progression and Virological Failure on Highly Active Antiretroviral Therapy in HIV-1 Patients : A Prospective Cohort Study. Lancet. 1999;353(3):863–8.

33. The CASCADE Collaboration. Determinants of survival following HIV-1 seroconversion after the introduction of HAART. 2003;362:1267–74.

34. Tan DH-S, Murphy K, Shah P, Walmsley SL. Herpes simplex virus type 2 and HIV disease progression: a systematic review of observational studies. BMC Infect Dis [Internet]. 2013;13(1):502. Available from: http://bmcinfectdis.biomedcentral.com/articles/10.1186/1471-2334-13-502

35. Deayton JR, Prof Sabin C a, Johnson M a, Emery VC, Wilson P, Griffiths PD. Importance of cytomegalovirus viraemia in risk of disease progression and death in HIV-infected patients receiving highly active antiretroviral therapy. Lancet. 2004;363:2116–21.

36. Van A. infection with concurrent multiple hepatitis C virus genotypes is associated with faster HIV disease progression. AIDS. 2004;18:2313–8.

37. Hall HI, McDavid K, Ling Q, Sloggett A. Determinants of Progression to AIDS or Death After HIV Diagnosis, United States, 1996 to 2001. Ann Epidemiol. 2006;16(11):824–33.

38. Grigoryan A, Hall HI, Durant T, Wei X. Late HIV Diagnosis and Determinants of Progression to AIDS or Death after HIV Diagnosis among Injection Drug Users, 33 US States, 1996-2004. PLoS One. 2009;4(2).

39. Sterling TR, Vlahov D, Astemborski J, Hoover DR, Margolick JB, Quinn TC. Initial Plasma HIV-1 RNA Levels and Progression to AIDS in Women and Men. N Engl J Med. 2001;344(10):720–5.

40. Mecha JO, Kubo EN, Nganga LW, Muiruri PN, Njagi LN, Mutisya IN, et al. Trends in clinical characteristics and outcomes of Pre-ART care at a large HIV clinic in Nairobi, Kenya: A retrospective cohort study. AIDS Res Ther. 2016;13(1):1–11.

41. The CASCADE Collaboration. Time from HIV-1 seroconversion to AIDS and death before widespread use of highly-active antiretroviral therapy: a collaborative re-analysis. Lancet [Internet]. 2000;355(9210):1131–7. Available

from: http://linkinghub.elsevier.com/retrieve/pii/S0140673600020614

42. Centers for Disease Control and Prevention (CDC). Surveillance Systems [Internet]. 2015. Available from: https://www.cdc.gov/hiv/statistics/surveillance/systems/index.html

43. Centers for Disease Control and Prevention (CDC). eHARS v4.9 Technical Reference Guide. 2017;(January).

44. US Census Bureau. North Carolina Becomes Ninth State With 10 Million or More People, Census Bureau Reports. 2015; Available from: https://www.census.gov/newsroom/press-releases/2015/cb15-215.html

45. Centers for Disease Control and Prevention (CDC). Guidelines for national human immunodeficiency virus case surveillance, including monitoring for human immunodeficiency virus infection and acquired immunodeficiency syndrome. MMWR. 1999;48.

46. Hammer SM, Saag MS, Schooley RT, Jacobsen DM, Thompson MA, Carpenter CCJ, et al. Treatment for Adult HIV Infection. 2006;296(7):827–43.

47. Hogg RS. Rates of Disease Progression by Baseline CD4 Cell Count and Viral Load After Initiating Triple-Drug Therapy. Jama [Internet]. 2001;286(20):2568. Available from: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.286.20.2568

48. Jiang H, Xie N, Cao B, Tan L, Fan Y, Zhang F, et al. Determinants of Progression to AIDS and Death Following HIV Diagnosis: A Retrospective Cohort Study in Wuhan,. PLoS One. 2013;8(12):1–11.

49. Dong Y, Peng CYJ. Principled missing data methods for researchers. Springerplus. 2013;2(1):1–17.

50. O'Brien TR, Blattner WA, Waters D, Eyster ME, Hilgartner MW, Cohen AR, et al. Serum HIV-1 RNA Levels and Time to Development of AIDS. JAMA. 1996;276(2):105–10.

51. Yerly S, Perneger T V., Hirchel B, Dubuis O, Matter L, Malinverni R, et al. A Critical Assessment of the Prognostic Value of HIV-1 RNA Levels and CD4+ Cell Counts in HIV-Infected Patients. Arch Intern Med. 1998;158:247–52.

52. Simon AK, Hollander GA, McMichael A. Evolution of the immune system in humans from infancy to old age. Proc R Soc B Biol Sci [Internet]. 2015;282(20143085). Available from: http://rspb.royalsocietypublishing.org/lookup/doi/10.1098/rspb.2014.3085