Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Michalina Rubin Date

Characterization of the Association between Route of HIV Transmission and the Development of Non-AIDS Defining Cancers

By

Michalina Rubin
Master of Public Health

Department of Epidemiology

_____

Jodie Guest, PhD
Committee Chair

.

Characterization of the Association between Route of HIV Transmission and the Development of Non-AIDS Defining Cancers

By

Michalina Rubin

B.S. Community Health
University of Maryland, College Park
2015

Thesis Committee Chair: Jodie Guest, PhD

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2017

Abstract

Characterization of the Association between Route of HIV Transmission and the Development of Non-AIDS Defining Cancers

By Michalina Rubin

Since the introduction of Highly Active Antiretroviral Therapy (HAART) as a treatment for Human Immunodeficiency Virus (HIV), the number of non-Acquired Immune Deficiency Syndrome (AIDS) defining cancers (NADCs) has increased in the HIV community. NADCs have an increased prevalence in the HIV-positive population compared to their HIV-negative counterparts and account for the majority of the mortality burden in HIV infected people. This study aims to discover if route of transmission of HIV is an independent risk factor for NADCs or merely a proxy measure for other closely associated risk factors. This study is a secondary analysis of a prospective cohort study using data from the HIV Atlanta Veterans Administration Cohort Study (HAVACS). The association between HIV transmission route and risk of NADC development was assessed using logistic and polytomous logistic regression. Models were also created for predicting NADC occurrence based on viral, immunologic, and life-style risk factors. The odds of having acquired HIV through male/male sexual (MSM) transmission in the HIV-positive veterans who have a NADC is 1.31 times the odds of having acquired HIV through heterosexual transmission in the HIV-positive HAVACS veterans, with a 95% confidence interval of (0.74, 2.30). The odds of having acquired HIV through intravenous drug use (IDU) in HIV-positive patients who developed any NADC is 1.00 times the odds of having acquired HIV through heterosexual transmission in HIV-positive veterans, with a 95% confidence interval of (0.51, 1.93). Predictive analyses produced c-statistics of 0.77 to 0.89 and showed that viremia, immunosuppression, oncogenic viral co-infections, high risk lifestyle factors, and route of HIV transmission are strong predictors of NADCs in HIV positive veterans. Though no significant associations were found due to sample size limitations, MSM route of HIV transmission appears to pose an increased risk of NADC development. Further studies should be done to validate this association.

Characterization of the Association between Route of HIV Transmission and the Development of Non-AIDS Defining Cancers

By

Michalina Rubin

B.S., Community Health
University of Maryland, College Park
2015

Thesis Committee Chair: Jodie Guest PhD

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Public Health
in Epidemiology
2017

Literature Review

NADCs & ADCs

In 1996, Highly Active Antiretroviral Therapy (HAART) became widely available in the United States as a treatment for Human Immunodeficiency Virus (HIV)/Acquired Immune Deficiency Syndrome (AIDS), and remains the most effective treatment to date. Because of the wide availability and use of HAART, the incidence of AIDS-defining illnesses has decreased dramatically [1-13]. Accordingly, the rates of morbidity and mortality related to AIDS-defining cancers (ADCs), Kaposi's sarcoma, non-Hodgkin's lymphoma, and invasive cervical cancer, have also declined considerably [9, 14, 15]. As the rates of ADCs have declined, there has been a steady increase in the number of cases of non-AIDS-defining cancers (NADCs) in HIV-positive persons [1-10, 15-18]. NADCs are now responsible for most of the cancer burden in the HIV-positive population, accounting for over 50% of cancer cases [4, 15, 19]. NADCs are so prevalent in the HIV community that the risk for these cancers in HIV-positive individuals exceeds the risk in the general population by two to threefold [3, 4, 9, 18], even when controlling for viral load, immunosuppression and other known cancer risk factors [20, 21]. With the increase in NADCs, cancer is now the leading cause of death in HIV-positive individuals [3, 10, 12, 22-26]. NADCs cover a wide range of cancers, including but not limited to anal, hepatocellular, oropharyngeal, lung, colorectal, vaginal, renal, and skin cancer, as well as Hodgkin's lymphoma, melanoma, leukemia [3, 8-11, 15, 16, 22, 24, 27, 28]. NADCs can be further subdivided into two groups, those with a known viral oncogenesis and those without a known viral oncogenesis. Non-virus related NADCs are melanoma, leukemia, lung, kidney, prostate and testicular cancer. Virus related NADCs are lymphocytic leukemia, Hodgkin's lymphoma, liver, nasopharyngeal, penial, anal, vaginal, vulvar, oral,

and colorectal cancers. Epstein-Barr virus (EBV) is associated with non-Hodgkin's lymphoma and nasopharyngeal cancer, hepatitis B (HBV) and C (HCV) cause hepatocellular cancer, and HPV is associated with colorectal, vulvar, vaginal, penial, anal, oropharyngeal, and tonsillar cancers [7, 16, 19, 29].

Diagnosis, Treatment, & Outcomes

HIV-positive individuals with a NADC are statistcially significantly younger at diagnosis than HIV-negative patients who are diagnosed with these same cancers even after controlling for race and time period of diagnosis [2, 5, 7, 11, 21]. Given that cancer is usually a disease correlated with advancing age, the younger onset of cancer in the HIV community is important and has implications for screening. NADCs tend to present at a more advanced stage of disease and with a more aggressive form of cancer in HIV-positive individuals [2, 3, 6, 7, 21, 28, 30].

HIV-positive patients are also less likely to receive any type of cancer treatment compared to HIV-negative patients [25]. Historically, HIV pateints have been excluded from chemotherapy clinical trials, so there are no specific clinical guidelines for administering treatment to HIV-positive individuals with cancer [31]. Clinicians are often wary to use chemotherapy with HIV-positive patients as the combination of HAART and chemotherapy can lead to toxcity, especially when the individual under treatment has a CD4 count of less than 200 cells/mm^3 [3, 6, 10, 17, 21, 25, 31], but treatment decisions are difficult to make without specific guidelines or experience. Chemotherapy and HAART may also have drug interactions that can cause decreased efficacy of one or both medications [3, 10, 21, 25]. Chemotherapy often leads to immunosuppression, already a

problem for HIV-positive patients, and can induce nausea and vomitting which may affect adherence [10, 21, 25]. For all these reasons, HIV-infected individuals often receive less cycles of chemotherapy than recommended in the general guidelines [32]. HIV-positive cancer patients are also generally considered poorer surgical candidiates. They experience surgical complications more often and have worse postoperative survival than HIV-negative patients [7, 17, 32]. Even when patients had similar cancer treatments, stage of cancer, and patient adherence, HIV-positive patients had poorer survival than HIV-negative patients [31, 32].

Before HAART, less than 10% of HIV-postive patient deaths were caused by cancer. In the HAART era, over a quarter of HIV-positive patient deaths are due to malignancies [31, 33]. NADCs increasingly contribute to the mortality of the HIV community. Decreased survival is linked to the problems of advanced stage of cancer and limitations in treatment options [2, 10, 12, 24-26, 32].

Hypotheses for the Cause of NADCs

Several researchers have supposited that immunosuppression, similar to the way that it causes ADCs, also causes NADCs [4, 6-8, 11, 16, 22, 24, 34]. Immunsuppression is a biologically plausible mechanism for affecting NADC development. For those infected with HIV, the immune system undergoes an aging process known as immunosenescence. As part of the immunosuppression and immunosenescence associated with HIV, the immune system has an impaired ability to sense and destroy cancer cells and oncogenic agents in the body [16, 31, 35]. However, there have been many inconsistencies in the relationship between immunsuppresion and NADCs, and its role remains controversial

[17, 27, 30, 36, 37]. The measures of immunosuppresion and their association with NADCs have differed by study. Some studies have described immunodeficiency as a risk factor for all NADCs [12, 16, 18, 34], while other studies have concluded that immunosuppression is only related to a subset of NADCs such as Hodgkin's lymphoma, multiple myeloma, connective tissue cancer, and anal cancer [19, 38]. Still, other studies are more specific with their definitions of immunsuppression. Long term low CD4 counts has been hypothesized as a risk factor for NADCs, showing a stronger association than nadir CD4 count or the CD4 count prior to cancer diagnosis [35, 39]. One study found nadir CD4 count has been associated with anal, colorectal, and lung cancer [8], while another found that low CD4/CD8 ratio was associated with increased risk for all NADCs [35]. Several studies have found no association between immunodeficieny and risk of NADCs [1, 3, 6, 10] including no association with CD4 count prior to cancer diagnosis [1] and no relationship between time varying CD4 count and NADC risk [27], directly contradicting the finding of other studies. At this point in time, the role of immunosuppression in the development of NADCs is unclear.

Another common theory as to the cause of the increase of NADCs in the HIV-positive population is the overrepresntation of oncogenic viruses. Acculumating evidence from several studies show an association between increasing prevalence of several co-infections and the high prevelance of NADCs [2, 22, 34, 36]. Specific viruses are thought to be the culprit behind the increase in NADCs, such as Epstein Barr Virus (EBV), human herpes virus 8, Hepatitis B Virus (HBV), Hepatitis C Virus (HCV), herpes simplex virus (HSV) and HPV [1, 2, 9, 15-18, 20, 37, 39, 40]. HPV is the most common

sexually transmitted disease and is responsible for five percent of cancers in men and ten percent in women [31]. HIV infected patients are more likely to be infected with HPV, have the more virulent, high risk strains of the HPV virus, and have multiple, concurrent HPV infections [8, 16, 33, 34]. HPV isassociated with the risk of anal, oropharyngeal, cervcial, vulvar, vaginal, penial, oral, and tonsillar cancers in HIV patients [15, 16, 18, 33]. HCV and HBV are also common co-infections with HIV because they share similar routes of transmission, sexual transmission and intravenous drug use (IDU). HCV and HBV are associated with increased risk of hepatocellular and anal cancer as well as Hodgkin's lymphoma [2, 4, 8, 12, 15-18]. Both viruses contribute to an accelaterated progression of cirrohsis leading to an increased risk of liver cancer [2, 17]. EBV is an endemic infection in the American population, with over 90% of the US infected with the virus. Even though EBV is generally ubiquitous in the HIV-positive population, infection with EBV is associated with an increased risk of Hodgkin's lymphoma, nasopharyngeal cancer, Burkitt's lymphoma, and anal cancer [7, 15-18]. Other viruses that have been found to be linked to an increased risk of NADCs, but have had less reaserch done on their association include human herpes virus 8, associated with primary effusion lymphoma, as well as cytomegalovirus and HSV, both associated with anal cancer [16, 17]. Immunosupression is a potential mechanism through which these viral co-infections might affect increased NADC incidence. Because HIV-positive pateints have a compromised immune system, immune surveillance is impaired, and therefore these viral co-infections can allow cancer growth to go unchecked in infected tissue [15, 16, 40]. However, viral co-infection does not explain the relationship between HIV and non-virus

related NADCs. Lastly, oncogenic co-infection can not fully explain the increased rate of NADCs in HIV patients [28].

The HIV virus hasbeen attributed as a direct cause of the increase of NADCs in HIV-positive individuals. Several studies have conducted research that supports this hypothsesis [4, 16, 33, 34, 40]. In particular, HIV viremia has been specifically associated with Hodgkin's lymphoma and anal cancer [34]. It has been hypothesized that the HIV virus may cause cancer through two paths. Firstly, the virus itself may, through chronic infection, directly cause tumorogensis in some tissues [16, 33]. Alternatively, HIV causes chronic inflammation in tissue, which in turn can stimulate cancer development. Chronic inflammation can still exist if the HIV patient is on HAART and virally suppressed [4, 34, 40]. However, even though this theory is biologically plausible, data is conflicting as to its validity. Several studies have found that HIV viremia and viral load had no effect on the development of NADCs [3, 10, 39]. The role of HIV viral load in the development of NADCs remains undetermined.

With the advent of HAART, the life expectancy of those infected with HIV has dramatically increased. This has also led to a shift in the demographics of those who are HIV-positive. In 2017, a quarter of the HIV-positive population is over the age of 50 [19]. Given that age is the leading risk factor for cancer, it is only natural that many researchers are postulating that the increase in NADCs is solely due to the increases in survival and age in this group [2, 6, 7, 12, 19, 24, 33, 37]. Multiple studies have found age to be a statistically signficantly associated with NADCs [11-13, 16, 17, 27, 34, 39],

though age alone is unlikely the only explanation for the growing prevalence of NADCs [16].

HIV acquisition is usually associated with a high risk lifestyle. This high risk lifestyle and the behavioral risk factors that accompany it are another hypothesis as to why NADC rates are higher in HIV-positive individuals [1, 11, 18, 34]. The three most common behavioral factors associated with increased risk of NADCs are smoking, alcohol consumption, and drug use. Smoking is most strongly associated with the risk of NADCs [4, 7, 15, 16, 20, 22, 26, 34, 36, 37]. People living with HIV have a higher prevalence of smoking than the general population, with about 45-75% being smokers compared to 17% smokers in the general US population [4, 7, 15, 26, 37]. Smoking not only increases the risk of lung kidney, laryngeal, oral, and stomach cancers in the HIV-positive population, but also increases their likelihood of dying from any solid tumor [2, 7, 15, 17]. Alcohol consumption is another behavioral risk factor that may be promoting NADC development in people living with HIV [13, 16, 20, 22, 34, 37]. Excessive alcohol consumption can cause cirrhosis, which can then lead to hepatocellular carcinoma. A higher prevalence of alcohol consumption has been found in the HIV-positive community and could explain the disparity between the HIV population and the general population regarding NADCs [4, 37]. Other studies contribute the rise in NADC prevalence to drug abuse, another high risk lifestyle factor prevalent in the HIV community [13, 16]. However, behavioral risk factors cannot fully explain of the increase in NADC prevalence [28].

HAART has also been linked with increased risk of NADCs in the HIV-positive population [1, 4, 16, 34]. The methods by which HAART would cause NADCs, however, is disputed in the literature. Several studies claim that HAART medications have direct tumurogenic effects by inducing micronuclei, chromosomal aberrations, the activity of latent oncogenic materials, sister chromatid exchange, and shortening telomeres in body cells [2, 33]. Other studies suggest that HAART may have indirect effects such as lengthing the life expectancy of HIV patients, thus increasing their risk of cancer. It may also be that nonadherence to HAART may allow the growth of cancer [33, 39]. However, data surrounding the effect of HAART on NADCs is conflicting at best. While some studies have found no association between HAART and NADCs [17], more research has found that HAART is chemopreventive against cancer [1, 7, 8, 17]. Further research is still needed to decipher the role of HAART in NADC development.

Screening and engagement in care are important aspects of healthcare for the HIV-positive population. Because increased vigilance and linkage to care is stressed in this population, cancer oversurveillance, also known as surveillance bias, has been postulated as causing a false increase in the number of NADCs [2, 10, 11, 17, 36, 37]. Once again, the data are conflicting. Some studies report that HIV-positive patients undergo more screening than their uninfected conterparts [36]. Other studies have found that cancer screening is conducted less regularly in the HIV-positive community [21]. More studies should be conducted to compare screening rates in the HIV infected and uninfected communities.

Lung Cancer

The risk of developing lung cancer for HIV patients is at least twofold higher than the general population, and the rates are rising with time [7, 19, 27, 32]. According to several studies, lung cancer is the most commonly occurring NADC in the HIV-positive community [3, 7, 9, 32, 33]. HIV patients have substantially worse lung cancer prognoses than their HIV-negative counterparts. Prognostic complications include greater genetic instability of cancer, more advanced stage, younger age of diagnosis, therapy intolerance, and greater risk of treatment toxicity [7, 11, 28, 30, 32]. These disparities hold true even when controlling for age, cancer stage, histology, treatment, and competing risks [30, 32]. Given these complications with lung cancer diagnosis, it is not surprising that lung cancer is the most frequent cause of cancer death among HIV infected patients [32]. The increased prevalence of lung cancer among HIV patients has been linked to high rates of smoking, with prevalence rates estimated to be between 45-75% [3, 19, 27, 28, 32, 33]. However, studies have shown that HIV patients may develop lung cancer with less smoking exposure than their HIV-negative counterparts implying that smoking cannot account fully for the increased rates of lung cancer [32]. After controlling for smoking history and duration, the risk of lung cancer for HIV-positive patients is at least three times higher than uninfected patients [9, 28]. Some studies suggest that repeated or chronic pulmonary lung infections may contribute to the development of lung cancer in HIV-positive persons [15, 18, 19, 32]. These infections can lead to acute inflammation that in turn can lead to lung cancer. Some studies have found that immunosuppression from low CD4 cell counts plays a role in lung cancer development [18, 28, 32] while others have found no similar relationship [28, 30]. The HIV virus has also been

postulated as a contributing factor to the development of lung cancer in one of two ways [32]. HIV could contribute to lung inflammation, leading to lung cancer, or the lung may be a compartment in which HIV clusters and in which systemic viral suppression is not reflected [32]. Other studies show that HIV viral load level is not associated with lung cancer [28]. Lastly, some studies have observed a positive association between HAART and lung cancer among HIV patients [18], while others have found that HAART is protective against lung cancer [32]. In summary, the explanation for the increased risk of lung cancer among HIV-positive persons is not well understood.

Hodgkin's Lymphoma

The risk of developing Hodgkin's lymphoma is up to ten times higher for HIV-positive persons compared to HIV-negative persons [8, 9, 19]. Like many other NADCs, Hodgkin's lymphoma has poor prognostic results. Examples of prognostic complications associated with Hodgkin's lymphoma include mixed cellularity subtype histology, more advanced stage, extranodal involvement, more aggressive cancer, and poorer therapeutic outcomes [3, 9, 12, 17]. Two main theories exist as to the increased prevalence of Hodgkin's lymphoma among HIV patients. Decreased immune function has been positively correlated with Hodgkin's lymphoma in some studies [1, 3, 8, 38], while others have not found a clear connection between immunosuppression and Hodgkin's lymphoma [9, 17, 33]. The other potential cause for the increase in Hodgkin's lymphoma prevalence is co-infection with EBV [3, 8, 19, 28, 30, 33]. While EBV is found in nearly 90% of the US population, all cases of Hodgkin's lymphoma in HIV-positive persons have been in patients with HIV and EBV co-infection [3]. This suggests that Hodgkin's lymphoma is caused by loss of control of EBV and the direct oncogenic effects of EBV.

Thus far, no studies contradicted the association between EBV and Hodgkin's lymphoma.

Anal Cancer

Anal cancer is the third most common NADC in the HIV-positive population [31]. The risk of developing anal cancer is elevated and rising in HIV patients compared to the general population, with at least a 32-fold increase in risk [1, 19, 27, 31]. The risk is even higher for HIV-positive men who have sex with men (MSM) [31]. The prognosis for anal cancer in the HIV community also differs from the general population. HIV-positive persons are more likely to be diagnosed with anal cancer at a younger age [11]. Several risk factors have been associated with the high risk of anal cancer in HIV patients [3, 19, 27, 31]. HPV may explain the disparity seen in anal cancer risk. HPV is the cause of 88% of anal cancers and because the prevalence of HPV among HIV-positive MSM is between 87% and 98%, this oncogenic virus is nearly endemic in the MSM subgroup [31]. HIV infection may weaken the immune system's ability to survey and destroy HPV. Other risk factors found to be positively associated with anal cancer risk are immunosuppression, low CD4 count [31, 38] and HAART use [28]. The increase may also be due to surveillance bias as the MSM community is likely to be more commonly screened for anal cancer compared to the general population [3].

Liver Cancer

Liver cancer, also known as hepatocellular cancer or hepatocellular carcinoma (HCC), is another common NADC. The risk for HCC is elevated and rising in the HIV-positive population with an estimated threefold increased risk compared to the general population [19, 41]. A more grave prognosis for HCC is common in HIV-positive patients compared

to the general population. HIV-positive persons are more likely to be diagnosed at a more advanced stage, have more symptomatic multifocal, infiltrative tumors, have a shorter interval between HCV infection and HCC diagnosis, and have increased cause-specific mortality [41]. The increased risk of HCC has been mostly attributed to high rates of co-infection of HBV and HCV in the HIV community [3, 19, 33, 41]. Thirty percent of HIV patients are co-infected with HCV, and HIV patients are at a fourfold increased risk of contracting HCV due to similar routes of transmission [3, 41]. Another theory is that the increased rates of alcohol consumption in the HIV-positive community could be associated with the increased risk of HCC [3, 41]. Excessive alcohol consumption can lead to cirrhosis, which can advance to HCC. Several studies have also found that immunosuppression is associated with increased risk of liver cancer [19, 41]. HIV can impact HCC directly in several ways, including accelerating hepatitis to cirrhosis and cirrhosis to HCC, increasing viral loads of HBV and HCV, disrupting the HCV-specific immune response, and potentially directly contributing to liver toxicity [41]. Lastly, HAART has the potential to cause liver toxicity and therefore may directly contribute to the development of HCC [33, 41].

Other NADCs

Another very common, but less studied NADC is skin cancer. Increased risk of basal cell carcimona and squamous cell carcinoma have been found in the HIV-positive population compared to the general population, even when controlling for age, race, and gender [1, 3, 36]. The prognosis for HIV-associated skin cancer cases is worse than the general population. HIV patients develop skin cancer at younger ages, have more aggressive cancer at mutliple sites or in non-sun-exposed areas, and develop increased rates of

cancer recurrence [3, 17]. Risk factors that have been found to be associated with skin cancer are exposure to a long period of immunosuprresion and HPV co-infection [1].

Oropharyngeal cancers, also known as head and neck cancers, have an increased incidence in the HIV community [3]. HIV pateints who are diagnosed with an oropharyngeal cancer generally have worse outcomes compared to the general population. They are more likely to present at a younger age, have more advanced disease, have increases risk of mortality, and experience severe chemotherapy side effects, such as malnutrition, xerostomia, oral infections, and severe mucositis [3, 26]. Tobacco smoking and HPV co-infection have been associated with the increased risk of oropharyngeal cancer diagnosis in HIV-positive persons [3, 26].

Prostate cancer is one of the most interesting NADCs. Research has found conflciting evidence as to the increasing prevalence of prostate cancer in the HIV-positive community. The general conensus, however, is that the incidence of prostate cancer has risen since HAART introduction [15, 33] though the literature is not consistent. Given that HIV lowers androgen levels, which in turn can decrease prostate cancer risk, some researchers suggest this is a biologically plausible mechanism for refuting this claim [8]. Other studies have found that even though there has been an apparent rise in prostate cancer diagnoses since the advent of HAART, HAART is protective against developing prostate cancer and that the increased incidence is a by-product of surveillance bias [33].

Veteran Population

Among HIV-positive US veterans, NADCs such as anal cancer, lung cancer, Hodgkin's lymphoma, and liver cancer have been found in excess rates according to several studies [36, 42]. This study will specifically be using data from the HIV Atlanta Veterans Affairs Cohort Study (HAVACS). Specifically, the dataset is a subset from the HAART era (1996 and forward) of HAVACS. HAVACS is a database that includes all HIV-positive patients seen or treated at the Atlanta VA since 1982, totaling 4334 retired servicemen. As of April 2016, 97% of the cohort was male, 72% African American, 26% Caucasian, and 1.5% Hispanic. The average age in the cohort is 52 years of age. Fifty-three percent contracted HIV through male/male sex, 16% through intravenous drug use (IDU), and 5% of individuals were exposed through heterosexual sexual contact. Seventeen percent of the cohort is co-infected with HCV. Sixty percent of retired servicemen in the cohort had undetectable viral loads and 55% had CD4 cell counts >500 cells/mm^3. Only 2.8% of patients have been lost to follow up and average follow up time per individual is 82 months [42].

Introduction

Since the advent of highly active anti-retroviral therapy (HAART) in 1996, the natural progression of disease for those infected with Human Immunodeficiency Virus (HIV)/ Acquired Immunodeficiency Syndrome (AIDS) has changed drastically. Before HAART, opportunistic infections and AIDS-defining cancers, such as Kaposi's sarcoma, non-Hodgkin's lymphoma (NHL), and invasive cervical cancer, caused the majority of morbidity and mortality related to HIV [9, 14, 15]. As HAART has became the standard of treatment for HIV, HIV patients are living longer and the incidence of AIDS-defining cancer has decreased dramatically while the incidence of non-AIDS defining cancers (NADCs) has increased [1-10, 15-18]. Many NADCs, such as Hodgkin's lymphoma, melanoma, leukemia, liver, lung, anus, oral, pharyngeal, vaginal, colorectal, and kidney cancers, have even been found at a higher incidence in HIV-positive individuals than in the HIV-negative population, even when controlling for cancer risk factors such as age [7, 9, 14, 17, 28, 36]. In addition, those with HIV who are diagnosed with a NADC are usually diagnosed at a younger age and with a more aggressive cancer at a more advanced stage of disease [2, 3, 5-7, 11, 21, 28, 30]. HIV-positive patients also experience more complications with cancer treatment than the general population. While both HAART and chemotherapy regimens have elements that are toxic to the body, HIV-positive patients face the issue of dual toxicity and therefore often need longer breaks between treatments of chemotherapy, which can be less effective in treating cancer [3, 6, 10, 17, 21, 25, 31]. HIV-positive individuals are already immunosuppressed, and the combination of HIV with chemotherapy allows for the cancer patient's immune system to become immensely compromised [10, 21, 25]. Lastly, a major treatment for cancer is

often surgery and HIV-positive patients canmake poor surgical candidates, potentially limiting their treatment options [7, 17, 32].

The cause for the increased risk and deadliness of these cancers among HIV-positive individuals is mostly unknown, though there are several theories that exists as to the biological mechanism. However, the majority of the literature is conflicting. Like AIDS-defining cancers, immunosuppression has been suggested as a potential cause of NADCs by decreasing the immune system's tumor surveillance ability and allowing tumors to grow unchecked [4, 6-8, 11, 16, 22, 24, 34]. However, data have not consistently shown that key markers of immunosuppression, such as CD4 count, are correlated with cancer development [1, 3, 6, 10]. A larger prevalence of traditional cancer risk factors such as sun exposure and smoking in the HIV-positive community have also been hypothesized as a contributing factor to the disparity in cancer rates [1, 11, 18, 34]. After controlling for these known risk factors of cancer, the rates among HIV-infected persons are still higher than the general population [28].

There are two distinct types of NADCs, cancers that have a known viral oncogenesis and cancers without a known viral oncogenesis. Six viruses are universally recognized to be cancer causing viruses. Human T-lymphotrophic virus-1 has been shown to cause lymphocytic leukemia and human herpes virus 8 (HHV8) causes lymphoma. Hepatitis B (HBV) and Hepatitis C (HCV) both led to liver cancer. Epstein-Barr virus (EBV) can cause lymphomas and nasopharyngeal cancers. Lastly, Human Papilloma virus (HPV) is associated with many cancers, including penial, anal, vaginal, vulvar, cervical,

oropharyngeal, and colorectal [7, 16, 19, 29]. Cancers such as melanoma, lung, kidney, bladder, pancreatic, thyroid, prostate, stomach, testicular, brain, eye, esophageal, bone, and soft tissue are not currently definitively linked with any viruses [7, 16, 19, 29]. People infected with HIV are often infected with other viruses, such as EBV, HBV, HCV, herpes, and HPV. It has been hypothesized that these viral co-infections are the culprits behind the virus related NADC disparity between HIV-positive and HIV-negative patients [1, 2, 9, 15-18, 20, 37, 39, 40], but there is still conflicting evidence of the validity of this claim [28]. Even though there is no evidence in humans that HAART causes tumorigenesis, HAART itself has also been theorized as a potential cause of the NADC increase by building on the evidence that HAART increases the risk of cardiovascular disease, neurocognitive disease, and neuroendocrine dysfunctions [1, 4, 16, 34], [1, 7, 8, 17]. Similarly, the HIV virus has been postulated as a direct cause of NADCs [4, 16, 33, 34, 40], but evidence is lacking as to its direct oncogenic effect [3, 10, 39]. Lastly, over-surveillance of HIV-positive individuals for cancer may have caused a false increase in cancer cases for HIV-positive individuals due to increased screening [2, 10, 11, 17, 36, 37] but these data are not consistent.Researchers have also found that routine cancer screening is less often performed in the HIV-positive population, so it may very well be under-reported [21]. There is no universally accepted theory for the cause of this cancer disparity. This far, there has been no research into whether route of HIV transmission has any effect on development of these non-AIDS defining malignancies in HIV-positive individuals, controlling for lifestyle and immunological factors. Often, HIV infection is the outcome of repeated high risk behaviors, such as intravenous drug use or unsafe sexual encounters. Repeat dosing of the HIV virus at the primary site of infection

and the virus's course throughout the rest of the body could affect the development of NADCs independent of viral co-infection, lifestyle factors, and immunological factors. This study seeks to identify if route of transmission of HIV is an independent risk factor for NADCs or merely a proxy measure for other closely associated risk factors.

Methods

Data Set

The dataset is a subset from the HAART era (1996 and forward) of the HIV Atlanta Veterans Affairs Cohort Study (HAVACS). HAVACS is a database that is includes all HIV-positive patients seen or treated at the Atlanta VA since 1982, totaling 4334 retired servicemen. The analysis subset only includes patients who were diagnosed with HIV and/or received their first treatment after 1996, totaling 2115 HIV-positive veterans. This was done to avoid issues of confounding from delayed HAART use. Seven hundred twenty-three more veterans were excluded due to having an unknown or uncommon route of HIV transmission. Only the three most common HIV transmission routes were retained, MSM, IDU, and HETERO. None of the women in the dataset had cancer and were therefore excluded (n=40). The final analysis dataset included 1352 veterans.

Analysis

This study is a secondary analysis of cohort study data. Four hundred and one observations had some pattern of missing variables, with the following variables missing in varying quantities and patterns: number of current HAART medication, category of HAART medications, nadir CD4 count, time on HAART, viral load, log of viral load, and years with AIDS. All variables included were not missing more than 21%. Data was found to be missing at random , so multiple imputation techniques were used to generate matrices of potential values for observations with missing values. Twenty-five imputations were done.

Two outcomes were examined, cancer diagnosis (Yes versus No) and oncongenic cause of the NADC (viral, non-viral, versus no cancer). Associative and predictive analyses were done for both outcomes with both imputed and non-imputed data. Variables for analysis were selected from important variables highlighted in the literature. Multicollinearity was assessed for the imputed and non-imputed datasets using condition indices and variance decomposition proportions. For the binary outcome, logistic regression was used to calculate odds ratios for the odd of developing cancer in those veterans who developed HIV due to MSM or IDU exposure compared to HETERO exposure. For the nominal outcome, polytomous logistic regression was used to estimate odds ratios for the odds of developing a cancer with a known viral oncogenesis and a cancer without a viral cause in those veterans who developed HIV due to MSM or IDU exposure compared to HETERO exposure. The most precise models were chosen.

For predictive modeling, the non-imputed data was randomly divided into two equalized groups, a development dataset and a validation dataset, each containing 676 observations. Statisical testing was done to determine if there was a statistically significant difference between the non-imputed development dataset and the validation dataset. Two sample t test for means, analysis of variance, and Fisher's exact test were the methods used to determines significance. Systematic random sampling was used to select 676 observation from the imputed dataset with no repeated HAVACS identification numbers from different imputations.. The remainder of observations in the imputed dataset became the validation dataset. Predicitive models were created for both the binary and nominal outcomes in the imputed and non-imputed data. Receiver operating characteristic (ROC)

curves were used to find the best predictive models. Those models with the highest area under the curve (AUC) in the development datasets were selected as the most predictive models. These models were then rerun in their corresponding validation dataset to check for overfitting and accuracy of the model.

Results

Descriptive Statistics

As of April 2016, the analysis dataset was comprised of 1352 males, 77.4% of which were African American, 20.5% Caucasian, and 2.1% other races. The average age in the cohort is 50 years old. Seventy-three-point-two percent contracted HIV through MSM, for 14.4% the probable route of HIV transmission was IDU, 12.4% of individuals were exposed through HETERO. Eighteen percent of the cohort is co-infected with HCV, and 7.9% is co-infected with HBV. The median nadir CD4 count and viral load of the cohort is 204.5 cells/mm^3 and 33497 copies of HIV per ml of blood. Fifty-two-point-three percent of participants have had an AIDS diagnosis. Lastly, 93.9% of veterans in this subcohort are currently on a HAART regimen. Table 1 contains the information on the missing variables within the dataset. These variables were imputed upon through multiple imputation. For further demographic information on about the cohort in the non-imputed data, see Table 2.

Bivariate Analysis

All variables were compared individually to the binary outcome using logistic regression. In the non-imputed dataset, patient age, AIDS status, race, nadir CD4 count, years with HIV, and years with AIDS are each bivariately associated with the risk of NADC. In the imputed dataset, patient age, AIDS status, years with HIV, race, nadir CD4 count, years with AIDS, and years on HAART are each associated with developing cancer. Polytomous logistic regression was used to assess bivariate correlations with NADCs associated with viruses and NADCs not associated with viruses. Patient age, AIDS status,

Table 1. Missing Variables in the HAVACS Data Subset

Variable	Number Missing	Percent Missing
Number of HAART Medications	91	6.7%
Category of HAART Medications	202	14.9%
Race	15	1.1%
Nadir CD4 Count	92	6.8%
Years on HAART	278	20.6%
Log of Viral Load	41	3.0%
Viral Load	41	3.0%
Years with AIDS	6	0.4%

race, nadir CD4 count, and years with AIDS are all bivariately associated with either viral or non-viral NADCs in both the non-imputed and imputed datasets.

Multivariate Analysis

The odds of having acquired HIV through MSM behaviors in the HIV-positive veterans who have a NADC is 1.31 times the odds of having acquired HIV through HETERO behaviors in the HIV-positive HAVACS veterans. This finding, however, was not statistically significant. The odds of having acquired HIV through IDU behaviors in HIV-positive patients who developed any NADC is 1.00 times the odds of having acquired HIV through HETERO behaviors in the HIV-positive veterans. This nearly null association was not found to be statistically significant. In the non-imputed data, measures of association were slightly higher for each route of transmission. 1.31 versus 1.39 for odds of having acquired HIV through MSM behaviors in the HIV-positive veterans who have a NADC, and 1.00 versus 1.03 for IDU transmission, in the imputed and non-imputed datasets respectively. Measures of association changed dramatically for MSM transmission when comparing viral and non-viral related NADC. The odds of

Table 2. Descriptive Statistics for the HAVACS Subcohort 1996-2016, n=1352

Variable	Mean or N (SD or Percent)
Patient Age	49.96 (11.69)
Race	
Caucasian	274 (20.5)
African American	1035 (77.4)
Other	28 (2.1)
Nadir CD4 Count	236.72 (203.08)
Log of Viral Load	2.37 (1.45)
Years with HIV	11.06 (5.82)
Route of Transmission	
MSM	989 (73.2)
IDU	195 (14.4)
HETERO	168 (12.4)
AIDS Diagnosis	707 (52.3)
Years with AIDS	4.86 (6.13)
Currently on HAART	1270 (93.9)
Number of HAART Medications	1.78 (1.20)
Years on HAART	10.11 (5.66)
HEART Medication Category	
PI	260 (22.6)
II	105 (9.1)
NNRTI	90 (7.8)
Other	695 (60.4)
Drug Use	
Alcohol Abuse	329 (24.3)
HCV Co-infection	260 (19.2)
HBV Co-infection	244 (18.0)
	107 (7.9)

having acquired HIV through MSM behaviors in the HIV-positive veterans who have a viral related NADC is 1.72 times the odds of having acquired HIV through HETERO behaviors in the HIV-positive veterans. Conversely, the odds of having acquired HIV through MSM behaviors in this population who have a non-viral NADC is only 1.23 times the odds compared to having acquired HIV through HETERO behaviors. Even though there is a sharp increase in the odds for MSM developing a viral NADC, both measures of association are not statistically significant. The change in measure of association for IDU transmission is not as significant. The odds of having acquired HIV

through IDU behaviors in this population who have a viral NADC is only 1.07 times the odds compared to having acquired HIV through HETERO behaviors. Similarly, the odds of having acquired HIV through IDU behaviors in HAVACS veterans who have a non-viral NADC is only 1.10 times the odds of having acquired HIV through HETERO behaviors. Again, these measures of association had no statistical significance. In the imputed and non-imputed datasets, odds ratios for MSM in the polytomous logistic regression did not differ, 1.72 for viral cancers and 1.23 for non-viral cancers in both dataset, respectively. Alternatively, the IDU odds ratios differed significantly between the imputed and non-imputed datasets. For viral cancers, the odds ratio for IDU transmission in the non-imputed dataset is 0.83 compared to 1.07 in the imputed dataset. Concerning non-viral cancers, 1.32 is the odds ratio for IDU transmission in the non-imputed dataset and 1.10 in the imputed dataset. Among all measures, imputed and non-imputed, no statistical significance was found, likely due to limited sample size. Table 3 contains more detailed information on the models and measures of association between routes of association and cancer outcomes.

Predictive Analysis

The development and validation datasets for non-imputed data each contain 676 observation. The groups are remarkably similar, with the only statistically significant differences being in years with HIV and years on HAART (p-value=0.02 for both variables). Both log of viral load and number of HAART medications have borderline statistical significance between the two group, with p-values of 0.08 and 0.07 for each variable, respectively. No other statistically significant differences were found between

Table 3. Measures of Association between Route of Transmission and Various Cancer Types, HAVACS Subcohort n=1352

Data	Outcome	Variables in Model	Route of Transmission	OR	95% CI
Non-Imputed	Cancer	Number of HAART medications, Patient age, Time on HAART, HCV Co-infection, Route of transmission	MSM	1.3855	(0.7377, 2.6024)
			IDU	1.0290	(0.4390, 2.4118)
	Viral Related Cancer	Number of HAART medications, Patient age, Race, Nadir CD4 Count, Time on HAART, HCV Co-infection, Years with HIV, Route of transmission	MSM	1.7233	(0.6712, 4.4246)
			IDU	0.8317	(0.2255, 3.0676)
	Non-Viral Related Cancer		MSM	1.2312	(0.5115, 2.9635)
			IDU	1.3187	(0.4226, 4.1155)
Imputed	Cancer	Patient age, AIDS status, Race, Route of Transmission	MSM	1.3061	(0.7419, 2.2996)
			IDU	0.9957	(0.5127, 1.9338)
	Viral Related Cancer	Patient age, AIDS status, Log of Viral Load, HCV Co-infection, Route of Transmission	MSM	1.7229	(0.7775, 3.8178)
			IDU	1.0672	(0.3778, 3.0144)
	Non-Viral Related Cancer		MSM	1.2300	(0.6011, 2.5171)
			IDU	1.1003	(0.4261, 2.8409)

the groups. Table 4 contains further information about the development and validation datasets. For the non-imputed dataset, the best model to predict any NADC contained the variables HAART medication category, patient age, AIDS status, nadir CD4 count, drug use, alcohol abuse, HPV co-infection, HAART nonadherence, HBV co-infection, HCV co-infection, viral load, years with AIDS, and route of HIV transmission. The AUC in the development dataset was 0.8178 and the AUC in the validation dataset was 0.8117. For the imputed dataset, the best predictive model for any NADC contained number of

HAART medications, category of HAART medications, patient age, AIDS status, race, nadir CD4 count, log of viral load, alcohol abuse, HPV co-infection, herpes co-infection, HCV/HBV co-infection, HAART non-adherence, H. Pylori co-infection, viral load, years with HIV, years with AIDS, and route of transmission. The development dataset's AUC was 0.8074 and the validation dataset's AUC was 0.7723 for any NADC in the imputed dataset. The model that best predicts development of a non-viral NADC in HIV-positive

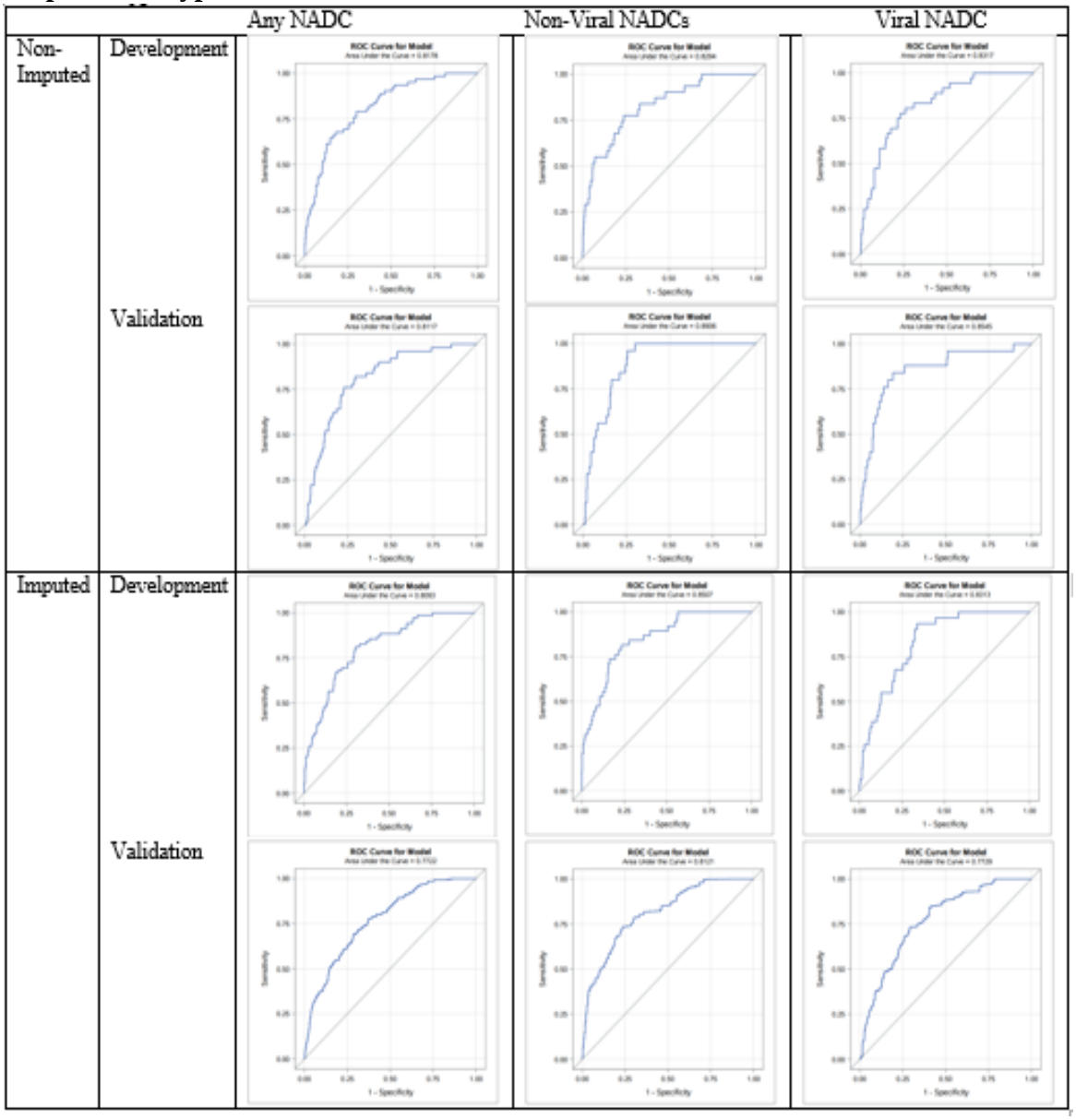
Table 4. Comparison of Descriptive Statistics between the Development and Validation Datasets for Predictive Modeling, n=676 for each

	Development Dataset, n=676		Validation Dataset, n=676		
Variable	**Mean or N**	**SD or Percent**	**Mean or N**	**SD or Percent**	**p-value**
Patient Age	49.82	11.93	50.10	11.44	0.6619
Race					0.3348
Caucasian	126	18.9	148	22.1	
African American	528	79.0	507	75.8	
Other	14	2.1	14	2.1	
Nadir CD4 Count	245.83	212.89	227.60	192.52	0.1112
Log of Viral Load	2.30	1.40	2.44	1.50	0.0828
Years with HIV	10.70	5.83	11.42	5.80	0.0234
Route of Transmission					0.6525
MSM	469	69.4	493	72.9	
IDU	101	14.9	94	13.9	
HETERO	106	15.7	89	13.2	
AIDS Diagnosis	346	51.2	315	46.6	0.4459
Years with AIDS	4.66	5.99	5.07	6.27	0.2198
Currently on HAART	637	94.2	633	93.6	0.7327
Number of HAART Medications	1.72	1.15	1.84	1.24	0.0683
Years on HAART	9.71	5.72	10.51	5.56	0.0195
HAART Medication Category					0.1285
PI	116	19.9	144	25.4	
II	56	9.6	49	8.6	
NNRTI	43	7.4	47	8.3	
Other	367	63.1	328	57.8	
Drug Use	167	24.7	162	24.0	0.7999
Alcohol Abuse	125	18.5	135	20.0	0.5346
HCV Co-infection	120	17.8	124	18.3	0.8320
HBV Co-infection	49	7.3	58	8.6	0.4204

*Two sample t test, ANOVA, and Fisher's Exact test used to test for significance.

veterans in the non-imputed dataset contains HAART medication category, patient age, AIDS status, race, years on HAART, log of viral load, drug use, alcohol abuse, HPV/HCV/HBV/H. Pylori co-infection, HAART nonadherence, viral load, years with HIV, years with AIDS, and route of transmission. The development AUC was 0.8294 and the validation AUC was 0.8906. In the non-imputed dataset for viral NADC outcome, number of HAART medications, patient age, AIDS status, nadir CD4 count, drug use, alcohol abuse, HPV/HCV/HBV/H. Pylori, HAART nonadherence, years with HIV, years with AIDS, and route of transmission were the variable that best predicted occurrence of viral NADCs. The AUC in the development dataset was 0.8317 compared to 0.8545 in the validation dataset. For the imputed dataset, the model that best predict non-viral NADCs in HIV-positive veterans contains all variables except number of HAART medications, HBV co-infection, and drug use. The development dataset had an AUC of 0.8515 and the validation dataset had an AUC of 0.8138. For viral NADCs, the imputed dataset's best predictive model contained all variables except category of HAART medication. The development AUC was 0.8257 and the validation AUC was 0.7729 for viral NADCs in the imputed dataset. Figure 1 displays the ROC curves for the development and validation data of varying imputation types and cancer outcomes.

Figure 1. ROC Curves of Development and Validation Dataset of Varying Imputation Types and Cancer Outcomes



Discussion

While route of HIV transmission was found to be associated with risk of developing NADCs, this was not statistically significant. Specifically, patients with MSM transmission were 30% more likely to develop NADCs compared to patients who had HETERO transmission, although this was not statistically significant. The risk of developing NADCs between HIV positive veterans who contracted the virus through IDU and those who contracted HIV through HETERO behaviors is comparable. HIV-positive patients with MSM transmission have a 72% increased odds of developing a viral NADC compared to those veterans who develop HIV through HETERO behaviors, but findings were not statistically significant. When comparing risk of developing non-viral NADCs, HIV-positive veterans with MSM transmission have a 23% increased odds compared to HETERO transmission veterans, though this finding did not prove to be statistically significant. The measures of association for veterans that contracted HIV through MSM behaviors did not change with imputation. However, the measures of association for veterans who became HIV infected through IDU behaviors varied greatly between the non-imputed and imputed datasets. HIV-positive veterans with IDU transmission were at 20% decreased odds of developing a viral NADC compared to veterans who contracted HIV through HETERO behaviors in the non-imputed data. This protective effect became harmful in the imputed data, but neither measure showed statistical significance. Conversely, veterans with IDU transmission with HIV have a 7% increased odds of developing a viral NADC compared to HETERO infected veterans. In the non-imputed dataset, there is a 32% increased odds of developing a non-viral NADC for veterans with IDU transmission that have HIV compared to veterans with HETERO

transmission of HIV. In the imputed data, the odds of developing a non-viral NADC are only 10% increased for veterans with IDU transmission compared to veterans with HETERO transmission. In this instance, while the measure of association remains harmful for imputed and non-imputed data, the odds decreased substantially.

The differences between the measures of association when comparing imputed and non-imputed data for the effect of IDU on the development of viral related NADCs can be attributed to the increased proportion of missing of variable among intravenous drug users. Veterans who contacted HIV through IDU behaviors are the smallest risk group in the study. Additionally, intravenous drug users are difficult to gather information and follow-up on due to the transitory nature of their lifestyle. Relatively low numbers of intravenous drug users to start with, combined with removal of IDU transmission observations that did not have complete data for logistic analysis, could explain the substantial difference in association between imputed and non-imputed data for intravenous drug users.

The best model for predicting any type of NADC varied substantially between the imputed and non-imputed dataset. In the non-imputed data, thirteen variables improve predictability of NADCs, including drug use, which is not predictive in the imputed model. For the imputed predictive model, eighteen variables predict NADC development, including number of HAART medications, race, log of viral load, Herpes co-infection, H. Pylori co-infection, and years with HIV, which are not predictive in the non-imputed data. Variables related to viruses and immunosuppression are more predictive in the imputed data. The non-imputed model for both development and validation produces the

higher AUCs compared to the imputed AUCs, but may be limited by increased missing variables. When non-viral NADCs are the outcome of interest, the predictive models for imputed and non-imputed data were similar. The non-imputed model contained seventeen variables and differed from the imputed model by including drug use and HBV co-infection. Eighteen variables were in the imputed model, including nadir CD4 count, years on HAART, and herpes co-infection, which were not included in the non-imputed predictive model. Overall, the models were very similar and forcing the models to be identical would not have drastically changed the AUCs in the development or validation of either dataset.

For viral NADC, there are many differences between the imputed and non-imputed predictive models. There are fourteen variables in the non-imputed model. Within the imputed model, nineteen variables are predictive of viral NADC, including all variables from the non-imputed model plus race, years on HAART, log of viral load, viral load, and Herpes co-infection. Similar to predicting overall risk of NADC, risk of viral NADCs in the imputed data incorporated more variables related to viruses and immunosuppression, reflecting the hypothesized causes of these cancers. Overall, the non-imputed models produced the higher AUCs for both development and validation, with the only exception being models containing non-viral NADC as the outcome. AUCs ranged from 0.77 to 0.89, showing that viremia, immunosuppression, oncogenic viral co-infections, high risk lifestyle factors, and route of transmission are strong predictors of NADCs in HIV positive veterans.

This study is not without its limitations. As it is a secondary analysis of a prospective cohort, the type of data collected and methods of data collection were not specific to the needs of this analysis. In addition, there were issues with sample size due to the small numbers cancer cases. Only 140 patients had cancer in the 1352 person dataset, representing only 10.4% of the study sample. This is a lower rate than expected given data about prevalence of NADCs previously reported in the literature Almost 90% of the study population was between 45-64 years of age. In this age group in the general US population, 9.3% of people have cancer [43]. In our dataset, 10.6% of the patients in this age range had cancer, a rate that is not statistically significantly higher than the general population ($p=0.17$).

Another limitation was the size of the population included for analysis. Only patients that were diagnosed in the HAART era were used because the rates of NADCs differ significantly pre and post introduction of HAART. Sample size was also limited by information on route of HIV transmission. As route of HIV transmission was the primary risk factor of interest, missing information on this variable was not imputed and therefore all patients who were missing this variable were removed from the analysis. There was not, however, a difference cancer rates between patients with known and unknown HIV transmission risk factors for those included in the analysis versus those who were excluded ($p=0.82$).

Furthermore, two potential confounders in the associative modeling and covariates in the predictive modeling could not be assessed. Smoking and EBV diagnosis are not consistently included in the HAVACS dataset and therefore their effects could not be

assessed. Smoking is not necessarily linked with all NADCs, but it is causally linked with several important cancers such as lung and oropharyngeal cancer. EBV has been linked with Hodgkin's lymphoma, especially in HIV patients. However, EBV can be found in 90% of adult US citizens [44], therefore its effect as a confounder is likely limited. Regardless, it should be included in future research.

A commonly accepted rule for predictive modeling is that the number of outcomes should be at least be equal to ten times the number of variables. This only holds true for one of the prediction models included in this analysis. Lastly, for validation of the predictive model, an independent sample was not used, so true validation was not achieved. The development and validation data are just subgroups of the imputed and non-imputed dataset and therefore not truly independent. However, doing validation in this way does correct for overfitting of the model to the data.

This study also has several strengths. This study has a long follow-up time for participants with an average over six years of follow-up data accumulated per veteran. This is ample time for cancer diagnosis in this older population. The study also has very low loss to follow up with a rate of attrition of less than 3% per year. Misclassification bias of either the outcome or the exposure is not an issue in this study as all cases of cancer were histologically confirmed at the VA and unknown/unsure routes of HIV transmission were removed from the analysis. Selection bias should also not be an issue for this study because all HAVACS participants that were eligible to participate (i.e. diagnosed during or after 1996 with a known route of transmission) were used in analysis. Lastly, veterans receive all their medical care at the VA including their HIV

care, which gives the HAVACS dataset and researchers full access to the participants' medical history.

In summary, even though it cannot be stated with any statistical certainty, risk of NADC development seems to vary by route of HIV transmission. While patients with HETERO and IDU HIV transmission risk carry similar odds for developing a NADC, patients with MSM transmission appear to have increased odds of developing NADCs. This study should be repeated with either a larger sample size to reach statistical significance or as a case-control study due to the limited number of outcomes. Study findings suggest that perhaps specifically HIV-positive MSM should be screened regularly for NADCs. Viral co-infections, immunosuppression proxy factors, and measures of viremia together significantly predict the risk of developing NADCs. These models can be used by clinicians to predict cancer risk in the HIV-positive community and to adjust cancer screening routines by HIV transmission route as needed.

Public Health Implications

This study has some public health implications. With the suggestion of differing risks for NADCs by HIV risk group, the idea of targeted surveillance and prevention measures is supported. For example, the study found that MSM veterans are at greater risk of developing any NADC, viral NADCs, and non-viral NADCs. With this information, we can increase screenings for NADC in patients who acquired HIV through MSM behaviors at earlier ages. By catching the cancer earlier, the patients may have better prognoses and survival ouctomes. Additionally, by creating a predictive model for NADC development among HIV-positive veterans, the first steps have been taken for making a risk assessment score for categorizing HIV-positive patients into NADC risk categories. By dividing HIV patients into transmission risk categories, we can screen and intervene earlier to improve cancer and health outcomes.

There are several future directions that this research can take. Due to the low number of cancer outcomes, recreating this study as a case control study may produce more statistically signifiant outcomes than the cohort study. With a larger cohort and more outcomes, researchers could identify if particular primary cancer sites are associated with different routes of HIV transmission. This could be extended into examining the relationship between HIV routes of transmission and histology types of cancer. Additionally, after using the predicitive model results to create a risk score assessment model, more specific risk score assessments can be made. Similar to the future directions the associative modeling can take, predictive risk score assessments can be made for both specific primary cancer sites and cancer histology types.

References

1. Bedimo, R., et al., *Trends in AIDS-Defining and Non—AIDS-Defining Malignancies among HIV-Infected Patients: 1989–2002.* Clinical infectious diseases, 2004. **39**(9): p. 1380-1384.
2. Kowalski, J., et al., *The Spectrum of Malignancies among Adult HIV Cohort in Poland between 1995 and 2012: A Retrospective Analysis of 288 Cases.* Contemp Oncol (Pozn), 2015. **19**(3): p. 226-235.
3. Layton, J.L. and J.J. Castillo, *Non-AIDS defining cancers.* Medicine and Health Rhode Island, 2010. **93**(10): p. 296.
4. Lesko, C.R., et al., *Association of injection drug use with incidence of HIV-associated non-AIDS-related morbidity by age, 1995–2014.* AIDS, 2016. **30**(9): p. 1447-1455.
5. Lin, C.-S., et al., *Cancer survival in patients with HIV/AIDS in the era of highly active antiretroviral therapy in Taiwan: A population-based cohort study.* Cancer epidemiology, 2013. **37**(5): p. 719-724.
6. Meernik, C., et al., *The changing pattern of ano-rectal cancer, squamous cell carcinoma of the eye, and Hodgkin's lymphoma as non-AIDS-defining cancers, by HIV status, in Tanzania over 11 years (2002-2012): a retrospective case-report study.* Infectious agents and cancer, 2014. **9**(1): p. 1.
7. Nguyen, M.L., K.J. Farrell, and C.J. Gunthel, *Non–AIDS-Defining Malignancies in Patients with HIV in the HAART Era.* Current infectious disease reports, 2010. **12**(1): p. 46-55.

8. Patel, P., et al., *Incidence of types of cancer among HIV-infected persons compared with the general population in the United States, 1992–2003.* Annals of internal medicine, 2008. **148**(10): p. 728-736.
9. Pinzone, M., et al., *Non-AIDS-defining cancers among HIV-infected people.* Eur Rev Med Pharmacol Sci, 2012. **16**(10): p. 1377-88.
10. Riedel, D.J., et al., *Virologic and Immunologic Outcomes in HIV-Infected Patients with Cancer.* AIDS Research and Human Retroviruses, 2016.
11. Shiels, M.S., et al., *HIV Infection, Immunosuppression, and Age at Diagnosis of Non-AIDS-Defining Cancers.* Clinical Infectious Diseases, 2016: p. ciw764.
12. Worm, S.W., et al., *Non-AIDS defining cancers in the D: A: D Study-time trends and predictors of survival: a cohort study.* BMC infectious diseases, 2013. **13**(1): p. 1.
13. Yang, J., et al., *Prevalence and mortality of cancer among HIV-infected inpatients in Beijing, China.* BMC infectious diseases, 2016. **16**(1): p. 1.
14. Shiels, M.S., et al., *A meta-analysis of the incidence of non-AIDS cancers in HIV-infected individuals.* Journal of acquired immune deficiency syndromes (1999), 2009. **52**(5): p. 611.
15. Shiels, M.S., et al., *Cancer burden in the HIV-infected population in the United States.* Journal of the National Cancer Institute, 2011. **103**(9): p. 753-762.
16. Calabresi, A., et al., *Incidence of AIDS-defining cancers and virus-related and non-virus-related non-AIDS-defining cancers among HIV-infected patients compared with the general population in a large health district of northern Italy, 1999–2009.* HIV medicine, 2013. **14**(8): p. 481-490.

17. Pantanowitz, L., H.P. Schlecht, and B.J. Dezube, *The growing problem of non-AIDS-defining malignancies in HIV.* Current Opinion in Oncology, 2006. **18**(5): p. 469-478.
18. Zlotorzynska, M., et al., *Retrospective cohort study of cancer incidence and mortality by HIV status in a Georgia, USA, prisoner cohort during the HAART era.* BMJ open, 2016. **6**(4): p. e009778.
19. Shiels, M.S. and E.A. Engels, *Evolving epidemiology of HIV-associated malignancies.* Current Opinion in HIV and AIDS, 2017. **12**(1): p. 6-11.
20. Burgi, A., et al., *Incidence and risk factors for the occurrence of non-AIDS-defining cancers among human immunodeficiency virus-infected individuals.* Cancer, 2005. **104**(7): p. 1505-1511.
21. Deeken, J.F., et al., *The rising challenge of non–AIDS-defining cancers in HIV-infected patients.* Clinical infectious diseases, 2012. **55**(9): p. 1228-1235.
22. Godbole, S.V., et al., *HIV and cancer registry linkage identifies a substantial burden of cancers in persons with HIV in India.* Medicine, 2016. **95**(37): p. e4850.
23. Hema, M.N., et al., *Low CD4/CD8 Ratio Is Associated with Non AIDS-Defining Cancers in Patients on Antiretroviral Therapy: ANRS CO8 (Aproco/Copilote) Prospective Cohort Study.* PloS one, 2016. **11**(8): p. e0161594.
24. Sachdeva, R.K., et al., *Spectrum of AIDS defining & non-AIDS defining malignancies in north India.* The Indian Journal of Medical Research, 2016. **143**(Suppl 1): p. S129.

25. Suneja, G., et al., *Cancer Treatment in Patients With HIV Infection and Non–AIDS-Defining Cancers: A Survey of US Oncologists.* Journal of Oncology Practice, 2015: p. JOP. 2014.002709.
26. Zucchetto, A., et al., *Non-AIDS Defining Cancer Mortality: Emerging Patterns in the Late HAART Era.* Journal of acquired immune deficiency syndromes (1999), 2016.
27. Castilho, J.L., et al., *HIV and cancer: a comparative retrospective study of Brazilian and US clinical cohorts.* Infectious agents and cancer, 2015. **10**(1): p. 1.
28. Engels, E.A., *Non-AIDS-defining malignancies in HIV-infected persons: etiologic puzzles, epidemiologic perils, prevention opportunities.* AIDS (London, England), 2009. **23**(8): p. 875-885.
29. Liao, J.B., *Viruses and Human Cancer.* The Yale Journal of Biology and Medicine, 2006. **79**(3-4): p. 115-122.
30. Lim, S.T. and A.M. Levine, *Non-AIDS-defining cancers and HIV infection.* Current HIV/AIDS Reports, 2005. **2**(3): p. 146-153.
31. Chia-ching, J.W., J. Sparano, and J.M. Palefsky, *Human Immunodeficiency Virus/AIDS, Human Papillomavirus, and Anal Cancer.* Surgical Oncology Clinics of North America, 2017. **26**(1): p. 131.
32. Sigel, K., A. Makinson, and J. Thaler, *Lung cancer in persons with HIV.* Current Opinion in HIV and AIDS, 2017. **12**(1): p. 31-38.
33. Cobucci, R.N.O., et al., *Assessing the impact of HAART on the incidence of defining and non-defining AIDS cancers among patients with HIV/AIDS: A systematic review.* Journal of infection and public health, 2015. **8**(1): p. 1-10.

34. Riedel, D.J., A.F. Rositch, and R.R. Redfield, *Patterns of HIV viremia and viral suppression before diagnosis of non-AIDS-defining cancers in HIV-infected individuals.* Infectious agents and cancer, 2015. **10**(1): p. 1.
35. Pacheco, Y., et al., *Increased risk of non-AIDS-related events in HIV subjects with persistent low CD4 counts despite cART in the CoRIS cohort.* Antiviral research, 2015. **117**: p. 69-74.
36. Bedimo, R.J., et al., *Incidence of Non-AIDS-Defining Malignancies in HIV-Infected Vs. Non-Infected Patients in the HAART Era: Impact of Immunosuppression.* Journal of acquired immune deficiency syndromes (1999), 2009. **52**(2): p. 203.
37. Cooley, T.P., *Non–AIDS-defining cancer in HIV-infected people.* Hematology/oncology clinics of North America, 2003. **17**(3): p. 889-899.
38. Grulich, A.E., et al., *Rates of non-AIDS-defining cancers in people with HIV infection before and after AIDS diagnosis.* Aids, 2002. **16**(8): p. 1155-1161.
39. Kesselring, A., et al., *Immunodeficiency as a risk factor for non-AIDS-defining malignancies in HIV-1-infected patients receiving combination antiretroviral therapy.* (1537-6591 (Electronic)).
40. Squillace, N., et al., *High-density lipoprotein-cholesterol levels and risk of cancer in HIV-infected subjects: Data from the ICONA Foundation Cohort.* Medicine, 2016. **95**(36).
41. El Dika, I., J.J. Harding, and G.K. Abou-Alfa, *Hepatocellular carcinoma in patients with HIV.* Current Opinion in HIV and AIDS, 2017. **12**(1): p. 20-25.

42. Guest, J.L., et al., *Cohort Profile: The HIV Atlanta Veterans Affairs Cohort Study (HAVACS).* International journal of epidemiology, 2016: p. dyw071.
43. National Center for Health Statistics, *Table A-3a. Age-adjusted percentages (with standard errors) of cancer among adults aged 18 and over, by selected characteristics: United States, 2014* Summary Health Statistics: National Health Interview Survey, Editor. 2014, Centers for Disease Control and Prevention: Atlanta.
44. Balfour, H.H., et al., *Age-specific prevalence of Epstein–Barr virus infection among individuals aged 6–19 years in the United States and factors affecting its acquisition.* Journal of Infectious Diseases, 2013. **208**(8): p. 1286-1293.