Published in final edited form as:

Pharmacoepidemiol Drug Saf. 2013 September ; 22(9): . doi:10.1002/pds.3476.

Validation of an Algorithm to Identify Antiretroviral-Naïve Status at Time of Entry into a Large, Observational Cohort of HIV-infected Patients

Neel R. Gandhi, Janet P. Tate, Maria C. Rodriguez-Barradas, David Rimland, Matthew Bidwell Goetz, Cynthia Gibert, Sheldon T. Brown, Kristin Mattocks, and Amy C. Justice
 Division of General Internal Medicine, Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, New York (NRG); Departments of Epidemiology, Global Health, & Infectious Diseases, Rollins School of Public Health, Emory University, Atlanta (NRG); Section of General Internal Medicine, Department of Medicine, Yale University School of Medicine and VA Connecticut Healthcare System, New Haven, Connecticut (JPT, ACJ); Infectious Disease Section and Department of Medicine, Michael E. DeBakey VA Medical Center and Baylor College of Medicine, Houston, TX (MCRB); Infectious Diseases Section, Department of Medicine, Emory University School of Medicine and Atlanta VA Medical Center, Atlanta, Georgia (DR); Section of Infectious Diseases, VA Greater Los Angeles Healthcare System and David Geffen School of Medicine at UCLA, Los Angeles, California (MBG); Section of Infectious Diseases, Washington DC VA Medical Center and George Washington University Medical Center, Washington, District of Columbia (CG); Infectious Disease Section, Department of Medicine, James J. Peters VA Medical Center and Mt. Sinai School of Medicine, New York, New York (STB).

Abstract

Purpose—Large, observational HIV cohorts play an important role in answering questions which are difficult to study in randomized trials; however, they often lack detailed information regarding previous antiretroviral treatment (ART). Knowledge of ART treatment history is important when ascertaining the long-term impact of medications, co-morbidities, or adverse reactions on HIV outcomes.

Methods—We performed a retrospective study to validate a prediction algorithm for identifying ART-naïve patients using the Veterans Aging Cohort Study's Virtual Cohort—an observational cohort of 40,594 HIV-infected veterans nationwide. Medical records for 3070 HIV-infected patients were reviewed to determine history of combination ART treatment. An algorithm using Virtual Cohort laboratory data was used to predict ART treatment status and compared to medical record review.

Results—Among 3070 patients' medical records reviewed, 1223 were eligible for analysis. Of these, 990 (81%) were ART naïve at cohort entry based on medical record review. The prediction algorithm's sensitivity was 86%, specificity 47%, positive predictive value (PPV) 87% and negative predictive value 45%, using a viral load threshold of <400 copies/ml. Sensitivity analysis revealed that PPV would be maximized by increasing the viral load threshold, whereas sensitivity would be maximized by lowering the viral load threshold.

Conclusions—A prediction algorithm using available laboratory data can be used to accurately identify ART-naïve patients in large, observational HIV cohorts. Use of this algorithm will allow

Corresponding Author: Neel R. Gandhi, MD Department of Epidemiology Rollins School of Public Health Emory University 1518 Clifton Rd NE, CNR 3031 Atlanta, GA 30322 neelgandhi@alumni.williams.edu 404 727-2317.

Conflict of Interest Statement: The authors have no conflicts of interest to report.

investigators to accurately limit analyses to ART-naïve patients when studying the contribution of ART to outcomes and adverse events.

Keywords

HIV-1; Antiretroviral Therapy, Highly Active; Cohort Studies; HIV infections/epidemiology; HIV infections/drug therapy

INTRODUCTION

Over the past two decades, HIV has been transformed from a universally fatal illness to a chronic disease, with the advent of combination or “highly-active” antiretroviral therapy (ART).¹ While randomized trials in carefully selected patients were key to determining the efficacy of ART,² data from observational studies have allowed us to learn the effectiveness and safety of ART use in a wider range of patients. Large cohorts have the advantage of being able to observe long-term impact of treatment, including adverse effects, and durability of treatment.³⁻⁸ Observational studies also include populations underrepresented in clinical trials, such as persons with pre-existent co-morbidities, older patients, minorities and women.⁹⁻¹³ Thus, insights from these studies are more generalizable to routinely-treated populations. Recognizing this potential, large collaborative cohorts have been organized, that link and pool data, providing sufficient statistical power to examine a variety of research questions which haven’t been addressed by randomized trials.^{4,8,9,14-19}

Knowledge of patients’ prior ART treatment history is essential to understanding the impact of various factors on outcomes and survival. Specifically, whether patients are ART-naïve or ART-experienced at entry into an observational cohort must be determined for several reasons. ART is usually more effective in those who have not had prior antiretroviral treatment.^{5,10} A patient’s first ART regimen is often the most durable.⁶ Because adverse events or toxicities from ART medications may be cumulative and irreversible,²⁰⁻²² knowledge of previous exposure to ART is needed to correctly attribute toxicities or the impact of comorbidities, and to account for potential biases in subsequent treatment choices. For these reasons, many studies attempt to restrict inclusion to ART-naïve patients to ensure the study population is observed from the beginning of their ART treatment. However, ART treatment status is often difficult to ascertain for patients pooled in large collaborative cohorts, since interviews of patients or review of physicians’ records may not be feasible.

In an effort to overcome this challenge, we validated an algorithm for predicting ART-naïve status using laboratory data readily available in a large observational cohort of HIV-infected individuals nationwide, and comparing it to the gold-standard of comprehensive chart review. The algorithm was originally developed using 216 patients from the Atlanta VA Medical Center.²³ Validation was performed in a subset of patients in the Veterans Aging Cohort Study Virtual Cohort who were receiving care at one of eight VA facilities.

METHODS

Setting and data sources

The Veterans Aging Cohort Study (VACS) is an observational study of HIV-infected Veterans, which consists of two complementary cohorts: the “Virtual Cohort” and an 8-site prospective study (VACS-8). The Virtual Cohort, described in detail previously,¹⁸ consists of 40,594 HIV-infected Veterans for whom data on inpatient and outpatient diagnoses, laboratory results, and pharmacy fills are available for all care received within the VA since the introduction of the electronic medical record in 1997. The Virtual Cohort includes nearly all HIV-infected Veterans in care at VA medical centers nationwide. The “date of entry”

into the Virtual Cohort is the date of the first HIV ICD-9 diagnostic code (e.g., 042 or V08) within any VA utilization database.

For the current study, we used a subset of Virtual Cohort patients who were also enrolled in VACS-8, a multi-center, prospective observational study of HIV-infected Veterans in care at VA medical centers in Atlanta, Baltimore, Bronx, Brooklyn/Manhattan, Houston, Los Angeles, Pittsburgh, and Washington DC.²⁴ In addition to the complete electronic medical record, VACS-8 collects detailed and extensive data from patient interviews, progress notes, and records received from other facilities. In this study, we performed a comprehensive review of the patient medical record and data collected for VACS-8 to determine the location and timing of a patients' first ART, whether in the VA or elsewhere.

Study population and data collection

Medical records of all HIV-infected Virtual Cohort patients enrolled in VACS-8 before 31 December 2005 were reviewed for eligibility. Patients were eligible for the current study if: their full medical record was available; it was possible to determine whether, and when, they ever received any ART; and their Virtual Cohort entry was on or after 1 January 1997. Patients were excluded if: they did not have an HIV viral load measured within 365 days of Virtual Cohort entry; or they received a mono- or dual-ART treatment regimen at any point before their first combination or "highly-active" ART regimen, based on medical record review. Trained reviewers manually abstracted medical records to determine the date and location (in VA or outside) of each patient's first combination ART regimen. Reviewers examined electronic and paper medical charts, at the patient's enrollment site and were instructed to search for electronic records from other VA sites where the patient had been seen. When an exact date was not available (e.g., March 1999, or "sometime in 1996") it was recorded as the 15th of the month or July 1 of the year.

We defined combination or "highly-active" ART as a regimen containing at least three antiretroviral medications from two different classes of therapy (e.g., nucleoside reverse transcriptase inhibitors, protease inhibitors, etc.) or the specific triple nucleoside regimen of zidovudine/lamivudine/abacavir. Regimens with fewer than three drugs were considered mono or dual-ART. When patients received antiretroviral medications in a fixed dose combination pill (e.g., Combivir), the components (i.e., zidovudine and lamivudine) were considered as individual medications to determine whether it was a combination ART regimen. Low-dose ritonavir used to boost levels of other protease inhibitors was not counted as one of the three antiretroviral medications needed to qualify as a combination ART regimen. Patients were considered "ART-naïve" if they had never received an ART medication prior to entry into the Virtual Cohort; they were considered "ART-experienced" if their date of ART initiation preceded their entry into the Virtual Cohort. All patients in this study who were naïve to combination ART were, by definition, also naïve to all ART medications, since analysis was restricted to patients who had not received mono or dual ART. Patients known to have started ART, but with unknown initiation date, were classified as ART-experienced at Virtual Cohort entry.

Algorithm

The goal of the algorithm was to identify whether individuals were ART-naïve at entry into the Virtual Cohort. The algorithm was initially derived with HIV-infected patients at the Atlanta VA Medical Center, in whom data from the Virtual Cohort were utilized to determine its predictive value, compared with medical record review.²³ Data regarding patients' first HIV test, CD4 count, viral load, HIV genotype, prescription of ART within the VA, date of HIV diagnosis, and ICD-9 diagnostics codes for HIV, AIDS, opportunistic illnesses, and comorbid conditions (e.g., hepatitis B & C, sexually transmitted disease), were

tested for inclusion in the algorithm. The optimal algorithm found in this initial derivation study, used the following criteria to declare a patient ART-naïve at cohort entry if 1) both CD4 count and HIV viral load were measured before receipt of the first ART prescription in the VA and 2) that the HIV viral load result at the time of Virtual Cohort entry was greater than 400 copies/ml.²³ The absence of a CD4 test, or the presence of a low viral load, suggested that the patient had already begun ART outside the VA, before beginning to receive medication from the VA. Among 216 derivation patients, the algorithm had a positive predictive value of 75% for ART-naïve status.

We sought to validate this algorithm using a larger patient sample from multiple, geographically-distributed VA sites. As in the previous study, patients' ART-naïve status at the time of entry into the Virtual Cohort was determined using the medical record as the "gold standard". Patients were considered ART-naïve if their first ART use occurred after the Virtual Cohort entry date. Patients with ART use before their Virtual Cohort entry date were considered ART-experienced.

Next, we applied the algorithm to each patient to determine whether their status was "ART-naïve" or "ART-experienced" using only Virtual Cohort data. We then compared the algorithm's designation to the gold standard medical record review to determine the sensitivity, specificity, positive and negative predictive values (PPV and NPV, respectively) of the algorithm. Additionally we tested the effect of each component of the algorithm individually, as well as various HIV viral load cut points (e.g., 50, 400, 1000, 5000 or 10,000 copies/ml), to determine their influences on the performance measures.

Finally, we examined sensitivity and PPV versus HIV viral load cut points to allow other investigators to determine an optimal threshold for their specific needs (e.g., highest PPV irrespective of loss in sensitivity; tolerance of low PPV in effort to maintain large sample size and high sensitivity). For our purposes, we sought to maximize PPV, while minimizing the loss in sensitivity, when searching for the optimal threshold. Nonetheless, the data and graph of multiple viral load cut points is presented to allow investigators to customize the threshold for their individual needs.

This study was approved by the Institutional Review Boards of Yale University and each VACS-8 site. SAS version 9.2 was used for all analyses.

RESULTS

We reviewed medical records for 3070 HIV-infected patients enrolled in both the Virtual Cohort and VACS-8. Of these, 1223 were eligible for analysis (1254 were excluded because their Virtual Cohort entry was before 1 January 1997, 10 had no HIV viral load measured within a year of Virtual Cohort entry, and 583 previously received a mono/dual-ART regimen ; Figure 1). The vast majority of patients were male (1190, 97%) with a median age of 45 years (IQR 39 – 51; Table 1). Seventy-two percent were African-American, 16% white and 9% Hispanic. Nearly half the patients entered the Virtual Cohort before 2000. The median CD4 count at Virtual Cohort entry was 284 cells/mm³ (IQR 96–470) and median HIV viral load was 21,752 copies/mL (IQR 1386 – 109,814).

ART status by chart review

At the time of entry into the Virtual Cohort, 990 (81%) patients were naïve to ART, based on medical record review, whereas 233 (19%) had already received an ART regimen prior to entry. The proportion of patients that were ART-naïve at Virtual Cohort entry was higher in 1997-1999 (87%) than in 2000-2005 (76%), and was higher at the Bronx (92%) and Baltimore (91%) sites than at the Atlanta (69%) and Pittsburgh (69%) sites (Figure 2). Of

the 990 ART-naïve patients, 606 (61%) had a CD4 count less than 350, 189 (19%) between 350 and 500, and 198 (20%) greater than 500 cells/mm³. Among 606 patients with CD4 counts less than 350 cell/mm³, 84% (n= 507) initiated ART within one year, while 11% (n=65) initiated ART more than 1 year later, and 6% (n=34) had not initiated ART by the end of the follow-up period.

Validation of prediction algorithm

Compared to the medical record review, the algorithm had 86% (95% CI 84%-89%) sensitivity and 47% (95% CI 40%-53%) specificity for identifying ART-naïve status, with PPV of 87% (95% CI 85%-89%) and NPV of 45% (95% CI 38%-51%) using an HIV viral load threshold of >400 copies/ml (Table 2). This is an improvement over the *de facto* 81% PPV without any algorithm. Sensitivity varied by site from 77% to 93%, as did PPV, 78% to 95%. Sensitivity was similar in 1997-1999 (87%) and 2000-2005 (86%) but PPV was higher in the earlier years (91%) than the later years (84%); there were no significant differences in sensitivity or PPV based on age, race or sex (data not shown).

The viral load threshold of >400 copies/ml appeared to be optimal for maximizing both PPV and sensitivity. An increase in the HIV viral load threshold (e.g., to VL>5000) would further improve the PPV, but would decrease sensitivity and thus sample size (Figure 3).

Decreasing the viral load threshold improved sensitivity and sample size, but decreased PPV. The distribution of viral loads among patients who are ART-naïve vs ART experienced is presented in Figure 4.

DISCUSSION

In this study, we performed medical record review to determine the proportion of patients who were ART-naïve at entry to a large, observational HIV cohort. We found that more than three-quarters of patients were ART-naïve at entry into the Virtual Cohort. Using an algorithm requiring CD4 count and HIV-1 viral load > 400, we were able to improve the predictive value of ART-naïve status to nearly 90 percent. Our findings validate use of this algorithm to identify a subset of ART-naïve patients when performing analyses in large, observational cohorts where detailed ART-use histories are not available. Moreover, we found that sensitivity or positive predictive value could be optimized to suit individual investigators' needs by adjusting the viral load threshold.

Similar to other large, observational HIV cohorts, the VACS Virtual Cohort contains data regarding pharmacy, laboratory, healthcare utilization and outcomes, for a large, diverse group of HIV-infected patients from a variety of healthcare and geographic settings.^{4,14,16,18} The size and diversity of such cohorts affords the opportunity and power to investigate clinical questions which are often difficult to examine in clinical trials, for reasons of complexity, generalizability, or low prevalence. An important limitation in the use of these large pooled cohorts, however, has been difficulty limiting analyses to individuals without a prior history of antiretroviral use. With validation of this algorithm for identifying ART-naïve individuals, we now have the ability to construct subgroups within the larger cohort, who have a higher likelihood of being ART-naïve. As a result, our ability to answer clinically relevant questions, such as toxicity with antiretroviral use and the impact of comorbidities on outcomes, will be substantially improved.

Our study was made possible with comprehensive medical record review on the more than 3000 patients who were also enrolled in the Veterans Aging Cohort Study's VACS-8 study, since VACS-8 combines patient interview and medical records, with the administrative data available for VACS virtual cohort. Also, the VA electronic medical record system allowed our study staff to examine records not only at the enrollment site, but also at other VA

medical centers, where patients may have accessed care. As a result of these resources, we were able to determine date of initial antiretroviral regimen received for nearly all patients, to serve as a gold standard against which to compare our algorithm results.

An additional strength of our study includes the sensitivity analysis performed, validating not only the previously-derived algorithm (viral load threshold of 400 copies/ml), but also exploring additional viral load thresholds and their respective PPVs and sensitivities. The various viral load thresholds afford investigators the opportunity to tailor the algorithm to their specific purposes. For example, if investigators wanted the highest assurance of ART-naïve status, and needed only a relatively small sample size (e.g., 100 participants), they could achieve a greater PPV using higher viral load thresholds, such as 5000 or 10,000 copies/ml. Conversely, for studies requiring a large sample size that could tolerate a larger proportion of “contamination” by ART experienced patients, using a low viral load threshold, such as 50 or 100 copies/ml, would be preferable. For our group, many future analyses would seek to improve the certainty that nearly all of our subjects are ART-naïve by maximizing the positive predictive value, while minimizing the impact on sensitivity and sample size. For these purposes, the threshold of 400 copies/ml appears ideal. Studies utilizing large observational data should also consider performing similar sensitivity analysis to ensure their findings are robust, regardless of the initial viral load threshold selected.

Finally, our cohort had the advantage of examining patients from geographically diverse HIV treatment sites, as well as from early and later periods in the antiretroviral treatment era. Thus, we were able to examine the performance of the algorithm among settings with varying ART-naïve prevalence rates. The sensitivity of the algorithm was high across all subgroups and, as expected, the PPV correlated with prevalence.

As with all retrospective, observational studies, our study is subject to a number of limitations. First, validation of the ART-naïve status of patients relied upon information collected in medical records regarding the timing and location of ART initiation. Although the VA electronic medical record facilitates identification of the first ART use within the VA from several sources (e.g., physician notes, pharmacy records, remote data from other VA clinics), determination of any ART use outside the VA depended on medication histories documented within physician or clinic notes. Thus, it is possible that we overestimated the number of patients who were ART-naïve, since some patients may have previously received ART outside the VA, but lacked documentation in the medical records reviewed. Second, since no consensus exists for the definition of combination or “highly-active” antiretroviral therapy, we created criteria based on the prevailing practices during the study period. Although our criteria may not include all possible ART combinations used at that time (e.g., a two protease inhibitor regimen), the vast majority of regimens would be captured. Similarly, we focused our analysis on patients who had never previously received mono- or dual-ART regimens. This decision was made to create a homogenous population which is generalizable to the majority of patients in observational cohorts today. Exclusion of mono- and dual-regimen patients did not substantially alter the algorithm performance; when they were included the algorithm had a sensitivity of 84% and a PPV of 81% (compared with 86% and 87% respectively, in the main analysis) at viral load threshold of 400 copies/ml. Finally, patients were excluded from our study if they did not have a viral load measured within 365 days of their Virtual Cohort entry. Limitations in our ability to extract viral load values may have unnecessarily excluded patients who had viral load measurements performed; however, these exclusions are unlikely to have introduced any systematic biases.

In summary, application of an algorithm using available laboratory data in a large observational cohort can facilitate identification of patients who were naïve to ART at the

time of entry into the cohort. By focusing on the subset of patients who are ART-naïve, analyses of treatment effectiveness, adverse events, and the impact of comorbidities can be carried out with greater precision within large, observational HIV cohorts.

Acknowledgments

We acknowledge the Veterans who participate in the Veterans Aging Cohort Study and the study coordinators and staff at each of our sites and at the West Haven Coordinating Center. Without the commitment and care of these individuals, this research would not be possible. We would also like to acknowledge the substantial in kind support we receive from the Veterans Affairs Healthcare System.

Funding: This work was financially supported by National Institutes of Health: NIAAA (U10-AA13566), and NIA (R01-AG029154; K23 AG024896). Dr. Gandhi is the recipient of the Doris Duke Charitable Foundation Clinical Scientist Development Award (2007070). Dr. Tate was supported by the Training Program in Environmental Epidemiology funded under grant no. T32 ES07069. This work was also facilitated by the Centers for AIDS Research at Albert Einstein/Montefiore (AI-51519) and at Emory University (P30 AI050409).

Reference List

1. Palella FJ Jr, Delaney KM, Moorman AC, et al. Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. HIV Outpatient Study Investigators. *N Engl J Med*. 1998; 338:853–60. [PubMed: 9516219]
2. Hammer SM, Squires KE, Hughes MD, et al. A controlled trial of two nucleoside analogues plus didanosine in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *N Engl J Med*. 1997; 337:725–33. [PubMed: 9287227]
3. Lucas GM, Eustace JA, Sozio S, et al. Highly active antiretroviral therapy and the incidence of HIV-1-associated nephropathy: a 12-year cohort study. *Aids*. 2004; 18:541–546. [PubMed: 15090808]
4. Egger M, May M, Chene G, et al. Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: a collaborative analysis of prospective studies. *Lancet*. 2002; 360:119–29. [PubMed: 12126821]
5. Sterne JA, Hernan MA, Ledergerber B, et al. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *Lancet*. 2005; 366:378–384. [PubMed: 16054937]
6. Chen RY, Westfall AO, Mugavero MJ, et al. Duration of highly active antiretroviral therapy regimens. *Clin Infect Dis*. 2003; 37:714–722. [PubMed: 12942406]
7. Fischer MJ, Wyatt CM, Gordon K, et al. Hepatitis C and the risk of kidney disease and mortality in veterans with HIV. *J Acquir Immune Defic Syndr*. 2010; 53:222–226. [PubMed: 20104121]
8. Friis-Moller N, Reiss P, Sabin CA, et al. Class of antiretroviral drugs and the risk of myocardial infarction. *N Engl J Med*. 2007; 356:1723–1735. [PubMed: 17460226]
9. Kesselring AM, Wit FW, Sabin CA, et al. Risk factors for treatment-limiting toxicities in patients starting nevirapine-containing antiretroviral therapy. *Aids*. 2009; 23:1689–1699. [PubMed: 19487907]
10. Althoff KN, Justice AC, Gange SJ, et al. Virologic and immunologic response to HAART, by age and regimen class. *Aids*. 2010; 24:2469–2479. [PubMed: 20829678]
11. Greenbaum AH, Wilson LE, Keruly JC, et al. Effect of age and HAART regimen on clinical response in an urban cohort of HIV-infected individuals. *Aids*. 2008; 22:2331–2339. [PubMed: 18981772]
12. Grabar S, Kousignian I, Sobel A, et al. Immunologic and clinical responses to highly active antiretroviral therapy over 50 years of age. Results from the French Hospital Database on HIV. *Aids*. 2004; 18:2029–2038. [PubMed: 15577624]
13. Choi AI, Rodriguez RA, Bacchetti P, et al. Racial differences in end-stage renal disease rates in HIV infection versus diabetes. *J Am Soc Nephrol*. 2007; 18:2968–2974. [PubMed: 17942954]

14. Gange SJ, Kitahata MM, Saag MS, et al. Cohort profile: the North American AIDS Cohort Collaboration on Research and Design (NA-ACCORD). *Int J Epidemiol.* 2007; 36:294–301. [PubMed: 17213214]
15. Zhou J, Kumarasamy N, Ditangco R, et al. The TREAT Asia HIV Observational Database: baseline and retrospective data. *J Acquir Immune Defic Syndr.* 2005; 38:174–179. [PubMed: 15671802]
16. McGowan CC, Cahn P, Gotuzzo E, et al. Cohort Profile: Caribbean, Central and South America Network for HIV research (CCASAnet) collaboration within the International Epidemiologic Databases to Evaluate AIDS (IeDEA) programme. *Int J Epidemiol.* 2007; 36:969–976. [PubMed: 17846055]
17. Lang S, Mary-Krause M, Cotte L, et al. Impact of individual antiretroviral drugs on the risk of myocardial infarction in human immunodeficiency virus-infected patients: a case-control study nested within the French Hospital Database on HIV ANRS cohort CO4. *Arch Intern Med.* 2010; 170:1228–1238. [PubMed: 20660842]
18. Fultz SL, Skanderson M, Mole LA, et al. Development and verification of a “virtual” cohort using the National VA Health Information System. *Med Care.* 2006; 44:25–30.
19. Braitstein P, Brinkhof MW, Dabis F, et al. Mortality of HIV-1-infected patients in the first year of antiretroviral therapy: comparison between low-income and high-income countries. *Lancet.* 2006; 367:817–824. [PubMed: 16530575]
20. Boubaker K, Flepp M, Sudre P, et al. Hyperlactatemia and antiretroviral therapy: the Swiss HIV Cohort Study. *Clin Infect Dis.* 2001; 33:1931–1937. [PubMed: 11692306]
21. Friis-Moller N, Sabin CA, Weber R, et al. Combination antiretroviral therapy and the risk of myocardial infarction. *N Engl J Med.* 2003; 349:1993–2003. [PubMed: 14627784]
22. Stenzel MS, Carpenter CC. The management of the clinical complications of antiretroviral therapy. *Infect Dis Clin North Am.* 2000; 14:851–78. vi. [PubMed: 11144642]
23. Gandhi NR, Skanderson M, Gordon KS, et al. Delayed presentation for human immunodeficiency virus (HIV) care among veterans: a problem of access or screening. *Med Care.* 2007; 45:1105–1109. [PubMed: 18049352]
24. Justice AC, Dombrowski E, Conigliaro J, et al. Veterans Aging Cohort Study (VACS): Overview and description. *Med Care.* 2006; 44:13–24.

Key Points

- Observational studies are critical for understanding drug safety and effectiveness in populations broader and larger than those enrolled in randomized trials.
- Databases pooling information from large cohorts of patients provide a valuable opportunity to examine clinical outcomes and adverse events.
- In HIV care, it is important to identify subjects who are treatment-naïve to fully ascertain the impact of antiretroviral medications; however, determination of treatment-naïve status can be challenging in large databases.
- In this study, we validated an algorithm to identify antiretroviral treatment-naïve subjects.
- The algorithm can be adapted to optimize the sensitivity or positive-predictive value to meet individual researchers' purposes.

Figure 1.
Description of Patients eligible for study

Figure 2.

Percent of patients ART naïve at Virtual Cohort entry according to medical record review by site and year. (Dashed lines indicate the mean percent of ART naïve patients for each time period: 87% for 1997-1999 and 76% for 2000-2005)

Viral Load Threshold	50	75	100	400	500	1000	5000	10,000
Sensitivity, %	93	92	92	86	84	81	73	65
Positive Predictive Value (PPV), %	84	84	84	87	88	89	90	90

Figure 3.
Performance of prediction algorithm compared to medical record review in determining ART status at Virtual Cohort entry by HIV-1 viral load cutpoint.

Figure 4. Distribution of baseline viral load values among patients who are ART-naïve vs. those who are ART-experienced.

Table 1

Characteristics of 1223 HIV-infected veterans at time of entry into the Veterans Aging Cohort Study Virtual Cohort

Age	<50	828 (68)
	50-64	356 (29)
	65+	39 (3.2)
	Median (IQR)	45 (39-51)
Male		1190 (97)
Race	Black	877 (72)
	White	194 (16)
	Hispanic	105 (8.6)
	Other/Unknown	47 (3.8)
VACS-8 site of care	Atlanta	203 (17)
	Baltimore	145 (12)
	Bronx	144 (12)
	Brooklyn/Manhattan	106 (8.7)
	Houston	215 (18)
	Los Angeles	195 (16)
	Pittsburgh	173 (14)
	Washington DC	42 (3.4)
Year of Virtual Cohort entry	1997-1999	548 (45)
	2000-2002	491 (40)
	2003-2005	184 (15)
CD4 cells/mm3	Median (IQR)	284 (96-470)
HIV-1 RNA copies/ml (not mutually exclusive)	Median (IQR)	21,752 (1386-109,814)
	>50	1168 (95.5)
	>75	1150 (94.0)
	>100	1145 (93.6)
	>400	1018 (83.2)
	>500	985 (80.5)
	>1000	939 (76.8)
	>5000	823 (67.3)
	>10000	723 (59.1)

Table 2

Performance of prediction algorithm compared to medical record review in determining ART status at Virtual Cohort entry (at a viral load threshold of 400 copies/ml).

ART Naïve at Virtual Cohort entry	Medical Record Review			
Prediction Algorithm		Yes	No	Total
	Yes	856	124	980
	No	134	109	243
	Total	990	233	1223

Sensitivity = $856/990 = 86\%$

Specificity = $109/233 = 47\%$

PPV = $856/980 = 87\%$

NPV = $109/243 = 45\%$