OPEN

Comparative Effectiveness of First Antiretroviral Regimens in Clinical Practice Using a Causal Approach

*Lise Cuzin, MD, Pascal Pugliese, MD, Clotilde Allavena, MD, Christine Katlama, MD, PhD,
Laurent Cotte, MD, Antoine Cheret, MD, PhD, André Cabié, MD, PhD, David Rey, MD,
Catherine Chirouze, MD, PhD, Firouze Bani-Sadr, MD, PhD,
and Philippe Flandre, PhD, for the Dat'AIDS Study group*

Abstract: The objective of this study was to estimate the cumulative incidences of failure by months 12 (M12) and 24 (M24) for the most prescribed first-line anti-retroviral regimens (ART).

It is retrospective analysis of a prospectively collected database.

All patients who initiated their first ART with the most prescribed regimens between 1st January 2004 and 30th June 2013 in 12 large HIV reference centers in France were included. The outcome was treatment failure—defined by any treatment modification for virological or tolerability reasons—and comparisons between regimens were carried out at M12 and M24. Adjusted and weighted methods via the propensity score (PS) were used to compare the effectiveness of the first antiretroviral regimens. Potential confounders of the treatment-outcome association were used to estimate PS with multinomial logistic regression.

Overall, 3128 and 2690 patients were included in the M12 and M24 analyses, respectively. Patients received 5 different regimens (ABC/3TC with ATV/r or DRV/r, TDF/FTC with ATV/r, DRV/r, or EFV). Failure was reported in 25% and 42% at M12 and M24, respectively. Patients who received TDF/FTC/EFV had a significantly higher proportion of failure at M12 by comparison with TDF/FTC with DRV/r (reference), but not at M24. Patients in the 3 other groups had a trend toward a higher proportion of failure at M12 although not statistically significant. No difference was found at M24.

Using data from a large prospective cohort, we found that boosted atazanavir and darunavir had comparable effectiveness, whatever the associated NRTIs, whereas efavirenz-based regimens were relatively less performing on the short term.

(*Medicine* 94(39):e1668)

Abbreviations: ABC/3TC = abacavir with lamivudine, ART = antiviral therapy, ATV/r = ritonavir boosted atazanavir, CNIL = French national commission on informatics and liberty, DR = Double Robust estimator, DRV/r = ritonavir boosted darunavir, EFV = efavirenz, eGFR = estimated glomerular filtration rate, HIV = human immunodeficiency virus, INSTI = integrase strand-transfer inhibitor, IPCW = Inverse Probability of Censoring Weighted, IPTW = Inverse Probability of Treatment Weighted, M12 = month 12 after treatment initiation, M24 = month 24 after treatment initiation, MSM = men who have sex with men, NNRTI = non-nucleosidic reverse transcriptase inhibitor, NRTI = nucleosidic reverse transcriptase inhibitor, OR = odd ratio, PS = propensity score, RCT = randomized clinical trial, TDF/FTC = tenofovir with emtricitabine.

INTRODUCTION

Currently, recommended first-line antiretroviral regimens (ART) include 2 nucleoside analog reverse transcriptase inhibitors (NRTIs) in association with either a non-nucleoside reverse transcriptase inhibitor (NNRTI), a ritonavir boosted protease inhibitor (PI/r), or an integrase strand-transfer inhibitor (INSTI).[1] Due to the number of different drugs in each class, multiple potential triple combinations can be prescribed. Randomized controlled trials (RCTs), considered as the methodological gold standard, have been designed to compare virological efficacy of some of the potential triple drug regimens. These studies provide the core knowledge for recommendations and guidelines,[1–3] but they have limits. Selection and volunteer biases may compromise the generalization of their results to the general patient population.[4,5]

Observational cohort studies offer a complementary research design by providing information on comparative effectiveness of different antiretroviral regimens used in clinical practice. These studies allow comparisons between strategies not evaluated by RCTs, and they are representative of treatment strategies used in routine care settings. Reasons for modifying or interrupting ART may be different in this setting by comparison with RCTs, due to a broader use of ART. It is well known that in observational studies, unlike in RCTs, characteristics of patients between the treatment groups are quite different. The potential for confounding by indication may strongly impact the outcome interpretation and reliability of study findings.[4–6]

Statistical methods have therefore been introduced to provide a causal approach of the analysis of observational data. Most of these methods are based on the propensity score (PS), which is the probability of receiving a treatment given some observed covariates.[7] The goal of propensity scores is to balance observed covariates between subjects from the treatment groups in order to mimic what happens in a randomized study.[8] In practice propensity scores are unknown and are estimated via regression models. Adjusting and weighting via the PS for estimation of causal treatment effects have been mostly used for comparisons of 2 groups. Recently, however, some studies have used the multiple PS for the comparison of >2 treatment arms.[9–11]

Using a large prospectively collected observational cohort, we analyzed the effectiveness of the most prescribed first-line regimens. The outcome—treatment failure—was defined as any modification of the regimen because of lack of efficacy or poor tolerability. Crude analyses were compared with analyses adjusted for multiple PS or weighted via marginal structural models and doubly robust estimators.[12–14]

METHODS

Subjects and Data Collection

The patients were HIV-1-infected adults receiving care in 12 large HIV reference centers from 11 distinct geographical regions in France. These centers maintain prospective cohorts of all HIV-infected patients who received care and provided written consent. The data collection has been approved by the French national commission on informatics and liberty (CNIL). The database collects demographic, clinical, antiretroviral history, viral load, and CD4+ T cell counts data at regular 3 to 6 months intervals during routine clinical assessment.[15] For the purpose of this study, we selected all patients who initiated their first ART between 1st January 2004 and 30th June 2013. We then restricted the population to patients receiving a

regimen, which has been used by at least 200 patients and which was still recommended as the first regimen in the most recent years (thus regimens including lopinavir were excluded).

Key confounders of the treatment-outcome association included continuous and categorical variables that were assessed at or before the initiation of ART. Continuous variables were age, baseline viral load ($\log_{10}$ copies/ml), baseline CD4+ T cell counts (cells/mm^3), and duration of known HIV infection (time in months from HIV diagnosis to ART initiation). Categorical variables were hepatitis B or C co-infection (yes/no), AIDS at ART initiation (yes/no), prior history of depression (yes/no), having a glomerular filtration rate estimated by the abbreviated Modification of Diet and Renal Disease (eGFR) <90 ml/min per 1.73 m^2 (yes/no), a combination of sex and route of transmission (men who have sex with men [MSM], other men and women) and year of ART initiation (2004–2008, 2009–2010, or 2011–2013).

The outcome was treatment failure, defined by any modification of the first regimen due to the lack of efficacy or to poor tolerability. This outcome was investigated by month 12 (M12) and month 24 (M24). All other reasons for regimen modification, such as treatment simplification, pregnancy (planned or current), clinical trial participation, or planned antiretroviral interruptions were not considered as treatment failures and were considered as censored. For patients treated with tenofovir/emtricitabine with efavirenz, simplification to the single-pill formulation was not considered as a treatment modification. When the reason for ART modification was not clear, we searched for the last viral load value before modification. If this last viral load was >2.6 $\log_{10}$ copies/ml, failure was considered as the reason for the ART modification. Otherwise, the observation was censored.

Two dataset were used for each of the M12 and M24 analyses. Patients who did not experience the outcome and who had a follow-up <12 months were excluded from the M12 analysis. Similarly patients who did not experience the outcome and who had a follow-up shorter than 24 months were excluded from the M24 analysis.

STATISTICAL METHODS

Unadjusted analysis and adjusted/weighted analyses based on the propensity score (PS) were used to compare the effectiveness of the first ART regimens at M12 and M24. Adjusted method used multiple PSs estimated by multivariable logistic regression. Each PS is the conditional probability of assigning a given subject to one of the treatment groups given the observed covariates. Selection for the PS model was an iterative process similar to that suggested previously.[10,11] Marginal structural models were also used in a weighting strategy. We used stabilized weights based on the product of Inverse Probability of Treatment Weighted (IPTW) and Inverse Probability of Censoring Weighted (IPCW).[14] The double robust (DR) estimator combines 2 approaches to estimate the causal effect of a treatment on an outcome.[13] The first is based on an outcome-model and the second is a model for the propensity score. DR produces a consistent estimate of the treatment effect if either of the 2 models has been correctly specified. Generalized doubly robust estimator for multiple treatments has been proposed recently.[16] We use the bootstrap method to estimate confidence intervals and *P* values of doubly robust estimators as recommended.[17]

Two sensitivity analyses were performed. A first sensitivity analysis excluded censored patients (patients who

modified their initial ART regimen due to another reason than lack of efficacy or tolerability). A second analysis included only patients having an estimated PS or stabilized weight between the 1st and 99th percentile. All analyses were done with SAS (version 9.3; SAS Institute Inc, Cary, NC).

RESULTS

On the basis of our inclusion criteria, we selected 3628 patients who initiated ART with 5 different regimens. The 5 regimens were the following: atazanavir /ritonavir (ATV/r) with abacavir/lamivudine (ABC/3TC) (N = 250); ATV/r with tenofovir/emtricitabine (TDF/FTC) (N = 958); TDF/FTC with efavirenz (EFV) (N = 721); darunavir/ritonavir (DRV/r) with ABC/3TC (N = 340); and DRV/r with TDF/FTC (N = 1259). Among the 3628 patients, 500 patients were excluded from the M12 analysis because of a follow-up of <12 months. Thus, the M12 analysis was based on 3128 patients. Similarly, the M24 analysis was based on 2690 patients.

Baseline characteristics of the 3128 patients included in the M12 analysis are summarized in Table 1, showing to what extent the treatment groups initially differed. The use of TDF/FTC with DRV/r was more likely among subjects with high baseline viral load, low baseline $CD4^+$ T cell count, in those with AIDS at ART initiation or among those who had been recently diagnosed. Use of ATV/r was more likely in patients who initiated ART in 2004 to 2008. Similar results were found for the patients included in the M24 analysis (data not shown). The unadjusted P value corresponds to the comparison before correction on multiple PSs. Confounders variables included in the propensity score model were markedly different between the 5 treatment groups with all P values < 0.001. The adjusted P values, corresponding to a balance-check after correction, indicate a good balance between the treatment groups since only the $CD4^+$ T cell count is slightly unbalanced after correction.

Treatment Failure at M12

Among the 3128 subjects, 339 (10.8%) patients were censored. Among the 2789 remaining patients, 706 (25.3%) experienced a treatment failure before M12. Fifty six percent of failures occurred during the first 3 months of antiretroviral therapy, and only 53 of all M12 failures (7.5%) were due to the lack of efficacy. Table 2 reports adjusted odds ratios (ORs) based on multivariable logistic regression of treatment failure on potential confounders. Several covariates were associated with increased treatment failure: women versus MSM (OR = 1.46; P < 0.01), having an eGFR <90 ml/min per 1.73 m^2 (OR = 1.26; P = 0.05); presence of AIDS at ART initiation (OR = 1.73; P < 0.01); and higher baseline viral load (OR = 1.16 per log_{10}; P < 0.01). Probability of treatment failure was lower in patients treated early in the 2000 s (OR = 0.57 in 2004–2008 vs 2011–2013; P < 0.01; OR = 0.68 in 2009–2010 vs 2011–2013; P < 0.01).

Patients who received TDF/FTC with DRV/r had the lowest probability of failure whatever the method used; thus we used them as the reference group. Table 3 shows estimates of the probability of treatment failure in the reference group and risk differences for the other treatment groups. A positive difference in probability indicates a higher estimate of treatment failure by comparison with the reference group. The unadjusted analysis showed no significant difference in proportions of failure between the 4 treatment groups by comparison with TDF/FTC with DRV/r. Among the 2789 uncensored patients with a follow-up >12 months, we estimated a 4.5% to 8.8% higher absolute probability of treatment failure at M12 in the 4 treatment groups by comparison with the TDF/FTC with the

TABLE 1. Patients Characteristics at The Time of ART Initiation in Each Treatment Group

Continuous variables, median (IQR)	ABC/3TC/ATV/r	TDF/FTC/ATV/r	TDF/FTC/EFV	ABC/3TC/DRV/r	TDF/FTC/DRV/r	*P*-Value Unadjusted	Adjusted
	N = 319	N = 895	N = 647	N = 239	N = 1028		
Age (years)	40.2 (31–48)	39.1 (32–47)	36.9 (30–44)	40.6 (32–49)	39.7 (32–47)	0.002	0.06
Baseline viral load (log10copies/ml)	4.5 (4.0–5.0)	4.7 (4.2–5.2)	4.7 (4.1–5.1)	4.6 (4.1–5.2)	5.0 (4.5–5.4)	<0.001	0.19
Baseline $CD4^+$ T cell count (cells/ml)	319 (232–393)	306 (215–406)	368 (280–469)	309 (204–414)	289 (121–421)	<0.001	0.02
Time since HIV diagnosis (months)	52 (10–135)	47 (6–157)	38 (7–122)	28 (5–108)	8 (2–96)	<0.001	0.34
Categorical variables, N (%)							
AIDS at initiation of ART, yes	19 (6)	63 (7)	33 (5.1)	27 (11.3)	175 (17)	<0.001	0.24
Hepatitis B or C coinfection, yes	22 (6.9)	137 (15.3)	75 (11.6)	19 (7.9)	129 (12.5)	0.007	0.21
eGFR < 90 ml/min per 1.73 m^2	84 (26.3)	164 (18.3)	94 (14.5)	67 (28.0)	210 (20.4)	<0.001	0.39
Depression, yes	31 (9.7)	57 (6.4)	15 (2.3)	12 (5.0)	50 (4.9)	<0.001	0.05
Gender and route of transmission, N (%)							
MSM	125 (39.2)	415 (46.5)	349 (53.9)	101 (42.3)	511 (49.7)	<0.001	0.09
Other men	101 (31.7)	266 (29.7)	176 (27.2)	79 (33.1)	314 (30.5)		
Female	93 (29.2)	214 (23.9)	122 (18.9)	59 (24.7)	203 (19.7)		
Year at initiation of ART							
2004–2008	120 (37.6)	318 (35.5)	27 (4.2)	10 (4.2)	25 (2.4)	<0.001	0.21
2009–2010	99 (31.0)	353 (39.4)	300 (46.6)	77 (32.2)	383 (37.3)		
2011–2013	100 (31.3)	224 (25.0)	320 (49.5)	152 (63.6)	620 (60.3)		

3TC = lamivudine, ABC = abacavir, AIDS = acquired immune deficiency syndrome, ART = antiviral therapy, ATV/r = ritonavir boosted atazanavir, DRV/r = ritonavir boosted darunavir, EFV = efavirenz, eGFR = glomerular filtration rate estimated by the abbreviated modification of diet and renal disease, FTC = emtricitabine, HIV = Human immunodeficiency virus, IQR = 25% interquartile, MSM = men who have sex with men, TDF = tenofovir.

TABLE 2. Adjusted Odd-Ratios Based on Multivariable Logistic Regression of Treatment Failure by M12 on Potential Confounders

	Treatment Failure at Month 12		
Potential Confounder	**Adjusted OR**	**95% CI**	***P*-Value**
Age (per 10 years)	1.05	0.96 to 1.15	0.25
Baseline viral load (per $1\log_{10}$copies/ml)	1.16	1.04 to 1.30	<0.01
Baseline CD4$^+$ T cell count (per 100 cells/ml)	1.03	0.98 to 1.09	0.22
Time since HIV diagnosis (per year)	0.99	0.97 to 1.01	0.40
AIDS at initiation of ART	1.73	1.26 to 2.37	<0.01
Hepatitis B or C co-infection	1.06	0.78 to 1.44	0.73
eGFR < 90 ml/min per 1.73 m^2	1.26	0.99 to 1.58	0.05
Depression	0.92	0.60 to 1.42	0.71
Other male versus MSM	0.86	0.68 to 1.09	0.22
Female versus MSM	1.46	1.15 to 1.86	<0.01
Treated in 2004–2008 versus 2011–2013	0.57	0.42 to 0.78	<0.01
Treated in 2009–2010 versus 2011–2013	0.68	0.56 to 0.84	<0.01

AIDS = acquired immune deficiency syndrome, CI = confidence interval, eGFR = glomerular filtration rate estimated by the abbreviated modification of diet and renal disease, HIV = Human immunodeficiency virus, MSM = men who have sex with men, OR = odds ratio.

DRV/r group. These estimates were higher than the unadjusted associations. In particular, patients receiving TDF/FTC with EFV had a significantly higher probability of failure than patients receiving TDF/FTC with DRV/r. Patients receiving ABC/3TC with ATV/r, TDF/FTC with ATV/r, or ABC/3TC with DRV/r had a trend toward a higher probability of failure with *P* values ∼ 0.08–0.10.

Treatment Failure at M24

Among the 2690 subjects of the M24 analysis, 536 (19.9%) patients were censored. Among the 2154 remaining patients 893 (41.5%) presented with treatment failure before M24. Most of the predictors of M12 treatment failure (Table 2) were also predictive of failure at M24 (Table 4), except for an eGFR <90 ml/min per 1.73 m^2. In addition, higher baseline CD4$^+$ T cell count (OR = 1.07 per 100 cells/ml; *P* = 0.03) was significantly associated with an increased probability of treatment failure at M24.

Probabilities of treatment failure varied from 41% to 46% in the TDF/FTC with DRV/r group (Table 5). In the unadjusted

analysis, patients receiving an ATV/r-containing regimen had a lower absolute probability of failure compared with the reference group (−8% for ABC/3TC with ATV/r; *P* = 0.008; −3.8% for TDF/FTC with ATV/r; *P* = 0.38). The use of appropriate statistical methods showed a quite different picture with a higher probability, although not statistically significant, of failure in patients receiving an ATV/r-containing regimen. Marginal structural models and double robust estimators showed no statistical difference in probability of treatment failure of the 4 treatment groups by comparison with the reference group.

All sensitivity analyses described in the method section provided similar results for both M12 and M24 analyses (data not shown).

DISCUSSION

We here provide a comparison of the effectiveness of the 5 most prescribed first-line ART regimens for HIV-infected patients in France between 2004 and 2013. The 2 most frequently used regimens were TDF/FTC with either EFV or DRV/

TABLE 3. Estimates of the Probability of Treatment Failure by M12 in the Reference Group and Risk Differences in the Other Groups

Treatment Group	Unadjusted Analysis		Adjusted on the Multiple PS		Marginal Structural Models		Doubly Robust	
Reference TDF/FTC/DRV/r	Treatment failure 24.6%		Treatment failure 21.4%		Treatment failure 20.8%		Treatment failure 21.5%	
	Difference in proportions of failure, % (95% CI)	*P*	Difference in failure probability, % (95% CI)	*P*	Difference in failure probability, % (95% CI)	*P*	Difference in failure probability, % (95% CI)	*P*
ABC/3TC/ATV/r	0.7 (−5.8 to 7.2)	0.83	6.3 (−0.6 to 13.2)	0.08	7.0 (−2.1 to 16.0)	0.13	7.4 (−0.02 to 16.2)	0.10
TDF/FTC/ATV/r	−0.8 (−5.2 to 3.6)	0.71	5.9 (0.9 to 10.9)	0.02	6.0 (−0.2 to 12.2)	0.06	4.5 (−0.01 to 10)	0.10
TDF/FTC/EFV	4.2 (−0.8 to 9.3)	0.10	7.5 (2.4 to 12.6)	0.004	8.8 (1.4 to 16.3)	0.02	8.0 (0.1 to 15.4)	0.04
ABC/3TC/DRV/r	4.9 (−2.2 to 12.0)	0.18	6.2 (−0.8 to 13.3)	0.08	8.1 (−0.8 to 16.9)	0.07	7.7 (−0.01 to 15.8)	0.07

/r = ritonavir boost, 3TC = lamivudine, ABC = abacavir, ATV = atazanavir, CI = confidence interval, DRV = darunavir, EFV = efavirenz, FTC = entricitabine, PS = propensity score, TDF = tenofovir.

TABLE 4. Adjusted Odd-Ratios Based on Multivariable Logistic Regression of Treatment Failure by M24 on Potential Confounders

	Treatment Failure at Month 24		
Potential confounder	Adjusted OR	95% CI	*P*-value
Age (per 10 years)	1.08	0.98 to 1.19	0.11
Baseline viral load (per $1\log_{10}$ copies/ml)	1.24	1.11 to 1.39	<0.01
Baseline CD4$^+$ T cell count (per 100 cells/ml)	1.07	1.01 to 1.14	0.03
Time since HIV diagnosis (per year)	0,99	0.97 to 1.01	0.46
AIDS at initiation of ART	1.47	1.05 to 2.08	0.03
Hepatitis B or C coinfection	0.98	0.71 to 1.36	0.92
eGFR < 90 ml/min per 1.73 m^2	1.07	0.84 to 1.37	0.59
Depression	0.77	0.50 to 1.20	0.25
Other male versus MSM	0.86	0.67 to 1.10	0.22
Female versus MSM	1.40	1.07 to 1.83	0.01
Treated in 2004–2008 versus 2011–2013	0.25	0.19 to 0.35	<0.01
Treated in 2009–2010 versus 2011–2013	0.28	0.22 to 0.34	<0.01

AIDS = acquired immune deficiency syndrome, eGFR = glomerular filtration rate estimated by the abbreviated modification of diet and renal disease, HIV = human immunodeficiency virus, MSM = men who have sex with men, OR = odds ratio.

r as recommended in international HIV treatment guidelines.[1–3] ABC/3TC with ATV/r was received by >300 patients although it was not a recommended regimen. By July 2013, the number of patients receiving INSTI-based regimens (not recommended as first-line in France before 2014) was not large enough to be included in the present analysis. Crude treatment failure rates at months 12 and 24 were 25 and 42%, respectively. Patients receiving TDF/FTC with DRV/r had the lowest risk of treatment failure at months 12 and 24. After controlling for confounding by propensity score modeling, marginal structural models and double robust we found that, at month 12, patients receiving TDF/FTC with EFV had a significantly higher probability of treatment failure. Patients receiving boosted atazanavir regimens and ABC/3TC with DRV/r had a nonsignificantly higher probability of treatment failure. At month 24, all regimens had a comparable effectiveness.

The observed crude failure rate, ~25% at 1 year, is in line with the finding that 47% of the patients of our cohort changed their first-line regimen before the end of the first year, whatever the reasons for treatment modification.[18] Thus, it is a reasonable estimation of treatment effectiveness. By comparison, the 1 and 2-year cumulative incidences of treatment modification in patients initiating ART in 2002 to 2009 were 25 and 39%, respectively, in a large collaboration of cohort studies.[19] As observed in RCTs, treatment modifications were mainly due to adverse events rather than to virological or clinical failures. Fifty percent of failing patients modified their regimen during the first 3 months. Early interruption of ART regimens due to short-term poor tolerability has already been described.[19,20]

Observational studies provide an insight in how different treatments will be used in ''real life'' by the patients and their physicians. We studied a large population seeking care in different centers, representative of the patients under care in France during the last 10 years. The large population of our cohort fulfills the necessary condition to allow the selection of regimens for which no data have been produced by RCTs.[21] Permanent assessment and control of the quality of the database[15] limits the errors that one can encounter with this type of studies. Appropriate analysis of carefully collected prospective observational data can then complement the findings of RCTs[22]

TABLE 5. Estimates of the Probability of Treatment Failure by M24 in the Reference Group and Risk Differences in the Other Groups

	Unadjusted Analysis		Adjusted on the Multiple PS		Marginal Structural Models		Doubly Robust	
Reference TDF/FTC/DRV/r	Treatment failure 45.6%		Treatment failure 41.1%		Treatment failure 41.9%		Treatment failure 44.1%	
Treatment group	Difference in failure probability, % (95% CI)	*P*	Difference in failure probability, % (95% CI)	*P*	Difference in failure probability, % (95% CI)	*P*	Difference in failure probability, % (95% CI)	*P*
ABC/3TC/ATV/r	−3.8 (−12.2 to 4.6)	0.38	8.5 (−0.4 to 17.4)	0.06	8.3 (−4.8 to 21.5)	0.22	4.3 (−8.2 to 16.7)	0.50
TDF/FTC/ATV/r	−8.0 (−13.8 to −2.1)	0.008	6.3 (−0.6 to 13.2)	0.07	3.3 (−7.2 to 13.8)	0.54	−1.2 (−12.8 to 10.4)	0.85
TDF/FTC/EFV	0.6 (−6.0 to 7.1)	0.86	5.1 (−1.7 to 11.9)	0.14	6.7 (−5.0 to 18.5)	0.26	−0.8 (−13.3 to 11.6)	0.89
ABC/3TC/DRV/r	5.5 (−3.9 to 14.9)	0.25	6.5 (−2.9 to 15.9)	0.18	3.0 (−10.8 to 16.9)	0.67	3.5 (−8.9 to 15.9)	0.58

/r = ritonavir boost, 3TC = lamivudine, ABC = abacavir, ATV = atazanavir, CI = confidence interval, DRV = darunavir, EFV = efavirenz, FTC = entricitabine, PS = propensity score, TDF = tenofovir.

to avoid that first-line regimens be chosen by the physicians on the basis of self-conviction, indirect comparisons, or other factors.

By choosing a pragmatic definition of treatment failure, we took into account some reasons for failure that may not be accounted for in RCTs. Most trials used virological failure as primary endpoints, but treatment modifications should also be considered as failures. The large number of planned visits in an RCT, especially during the first year, and strict protocol rules are 2 important differences with observational studies. One can suspect that clinicians modified patients' regimen more easily and early in clinical practice than in an RCT. The surprisingly high probability of failure rate in the most recent calendar years can be explained by the availability of an increasing number of potent drugs leading to treatment modifications for minor toxicities. Comparative effectiveness of initial antiretroviral therapy regimens based on virological failure between RCTs and observational studies have been made and showed a good agreement.[23]

The methods we used were designed to limit the indication bias, taking into account the major patients characteristics that drive a physician decision while selecting ART. A key feature of the methods used here is the variable selection for propensity score models or for inverse probability weights.[24,25] There is trade-off between reducing confounding bias and increasing bias and variance due to a large selection of variables in the propensity model.[25] We followed the current recommendations to construct both propensity scores and stabilized weights.[10,25] Similar statistical methods have been used to compare the virological efficacy of boosted double versus boosted single protease inhibitor regimens.[26]

A limitation of our analyses is that some unmeasured confounding may persist, as some socio-economic characteristics were not recorded in the database. We think, however, that major confounders, including baseline viral load and the year of ART initiation, were included in the models used.

In conclusion, using data from a large prospective cohort of patients seeking care in France in the recent years, we found that TDF/FTC with DRV/r had the lowest probability of treatment failure at months 12 and 24. At month 12, patients receiving TDF/FTC with EFV had a significant higher probability of treatment failure whereas no difference was found for the patients receiving boosted atazanavir regimens and ABC/3TC with DRV/r. At month 24, all regimens had a comparable effectiveness. This approach allowed us to address some questions that have not been and will probably not be considered in RCTs, adding important information for physicians and patients that will be making decisions on the choice of the first ART regimen.

ACKNOWLEDGMENTS

REFERENCES

1. Gunthard HF, Aberg JA, Eron JJ, et al. Antiretroviral treatment of adult HIV infection: 2014 recommendations of the International Antiviral Society—USA Panel. *JAMA.* 2014;312:410–425.

2. Department of Health and Human Services. Panel on Antiretroviral Guidelines for Adults and Adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. http://aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf. Accessed January 4, 2015.

3. European AIDS Clinical Society. Guidelines, version 7.1, November 2014. http://www.eacsociety.org/guidelines/eacs-guidelines/eacs-guidelines.html. Accessed January 5, 2015.

4. D'Agostino RB Jr, D'Agostino RB Sr. Estimating treatment effects using observational data. *JAMA.* 2007;297:314–316.

5. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med.* 2000;342:1907–1909.

6. Wood E, Hogg RS, Heath KV, et al. Provider bias in the selection of non-nucleoside reverse transcriptase inhibitor and protease inhibitor-based highly active antiretroviral therapy and HIV treatment outcomes in observational studies. *AIDS.* 2003;17:2629–2634.

7. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika.* 1983;70:41–55.

8. Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. *Am J Epidemiol.* 1999;150:327–333.

9. Frisco ML, Muller C, Frank K. Parents' union dissolution and adolescents' school performance: comparing methodological approaches. *J Marriage Fam.* 2007;69:721–741.

10. Spreeuwenberg MD, Bartak A, Croon MA, et al. The multiple propensity score as control for bias in the comparison of more than two treatment arms: an introduction from a case study in mental health. *Med Care.* 2010;48:166–174.

11. Zanutto E, Lu B, Hornik R. Using propensity score subclassification for multiple treatment doses to evaluate a National Antifrug Media campaign. *J Educ Behav Stat.* 2005;30:59–73.

12. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology.* 2000;11:561–570.

13. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med.* 2004;23:2937–2960.

14. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology.* 2000;11:550–560.

15. Pugliese P, Cuzin L, Cabie A, et al. A large French prospective cohort of HIV-infected patients: the Nadis Cohort. *HIV Med.* 2009;10:504–511.

16. Tu C, Koh WY, Jiao S. Using generalized doubly robust estimator to estimate average treatment effects of multiple treatments inobservational studies. *J Stat Comput Simulation.* 2013;83:1518–1526.

17. Funk MJ, Westreich D, Wiesen C, et al. Doubly robust estimation of causal effects. *Am J Epidemiol.* 2011;173:761–767.

18. Keita M, Perbost I, Pugliese-Wehrlen S, et al. Incidences and risk factors of first-line HAART discontinuation: a limitation to the success of the ''seek, test, treat, and retain'' strategy? *AIDS Care.* 2014;26:1058–1069.

19. Abgrall S, Ingle SM, May MT, et al. Durability of first ART regimen and risk factors for modification, interruption or death in HIV-positive patients starting ART in Europe and North America 2002–2009. *AIDS.* 2013;27:803–813.

20. Willig JH, Abroms S, Westfall AO, et al. Increased regimen durability in the era of once-daily fixed-dose combination antiretroviral therapy. *AIDS.* 2008;22:1951–1960.

21. Phillips AN, Grabar S, Tassie JM, et al. Use of observational databases to evaluate the effectiveness of antiretroviral therapy for HIV infection: comparison of cohort studies with randomized trials. EuroSIDA, the French Hospital Database on HIV and the Swiss HIV Cohort Study Groups. *AIDS.* 1999;13:2075–2082.

22. Hlatky MA, Califf RM, Harrell FE Jr et al. Comparison of predictions based on observational data with the results of randomized controlled clinical trials of coronary artery bypass surgery. *J Am Coll Cardiol.* 1988;11:237–245.

23. Mugavero MJ, May M, Ribaudo HJ, et al. Comparative effectiveness of initial antiretroviral therapy regimens: ACTG 5095 and 5142 clinical trials relative to ART-CC cohort study. *J Acquir Immune Defic Syndr.* 2011;58:253–260.

24. Brookhart MA, Schneeweiss S, Rothman KJ, et al. Variable selection for propensity score models. *Am J Epidemiol.* 2006;163:1149–1156.

25. Cole SR, Hernan MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol.* 2008;168:656–664.

26. Petersen ML, Wang Y, van der Laan MJ, et al. Virologic efficacy of boosted double versus boosted single protease inhibitor therapy. *AIDS.* 2007;21:1547–1554.