Effect Modification by Sex and Baseline CD4+ Cell Count Among Adults Receiving Combination Antiretroviral Therapy in Botswana: Results from a Clinical Trial

C. William Wester,[1–3] Ori M. Stitelman,[4] Victor deGruttola,[5] Hermann Bussmann,[1,6]
Richard G. Marlink,[1,6] and Mark J. van der Laan[4]

Abstract

The Tshepo study was the first clinical trial to evaluate outcomes of adults receiving nevirapine (NVP)-based versus efavirenz (EFV)-based combination antiretroviral therapy (cART) in Botswana. This was a 3 year study ($n = 650$) comparing the efficacy and tolerability of various first-line cART regimens, stratified by baseline CD4+: <200 (low) vs. 201-350 (high). Using targeted maximum likelihood estimation (TMLE), we retrospectively evaluated the causal effect of assigned NNRTI on time to virologic failure or death [intent-to-treat (ITT)] and time to minimum of virologic failure, death, or treatment modifying toxicity [time to loss of virological response (TLOVR)] by sex and baseline CD4+. Sex did significantly modify the effect of EFV versus NVP for both the ITT and TLOVR outcomes with risk differences in the probability of survival of males versus the females of approximately 6% ($p = 0.015$) and 12% ($p = 0.001$), respectively. Baseline CD4+ also modified the effect of EFV versus NVP for the TLOVR outcome, with a mean difference in survival probability of approximately 12% ($p = 0.023$) in the high versus low CD4+ cell count group. TMLE appears to be an efficient technique that allows for the clinically meaningful delineation and interpretation of the causal effect of NNRTI treatment and effect modification by sex and baseline CD4+ cell count strata in this study. EFV-treated women and NVP-treated men had more favorable cART outcomes. In addition, adults initiating EFV-based cART at higher baseline CD4+ cell count values had more favorable outcomes compared to those initiating NVP-based cART.

Introduction

A SIGNIFICANT PROPORTION of the 6.6 million persons receiving combination antiretroviral therapy (cART) in low and middle income countries of the world reside in sub-Saharan Africa.[1] Large numbers of national initiatives offering public nonnucleoside reverse transcriptase inhibitor (NNRTI)-based cART have commenced in the region with preliminary data documenting impressive efficacy outcomes among the vast majority of cART-treated adults.[2–6] In resource-rich settings, based on available data from numerous clinical trials,[7–13] efavirenz (EFV) is the NNRTI of choice, and is "*preferred*" for first-line cART, along with the NRTIs teno-

fovir (TDF) and emtricitabine (FTC).[7] This recommendation is based on efficacy and more favorable tolerability data.[7–13] In resource-limited limited settings, however, the majority of cART-treated adults are female[2–6,14–18] and have been prescribed nevirapine (NVP)-based cART regimens due to the potential teratogenic effects of EFV. Family planning considerations in sub-Saharan Africa also strongly influence the choice of NNRTI, especially as pregnancy rates among cART-treated women are high and EFV is limited to women committing to using at least two reliable contraceptive methods.

The 2NN trial,[19] a large adult randomized trial, compared 1216 adults receiving stavudine (d4T) plus lamuvidine (3TC) with either NVP or EFV in North and South America,

---

[1]Harvard School of Public Health, Boston, Massachusetts.
[2]Department of Medicine, Division of Infectious Diseases, Vanderbilt University School of Medicine, Nashville, Tennessee.
[3]Vanderbilt Institute for Global Health (VIGH), Nashville, Tennessee.
[4]Division of Biostatistics, University of California at Berkeley, Berkeley, California.
[5]Harvard School of Public Health, Center for Biostatistics in AIDS Research (CBAR), Boston, Massachusetts.
[6]Botswana-Harvard School of Public Health AIDS Initiative Partnership for HIV Research and Education (BHP), Gaborone, Botswana.

Australia, Europe, South Africa, and Thailand. They found non-inferiority (NVP vs. EFV) in their primary outcome of virologic failure. Additional 2NN analyses,[19] however, showed an association between NVP and higher rates of serious toxicity. The CPCRA 058 and INSIGHT study[20] team, reporting randomized clinical trial data from NNRTI-treated adults in the United States and Western Europe, however, did show inferiority in primary endpoint, namely higher rates of virologic failure with and without resistance among NVP-treated vs. EFV-treated patients. Wester *et al.*[14] performed an analysis of the Adult Antiretroviral Treatment and Drug Resistance ("*Tshepo*") study in Botswana using Cox proportional hazards analysis and concluded that there was no significant difference by assigned NNRTI in time to death or virologic failure. Women receiving NVP-based cART, however, trended toward higher virological failure rates when compared to EFV-treated women, Holm-corrected log-rank *p*-value = 0.072.[14] There were no differences among men.[14] Furthermore, they concluded that individuals treated with NVP had significantly shorter times to treatment modifying toxicity when compared to those receiving EFV-based cART.[14]

Current methods used to evaluate effect modification (often referred to as statistical interaction) in time to event data, such as the Cox proportional hazards model, posit highly restrictive parametric models and attempt to estimate parameters that are specific to the model proposed.[21] These methods tend to be biased and force providers to estimate parameters out of convenience rather that what they are actually interested in. The targeted maximum likelihood estimation (TMLE) methodology, originally proposed by van der Laan and Rubin,[21,22] and applied to time to event outcomes by Moore and van der Laan[23] improves on the currently implemented methods in both robustness (its ability to provide unbiased estimates) and flexibility (allowing providers to estimate parameters of direct interest to them).

Evaluating data from the recently completed *Tshepo* study, a clinical trial comparing the efficacy and tolerability of various first-line cART regimens in Botswana, we compared TMLE results to results obtained from conventional Cox proportional hazards analyses. In doing so, we aimed to definitively evaluate the causal effect of NNRTI treatment and effect modification by sex and baseline CD4$^+$ cell count on the time to virologic failure or death [intent-to-treat (ITT)] and the time to minimum of virologic failure, death, or treatment modifying toxicity [time to loss of virological response (TLOVR)], the preferred FDA outcome in this unique sub-Saharan African clinical trial population.

Materials and Methods

Study design

Utilizing outcomes data from the recently completed Adult Antiretroviral Treatment and Drug Resistance ("*Tshepo*") study, a clinical trial comparing the efficacy and tolerability of various first-line cART regimens in Botswana in which patients were stratified by baseline CD4$^+$ cell count [< 200 cells/mm^3 (low) versus 201–350 cells/mm^3 (high)], we retrospectively compared TMLE results to results obtained from conventional Cox proportional hazards analyses.

The *Tshepo* study was an open-label, randomized, 3 × 2 × 2 factorial design study conducted in Gaborone, Botswana to evaluate the efficacy, tolerability, and development of drug resistance of six different first-line cART regimens: zidovudine (ZDV), lamuvidune (3TC), and nevirapine (NVP) (Arm A); zidovudine (ZDV), lamuvidune (3TC), and efavirenz (EFV) (Arm B); zidovudine (ZDV), didanosine (ddI), and nevirapine (NVP) (Arm C); zidovudine (ZDV), didanosine (ddI), and efavirenz (EFV) (Arm D); stavudine (d4T), lamuvidine (3TC), and nevirapine (NVP) (Arm E); and stavudine (d4T), lamuvidine (3TC), and efavirenz (EFV) (Arm F). The study also compared two different adherence strategies: standard-of-care (SOC) versus SOC plus community-based supervision (Com-DOT) to determine the optimal means of promoting adherence among adults receiving first-line cART.

Participants were assigned in equal proportions (in an open-label, unblinded fashion) to one of six initial treatment arms and one of two adherence arms using permuted block randomization. Randomization was stratified by CD4$^+$ cell count (less than 200 cells/mm^3, 201–350 cells/mm^3) and by whether the participant had an adherence assistant. Half of the participants were enrolled in each CD4$^+$ cell count stratum, but there were no restrictions on whether they had an adherence assistant prior to study enrollment.

The primary endpoints of the study were development of virologic failure with genotypic drug resistance and development of treatment-related toxicity, as defined by first incidence of a grade 3 or higher adverse event. Secondary endpoints were death for any reason and time to non-adherence, as estimated by an adherence rate of less than 90%. For additional study details, please refer to the previously published article by Wester *et al.*[14]

The study was approved by the institutional review boards of the Botswana Ministry of Health (Health Research Development Committee) and the Harvard School of Public Health (Human Subjects Committee) and written informed consent was obtained from all participants.

Study population

Adult (≥18 years of age), HIV-1 infected, cART-naive Botswana citizens who attended one of the five ART screening clinics in Gaborone were approached for possible enrollment. All potentially eligible adults had to qualify for cART based on existing Botswana national ARV treatment guidelines[24–26] of having an AIDS-defining illness and/or CD4$^+$ cell count ≤200 cells/mm^3 or meet the study's eligibility criteria of a CD4$^+$ cell count between 201 and 350 cells/mm^3 with a plasma HIV-1 RNA level greater than 55,000 copies/ml. Inclusion criteria were hemoglobin value > 8.0 g/dl; absolute neutrophil count ≥1.0×10^3/mm^3; aminotransferase levels less than five times the upper limit of the normal; and for women of child-bearing potential, a willingness to maintain active contraception throughout the duration of the study and a negative urine pregnancy test within 14 days of study enrollment. Exclusion criteria were poor Karnofsky performance score (40 or below); an AIDS-related malignancy other than mucocutaneous Kaposi's sarcoma; grade 2 or higher peripheral neuropathy; major psychiatric illness; and for women, actively breastfeeding or less than 6 months postpartum.

Data collection and follow-up

Clinical and adherence assessments were done monthly at the study clinic. To monitor treatment efficacy, CD4$^+$ cell

counts (*FACS Calibur* flow cytometer, Becton Dickinson, San Jose, CA) and plasma HIV-1 RNA levels (*Amplicor* HIV-1 Monitor test, version 1.5 Roche Diagnostics Systems, Branchburg, NJ) were obtained at enrollment and then every 2 months for the duration of the study. Laboratory safety monitoring included comprehensive chemistry and full blood count specimens at study enrollment, then every month for the first 6 months of the study, every 2 months during months 6–12 of study participation, and every 4 months during the remainder of participation. In addition, all patients had lipid chemistries performed at baseline and then every 6 months. Comprehensive care for study participants was provided in accordance with existing national policy and was free of charge.[25,26]

Statistical considerations

Conventional time-to-event results, performed utilizing Cox proportional hazards models, were compared to results obtained utilizing the TMLE methodology.[21]

The TMLE estimates that are presented represent the mean difference in the marginal additive risk in the probability of survival and the mean difference in marginal log relative hazard between levels of the possible effect modifier; which was taken over the first 34 months following randomization.[21] The difference in the additive risk at the final time point, 34 months, was also calculated (data not shown). For all computations, the treatment variable A was set as for analyzing the effect of NVP versus EFV treatment. The effect modifier variable, V, equaled 1 for males and 0 for females. Thus, a negative mean difference in the marginal difference in the log relative hazard and a positive mean difference in the additive risk indicated that EFV was having a larger beneficial causal effect for females.[21]

The estimates of causal parameters for numerous outcomes of interest were performed. For purposes of these analyses, we focused on the following two outcomes: (1) the time to minimum of virologic failure or death censored by the end of study [which we will refer to as the "intent-to-treat" (ITT) outcome] and (2) the time to minimum of virologic failure or death or treatment modification censored by end of study, which is the preferred FDA outcome for evaluating the efficacy and safety of a particular treatment (which is commonly referred to as the "time to loss in virologic response," which we will refer to as "TLOVR" outcome). For each outcome, we addressed the following causal questions: (1) Was the effect of NNRTI-based cART different by sex? and (2) Was there causal effect modification of NNRTI-based cART by baseline CD4$^+$ cell count?

All statistical analyses were conducted using R statistical software.

Results

Baseline characteristics

Overall, 650 adults were enrolled, 451 (69.4%) of whom were female. The median age was 33.3 years [IQR 28.9–38.7]. Forty-three percent had advanced WHO clinical

TABLE 1. BASELINE CHARACTERISTICS OF STUDY POPULATION BY NNRTI ASSIGNMENT [NEVIRAPINE (NVP) VERSUS EFAVIRENZ (EFV)-BASED COMBINATION ANTIRETROVIRAL THERAPY]

Characteristic		NVP	EFV	Total
Number of participants		325	325	650
Age of participant	Median	33.2	33.7	33.3
	[IQR]	29.0, 38.3	28.8, 39.1	28.9, 38.7
	Age ≤ 40	261 (80.3%)	255 (78.5%)	516 (79.4%)
	Age ≥ 40	64 (19.7%)	70 (21.5%)	134 (20.6%)
Gender	Male	95 (29.2%)	104 (32.0%)	199 (30.6%)
	Female	230 (70.8%)	221 (68.0%)	451 (69.4%)
Weight	Median	57.50	57.0	57.0
	[IQR]	51, 66	50.25, 65.50	51, 66
Baseline BMI	Median	21.2	21.4	21.3
	[IQR]	19.2, 24.3	19.2, 24.1	19.2, 24.3
	< 18.5	61 (18.8%)	60 (18.5%)	121 (18.6%)
	18.5–24.99	187 (57.5%)	199 (61.2%)	386 (59.4%)
	25–29.99	52 (16.0%)	46 (14.2%)	98 (15.1%)
	≥ 30	22 (6.8%)	19 (5.8%)	41 (6.3%)
Baseline HIV-1 RNA (1000s)	Median	183	204	195
	[IQR]	63, 466	85, 499	70, 477
Baseline CD4$^+$ cell count	Median	199	199	199
	[IQR]	138, 243	131, 260	136, 252
	CD4 < 201	165 (50.8%)	165 (50.8%)	330 (50.8%)
	CD4 = 201–350	160 (49.2%)	160 (49.2%)	320 (49.2%)
WHO clinical stage	Stage 1	90 (27.7%)	108 (33.2%)	198 (30.5%)
	Stage 2	84 (25.8%)	77 (23.7%)	161 (24.8%)
	Stage 3	117 (36.0%)	99 (30.5%)	216 (33.2%)
	Stage 4	25 (7.7%)	33 (10.2%)	58 (8.9%)
Baseline diagnostic history	Pulmonary TB	27 (8.3%)	32 (9.8%)	59 (9.1%)

BMI, body mass index; IQR, interquartile range; TB, tuberculosis.

disease (Stage 3 or 4) at the time of enrollment. Of patients, 330 (50.9%) were enrolled in the lower CD4$^+$ stratum with a median CD4$^+$ of 137 cells/mm^3, and 320 (49.1%) were enrolled in the upper CD4$^+$ cell count stratum (CD4$^+$ cell count value between 201 and 350 cells/mm^3 and plasma HIV-1 RNA >55,000 copies/ml) with a median CD4$^+$ of 252 cells/mm^3. Baseline characteristics of patients in the NVP vs. EFV arms were evenly balanced at entry, with 325 patients randomized to each NNRTI arm (Table 1). Females differed significantly from enrolled males in that they were younger, had lower body weights (and body mass indexes), as well as lower hemoglobin values, as expected (Table 2). Baseline characteristics of the entire study population by baseline CD4$^+$ cell count strata (low versus high) are shown in Table 3.

Causal effect modification by sex

Sex did significantly modify the effect of NVP versus EFV for the ITT and TLOVR outcomes.

For the ITT outcome, namely, the time to minimum of virologic failure or death censored by the end of study, the Cox proportional estimate was −1.16, with a corresponding standard error of 0.45; p-value = 0.011; and using TMLE, the mean marginal log relative hazard was −1.13 (standard error = 0.49; p-value = 0.021) and risk difference in probability of survival = 0.062 (standard error = 0.025; p-value = 0.015).

For the TLOVR outcome, namely, the time to minimum of virologic failure or death censored by treatment modification or the end of study, the Cox proportional estimate was −0.95, with a corresponding standard error of 0.35; p-value = 0.007; and using TMLE, the mean marginal log relative hazard was −0.82 (standard error = 0.33; p-value = 0.013) and risk difference in probability of survival = 0.116 or approximately 12% (standard error = 0.035; p-value = 0.001).

Causal effect modification by baseline CD4$^+$ cell count

We then sought to determine whether there was causal effect modification due to baseline CD4$^+$ cell count level [low

TABLE 2. BASELINE CHARACTERISTICS OF STUDY POPULATION BY SEX

	Male (n = 199)	Female (n = 451)	Combined (n = 650)	p-value[a]
Age[b]	34.8 (30.7, 40.6)	32.5 (28.4, 37.4)	33.3 (28.9, 38.6)	<0.01
Age (dichotomized), n (%)				<0.01
<40	143 (71.9%)	373 (82.7%)	516 (79.4%)	
≥40	56 (28.1%)	78 (17.3%)	134 (20.6%)	
Weight[b]	60 (54.2, 66)	56 (49, 66)	57 (51, 66)	<0.01
Missing weight, n (%)	0 (0.0%)	1 (0.2%)	1 (0.2%)	
BMI[b]	20.7 (18.9, 22.7)	21.8 (19.4, 25.3)	21.2 (19.1, 24.3)	<0.01
Missing BMI, n (%)	0 (0.0%)	5 (1.1%)	5 (0.8%)	
BMI (categorized), n %(%)				<0.01
<18.5	42 (21.1%)	78 (17.5%)	120 (18.6%)	
18.5–24.99	137 (68.8%)	249 (55.8%)	386 (59.8%)	
25–29.99	18 (9.0%)	81 (18.2%)	99 (15.3%)	
≥30	2 (1.0%)	38 (8.5%)	40 (6.2%)	
Hemoglobin[b]	13.2 (11.6, 14.2)	11.1 (10.1, 12.2)	11.6 (10.4, 12.9)	<0.01
Missing hemoglobin, n (%)	1 (0.5%)	12 (2.7%)	13 (2.0%)	
HIV-1 RNA level[b]	224,000 (88,300, 528,000)	177,000 (68,000, 426,500)	195,000 (70,300, 475,000)	0.02
HIV-1 RNA level (dichotomized) n (%)				0.24
≤500,000	146 (73.4%)	351 (78.0%)	497 (76.6%)	
>500,000	53 (26.6%)	99 (22.0%)	152 (23.4%)	
CD4 cell count[b]	203 (140.5, 253.5)	196 (132.5, 251.5)	199 (136, 252)	0.28
CD4 (dichotomized), n (%)				0.44
≤200	96 (48.2%)	234 (51.9%)	330 (50.8%)	
>200	103 (51.8%)	217 (48.1%)	320 (49.2%)	
WHO clinical stage, n (%)				0.31
Stage 1	67 (34.7%)	131 (29.8%)	198 (31.3%)	
Stage 2	42 (21.8%)	119 (27.0%)	161 (25.4%)	
Stage 3	63 (32.6%)	153 (34.8%)	216 (34.1%)	
Stage 4	21 (10.9%)	37 (8.4%)	58 (9.2%)	
NRTI treatment arm, n (%)				0.37
ZDV/3TC	73 (36.7%)	145 (32.2%)	218 (33.5%)	
ZDV/ddI	59 (29.6%)	157 (34.8%)	216 (33.2%)	
d4T/3TC	67 (33.7%)	149 (33.0%)	216 (33.2%)	

[a]To compare the distribution of study characteristics by sex, we employ chi-square tests. Similarly, we use a two-sample rank sum test for continuous variables by sex.
[b]Continuous variables are reported as medians (interquartile range).

TABLE 3. BASELINE CHARACTERISTICS OF STUDY POPULATION BY BASELINE CD4$^+$ CELL COUNT STRATA
[NAMELY, LOW (CD4$^+$ CELL COUNT ≤200) VERSUS HIGH (CD4$^+$ CELL COUNT 201–350)]

	≤ 200 cells/μl (n = 330)	> 200 cells/μl (n = 320)	Combined (n = 650)	p-value[a]
Age[b]	33.1 (29, 37.8)	33.4 (28.9, 39.3)	33.3 (28.9, 38.6)	0.45
Age (dichototmized), n (%)				0.28
<40	268 (81.2%)	248 (77.5%)	516 (79.4%)	
≥40	62 (18.8%)	72 (22.5%)	134 (20.6%)	
Sex, n (%)				0.44
Male	96 (29.1%)	103 (32.2%)	199 (30.6%)	
Female	234 (70.9%)	217 (67.8%)	451 (69.4%)	
Weight[b]	57 (50, 65.5)	58.8 (52, 66.2)	57 (51, 66)	0.13
Missing weight, n (%)	1 (0.3%)	0 (0.0%)	1 (0.2%)	
BMI[b]	21.1 (18.9. 24.1)	21.3 (19.6, 24.3)	21.2 (19.1, 24.3)	0.30
Missing BMI, n (%)	4 (1.2%)	1 (0.3%)	5 (0.8%)	
BMI (categorized), n (%)				0.72
<18.5	66 (20.2%)	54 (16.9%)	120 (18.6%)	
18.5–24.99	191 (58.6%)	195 (61.1%)	386 (59.8%)	
25–29.99	48 (14.7%)	51 (16.0%)	99 (15.3%)	
≥30	21 (6.4%)	19 (6.0%)	40 (6.2%)	
Hemoglobin[b]	11.5 (10.3, 12.7)	11.8 (10.5, 13)	11.6 (10.4, 12.9)	0.04
Missing hemoglobin, n (%)	6 (1.8%)	7 (2.2%)	13 (2.0%)	
HIV-1 RNA level[b]	230,000 (65,000, 607,000)	177,500 (73,520, 357,500)	195,000 (70,300, 475,000)	0.02
HIV-1 RNA level (dichotomized), n (%)				<0.01
≤500,000	225 (68.4%)	272 (85.0%)	497 (76.6%)	
>500,000	104 (31.6%)	48 (15.0%)	152 (23.4%)	
WHO clinical stage, n (%)				<0.01
Stage 1	82 (25.6%)	116 (37.1%)	198 (31.3%)	
Stage 2	71 (22.2%)	90 (28.8%)	161 (25.4%)	
Stage 3	130 (40.6%)	86 (27.5%)	216 (34.1%)	
Stage 4	37 (11.6%)	21 (6.7%)	58 (9.2%)	
NRTI, n (%)				0.99
ZDV/3TC	110 (33.3%)	108 (33.8%)	218 (33.5%)	
ZDV/ddI	110 (33.3%)	106 (33.1%)	216 (33.2%)	
d4T/3TC	110 (33.3%)	106 (33.1%)	216 (33.2%)	

[a]To compare the distribution of study characteristics by CD4 cell count, we employ chi-square tests. Similarly, we use a two-sample rank sum test for continuous variables by CD4 cell count.
[b]Continuous variables are reported as medians (interquartile range).

(<200 cells/mm^3) versus high (201–350 cells/mm^3] for our two specified outcomes.

Evaluating the ITT outcome, the Cox proportional estimate was 0.403, with a corresponding standard error of 0.402; p-value = 0.320; and using TMLE, the mean difference in marginal log relative hazard was 0.859 (standard error = 0.468; p-value = 0.066) and mean difference in additive risk in probability of survival = –0.053 (standard error = 0.028; p-value = 0.089) (Table 2).

For the TLOVR outcome, the Cox proportional estimate was – 0.675, with a corresponding standard error of 0.317; p-value = 0.033; and using TMLE, the mean difference in marginal log relative hazard was – 0.829 (standard error = 0.356; p-value = 0.020) and mean difference in additive risk of the probability of survival = –0.115 (standard error = 0.051; p-value = 0.023). For the TLOVR outcome, the Cox proportional hazard estimate was significantly different from zero but the TMLE had a more significant p-value. Specifically, for the TLOVR outcome, the Cox proportional estimate was 0.675, with a corresponding standard error of 0.317; p-value = 0.033; and using TMLE, the mean difference in marginal log relative

hazard was 0.829 (standard error = 0.356; p-value = 0.020) and mean difference in additive risk of probability of survival = –0.115 (standard error = 0.051; p-value = 0.023).

Figure 1 depicts the significant effect modification by baseline CD4$^+$ cell count (high vs. low) observed for the TLOVR outcome. Not only is the effect of treatment, EFV vs. NVP, among randomized study subjects having high baseline CD4$^+$ cell counts (201–350 cells/mm^3) different than the effect seen when evaluating randomized study subjects having low (<201 cells/mm^3) baseline CD4$^+$ cell counts, the effects are in opposite directions. Adults with high baseline CD4$^+$ cell counts who receive NVP-based cART have a higher probability of experiencing virologic failure, death, and/or treatment modification (TLOVR outcome). Conversely, adults with low baseline CD4$^+$ cell counts who receive EFV-based cART have a probability of experiencing these same adverse outcomes. Furthermore, these figures depict the double robustness of the TMLE methodology. Figure 1A used data adaptive methods to fit the initial hazard regression for all combinations of baseline CD4$^+$ and treatment levels while Fig. 1B used an intentionally misspecified initial estimate that
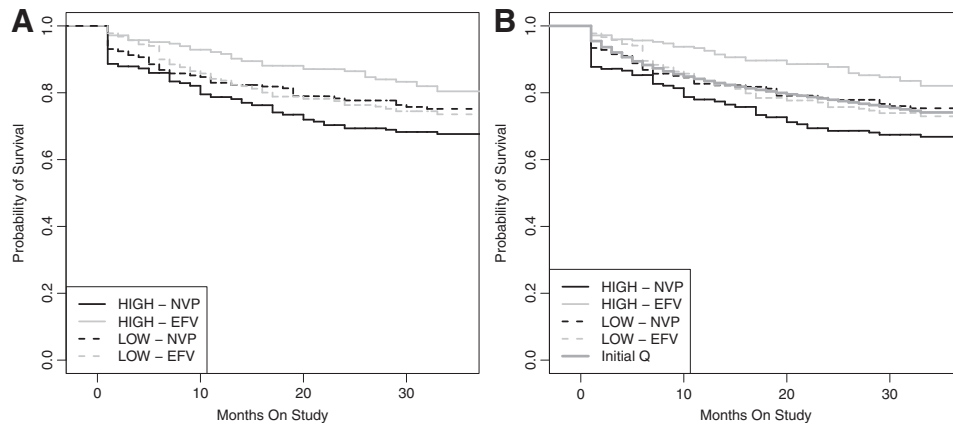
FIG. 1. Causal effect modification by baseline CD4$^+$ cell count: low (CD4$^+$ cell count <201 cells/mm^3) versus high (CD4$^+$ cell count 201–350 cells/mm^3) for the TLOVR Outcome (Time to Virologic Failure, Death, or Treatment Modifying Toxicity). The survival curve in **(A)** was estimated using data-adaptive methods to estimate the initial outcome hazards as well as possible, while the curve in **(B)** was estimated with an intentionally misspecified initial hazard that was the same for all four survival curves. The solid gray line in **(B)** shows the initial estimate of all four survival curves when the initial hazard was intentionally misspecified. Super-learner, a data-adaptive machine learning algorithm,[26] was then used to estimate the treatment distribution as well as possible, and since the only censoring event was the end of study, so that censoring was known to be independent of the baseline covariates, Kaplan–Meier estimates were used to estimate the censoring process.

was the same for all four groups. Yet, Fig. 1B shows that the effect modification was still detected since the censoring mechanism and propensity scores were known to be correctly specified.

Discussion

Our study compared the effectiveness of NVP-based versus EFV-based cART among a large group of adults in Botswana. Published outcomes using conventional Cox proportional hazards showed that females receiving NVP-based cART trended toward having higher virologic failure rates compared to EFV-treated women.[14] This was most likely due to the fact that NVP-treated adults tended to modify treatment sooner than EFV-treated adults due to the significantly higher treatment-modifying toxicity rates among NVP-treated (vs. EFV-treated) patients.[14]

Using TMLE, we further evaluated study outcomes, specifically evaluating the causal effect of assigned NNRTI on the time to virologic failure or death (ITT outcome) and the time to minimum of virologic failure, death, or treatment modifying toxicity (TLOVR outcome) by sex and baseline CD4$^+$ cell count [high (201–350 cells/mm^3) versus low (<201 cells/mm^3)].

Evaluating for possible effect modification by sex, we found that sex does modify the effect of EFV versus NVP on the time to death and the time until TLOVR. EFV-treated women tended to have more favorable outcomes in contrast to NVP-treated males having more favorable outcomes. For the TLOVR outcome, the average causal risk difference between the effect in males versus the females was 12% (*p*-value = 0.001). These results substantiate the borderline statistically significant results obtained among NNRTI-treated females with virologic failure using conventional Cox proportional hazards techniques. The main reason for this was that NVP-treated females tended to modify their cART regimens at significantly higher rates than the other treated

groups. The major difference in treatment modification rates tended to occur almost immediately post initiating NVP-based cART. However, treatment modification does not explain the entire effect modification that was observed because both outcomes including virologic failure and death were also statistically significant in terms of an effect in males versus females. In fact, for the outcomes including both virologic failure and death, the average risk difference between the effect in males versus females was approximately 6% (*p*-value = 0.015).

Evaluating for possible effect modification by baseline CD4$^+$ cell count, we did find that baseline CD4$^+$ did modify the effect of NVP versus EFV on the time until death and the TLOVR outcome. The effect modification for the outcome that included both virologic failure and death was close to significant. EFV tended to be favorable compared to NVP among cART-treated adults having higher (201–350 cells/mm^3) baseline CD4$^+$ cell count values. Among randomized patients having low baseline CD4$^+$ cell count values (<200 cells/mm^3), there was no significant difference in the treatment survival curves. For the TLOVR outcome, the average risk difference between the effect in the high CD4$^+$ cell count strata versus the low CD4$^+$ strata was 12% (*p*-value = 0.023).

One possible explanation for this is that NVP-treated adults have higher overall rates of treatment modifying toxicities, which occur along a full continuum of CD4$^+$ cell count ranges, particularly high (>250 cells/mm^3) CD4$^+$ cell count values, some of which may be life-threatening (i.e., hepatotoxicity and/or cutaneous hypersensitivity reactions). Such toxicities may be particularly problematic among persons with higher CD4$^+$ cell counts, resulting in higher rates of treatment modifying toxicities as such persons generally feel healthier and experience far fewer comorbidities and are therefore more likely to report earlier grade toxicities and have lower thresholds to request a change in their treatment.

Limitations include the fact that there were very few deaths among EFV-treated adults having high baseline CD4$^+$ cell

count values. NVP-treated patients having less pronounced immunosuppression, namely, having high baseline $CD4^+$ cell counts (201–350 cells/mm^3), however, tended to die at a higher frequency and sooner compared to those having more advanced baseline immunosuppression (i.e., having low baseline $CD4^+$ cell counts, in the less than 201 cells/mm^3 range).

Conclusions

This is the first study to date comparing time to event parameter estimates obtained using the conventional Cox proportional hazards model to TMLE among clinical trial participants in sub-Saharan Africa. As we previously indicated, TMLE is an efficient and robust method for estimating causal effect parameters that directly answer a specific scientific question of interest. TMLE allows for the clinically meaningful delineation and interpretation of the causal effect of NNRTI treatment and effect modification by sex and baseline $CD4^+$ cell count strata in this recently completed study. As the majority of cART-treated adults in resource-limited settings are female, which is in sharp contrast to male predominant U.S. and Western European cohorts, it will be of paramount importance to continue to evaluate for possible effect modification by relying on a statistical methodology that provides desired and more importantly readily interpretable parameter estimates.

In our urban Botswana setting, EFV-treated women and NVP-treated men had more favorable cART outcomes. In addition, adults initiating EFV-based cART at higher baseline $CD4^+$ cell count values had more favorable outcomes compared to those initiating NVP-based cART. Based on these findings, policymakers in sub-Saharan Africa may want to consider restricting NVP use to adult males having baseline $CD4^+$ cell counts less than 200 cells/mm^3. Additional studies including longer-term follow-up of larger numbers of patients are clearly warranted, as such information will greatly inform such potential policy decisions.

Acknowledgments

Author Disclosure Statement

No competing financial interests exist.

References

1. Towards universal access: Scaling up priority HIV/AIDS interventions in the health sector. Progress report 2010 (WHO, UNICEF, UNAIDS), p. 53.
2. Nglazi MD, Lawn SD, Kaplan R, *et al.:* Changes in programmatic outcomes during 7 years of scale-up at a community-based antiretroviral treatment service in South Africa. J Acquir Immune Defic Syndr 2011;56(1):e1–8.
3. Boulle A, Van Cutsem G, Hilderbrand K, *et al.:* Seven-year experience of a primary care antiretroviral treatment programme in Khayelitsha, South Africa. AIDS 2010;24(4):563–572.
4. Chi BH, Mwango A, Giganti M, *et al.:* Early clinical and programmatic outcomes with tenofovir-based antiretroviral therapy in Zambia. J Acquir Immune Defic Syndr 2010;54(1):63–70.
5. Bussmann H, Wester CW, Ndwapi N, *et al.:* Five-year outcomes of initial patients treated in Botswana's National Antiretroviral Treatment Program. AIDS 2008;22(17):2303–2311.
6. Coetzee D, Hildebrand K, Boulle A, *et al.:* Outcomes after two years of providing antiretroviral treatment in Khayelitsha, South Africa. AIDS 2004;18(6):887–895.
7. Panel on Antiretroviral Guidelines for Adults and Adolescents: Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. Department of Health and Human Services. January 10, 2011, 1–174. Available at http://www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentGL.pdf. Accessed 19 April 2011; pp. 41–62.
8. Staszewski S, Morales-Ramirez J, Tashima KT, *et al.:* Efavirenz plus zidovudine and lamuvidine, efavirenz plus indinavir, and indinavir plus zidovudine and lamuvidine in the treatment of HIV-infection in adults. Study 006 Team. N Engl J Med 1999;349(25):1865–1873.
9. Riddler SA, Haubrich R, DiRienzo AG, *et al.:* Class-sparing regimens for initial treatment of HIV-1 infection. N Engl J Med 2008;358(20):2095–2106.
10. Squires K, Lazzarin A, Gatell JM, *et al.:* Comparison of once-daily atazanavir with efavirenz, each in combination with fixed-dose zidovudine and lamuvidine, as initial therapy for patients infected with HIV. J Acquir Immune Defic Syndr 2004;36(5):1011–1019.
11. Lennox JL, DeJesus E, Lazzarin A, *et al.:* Safety and efficacy of raltegravir-based versus efavirenz-based combination therapy in treatment-naïve patients with HIV-1 infection: A multicentre, double-blind randomized controlled trial. Lancet 2009;374(9692):796–806.
12. Cooper DA, Heera J, Goodrich J, *et al.:* Maraviroc versus efavirenz, both in combination with zidovudine-lamuvidine, for the treatment of antiretroviral-naïve subjects with CCR5-tropic HIV-1 infection. J Infect Dis 2010;201(60):803–813.
13. Robbins GK, DeGruttola V, Shafer RW, *et al.:* Comparison of sequential three-drug regimens as initial therapy for HIV-1 infection. N Engl J Med 2003;349(24):2293–2303.
14. Wester CW, Thomas AM, Bussmann H, *et al.:* Non-nucleoside reverse transcriptase inhibitor outcomes among combination antiretroviral therapy-treated adults in Botswana. AIDS 2010;24(Suppl 1):S27–S36.
15. Beck EJ, Vitoria M, Mandalia S, *et al.:* National adult antiretroviral therapy guidelines in resource-limited countries: Concordance with 2003 WHO guidelines? AIDS 2006;20(11):1497–1502.
16. WHO/UNAIDS report: Treating 3 million by 2005. Making it happen, the WHO strategy. 2003. www.unaids.org (accessed 19 April 2011).

17. Severe P, Juste MAJ, Ambroise A, *et al.:* Early versus standard antiretroviral therapy for HIV-infected adults in Haiti. N Engl J Med 2010;363:257–265.

18. World Health Organization (WHO): Rapid Advice. "Antiretroviral therapy for HIV infection in adults and adolescents. November 2009. pp. 1–28. http://www.who.int/hiv/pub/arv/rapid_advice_art.pdf. Accessed 19 April 2011.

19. van Leth F, Phanuphak P, Ruxrungtham K, *et al.:* Comparison of first-line antiretroviral therapy with regimens including nevirapine, efavirenz, or both drugs, plus stavudine and lamivudine: A randomised open-label trial, the 2NN Study. Lancet 2004;363:1253–1263.

20. van den Berg-Wolf M, Hullsiek KH, Peng G, *et al.:* Virologic, immunologic, clinical, safety, and resistance outcomes from a long-term comparison of efavirenz-based versus nevirapine-based antiretroviral regimens as initial therapy in HIV-1 infected persons. HIV Clin Trials 2008;9(5):324–336.

21. Stitelman OM, Wester CW, DeGruttola V, *et al.:* Targeted maximum likelihood estimation of effect modification parameters in survival analysis. Int J Biostat 2011;7:Issue 1, Article 19. DOI: 10.2202/1557-4679.1307. Available at http://www.bepress.com/ijb/vol7/iss1/19.

22. van der Laan MJ and Rubin D: Targeted maximum likelihood learning. Int J Biostat 2006;2(1). Article 11.

23. Moore K and van der Laan MJ: Application of time-to-event methods in the assessment of safety in clinical trials. In *Design, Summarization, Analysis, & Interpretation of Clinical Trials with Time-to-Event Endpoints* (Pearce KE, ed.). Chapman & Hall, New York, 2009.

24. Botswana Guidelines on Antiretroviral Treatment, 2002 version. 2002, Ministry of Health, Botswana: Gaborone, Botswana.

25. Botswana Guidelines on Antiretroviral Treatment. 2005 version. 2005, Ministry of Health, Botswana: Gaborone, Botswana.

26. van der Laan MJ, Polley E, and Hubbard A: "Super learner." Stat Appl Genet Mol Biol 2007:6. Article 25.

Address correspondence to:
C. William Wester
Vanderbilt University School of Medicine
Department of Medicine
Division of Infectious Diseases
Vanderbilt Institute for Global Health (VIGH)
1161 21st Avenue South
A-2200, Medical Center North
Nashville, Tennessee 37232

E-mail: william.wester@vanderbilt.edu