Title

Transportability From Randomized Trials to Clinical Care: On Initial HIV Treatment With Efavirenz and Suicidal Thoughts or Behaviors.

Journal

American Journal of Epidemiology, 190(10)

ISSN

0002-9262

Authors

Mollan, Katie R
Pence, Brian W
Xu, Steven
et al.

Publication Date

2021-10-01

DOI

10.1093/aje/kwab136

Peer reviewed

American Journal of Epidemiology

Vol. 190, No. 10
https://doi.org/10.1093/aje/kwab136
Advance Access publication:
May 10, 2021

Original Contribution

Transportability From Randomized Trials to Clinical Care: On Initial HIV Treatment With Efavirenz and Suicidal Thoughts or Behaviors

Katie R. Mollan*, Brian W. Pence, Steven Xu, Jessie K. Edwards, W. Christopher Mathews, Conall O'Cleirigh, Heidi M. Crane, Ellen F. Eaton, Ann C. Collier, Ann Marie K. Weideman, Daniel Westreich, Stephen R. Cole, Camlin Tierney, and Angela M. Bengtson* for the CFAR Network of Integrated Clinical Systems and the AIDS Clinical Trials Group

* Correspondence to Katie R. Mollan, Center for AIDS Research, School of Medicine, University of North Carolina at Chapel Hill, 3126 McGavran-Greenberg Hall, Campus Box 7420, Chapel Hill, NC 27599 (e-mail: kmollan@unc.edu); or Dr. Angela M. Bengtson, Department of Epidemiology, School of Public Health, Brown University, Box G-S121-2, 121 South Main Street, Providence, RI 02903 (e-mail: angela_bengtson@brown.edu).

Initially submitted August 9, 2020; accepted for publication May 3, 2021.

In an analysis of randomized trials, use of efavirenz for treatment of human immunodeficiency virus (HIV) infection was associated with increased suicidal thoughts/behaviors. However, analyses of observational data have found no evidence of increased risk. To assess whether population differences might explain this divergence, we transported the effect of efavirenz use from these trials to a specific target population. Using inverse odds weights and multiple imputation, we transported the effect of efavirenz on suicidal thoughts/behaviors in these randomized trials (participants were enrolled in 2001–2007) to a trials-eligible cohort of US adults initiating antiretroviral therapy while receiving HIV clinical care at medical centers between 1999 and 2015. Overall, 8,291 cohort participants and 3,949 trial participants were eligible. Prescription of antidepressants (19% vs. 13%) and injection drug history (16% vs. 10%) were more frequent in the cohort than in the trial participants. Compared with the effect in trials, the estimated hazard ratio for efavirenz on suicidal thoughts/behaviors was attenuated in our target population (trials: hazard ratio (HR) = 2.3 (95% confidence interval (CI): 1.2, 4.4); transported: HR = 1.8 (95% CI: 0.9, 4.4)), whereas the incidence rate difference was similar (trials: HR = 5.1 (95% CI: 1.6, 8.7); transported: HR = 5.4 (95% CI: −0.4, 11.4)). In our target population, there was greater than 20% attenuation of the hazard ratio estimate as compared with the trials-only estimate. Transporting results from trials to a target population is informative for addressing external validity.

benzoxazines; efavirenz; HIV; inverse odds weights; multiple imputation; new user design; suicidal ideation; transportability

Abbreviations: ACTG, AIDS Clinical Trials Group; AIDS, acquired immunodeficiency syndrome; ART, antiretroviral therapy; CI, confidence interval; CNICS, Centers for AIDS Research Network of Integrated Clinical Systems; HIV, human immunodeficiency virus; HIV-1, human immunodeficiency virus type 1; HR, hazard ratio; IOPW, inverse odds of participation weights; IR, incidence rate; IRD, incidence rate difference; MI, multiple imputation; PHQ-9, Patient Health Questionnaire-9; RCT, randomized controlled trial.

For over 15 years, efavirenz was the nonnucleoside reverse transcriptase inhibitor of choice for first-line antiretroviral therapy (ART) in the treatment of human immunodeficiency virus (HIV) disease in the United States (1). While newer agents are available, many people living with HIV remain on efavirenz (1). Globally, efavirenz continues to be widely used, and the World Health Organization recommends efavirenz-containing ART as an alternative first-line regimen, with the recommended dose lowered from 600 mg to 400 mg (2, 3).

Controversy over a possible link between efavirenz use and suicidal thoughts/behaviors has been cause for ongoing

clinical concern (4, 5), and disparate findings between randomized and observational studies have led to a lack of clarity. In several analyses of randomized controlled trials (RCTs), initiating efavirenz increased the risk of suicidal thoughts/behaviors (6, 7), including a pooled analysis of 4 RCTs from the AIDS Clinical Trials Group (ACTG) which found an increase in the risk of suicidal thoughts/behaviors reported as adverse events (hazard ratio (HR) = 2.3, 95% confidence interval (CI): 1.2, 4.4). However, these findings were not confirmed by several large observational studies of adults living with HIV (8–10). In the Centers for AIDS Research Network of Integrated Clinical Systems (CNICS) observational cohort, the estimated association between initiating efavirenz and suicidal thoughts, measured by Patient Health Questionnaire-9 (PHQ-9) (11), was closer to the null value of 1 (HR = 1.2, 95% CI: 0.7, 2.3) (10). For clinicians considering prescribing efavirenz-containing ART to ART-naive individuals, the risk of serious psychiatric side effects in practice remains unclear (4).

There is growing interest in transporting effects from RCTs to target populations to evaluate the external validity of trial results and to understand discrepant results between randomized and observational studies (12–15). Results from trials and target populations will differ if the distribution of covariates that modify the treatment effect measure differ between the trial and the target population (16, 17). Many applications will involve missing data and a nonnested design, where the RCT is not embedded inside a sample of the target population (14, 18). We used individual-level data to transport the effect of initiating efavirenz upon suicidal thoughts/behaviors from 4 ACTG RCTs to a CNICS observational cohort sample. The CNICS sample was a target population of US adults living with HIV who were receiving care at a medical center and initiated ART between 1999 and 2015 (19), the years in which efavirenz was recommended as a first-line therapy in the United States (1). We evaluated what the effect of initiating efavirenz on suicidal thoughts/behaviors might have been had the trials been conducted in this target population.

METHODS

Our analysis examined transportability of the findings from the 4 aforementioned trials to a specific target population. We harmonized and combined participant-level data from 4 ACTG trials (RCT sample) and the CNICS cohort (observational, nonrandomized sample) and applied inverse odds of participation weights (IOPW) to estimate a target population hazard ratio and incidence rate difference (IRD). Baseline covariates from RCT and nonrandomized participants were used to construct IOPW. Efavirenz-containing regimens, the exposure of interest, were randomly assigned in each trial. Outcomes (suicidal thoughts/behaviors) from RCT participants were analyzed. Each trial required that participants be ART-naive at randomization; thus, there was no co-enrollment among the 4 RCTs. The RCT and observational samples were not nested (13, 14). Missing baseline covariate data were handled using multiple imputation (MI), and eligibility criteria were not imputed.

Analyses were conducted in SAS, version 9.4 (SAS Institute, Inc., Cary, North Carolina); Linux R, version 3.3.1 (R Foundation for Statistical Computing, Vienna, Austria) (20); and Windows R, version 3.6.2 (20). Participants in the ACTG trials and the CNICS cohort provided written informed consent. The institutional review board of the University of North Carolina at Chapel Hill provided ethical approval for this analysis.

Randomized controlled trials

The 4 ACTG RCTs enrolled participants between 2001 and 2007 across 68 US sites, and they had similar operating procedures and eligibility criteria across study protocols. Consenting RCT participants were eligible to enroll if they were ART-naive, at least 18 years of age, did not have substantially abnormal laboratory values, and were judged by trial investigators as able to participate in the study and comply with study medications. Eligibility criteria for each of the 4 RCTs were similar, and history of suicidal thoughts/behaviors was not an exclusion criterion. Participants were randomly assigned to initiate use of either an efavirenz-containing regimen or an efavirenz-free regimen as first-line ART. Efavirenz was open-label in 3 of the 4 RCTs and was administered as a once-daily 600-mg dose (with 2 of 4 trial protocols directing participants to take the medication at bedtime). Further details regarding eligibility criteria, ART regimens, and study follow-up have been published elsewhere (6).

Each RCT required reporting of all deaths, severe and life-threatening signs/symptoms, and any sign/symptom that led to modification of antiretroviral treatment; 3 of 4 trials required reporting of moderate central nervous system symptoms. The outcome in the trials was a composite of time to suicidal ideation or attempted or completed suicide as identified from signs/symptoms, diagnoses, adverse events, and death data via Medical Dictionary for Regulatory Activities coded records (6). In the previous pooled RCT analysis, 2 statisticians separately coded the suicidal thoughts/behaviors outcome, and causes of death were reviewed by clinical investigators blinded to efavirenz exposure. Suicidal events coded according to the Medical Dictionary for Regulatory Activities were manually compared with free-text descriptions, and free-text adverse event descriptions containing the string "suic" were manually reviewed prior to analyses.

Following harmonization of covariate and outcomes coding and data-set structure across the ACTG trials, RCT data were concatenated by column to create 1 data set (6). We restricted our analysis to US participants because our target population resided in the United States (n = 1,381 non-US participants were excluded; Figure 1A). Intention-to-treat analyses were conducted throughout, with follow-up in 2 of the RCTs censored at the release of data and safety monitoring boards' recommendations pertaining to efavirenz. Within each trial, the median duration of follow-up was similar between the efavirenz-containing and efavirenz-free regimen groups (see Web Table 1, available at https://doi.org/10.1093/aje/kwab136) (6). Right-censored follow-up in the trials was handled as noninformative, and a trials-only

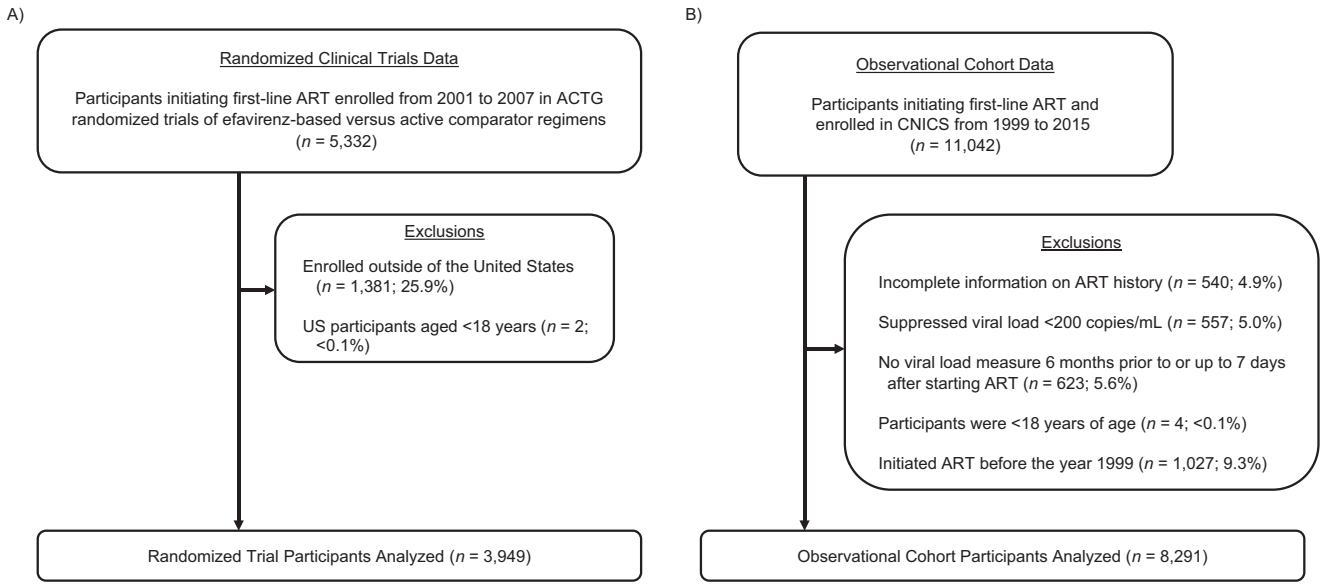

Figure 1. Study eligibility of antiretroviral therapy (ART)-naive adults living with human immunodeficiency virus in the United States who initiated first-line ART between 1999 and 2015. A total of 3,949 randomized participants from 4 AIDS Clinical Trials Group (ACTG) randomized controlled trials and 8,291 observational participants from the Centers for AIDS Research Network of Integrated Clinical Systems (CNICS) cohort were included. AIDS, acquired immunodeficiency syndrome.

sensitivity analysis was conducted using inverse probability of censoring weights (Web Appendix 1) (21, 22).

Observational cohort

The CNICS cohort includes over 31,000 adults living with HIV who are receiving clinical care at 8 academic medical center sites in the United States (19). To our knowledge, no enumerated sample of HIV-positive adults receiving care in the United States exists; thus, we cannot evaluate whether CNICs represents a random sample of our target population. Nonetheless, CNICS provides diverse patient representation with low refusal rates and contains rich individual-level data (19, 23). CNICS captures comprehensive clinical data that includes standardized demographic, diagnosis, medication, laboratory, and mortality information collected through electronic medical records and institutional data systems.

We defined a target population that met measured inclusion criteria for the ACTG randomized trials. Participants had to be at least 18 years of age, previously ART-naive (i.e., new users), and initiating a first-line combination ART regimen between 1999 and 2015. Baseline was defined as the date of ART initiation at a CNICS site. Comprehensive data on possible ART use before entry into the CNICS cohort was not available for all CNICS participants. Therefore, we excluded patients without complete ART information ($n = 540$; 5% of 11,042) and did not impute inclusion criteria. We also excluded participants who, in the 6 months prior to and up to 7 days after ART initiation, had a suppressed HIV RNA viral load (defined as <200 copies/mL; $n = 557$) or did not have any viral load measurements ($n = 623$) (Figure 1B). No additional laboratory measures (e.g., creatinine clearance) were used to restrict our target population.

Inverse odds of participation weights (W_i)

We applied IOPW in marginal structural models to account for measured factors potentially related to both selection into the trials and suicidal thoughts/behavior outcomes (13, 24). For the ith participant, let $S_i = 1$ denote participation in the RCT sample and $S_i = 0$ participation in the observational, nonrandomized sample, where $i = 1, 2, \ldots, n, n+1, \ldots N$ indexes participants in the combined samples, with n participants in the RCTs and $N - n$ participants in the nonrandomized sample. Let A_i indicate exposure to an efavirenz-containing regimen ($a = 1$) versus an efavirenz-free regimen ($a = 0$). The vector $\boldsymbol{Z_i}$ contains 9 measured baseline covariates related to both the outcome risk and selection S_i (13). IOPW were constructed as

$$\hat{W}_i = \frac{\Pr(S_i=1)/\Pr(S_i=0)}{\hat{\Pr}(S_i=1|\boldsymbol{Z_i})/\hat{\Pr}(S_i=0|\boldsymbol{Z_i})} = \frac{n/(N-n)}{\hat{\Pr}(S_i=1|\boldsymbol{Z_i})/\hat{\Pr}(S_i=0|\boldsymbol{Z_i})}$$

for $S_i = 1$; and $\hat{W}_i = 0$ for $S_i = 0$ (13). The numerator of the IOPW is the marginal odds of being in the RCT sample, and the denominator is the predicted odds of being in the RCT sample versus the observational sample, conditional on covariates $\boldsymbol{Z_i}$.

Estimation of IOPW proceeded in several steps. After harmonizing covariate data $\boldsymbol{Z_i}$, the randomized and observational samples were combined (i.e., stacked). Nine baseline participant characteristics, described in the previous trials-only analysis (6), were used to estimate IOPW: sex

(self-reported), race/ethnicity (self-reported), age (years), CD4 cell count (cells/μL), number of copies of human immunodeficiency virus type 1 (HIV-1) RNA per mL at ART initiation, history of acquired immunodeficiency syndrome (AIDS)-defining illness, history of injection drug use, indication of previous chronic viral hepatitis B or C infection, and prescription of antidepressants. Covariate main effects and all 2-way interaction terms (with linear-only interactions for age, CD4 cell count, and HIV-1 viral RNA load) were included in a logistic regression model with outcome variable S_i and covariates $\boldsymbol{Z_i}$. There were no additional known effect measure modifiers measured in the RCTs and not measured in the CNICS cohort (25), thus our analysis focuses on the 9 measured covariates $\boldsymbol{Z_i}$.

Hepatitis B was defined as testing positive for hepatitis B surface antigen, and hepatitis C was defined as testing positive for hepatitis C antibody. Prescription of antidepressants was defined as reported use of antidepressant medication within 30 days before ART initiation. Baseline measurements were taken near and prior to ART initiation. Additional mental health covariates, substance use, and body weight were not adequately measured and therefore could not be included. Continuous covariates (age, CD4 cell count, and HIV-1 viral RNA load) were fitted flexibly using restricted quadratic splines with 4 knots placed at the 20, 40, 60, and 80th percentiles (26). Categorical covariates were fitted using indicator variables.

Marginal structural models

Our main objective was to transport the randomized trials hazard ratio to a target population sample from CNICS. We also estimated incidence rates (IRs) and an IRD as a secondary analysis to put the results into the context of absolute risk (27) and to compare the transportability of the hazard ratio and IRD. Let T_i^a denote the potential time to suicidal thoughts/behavior had participant i received treatment a. A marginal structural Cox model was constructed: $h(T_i^a = t|S_i = 0) = h_0(t)\exp(\beta a)$, where $h_0(t)$ is an unspecified baseline hazard for the survival times $T_i^{a=0}$ and the hazard ratio $\exp(\beta)$ is our target estimand.

Given that A_i was randomly assigned in the RCTs, we assume exchangeability $T_i^a \perp A_i \mid S_i = 1$ for $a = 0$, 1 for $i = 1, 2, \ldots n$. By causal consistency, $T_i^a = T_i$ when $A_i = a$. We assume no interference and treatment version irrelevance for versions of $A_i = 1$ and versions of $A_i = 0$ (28). We further assume $T_i^a \perp S_i \mid \boldsymbol{Z_i}$ for $a = 0, 1$ (potential outcomes are exchangeable across the RCT sample and nonrandomized sample and across calendar time 1999–2015, conditional upon covariates $\boldsymbol{Z_i}$). Given our intention-to-treat approach, a component of this assumption is that ART compliance patterns in the trials are similar to ART compliance patterns in the target population. Further, we assume effect measure modifier coverage, that is, $\Pr(S_i = s|\boldsymbol{Z_i} = z) > 0$ for $s = 0, 1$ if $\Pr(\boldsymbol{Z_i} = z) > 0$ in the target population (25).

Let C_i denote censoring time and $Y_i \equiv \min(T_i, C_i)$ with $\delta_i = 1$ for an observed event (i.e., $T_i \leq C_i$), and $\delta_i = 0$ if right-censored. Lastly, we assume that $(\boldsymbol{Z_i}, S_i, A_i, Y_i, \delta_i)$ in the RCT sample ($i = 1, 2, \ldots, n$), and $(\boldsymbol{Z_i}, S_i)$ in the nonrandomized sample ($i = n + 1, \ldots, N$) are measured without error. We also assume well-specified marginal structural models and imputation models (25). Effect measure modifier coverage was assessed visually using predicted probabilities of RCT participation conditional on baseline covariates (Web Figure 1). The proportional hazards assumption in our marginal structural Cox model was evaluated by testing for a statistical interaction between efavirenz exposure and natural-log–transformed time.

An IOPW semiparametric Cox model was fitted to estimate a hazard ratio comparing time to suicidal thoughts/behaviors on an efavirenz-containing regimen with time to suicidal thoughts/behaviors on an efavirenz-free regimen (23, 29). Efron's method (30) was used to handle ties, and the baseline hazard was allowed to differ for each RCT. An IOPW Poisson regression with a natural-logarithm link was fitted to estimate the IR in the efavirenz-containing and efavirenz-free groups, and an IRD was estimated; RCT was handled as a stratum variable using the "svyglm" command in the R package "survey" (31). The probability of suicidal thoughts/behaviors was estimated using an IOPW Kaplan-Meier approach with MI.

Bootstrap procedures

To account for sampling variability from both the CNICS sample and RCT samples, bootstrap 95% confidence intervals for the transportability analyses were constructed by resampling with replacement from each of the 5 data sources. Each bootstrap resample was drawn to maintain the samples sizes of n for $S = 1$ and $N - n$ for $S = 0$. Within $S = 1$, we resampled from each RCT to maintain the sample size of each trial and for uniformity with our MI approach. This procedure is known as a stratified bootstrap. We generated $B = 10,000$ resampled data sets containing missingness. To handle missing data, we used MI within each bootstrap iteration and constructed nonparametric, percentile-based bootstrap 95% confidence intervals for ln(HR), ln(IR), and IRD (details are provided in Web Appendix 2). Proper (nominal frequentist) confidence interval coverage for this method (referred to as "Boot MI") has been demonstrated in simulation studies (32). For complete-case analyses, nonparametric bootstrap 95% confidence intervals were also computed using the same bootstrap procedure.

Multiple imputation

Missingness in 1 or more baseline covariates $\boldsymbol{Z_i}$ was rare in the RCT sample (0.3%) and was a minority of the nonrandomized sample (9%). To account for missing baseline covariate data, we employed a missing-at-random assumption, and MI was applied. Under a missing-at-random assumption, MI has been shown to be empirically unbiased in marginal structural models (33–35). We conducted MI using the "mice" package in R (36, 37), with $M = 30$ imputed data sets constructed using predictive mean matching and the random forest method to impute continuous and categorical variables, respectively (38, 39). The baseline

covariates $\boldsymbol{Z_i}$ were included in the imputation models (33), and data were imputed separately for the CNICS observational sample and for each RCT; that is, covariate distributions were not borrowed across the trials and cohort samples for imputation to avoid making the cohort distribution of $\boldsymbol{Z_i}$ artificially similar to the distribution of $\boldsymbol{Z_i}$ in the trials. Interactions and spline terms (i.e., transformations) were computed after imputation. The suicidal thoughts/behavior outcome indicator (δ_1) was included in the imputation model for the 0.3% missing trials data (40). Both MI and complete-case analyses were conducted.

RESULTS

We included 3,949 randomized trial and 8,291 observational cohort participants (Figure 1). In the observational cohort sample, 18% were women, and the median age was 38 years (range, 18–78; Table 1). The distributions of sex, age, viral load, and history of AIDS-defining illnesses were similar between the randomized and observational participants. Hispanic race/ethnicity was more frequent in the RCTs (22% vs. 14%), non-Hispanic White race/ethnicity was less frequent in the RCTs (39% vs. 44%), and non-Hispanic Black race/ethnicity was similarly represented in the RCTs and the cohort (36% vs. 37%). Prescription of antidepressants (13% vs. 19%), a history of injection drug use (10% vs. 16%), and hepatitis B or C virus infection (12% vs. 17%) were less common among randomized participants than in the observational cohort. Among the 3,949 randomized trial participants, 59% (n = 2,323) were assigned to receive efavirenz-containing ART and 41% (n = 1,626) were assigned to receive an efavirenz-free ART regimen. In the observational cohort (years 1999–2015), 45% of ART-naive participants initiated ART with an efavirenz-containing regimen. Overall, among trial participants, the median length of follow-up was 105 weeks (interquartile range, 56–144).

In the efavirenz-containing ART group, 2,323 trial participants contributed 4,345 person-years (PY) at risk and 39 composite events of suicidal ideation, suicide attempt, or death by suicide (Table 2). The estimated incidence of suicidal thoughts/behaviors in the trials was 9.0 (95% CI: 6.5, 12.3) per 1,000 PY in the efavirenz-containing group. In the efavirenz-free ART group, 1,626 trial participants contributed 3,352 PY and 13 events, and the estimated incidence of suicidal thoughts/behaviors was 3.9 (95% CI: 2.2, 6.7) per 1,000 PY. When the randomized trial results were transported to our target population, the estimated IRs of suicidal thoughts/behavior were higher in both groups, with 11.3 (95% CI: 7.0, 16.3) events per 1,000 PY in the efavirenz-containing group versus 5.9 (95% CI: 2.6, 10.0) in the efavirenz-free group (see Table 2, MI analysis).

The estimated IRD was 5.1 (95% CI: 1.6, 8.7) per 1,000 PY in the RCT analysis, indicating an increased incidence of suicidal thoughts/behaviors following initial ART with an efavirenz-containing regimen. The transported IRD estimate was 5.4 (95% CI: −0.4, 11.4) per 1,000 PY in the MI analysis and 6.0 (95% CI: −0.2, 12.6) per 1,000 PY in the complete-case analysis. In the trials-only analysis, the efavirenz-containing ART group had 2.3 times the estimated risk of suicidal thoughts/behaviors compared with the efavirenz-free ART group (HR = 2.3, 95% CI: 1.2, 4.4). When transported to our target population, the estimated hazard ratio was somewhat attenuated in both the MI analysis (HR = 1.8, 95% CI: 0.9, 4.4) and the complete-case analysis (HR = 1.9, 95% CI: 1.0, 4.8), as compared with the trials-only analysis (Table 2, Figure 2, Web Figure 3).

DISCUSSION

We evaluated the effect of efavirenz on suicidal thoughts/behaviors in a target population. Results from 4 pooled RCTs were transported to a target population of US adults living with HIV who were receiving care at a medical center and initiated ART between 1999 and 2015. Many participant characteristics from the RCTs were similar to those of the observational cohort. However, baseline antidepressant prescriptions and injection drug use are associated with suicidal thoughts/behavior and were more common in our target population than in the trials. Underrepresentation of higher-risk individuals in trials can result in lower risk estimates in the trials compared with a target population, which leads to nontransportability on a ratio or difference scale (24).

The presence of a causal effect of efavirenz use upon suicidal thoughts/behaviors—an uncommon but hazardous adverse event—was supported by our results, although the 95% confidence interval for the transported result crossed the null value. IRs of suicidal thoughts/behaviors were higher in the target population than in the randomized trials similarly for the efavirenz-containing and efavirenz-free groups. The IRD estimate was essentially unchanged when quantitatively transported from RCTs to the target population, yet the 95% confidence interval was wider. On the hazard ratio scale, higher IR estimates in the target population for both the efavirenz-containing and efavirenz-free groups resulted in a hazard ratio estimate that was 20% attenuated compared with the RCTs-only result, but it still reflected nearly a 2-fold increase in the hazard of suicidal thoughts/behaviors in our target population (HR = 1.8, 95% CI: 0.9, 4.4). The hazard ratio differed in the target population compared with RCTs, demonstrating the importance of conducting quantitative transportability analyses on each effect measure scale of interest.

Our findings corroborate the Strategic Timing of Antiretroviral Treatment (START) trial results, where people who initiated efavirenz in the immediate ART group had 3.3 times the risk of suicidal behavior as those in the deferred ART group (HR = 3.3, 95% CI: 1.1, 9.9) (7). In the START trial, the efavirenz-free group was deferred ART, whereas in the 4 ACTG RCTs the efavirenz-free group consisted of active comparator regimens, such as protease inhibitor-based treatment with nucleoside reverse transcriptase inhibitor backbones (6). Conversely, several large observational studies did not detect an association between efavirenz and suicidal thoughts/behaviors (8, 41). In previously reported CNICS observational outcomes analysis, without linkage to trials, Bengtson et al. (10) estimated a nearly null relationship between first-line efavirenz and suicidal thoughts (HR = 1.2,

Table 1. Baseline Characteristics of ART-Naive Adults Living With Human Immunodeficiency Virus and Starting Their First ART Regimen While Receiving Clinical Care, United States, 1999–2015

Characteristic	CNICS Cohort Target Sample (*n* = 8,291)		Randomized Clinical Trials Sample (*n* = 3,949)	
	No.	%	No.	%
Sex				
Female	1,451	18	716	18
Male	6,840	82	3,233	82
Race/ethnicity				
Hispanic	1,138	14	880	22
Non-Hispanic Black	3,010	37	1,408	36
Non-Hispanic White	3,621	44	1,544	39
Other[a]	446	5	112	3
Missing data[b]	76	0.9	5	0.1
Age, years				
Median[c]	38 (30–45)		38 (31–44)	
Range[d]	18–78		18–77	
Pretreatment CD4 cell count, cells/μL				
Median[c]	251 (95–394)		212 (76–324)	
Range[d]	0–1,670		0–1,336	
Missing data[b]	99	1.2	4	0.1
HIV-1 RNA load, $\log_{10}$ copies/mL				
Median[c]	4.78 (4.21–5.31)		4.72 (4.38–5.22)	
Range[d]	2.30–7.85		2.26–7.04	
Missing data[b]	0	0	1	0.03
History of AIDS diagnosis				
Yes	1,642	20	694	18
No	6,649	80	3,255	82
History of injection drug use				
Yes	1,306	16	380	10
No	6,860	84	3,569	90
Missing data[b]	125	1.5	0	0
Positive for HBV or HCV				
Yes	1,430	17	461	12
No	6,861	83	3,488	88
Prescription for antidepressants				
Yes	1,464	19	509	13
No	6,363	81	3,440	87
Missing data[b]	464	5.6	0	0

Abbreviations: AIDS, acquired immunodeficiency syndrome; ART, antiretroviral therapy; CNICS, Centers for AIDS Research Network of Integrated Clinical Systems; HBV, hepatitis B virus; HCV, hepatitis C virus; HIV-1, human immunodeficiency virus type 1.

[a] Asian, Native American (American Indian), Alaska Native, Asian/Pacific Islander, Pacific Islander, multiracial (>1 race), and those who reported their race as "other."

[b] Missing data were not included in the category percentage calculations.

[c] Values are presented as median (interquartile range).

[d] Values are presented as range (minimum–maximum).

95% CI: 0.7, 2.3), with PHQ-9 measurements introduced over calendar time such that 597 participants had evaluable data. The analysis herein used CNICS baseline covariate data and did not use CNICS PHQ-9 outcomes.

Table 2. Association of Suicidal Thoughts/Behaviors With Efavirenz-Containing Treatment Regimens Versus Efavirenz-Free Regimens Among Adults Living With Human Immunodeficiency Virus and Receiving Clinical Care, United States, 1999–2015[a]

Analysis Type and Treatment Group	Crude No. of Events	No. of PY	IR (No. of Events/ 1,000 PY)	95% CI	IRD (No. of Events/ 1,000 PY)	95% CI	HR	95% CI
Randomized trials only[b]								
EFV-containing group	39	4,345	9.0	6.5, 12.3	5.1	1.6, 8.7	2.3	1.2, 4.4
EFV-free group	13	3,352	3.9	2.2, 6.7	0		1.0	Referent
Transported from trials to cohort[c]								
Multiple imputation								
EFV-containing group			11.3	7.0, 16.3	5.4	−0.4, 11.4	1.8	0.9, 4.4
EFV-free group			5.9	2.6, 10.0	0		1.0	Referent
Complete-case analysis[d]								
EFV-containing group			11.8	7.3, 17.5	6.0	−0.2, 12.6	1.9	1.0, 4.8
EFV-free group			5.9	2.5, 10.0	0		1.0	Referent

Abbreviations: CI, confidence interval; EFV, efavirenz; HIV, human immunodeficiency virus; HR, hazard ratio; IR, incidence rate; IRD, incidence rate difference; PY, person-years.

[a] There were 2,323 participants in the EFV-containing group and 1,626 participants in the EFV-free group. The EFV-free group was the referent in all comparisons.

[b] For analysis of randomized trials only, Wald-based 95% CIs were calculated for the IR and HR and a nonparametric bootstrap 95% CI was calculated for the IRD; the Cox model proportional hazards assumption was not violated (Wald $P = 0.4$).

[c] Inverse odds weights were applied to estimate the transported IR, IRD, and HR; 95% CIs were constructed using a nonparametric bootstrap. Poisson and Cox models were fitted with AIDS Clinical Trials Group study data included as a stratification variable (to allow a separate baseline hazard function for each of the 4 trials).

[d] In the complete-case analysis, the inverse odds weights had mean $\hat{E}(W_i) = 0.99$ (standard deviation, 0.88), and W_i ranged from 0.15 to 13 (Web Figure 2); the proportional hazards assumption was not violated (bootstrap $P = 0.2$).

For the RCTs, recruitment strategies, particularly pertaining to mental health, substance use, and compliance, may have reduced external validity with respect to estimating the impact of efavirenz on suicidal thoughts/behaviors in a target population of interest to US prescribers. In contrast, the internal validity of observational data analyses may be reduced by unmeasured channeling bias, which may result from prescribers using patients' mental health information and in-person behaviors to inform treatment decisions, such as prescription of efavirenz, without the underlying treatment decision mechanism being well-captured by measured covariates. Mental health history may be unmeasured, due to lack of mental health care or stigma, or may be unreliable due to potential under- or overdiagnosing (i.e., mental health status measured with error). Transportability analysis, however, combines the internal validity strength of randomized trials with the external validity strength of observational data to estimate externally valid causal effects.

Unmeasured confounding in nonrandomized samples and lack of generalizability in trials are not the only possible reasons why the hazard ratio results differed substantially among prior studies. Measurement of suicidal thoughts/behavior outcomes is challenging, and ascertainment of suicidal outcomes was inherently different between the ACTG trials and the CNICS observational cohort (6, 10). In each of the 4 ACTG trials, suicidal thoughts/behaviors were measured using adverse event reports—a process involving study staff and participant reporting, or reporting by proxy (family contact or autopsy results) for death by suicide. Proactive assessment of suicidal thoughts was not systematically conducted in these trials—reported suicidal thoughts/behaviors were those which rose to the level of clinical attention or study adverse reporting. In the CNICS cohort, passive and active suicidal thoughts were collected systematically as patient-reported outcomes using the PHQ-9 questionnaire (10, 11), whereas attempted suicide and death by suicide were not ascertained in the CNICS. Importantly, participants need to be alive and receiving care to complete a PHQ-9.

We successfully harmonized and conditioned upon 9 patient baseline characteristics. Still, antidepressant medication use likely only captures a fraction of all depression cases (42, 43), and additional mental health covariates (e.g., full history of a psychiatric diagnosis, anxiety medications) and substance use were not measured or could not be harmonized. This is a drawback of retrospectively combining data sources. In the future, efforts to establish target validity will require a priori planning for integration of data from multiple sources (44). Unmeasured participant characteristics that have a causal relationship with outcome risk and that differed between our trials and cohort populations may have compromised the external validity of our analysis (25).

Transportability analysis using inverse odds weighting involves an inherent loss in precision through projection onto a new population, as observed here (16). Alternative methods exist for transportability or generalizability

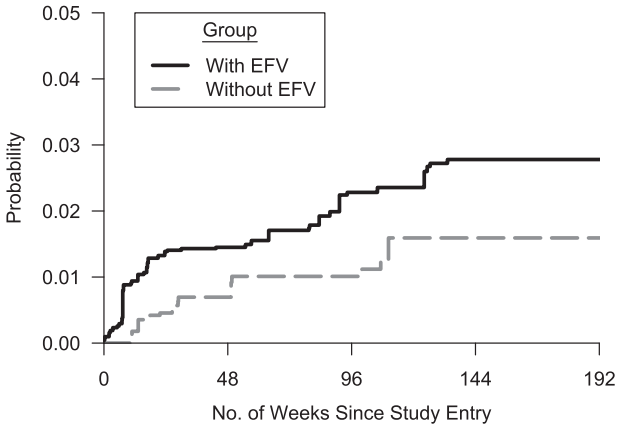

Figure 2. Cumulative probability of suicidal thoughts or behaviors in a target population of US adults living with human immunodeficiency virus who initiated first-line antiretroviral therapy between 1999 and 2015. The probability of suicidal thoughts/behaviors was estimated as 1 minus the inverse odds–weighted Kaplan-Meier estimate. The *y*-axis is truncated at 5%. Thirty imputed data sets were constructed, and the resulting 30 weighted Kaplan-Meier estimates were averaged at each time point. Crude number of events and number of participants at risk (*n*)—for the "with EFV" group (39 suicidal thoughts/behavior events): week 0, *n* = 2,323; week 48, *n* = 1,786; week 96, *n* = 1,238; week 144, *n* = 530; week 192, *n* = 59; for the "without EFV" group (13 events): week 0, *n* = 1,626; week 48, *n* = 1,326; week 96, *n* = 1,019; week 144, *n* = 491; week 192, *n* = 71. EFV, efavirenz.

analysis, including augmented estimators (12, 15, 45, 46). Yet, with a rare outcome, covariate-rich outcome modeling of suicidal thoughts/behaviors was infeasible. To use an augmented estimator or G-formula, we anticipate that reductions would be needed in the set of baseline covariates (and such reductions may violate the assumption that potential outcomes are conditionally exchangeable between the RCT and nonrandomized samples). In contrast, application of IOPW allowed us to condition upon 9 measured covariates and their 2-way interactions, particularly because both the RCT and observational samples were large.

We provide additional evidence about the safety profile of efavirenz and the feasibility of applied transportability analyses. Major strengths of our study include randomization and accurate measurement of efavirenz exposure, long-term follow-up for adverse events, adjustment for 9 measured baseline covariates, and handling of missing covariate data using MI. A practical drawback of the applied methodology was the large computational time needed to conduct MI with bootstrap resamples using the "boot MI" approach; we used job arrays on a Linux cluster to navigate this bottleneck, and use of a different approach called "MI boot" by Schomaker et al. (32) would also reduce computational time. We recommend future work to construct a closed-form variance estimator for IOPW, as exists for inverse probability of sampling weights (47).

For adults living with HIV and their health-care providers, the effect of efavirenz on suicidal thoughts/behaviors has remained an important question, with conflicting findings between randomized and nonrandomized evidence. In this analysis, when the effect of initiating efavirenz on combined suicidal thoughts and behaviors was transported from RCTs to a target population of adults engaged in HIV care, we observed evidence that was mostly consistent with an increase in the risk of suicidal thoughts/behaviors with efavirenz initiation. By combining participant-level randomized trials exposure and outcomes data with observational data on preexposure participant characteristics and handling missing covariate data with MI, we addressed internal and external validity in an effort to move towards target validity (48). When transportability analysis is feasible, one can formally quantify the extent to which population differences impact study results and provide researchers with a method for estimation of externally valid causal effects in a specific target population.

ACKNOWLEDGMENTS

Author affiliations: Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States (Katie R. Mollan, Brian W. Pence, Jessie K. Edwards, Daniel Westreich, Stephen R. Cole); Center for AIDS Research, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States (Katie R. Mollan, Ann Marie K. Weideman); Department of Statistics, College of Sciences, North Carolina State University (Steven Xu); Department of Medicine, School of Medicine, University of California, San Diego, La Jolla, California, United States (W. Christopher Mathews); Department of Psychiatry, Massachusetts General Hospital, Harvard Medical School, and Fenway Institute, Boston, Massachusetts, United States (Conall O'Cleirigh); Division of Allergy and Infectious Diseases, Department of Medicine, School of Medicine, University of Washington, Seattle, Washington, United States (Heidi M. Crane, Ann C. Collier); Division of Infectious Disease, School of Medicine, University of Alabama at Birmingham, Birmingham, Alabama, United States (Ellen F. Eaton); Department of Biostatistics, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, United States (Ann Marie K. Weideman); Center for Biostatistics in AIDS Research, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, United States (Camlin Tierney); Department of Epidemiology, School of Public Health, Brown University, Providence, Rhode Island, United States (Angela M. Bengtson).

This work was supported by the National Institute of Allergy and Infectious Disease (grants P30AI050410, R25AI140495, AI069481, U01AI068636, K01AI125087) and the National Institute of Mental Health (grants R01MH100970 and R00MH112413). Data were provided by the AIDS Clinical Trials Group (grants AI38858 and AI68634) and the Centers for AIDS Research Network of Integrated Clinical Systems (grant R24AI067039).

Upon approval from both the AIDS Clinical Trials Group and the Centers for AIDS Research Network of Integrated Clinical Systems, deidentified participant-level data may be shared given that the investigator who requests the data has approval from an institutional review board, independent ethics committee, or research ethics board, as applicable, and executes the required data-use/-sharing agreement(s). Software code for the transportability analysis is available on GitHub (49).

We thank Drs. Michael G. Hudgens, Joseph J. Eron, and Bonnie Shook-Sa for their helpful advice and Dr. Mark Reed and the University of North Carolina Research Computing group for providing Linux computational resources/support.

K.R.M. has received research support from grants or contracts awarded to the University of North Carolina from Ridgeback Biotherapeutics (Miami, Florida) and Gilead Sciences, Inc. (Foster City, California). The University of Washington has received a grant from ViiV Healthcare (London, United Kingdom) on behalf of H.M.C. E.F.E. was the recipient of a Bristol-Myers Squibb (New York, New York) virology fellows research grant. Grants from Merck & Co., Inc. (Kenilworth, New Jersey) have been received by the University of Alabama at Birmingham on behalf of E.F.E. The remaining authors have no conflicts of interest to disclose.

REFERENCES

1. Bengtson AM, Pence BW, Eaton EF, et al. Patterns of efavirenz use as first-line antiretroviral therapy in the United States: 1999–2015. *Antivir Ther*. 2018;23(4):363–372.
2. World Health Organization. Update of recommendations on first- and second-line antiretroviral regimens. https://apps.who.int/iris/bitstream/handle/10665/325892/WHO-CDS-HIV-19.15-eng.pdf. Published 2019. Accessed February 8, 2020.
3. Raffi F, Pozniak AL, Wainberg MA. Has the time come to abandon efavirenz for first-line antiretroviral therapy? *J Antimicrob Chemother*. 2014;69(7):1742–1747.
4. Kryst J, Kawalec P, Pilc A. Efavirenz-based regimens in antiretroviral-naive HIV-infected patients: a systematic review and meta-analysis of randomized controlled trials. *PLoS One*. 2015;10(5):e0124279.
5. Bristol-Myers Squibb. Efavirenz [package insert]. Princeton, NJ: Bristol-Myers Squibb; 2019.
6. Mollan KR, Smurzynski M, Eron JJ, et al. Association between efavirenz as initial therapy for HIV-1 infection and increased risk for suicidal ideation or attempted or completed suicide: an analysis of trial data. *Ann Intern Med*. 2014; 161(1):1–10.
7. Arenas-Pinto A, Grund B, Sharma S, et al. Risk of suicidal behavior with use of efavirenz: results from the strategic timing of antiretroviral treatment trial. *Clin Infect Dis*. 2018; 67(3):420–429.
8. Napoli AA, Wood JJ, Coumbis JJ, et al. No evident association between efavirenz use and suicidality was identified from a disproportionality analysis using the FAERS database. *J Int AIDS Soc*. 2014;17(1):19214.
9. Smith C, Ryom L, d'Arminio Monforte A, et al. Lack of association between use of efavirenz and death from suicide: evidence from the D:A:D study. *J Int AIDS Soc*. 2014; 17(4 suppl 3):19512.
10. Bengtson AM, Pence BW, Mollan KR, et al. The relationship between efavirenz as initial antiretroviral therapy and suicidal thoughts among HIV-infected adults in routine care. *J Acquir Immune Defic Syndr*. 2017;76(4):402–408.
11. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001; 16(9):606–613.
12. Lesko CR, Buchanan AL, Westreich D, et al. Generalizing study results: a potential outcomes perspective. *Epidemiology*. 2017;28(4):553–561.
13. Westreich D, Edwards JK, Lesko CR, et al. Transportability of trial results using inverse odds of sampling weights. *Am J Epidemiol*. 2017;186(8):1010–1014.
14. Dahabreh IJ, Haneuse SJPA, Robins JM, et al. Study designs for extending causal inferences from a randomized trial to a target population [published online ahead of print December 16, 2020]. *Am J Epidemiol*. (doi:10.1093/aje/kwaa270).
15. Dahabreh IJ, Robertson SE, Hernán MA. On the relation between G-formula and inverse probability weighting estimators for generalizing trial results. *Epidemiology*. 2019; 30(6):807–812.
16. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: the ACTG 320 trial. *Am J Epidemiol*. 2010;172(1):107–115.
17. Stuart EA, Bradshaw CP, Leaf PJ. Assessing the generalizability of randomized trial results to target populations. *Prev Sci*. 2015;16(3):475–485.
18. Wang C, Mollan KR, Hudgens MG, et al. Generalisability of an online randomised controlled trial: an empirical analysis. *J Epidemiol Community Health*. 2018;72(2): 173–178.
19. Kitahata MM, Rodriguez B, Haubrich R, et al. Cohort profile: the Centers for AIDS Research Network of Integrated Clinical Systems. *Int J Epidemiol*. 2008;37(5):948–955.
20. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019. https://www.R-project.org/. Accessed July 10, 2021.
21. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 2000;56(3):779–788.
22. Buchanan AL, Hudgens MG, Cole SR, et al. Worth the weight: using inverse probability weighted Cox models in AIDS research. *AIDS Res Hum Retroviruses*. 2014;30(12): 1170–1177.
23. Lesko CR, Cole SR, Hall HI, et al. The effect of antiretroviral therapy on all-cause mortality, generalized to persons diagnosed with HIV in the USA, 2009–11. *Int J Epidemiol*. 2016;45(1):140–150.
24. Westreich D. *Epidemiology by Design*. New York, NY: Oxford University Press; 2019.
25. Nguyen TQ, Ackerman B, Schmid I, et al. Sensitivity analyses for effect modifiers not observed in the target population when generalizing treatment effects from a randomized controlled trial: assumptions, models, effect scales, data scenarios, and implementation details. *PLoS One*. 2018;13(12):1–17.
26. Howe CJ, Cole SR, Westreich DJ, et al. Splines for trend analysis and continuous confounder control. *Epidemiology*. 2011;22(6):874–875.

27. Poole C. On the origin of risk relativism. *Epidemiology*. 2010;21(1):3–9.
28. VanderWeele TJ, Hernán MA. Causal inference under multiple versions of treatment. *J Causal Inference*. 2013;1(1): 1–20.
29. Hernán 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(5): 561–570.
30. Efron B. The efficiency of Cox's likelihood function for censored data. *J Am Stat Assoc*. 1977;72(359):557–565.
31. Lumley T. Analysis of complex survey samples. *J Stat Softw*. 2004;9(8):1–19.
32. Schomaker M, Heumann C. Bootstrap inference when using multiple imputation. *Stat Med*. 2018;37(14):2252–2266.
33. Moodie EEM, Delaney JAC, Lefebvre G, et al. Missing confounding data in marginal structural models: a comparison of inverse probability weighting and multiple imputation. *Int J Biostat*. 2008;4(1):Article 13.
34. Leyrat C, Seaman SR, White IR, et al. Propensity score analysis with partially observed covariates: how should multiple imputation be used? *Stat Methods Med Res*. 2019; 28(1):3–19.
35. Leyrat C, Carpenter JR, Bailly S, et al. Common methods for handling missing data in marginal structural models: what works and why. *Am J Epidemiol*. 2021;190(4):663–672.
36. White IR, Royston P, Wood AM. Multiple imputation using chained equations: issues and guidance for practice. *Stat Med*. 2011;30(4):377–399.
37. van Buuren S, Groothuis-Oudshoorn K. mice: Multivariate imputation by chained equations in R. *J Stat Softw*. 2011; 45(3):1–67.
38. Morris TP, White IR, Royston P. Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Med Res Methodol*. 2014;14:Article 75.
39. Shah AD, Bartlett JW, Carpenter J, et al. Comparison of random forest and parametric imputation models for imputing missing data using MICE: a CALIBER study. *Am J Epidemiol*. 2014;179(6):764–774.
40. Moons KGM, Donders RART, Stijnen T, et al. Using the outcome for imputation of missing predictor values was preferred. *J Clin Epidemiol*. 2006;59(10):1092–1101.
41. Nkhoma ET, Coumbis J, Farr AM, et al. No evidence of an association between efavirenz exposure and suicidality among HIV patients initiating antiretroviral therapy in a retrospective cohort study of real world data. *Medicine (Baltimore)*. 2016;95(3):1–8.
42. Asch SM, Kilbourne AM, Gifford AL, et al. Underdiagnosis of depression in HIV: who are we missing? *J Gen Intern Med*. 2003;18(6):450–460.
43. Pence BW, O'Donnell JK, Gaynes BN. Falling through the cracks: the gaps between depression prevalence, diagnosis, treatment, and response in HIV care. *AIDS*. 2012;26(5): 656–658.
44. Bareinboim E, Pearl J. Causal inference and the data-fusion problem. *Proc Natl Acad Sci U S A*. 2016;113(27): 7345–7352.
45. Dahabreh IJ, Robertson SE, Tchetgen EJ, et al. Generalizing causal inferences from individuals in randomized trials to all trial-eligible individuals. *Biometrics*. 2019;75(2):685–694.
46. Dahabreh IJ, Robertson SE, Steingrimsson JA, et al. Extending inferences from a randomized trial to a new target population. *Stat Med*. 2020;39(14):1999–2014.
47. Buchanan AL, Hudgens MG, Cole SR, et al. Generalizing evidence from randomized trials using inverse probability of sampling weights. *J R Stat Soc Ser A Stat Soc*. 2018;181(4): 1193–1209.
48. Westreich D, Edwards JK, Lesko CR, et al. Target validity and the hierarchy of study designs. *Am J Epidemiol*. 2019; 188(2):438–443.
49. University of North Carolina at Chapel Hill Center for AIDS Research (CFAR) Biostatistics Core. efv-transportability. https://github.com/unc-cfar-bios/efv-transportability. Published April 30, 2021. Accessed July 10, 2021.