Web Material

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 (CNICS) and AIDS Clinical Trials Group (ACTG)

Table of Contents

- Web Table 1. Follow-up time and drop-out in the RCTs by study arm
- Web Appendix 1. Sensitivity analysis applying inverse probability of censoring weights in a trials-only analysis
- Web Figure 1. Predicted probability of RCT participation conditional on baseline covariates
- Web Appendix 2. Algorithm for Boot MI
- Web Figure 2. Distribution of inverse odds weights
- Web Figure 3. Sampling distributions for incidence rate difference and log hazard ratio

Web Table 1. Follow-up time and drop-out in the RCTs by study arm[a]

		Median follow-up (Q1, Q3)		Study drop-out	
Study	**Randomization ratio for efavirenz**	**Efavirenz-containing**	**Efavirenz-free**	**Efavirenz-containing**	**Efavirenz-free**
A5095 (*n* = 1,147)	2:1	48 (29, 72)	48 (28, 72)	10%	12%
A5142 (*n* = 735)	2:1	112 (96, 129)	112 (96, 129)	18%	21%
A5175 (*n* = 210)	2:1	139 (95, 150)	140 (96, 151)	23%	23%
A5202 (*n* = 1,857)	1:1	137 (107, 168)	138 (106, 169)	19%	20%

[a] Additional details are published in Mollan (2014, *Annals of Internal Medicine,* Table 1).[1]

Web Appendix 1. **Sensitivity analysis applying inverse probability of censoring weights in a trials-only analysis**

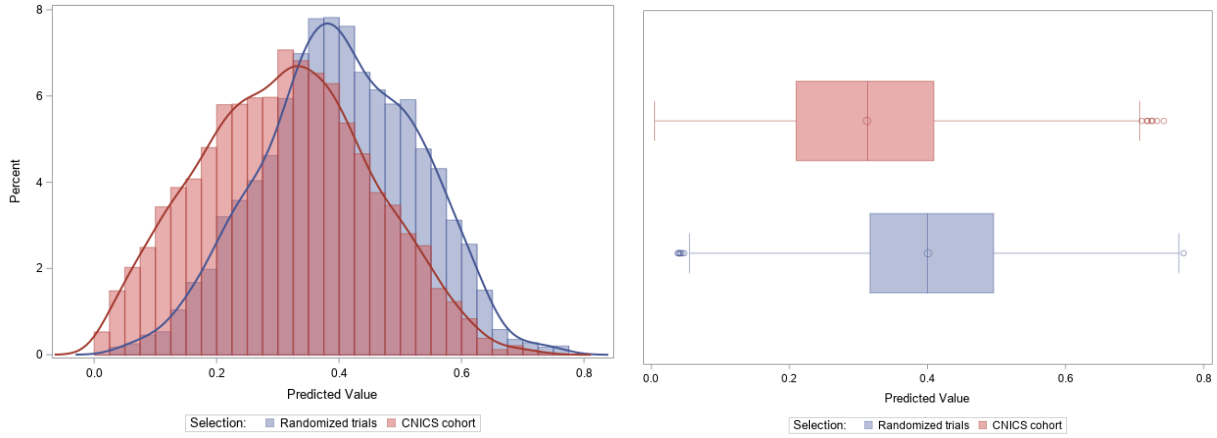
We conducted a trials-only *post hoc* sensitivity analysis where RCT dropout was accounted for using inverse probability of censoring weights (IPCW). This sensitivity analysis was conducted to assess our non-informative censoring assumption, which pertains to the internal validity of the RCT results. Let A indicate exposure to an efavirenz-containing regimen versus efavirenz-free regimen. Let $\boldsymbol{L}$ represent a vector of ten baseline covariates that may be associated with premature study dropout. For the IPCW analysis, we also included the ACTG study in the vector $\boldsymbol{L}$ plus the nine baseline covariates used for the transportability analysis, with the same functional forms as described in Methods section. Post-baseline (time-varying) covariates were not available in our data transfer and there could also be unmeasured covariates associated with premature study dropout that we were not able to account for here.

Let D be an indicator of event-free, premature dropout: $D = 1$ for dropout, and $D = 0$ if the planned or possible study follow-up was completed (or a suicidal thoughts or behavior event was observed). Participants who died during study follow-up were coded as $D = 0$, and administrative censoring was coded as $D = 0$. We constructed censoring weights: $IPCW = \Pr(D = 0|A)/Pr(D = 0|\boldsymbol{L}, A)$ which remove the causal arrows on a directed acyclic graph from $\boldsymbol{L}$ into D in order to alleviate selection bias from restricting analyses to RCT participants with observed follow-up visits. Pooled logistic regression was used to estimate the censoring weights, with follow-up time binned into 4-week intervals and fit using a restricted quadratic spline. Censoring weights were applied in a weighted Cox model using a robust variance estimator, and the unspecified baseline hazard was allowed to differ for each RCT.

We included n=3,939 participants in the trials-only analysis; 10 (0.3% of 3,949) participants who were missing some element of the vector $\boldsymbol{L}$ were excluded. Overall in the trials, 17% of participants had event-free dropout. In our unweighted, trials-only analysis the estimated hazard ratio was HR=2.3 (95% CI: 1.2, 4.4), n=3,939. In our IPCW sensitivity analysis, the mean of the weights was 0.997 (SD=0.103) and the estimated HR was 2.1 (95% CI: 1.1, 4.0).

We maintained our *a priori* assumption of non-informative censoring in the main transportability analyses. Importantly, there could still be unmeasured factors not accounted for by $\boldsymbol{L}$ that led to premature study dropout.

Web Figure 1. Predicted probability of RCT participation conditional on baseline covariates



$\widehat{\Pr}(S_i = 1|\boldsymbol{Z_i})$ is plotted for $S = 1$ (RCTs) in blue and $S = 0$ (CNICS cohort, non-randomized) in red to visually assess the effect measure modifier (EMM) coverage assumption, i.e., $Pr(S_i = s|\boldsymbol{Z_i} = \boldsymbol{z}) > 0$ for $s = 0,1$ if $Pr(\boldsymbol{Z_i} = \boldsymbol{z}) > 0$ in the target population. Predicted probabilities (propensity scores) are the propensity of being selected into the RCT sample conditional upon baseline covariates $\boldsymbol{Z_i}$.

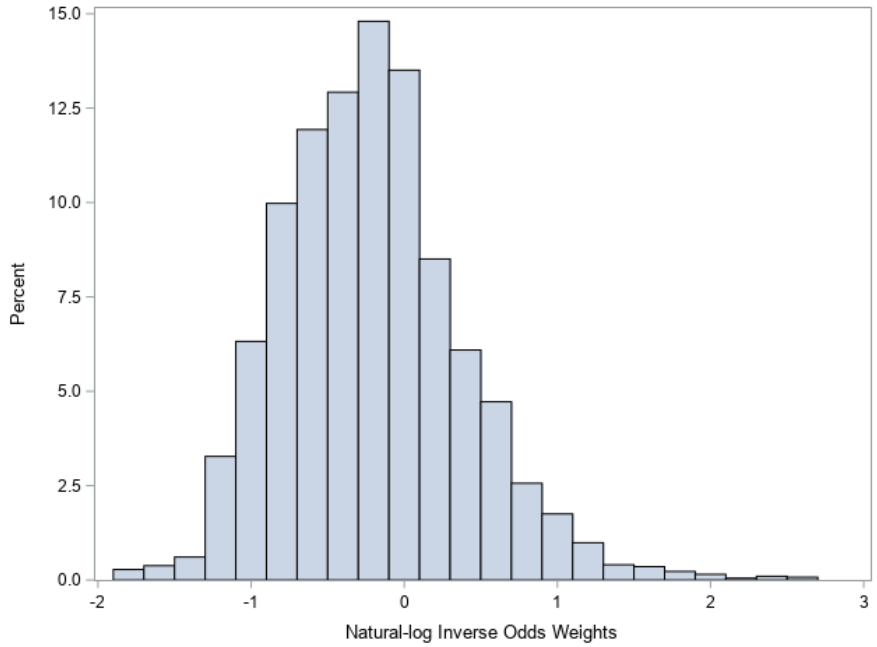
Web Appendix 2. Algorithm for Boot MI

We assumed the CNICS observational cohort is a representative sample of our arbitrarily large target population (US adults living with HIV who were engaged in care at a medical center and initiated ART between 1999–2015), and that there was no overlap between the CNICS sample and the RCT samples. The size of our US target population is not enumerated.

To handle missing data, we used multiple imputation (MI) on the observed data set and within each bootstrap resample. Let $\mathcal{D}$ denote the concatenated data set, containing observations from CNICS and the four RCTs, from which $B = 10{,}000$ bootstrap resamples $\mathcal{D}_1^*, \ldots, \mathcal{D}_B^*$ including observed and missing data were drawn. At each bootstrap iteration, simple random samples with replacement were drawn independently from CNICS and each RCT before concatenating the CNICS and RCT resamples to form $\mathcal{D}_1^*, \ldots, \mathcal{D}_B^*$. Multiple imputation (MI) was applied to each bootstrap resample (see Methods for details), generating $M = 30$ imputed data sets $\mathcal{D}_{b,1}^*, \ldots, \mathcal{D}_{b,M}^*$ for each resample $\mathcal{D}_b^*$, $b = 1, \ldots, B$.

Point estimates for the $\ln(HR)$, $\ln(IR)$, and IRD were calculated after applying MI to $\mathcal{D}$ (the observed CNICS and RCTS data). To construct bootstrap 95% CIs we employed the following algorithm referred to as "Boot MI" by Schomaker and colleagues (2018): From each imputed resample $\mathcal{D}_{b,m}^*$, $m = 1, \ldots, M$, we obtained IOPW estimates for the $\ln(HR)$, $\ln(IR)$, and IRD. Within each bootstrap iteration, the $M = 30$ MI-based estimates were pooled using Rubin's rule for each estimand. The resulting $B = 10{,}000$ estimates comprise the sampling distribution (plotted in Figure 3) and were used to construct 95% CIs for $\ln(HR)$, $\ln(IR)$, and IRD using the percentile method (i.e., the 2.5th and 97.5th percentiles of the sampling distribution). Rubin's MI rule for variance was not used herein given our reliance on the non-parametric, percentile-based bootstrap CI with MI conducted within each bootstrap iteration. We used a large number of bootstrap resamples ($B = 10{,}000$) to ensure stability of the 2.5th and 97.5th percentiles from the tails of each sampling distribution.

Web Figure 2. Distribution of inverse odds weights

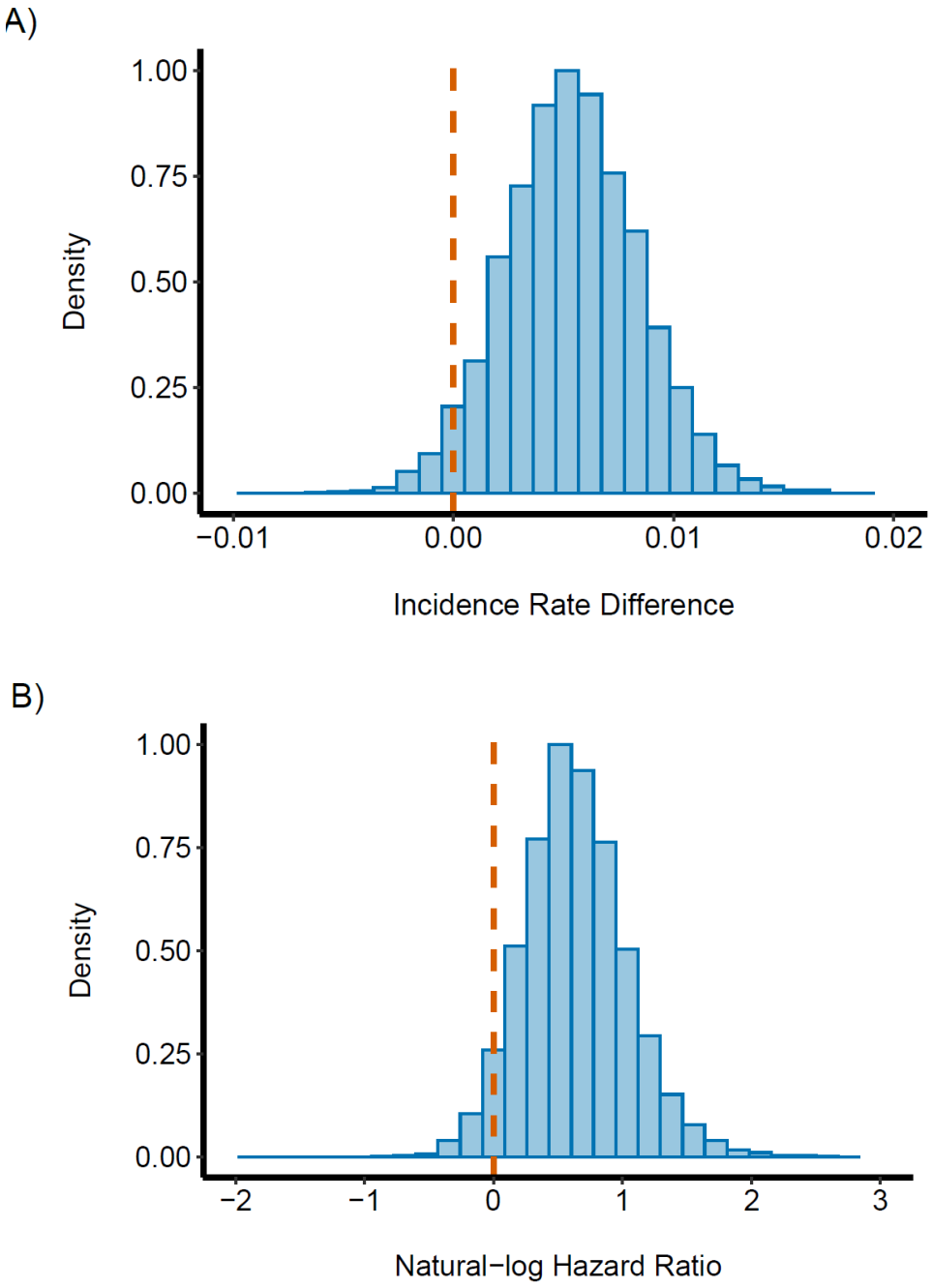


Summary statistics for inverse odds weights from complete-case analysis

Sample	*n* =	Mean	Standard Deviation	Minimum	Maximum
$S = 0$	7,555	0	0	0	0
$S = 1$	3,939	0.99	0.88	0.15	13.11

Web Figure 3. Sampling distributions for incidence rate difference and log hazard ratio

From the combined data set of trial and cohort participants, 10,000 resampled data sets containing missingness were generated to construct a nonparametric bootstrap 95% CI for each estimator. Each of 10,000 estimates was calculated from 30 multiple imputation data sets using Rubin's rule. Panel A shows the incidence rate difference. Panel B shows the natural-log hazard ratio.



Web Reference

1. 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. doi:10.7326/M14-0293