Supporting Information to "Generalizing the intention-to-treat effect of an active control from historical placebo-controlled trials: A case study of the efficacy of daily oral TDF/FTC in the HPTN 084 study"

Web Appendix A. Proof of Theorem 1

Proof. Consider the observed data functional

$$\psi = \int \delta\left(\mathbf{x}; h_1\right) \delta^D\left(\mathbf{x}; h_2\right) dF(\mathbf{x} \mid S = t)$$

where

$$\delta(\mathbf{X}; s) = \frac{\delta^Y(\mathbf{X}; s)}{\delta^D(\mathbf{X}; s)}.$$

To find the efficient influence function for ψ, we need to identify a random variable G with mean zero that satisfies

$$\left.\frac{\partial \psi_\tau}{\partial \tau}\right|_{\tau=0} = \mathbb{E}\{Gm(\mathbf{O}; \tau)\}\big|_{\tau=0}$$

where $\mathbf{O} = (\mathbf{X}, Z, D, Y, S)$, $m(\mathbf{O}) = \partial \log f(\mathbf{O}; \tau)/\left.\partial\tau\right|_{\tau=0}$ is a score function and ψ_τ is the target parameter under a regular parametric submodel indexed by τ. By the product rule,

$$\begin{aligned}
\left.\frac{\partial \psi_\tau}{\partial \tau}\right|_{\tau=0} =& \int \delta\left(\mathbf{x}; h_1\right) \delta_\tau^D\left(\mathbf{x}; h_2\right) \frac{\partial}{\partial \tau} dF_\tau(\mathbf{x} \mid S = t)\Big|_{\tau=0} \\
&+ \int \delta\left(\mathbf{x}; h_1\right) \frac{\partial}{\partial \tau} \delta_\tau^D\left(\mathbf{x}; h_2\right)\Big|_{\tau=0} dF(\mathbf{x} \mid S = t) \\
&+ \int \frac{\partial}{\partial \tau} \delta_\tau\left(\mathbf{x}; h_1\right)\Big|_{\tau=0} \delta^D\left(\mathbf{x}; h_2\right) dF(\mathbf{x} \mid S = t) \\
=& (i) + (ii) + (iii).
\end{aligned}$$

Turning first to (i),

$$
\begin{aligned}
(i) &= \mathbb{E}\left\{\delta\left(\mathbf{X};h_1\right)\delta^D\left(\mathbf{X};h_2\right)m(\mathbf{X}\mid S=t)\mid S=t\right\}\\
&= \mathbb{E}\left\{I(S=t)\delta\left(\mathbf{X};h_1\right)\delta^D\left(\mathbf{X};h_2\right)m(\mathbf{X}\mid S)\right\}/\kappa\\
&= \mathbb{E}\left(I(S=t)\delta\left(\mathbf{X};h_1\right)\delta^D\left(\mathbf{X};h_2\right)\left[\mathbb{E}\{m(\mathbf{O})\mid \mathbf{X},S\}-\mathbb{E}\{m(\mathbf{O})\mid S\}\right]\right)/\kappa\\
&= \mathbb{E}\left[I(S=t)\left\{\delta\left(\mathbf{X};h_1\right)\delta^D\left(\mathbf{X};h_2\right)-\psi\right\}m(\mathbf{O})\right]/\kappa.
\end{aligned}
$$

Then,

$$
\begin{aligned}
(ii) &= \mathbb{E}\left[f(S=t\mid\mathbf{X})\delta\left(\mathbf{X};h_1\right)\frac{\partial}{\partial t}\delta_t^D\left(\mathbf{X};h_2\right)\Big|_{t=0}\right]/\kappa\\
&= \mathbb{E}\left[\frac{(2Z-1)I\left(S=h_2\right)}{f\left(Z\mid\mathbf{X},S=h_2\right)}\frac{f(S=t\mid\mathbf{X})}{f\left(S=h_2\mid\mathbf{X}\right)}\delta\left(\mathbf{X};h_1\right)\{D-\mathbb{E}(D\mid\mathbf{X},S,Z)\}m(\mathbf{O})\right]/\kappa,
\end{aligned}
$$

where we apply the following results obtained e.g. in the proof of Theorem 5 in Wang and Tchetgen Tchetgen (2018):

$$
\frac{\partial}{\partial\tau}\mathbb{E}_\tau\left(D\mid\mathbf{X},S=h_2,Z=z\right)\Big|_{\tau=0} = \mathbb{E}\left[\left\{D-\mathbb{E}\left(D\mid\mathbf{X},S=h_2,Z=z,\right)\right\}m\left(D,S=h_2,Z=z\mid\mathbf{X}\right)\right],
$$

$$
\frac{\partial}{\partial\tau}\delta_\tau^D\left(\mathbf{X};h_2\right)\Big|_{\tau=0} = \mathbb{E}\left[\frac{(2Z-1)}{f(Z\mid\mathbf{X},S=h_2)}\frac{I\left(S=h_2\right)}{f(S=h_2\mid\mathbf{X})}\{D-\mathbb{E}(D\mid\mathbf{X},S,Z)\}m(\mathbf{O})\mid\mathbf{X}\right].
$$

Finally, again applying results from Wang and Tchetgen Tchetgen (2018),

$$
\begin{aligned}
(iii) &= \mathbb{E}\left[f(S=t\mid\mathbf{X})\left\{\frac{\frac{\partial}{\partial\tau}\delta_\tau^Y\left(\mathbf{X};h_1\right)\Big|_{\tau=0}\delta^D\left(\mathbf{X};h_1\right)-\frac{\partial}{\partial\tau}\delta_\tau^D\left(\mathbf{X};h_1\right)\Big|_{\tau=0}\delta^Y\left(\mathbf{X};h_1\right)}{\delta^D\left(\mathbf{X};h_1\right)^2}\right\}\delta^D\left(\mathbf{X};h_2\right)\right]/\kappa\\
&= \mathbb{E}\Bigg[\frac{(2Z-1)\,I\left(S=h_1\right)}{f\left(Z\mid\mathbf{X},S=h_1\right)}\frac{f\left(S=t\mid\mathbf{X}\right)}{f\left(S=h_1\mid\mathbf{X}\right)}\frac{\delta^D\left(\mathbf{X};h_2\right)}{\delta^D\left(\mathbf{X};h_1\right)}\\
&\qquad\times\left[\{Y-\mathbb{E}(Y\mid\mathbf{X},S,Z)\}-\{D-\mathbb{E}(D\mid\mathbf{X},S,Z)\}\delta\left(\mathbf{X};h_1\right)\right]m(\mathbf{O})\Bigg]/\kappa\\
&= \mathbb{E}\Bigg[\frac{(2Z-1)I\left(S=h_1\right)}{f\left(Z\mid\mathbf{X},S=h_1\right)}\frac{f(S=t\mid\mathbf{X})}{f\left(S=h_1\mid\mathbf{X}\right)}\frac{\delta^D\left(\mathbf{X};h_2\right)}{\delta^D\left(\mathbf{X};h_1\right)}\\
&\qquad\times\left[Y-\mu_0^Y\left(\mathbf{X};h_1\right)-\left\{D-\mu_0^D\left(\mathbf{X};h_1\right)\right\}\delta\left(\mathbf{X};h_1\right)\right]m(\mathbf{O})\Bigg]/\kappa
\end{aligned}
$$

Therefore, the efficient influence function is equal to

$$\frac{1}{\kappa}\frac{(2Z-1)I\left(S=h_1\right)}{f\left(Z\mid\mathbf{X},S=h_1\right)}\frac{f(S=t\mid\mathbf{X})}{f\left(S=h_1\mid\mathbf{X}\right)}\frac{\delta^D\left(\mathbf{X};h_2\right)}{\delta^D\left(\mathbf{X};h_1\right)}\left[Y-\mu_0^Y\left(\mathbf{X};h_1\right)-\left\{D-\mu_0^D\left(\mathbf{X};h_1\right)\right\}\delta\left(\mathbf{X};h_1\right)\right]$$
$$+\frac{1}{\kappa}\frac{(2Z-1)I\left(S=h_2\right)}{f\left(Z\mid\mathbf{X},S=h_2\right)}\frac{f(S=t\mid\mathbf{X})}{f\left(S=h_2\mid\mathbf{X}\right)}\delta\left(\mathbf{X};h_1\right)\left\{D-\mu_0^D\left(\mathbf{X};h_2\right)-\delta^D\left(\mathbf{X};h_2\right)Z\right\}$$
$$+\frac{1}{\kappa}I(S=t)\left\{\delta\left(\mathbf{X};h_1\right)\delta^D\left(\mathbf{X};h_2\right)-\psi\right\}$$

□

Web Appendix B: Partial identification of $CATE(\mathbf{X};S=h_1)$

Partial identification of the conditional average treatment effect in a randomized trial with non-compliance has been extensively discussed in the literature. Different partial identification intervals exist under different identification assumptions. Two extreme cases are (1) bounds under minimal, core IV assumptions and (2) point identification (i.e., a partial identification interval collapsing to a point) under no-interaction-type assumptions. We have discussed in detail no-interaction-type assumptions that permit point identification. Below, we provide a brief overview of partial identification bounds under minimal IV assumptions for completeness.

B.1: Balke-Pearl bounds

For a binary outcome, if we assume that core IV assumptions hold within strata defined by observed covariates $\mathbf{X}$, then the conditional average treatment effect in the historical trial $S=h_1$, i.e., $\mathbb{E}[Y(D=0)\mid\mathbf{X};S=h_1]$, is lower bounded by

$$\max\left\{\begin{array}{l} p_{1,0|\mathbf{X},1} \\ p_{1,0|\mathbf{X},S=h_1,0} \\ p_{1,0|\mathbf{X},S=h_1,0}+p_{1,1|\mathbf{X},S=h_1,0}-p_{0,0|\mathbf{X},S=h_1,1}-p_{1,1|\mathbf{X},S=h_1,1} \\ p_{0,1|\mathbf{X},S=h_1,0}+p_{1,0|\mathbf{X},S=h_1,0}-p_{0,0|\mathbf{X},S=h_1,1}-p_{0,1|\mathbf{X},S=h_1,1} \end{array}\right\},$$

and upper bounded by

$$\min\left\{\begin{array}{l} 1-p_{0,0|\mathbf{X},S=h_1,1} \\ 1-p_{0,0|\mathbf{X},S=h_1,0} \\ p_{0,1|\mathbf{X},S=h_1,0}+p_{1,0|\mathbf{X},S=h_1,0}+p_{1,0|\mathbf{X},S=h_1,1}+p_{1,1|\mathbf{X},S=h_1,1} \\ p_{1,0|\mathbf{X},S=h_1,0}+p_{1,1|\mathbf{X},S=h_1,0}+p_{0,1|\mathbf{X},S=h_1,1}+p_{1,0|\mathbf{X},S=h_1,1} \end{array}\right\}.$$

Analogously, $\mathbb{E}[Y(D=1) \mid \mathbf{X}; S=h_1]$ is lower bounded by

$$\max\left\{\begin{array}{l} p_{1,1|\mathbf{X},S=h_1,0} \\ p_{1,1|\mathbf{X},S=h_1,1} \\ -p_{0,0|\mathbf{X},S=h_1,0}-p_{0,1|\mathbf{X},S=h_1,0}+p_{0,0|\mathbf{X},S=h_1,1}-p_{1,1|\mathbf{X},S=h_1,1} \\ -p_{0,1|\mathbf{X},S=h_1,0}-p_{1,0|\mathbf{X},S=h_1,0}+p_{1,0|\mathbf{X},S=h_1,1}+p_{1,1|\mathbf{X},S=h_1,1} \end{array}\right\},$$

and upper bounded by

$$\min\left\{\begin{array}{l} 1-p_{0,1|\mathbf{X},S=h_1,1} \\ 1-p_{0,1|\mathbf{X},S=h_1,0} \\ p_{0,0|\mathbf{X},S=h_1,0}+p_{1,1|\mathbf{X},S=h_1,0}+p_{1,0|\mathbf{X},S=h_1,1}+p_{1,1|\mathbf{X},S=h_1,1} \\ p_{1,0|\mathbf{X},S=h_1,0}+p_{1,1|\mathbf{X},S=h_1,0}+p_{0,0|\mathbf{X},S=h_1,1}+p_{1,1|\mathbf{X},S=h_1,1} \end{array}\right\}.$$

where $p_{y,d|\mathbf{X},S=h_1,z}$ is a shorthand for the conditional mean $P(Y=y, D=d \mid \mathbf{X}, S=h_1, Z=z)$ for $Z \in \{0,1\}$ and $D \in \{0,1\}$ in the historical dataset $S=h_1$. All conditional means involved above may be estimated via fitting parametric models, e.g., multinomial logistic regressions, or using flexible machine learning methods. The bounds on the $CATE(\mathbf{X}; S=h_1) = \mathbb{E}[Y(D=1) \mid \mathbf{X}; S=h_1] - \mathbb{E}[Y(D=0) \mid \mathbf{X}; S=h_1]$ then follows from upper and lower bounds on $\mathbb{E}[Y(D=1) \mid \mathbf{X}; S=h_1]$ and $\mathbb{E}[Y(D=0) \mid \mathbf{X}; S=h_1]$ (Balke and Pearl, 1997).

B.2: Manski-Pepper bounds

For a bounded outcome, Manski (1990) derived the following "minimal-assumptions" partial identification bounds. Assume $Y \in [K_0, K_1]$ almost surely. Note that

$$\begin{aligned} & \mathbb{E}[Y(D=1) \mid Z=\mathbf{X}, S=h_1, 1] \\ = \; & \mathbb{E}[Y(D=1) \mid D=1, Z=\mathbf{X}, S=h_1, 1] \cdot P(D=1 \mid Z=\mathbf{X}, S=h_1, 1) \\ + \; & \mathbb{E}[Y(D=1) \mid D=0, Z=\mathbf{X}, S=h_1, 1] \cdot P(D=0 \mid Z=\mathbf{X}, S=h_1, 1), \end{aligned} \tag{1}$$

where $\mathbb{E}[Y(D=1) \mid D=1, Z=\mathbf{X}, S=h_1, 1] = \mathbb{E}[Y \mid D=1, Z=\mathbf{X}, S=h_1, 1]$ and $P(D=1 \mid Z=\mathbf{X}, S=h_1, 1)$ are both identified from observed data. The censored potential outcome term $\mathbb{E}[Y(D=1) \mid D=0, Z=\mathbf{X}, S=h_1, 1]$ is not identified from data, but is bounded between $[K_0, K_1]$. Therefore, we have the following "minimal-assumptions" bounds on $\mathbb{E}[Y(D=1) \mid Z=\mathbf{X}, S=h_1, 1]$:

$$\begin{aligned} & \psi(Z=1, D=1, \mathbf{X}; K_0, h_1) \\ & := \mathbb{E}[Y \mid D=1, Z=\mathbf{X}, S=h_1, 1] \cdot P(D=1 \mid Z=\mathbf{X}, S=h_1, 1) + K_0 \cdot P(D=0 \mid Z=\mathbf{X}, S=h_1, 1) \\ & \leq \mathbb{E}[Y(D=1) \mid Z=\mathbf{X}, S=h_1, 1] \leq \\ & \psi(Z=1, D=1, \mathbf{X}; K_1, h_1) \\ & := \mathbb{E}[Y \mid D=1, Z=\mathbf{X}, S=h_1, 1] \cdot P(D=1 \mid Z=\mathbf{X}, S=h_1, 1) + K_1 \cdot P(D=0 \mid Z=\mathbf{X}, S=h_1, 1). \end{aligned} \tag{2}$$

We also have similar bounds on $\mathbb{E}[Y(D=1) \mid Z=\mathbf{X}, S=h_1, 0]$:

$$\psi(Z=0, D=1, \mathbf{X}; K_0, h_1) \leq \mathbb{E}[Y(D=1) \mid Z=\mathbf{X}, S=h_1, 0] \leq \psi(Z=0, D=1, \mathbf{X}; K_1, h_1). \tag{3}$$

Marginalizing over Z and we have the following "minimal-assumptions" bounds on $\mathbb{E}[Y(D=1) \mid \mathbf{X}; S=h_1]$:

$$\psi(Z=1,D=1,\mathbf{X};K_0,h_1)\cdot P(Z=1\mid \mathbf{X};S=h_1)+\psi(Z=0,D=1,\mathbf{X};K_0,h_1)\cdot P(Z=0\mid \mathbf{X};S=h_1)$$
$$\leq \mathbb{E}[Y(D=1)\mid \mathbf{X};S=h_1]\leq$$
$$\psi(Z=1,D=1,\mathbf{X};K_1,h_1)\cdot P(Z=1\mid \mathbf{X};S=h_1)+\psi(Z=0,D=1,\mathbf{X};K_1,h_1)\cdot P(Z=0\mid \mathbf{X};S=h_1). \tag{4}$$

The bounds on the other conditional potential outcome $\mathbb{E}[Y(D=0) \mid \mathbf{X}; S=h_1]$ can be obtained similarly by replacing $D=1$ in each $\psi(\cdot)$ expression with $D=0$.

B.3: Additional partial identification bounds

There is a wide spectrum between "minimal assumptions" bounds like Balke-Pearl bounds and Manski-Pepper bounds, and point identification results in Wang and Tchetgen Tchetgen (2018). Researchers may impose additional assumptions on the data-generating process and obtain bounds that are considerably narrower compared to "minimal-assumptions" bounds. Some useful additional assumptions and corresponding bounds for a bounded outcome include *monotone instrumental variable*, *monotone treatment selection*, *monotone treatment response*, among many others (Manski and Pepper, 2000).

Web Appendix C: Additional details on estimation

C.1: Variants of the regression-based estimator

There are multiple variants of the simple regression-based estimator. When two historical trials collapse to one, i.e., $h_1 = h_2$, then $\widehat{ITT}_{\text{full, reg}}$ simplifies to:

$$\widehat{ITT}_{\text{full, reg}} = \frac{1}{|\mathcal{D}|}\sum_{i=1}^{|\mathcal{D}|}\mathbb{1}\{S_i=t\}\times\hat{\delta}^Y(\mathbf{X}_i;h_1), \tag{5}$$

which is the same estimator one would obtain under the conditional constancy assumption (Zhang, 2009). In other words, the conditional constancy assumption is implied by Assumption 3 imposed on the hypothetical placebo-controlled trial concerning the target population plus Assumption 3,

Assumption 4, and Assumption 6 imposed on the historical trial $S = h_1$.

When we have compliance data from the active-controlled trial (i.e., $\mathcal{D}_{\text{target}} = \{(\mathbf{X}_i, Z_i, D_i) : i = 1, \ldots, N\}$ as in the case study), then the term $CC_{AC}(\mathbf{X})$ is directly estimable from $\mathcal{D}_{\text{target}}$. Indeed, suppose we fit parametric models for $\mu_{D,1}(\mathbf{X}, t) = E(D \mid Z = 1, S = t, \mathbf{X})$ and $\mu_{D,0}(\mathbf{X}, h_2) = E(D \mid Z = 0, S = h_1, \mathbf{X})$, from which we obtain estimates $\hat{\mu}_{D,1}(\mathbf{X}, t) = E(D \mid Z = 1, S = t, \mathbf{X})$ and $\hat{\mu}_{D,0}(\mathbf{X}, h_2) = E(D \mid Z = 0, S = h_1, \mathbf{X})$. The regression-based estimator then becomes

$$\widehat{ITT}_{\text{full, reg}} = \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \mathbb{1}\{S_i = t\} \times \left\{ \frac{\hat{\delta}^Y(\mathbf{X}_i; h_1)}{\hat{\delta}^D(\mathbf{X}_i; h_1)} \times \left\{ \hat{\mu}_{D,1}(\mathbf{X}_i, t) - \hat{\mu}_{D,0}(\mathbf{X}_i; h_2) \right\} \right\}, \tag{6}$$

where the NI-trial-in-sample estimate $\hat{\mu}_{D,1}(\mathbf{X}, t)$ now replaces $\hat{\mu}_{D,1}(\mathbf{X}; h_2)$. As described in the main manuscript, if the parametric regression models postulated for $\delta^Y(\mathbf{X}; h_1)$, $\delta^D(\mathbf{X}; h_1)$, $\mu_{D,1}(\mathbf{X}, t)$ and $\mu_{D,0}(\mathbf{X}, h_2)$ are correctly specified, with parameters estimated using standard likelihood or M-estimation methods, then the resulting estimator is consistent and asymptotically normal under standard regularity conditions.

Under point identification assumptions discussed in the main article, the conditional average treatment $CATE(\mathbf{X}; h_1)$ can also be estimated by directly imposing a parametric model $\delta(\mathbf{X}; h_1, \alpha)$ on $\delta(\mathbf{X}; h_1) := \mathbb{E}[Y(D = 1) - Y(D = 0) \mid S = h_1, \mathbf{X}]$ and estimating the model parameters α by the $\hat{\alpha}$ that solves the following estimating equation (Robins, 1994):

$$\frac{1}{N_1} \sum_{i=1}^{N_1} h(\mathbf{X}) \{Y_i - D_i \delta(\mathbf{X}; \alpha)\} \frac{2Z_i - 1}{f(Z_i \mid \mathbf{X})} = 0,$$

where $h(\mathbf{X})$ is any vector function of the same dimension as α, e.g., $h(\mathbf{X}) = \mathbf{X}$, and $f(Z_i \mid \mathbf{X})$ is a known treatment assignment probability. $\delta(\mathbf{X}; \hat{\alpha})$ is then an estimator for $CATE(\mathbf{X}; h_1)$ and can be plugged directly into equation (5) in the main article. One advantage of this approach is that by choosing an appropriate model for $\delta(\mathbf{X})$, one can ensure that the estimator $\delta(\mathbf{X}; \hat{\alpha})$ is in-sample bounded; see for example the proposal of Wang and Tchetgen Tchetgen (2018). Okui et al. (2012) and Tan (2006) discussed other ways to impose models to estimate $CATE(\boldsymbol{X}; h_1)$.

C.2: No crossover in the planned active-controlled trial

If the hypothetical placebo-controlled trial in the target population precludes crossover, i.e., $Z = 0$ implies $D(Z) = 0$, then the population consists of only compliers and never-takers (Frangakis and Rubin, 1999, 2002). In this special case, the conditional intention-to-treat effect becomes:

$$\begin{aligned}
&\mathbb{E}[Y(Z=1) - Y(Z=0) \mid \mathbf{X}] \\
=\ &\underbrace{\mathbb{E}[Y(Z=1) - Y(Z=0) \mid D(1)=1, D(0)=0, \mathbf{X}]}_{\text{Term I}} \times \underbrace{P(D(1)=1, D(0)=0 \mid \mathbf{X})}_{\text{Term II}} \\
+\ &\underbrace{\mathbb{E}[Y(Z=1) - Y(Z=0) \mid D(1)=0, D(0)=0, \mathbf{X}]}_{\text{Term III}} \times \underbrace{P(D(1)=0, D(0)=0 \mid \mathbf{X})}_{\text{Term IV}} \\
=\ &\mathbb{E}[Y(Z=1) - Y(Z=0) \mid D(1)=1, D(0)=0, \mathbf{X}] \times P(D(1)=1, D(0)=0 \mid \mathbf{X}),
\end{aligned} \tag{7}$$

because Term III = 0 by definition. Term I is the average intention-to-treat effect in the subpopulation of compliers conditional on $\mathbf{X}$, which is equal to the conditional complier average treatment effect $\mathbb{E}[Y(D=1) - Y(D=0) \mid D(1)=1, D(0)=0, \mathbf{X}]$.

Assumption 4' is a variation of the mean generalizability assumption, and help identify Term I using data from a selected historical trial.

Assumption 4' (Mean generalizability among compliers)**.** $\mathbb{E}[Y(D=1) - Y(D=0) \mid D(1)=1, D(0)=0, \mathbf{X}] = \mathbb{E}[Y(D=1) - Y(D=0) \mid D(1)=1, D(0)=0, \mathbf{X}, S=h]$.

Term II characterizes the proportion of compliers conditional on $\mathbf{X}$. Observe that:

$$\begin{aligned}
P(D(1)=1, D(0)=0 \mid \mathbf{X}) &= P(D(1)=1 \mid \mathbf{X}) - P(D(1)=1, D(0)=1 \mid \mathbf{X}) \\
&= P(D=1 \mid \mathbf{X}, Z=1),
\end{aligned} \tag{8}$$

where the first equality is by definition and the second equality is because we exclude always-takers by assuming $D(Z=0) = 0$. As discussed before, $P(D=1 \mid Z=1, \mathbf{X})$ is identified from the active-controlled trial when partial data $\{\mathbf{X}, D, Z\}$ is available, or can be inferred from a historical trial under a one-arm compliance generalizability assumption analogous to Assumption 5. Using previous notation, an estimate of ITT in the no-crossover setting is given by:

$$\widehat{ITT}_{\text{no crossover, reg}} = \frac{1}{|\mathcal{D}|} \sum_{i=1}^{|\mathcal{D}|} \mathbb{1}\{S_i = t\} \times \left\{ \frac{\hat{\delta}^Y(\mathbf{X}_i; h_1)}{\hat{\delta}^D(\mathbf{X}_i; h_1)} \times \hat{\mu}_{D,1}(\mathbf{X}_i) \right\}. \tag{9}$$

C.3: Sensitivity analysis relaxing the no-interaction assumption

As summarized in Figure 1 in the main article, sensitivity analyses are useful to examine various identification assumptions. Relaxing Assumption 4, 5, and 6 is straightforward: Researchers may relax the equality constraint and allow $CATE(\mathbf{X})$ and $CC(\mathbf{X})$ to differ systematically from their historical-data-identified counterparts $CATE(\mathbf{X}; h)$ and $CC(\mathbf{X}; h)$. Assumption 3 states that there is no more common modifiers of the compliance behavior and treatment effect beyond those contained in $\mathbf{X}$. As a sensitivity analysis, one may consider the following parametrization that relaxes Assumption 3. Let U denote unmeasured common effect modifiers. We scale U so that $\mathbb{E}[U \mid \mathbf{X}] = 0$ and $\mathbb{E}[U^2 \mid \mathbf{X}] = 1$. We consider the following parametrization

$$\begin{aligned}
\mathbb{E}[Y(D=1) - Y(D=0) \mid \mathbf{X}, U] &= \mathbb{E}[Y(D=1) - Y(D=0) \mid \mathbf{X}] + \lambda_1 U, \\
\mathbb{E}[D(Z=1) - D(Z=0) \mid \mathbf{X}, U] &= \mathbb{E}[D(Z=1) - D(Z=0) \mid \mathbf{X}] + \lambda_2 U,
\end{aligned}$$

which holds, for instance, in the following semiparametric models for $Y(D)$ and $D(Z)$:

$$\begin{aligned}
\mathbb{E}[Y(D) \mid \mathbf{X}, U] &= f_1(\mathbf{X}, U) + g_1(\mathbf{X}) \cdot D + \lambda_1 U D, \\
\mathbb{E}[D(Z) \mid \mathbf{X}, U] &= f_2(\mathbf{X}, U) + g_2(\mathbf{X}) \cdot D + \lambda_2 U D.
\end{aligned}$$

Under this parametrization, the true conditional intention-to-treat effect $ITT(\mathbf{X})$ in the NI trial admits the following simple decomposition:

$$\begin{aligned}
ITT(\mathbf{X}) &= \mathbb{E}[Y(D=1) - Y(D=0) \mid \mathbf{X}] \times \mathbb{E}[D(Z=1) - D(Z=0) \mid \mathbf{X}] + \lambda_1 \lambda_2 \\
&= CATE(\mathbf{X}) \times CC(\mathbf{X}) + \lambda_1 \lambda_2.
\end{aligned} \tag{10}$$

Let $\Lambda_1 \times \Lambda_2$ denote a plausible sensitivity region for (λ_1, λ_2) and $\widehat{ITT}(\lambda_1, \lambda_2)$ a bias-corrected ITT estimate for $(\lambda_1, \lambda_2) \in \Lambda_1 \times \Lambda_2$. A level-$\alpha$ sensitivity interval may then be formed by taking a union over all level-α confidence intervals of $\widehat{ITT}(\lambda_1, \lambda_2)$ over the sensitivity region $\Lambda_1 \times \Lambda_2$.

Web Appendix D: Additional simulation details

D.1: Sampling distributions of 6 estimators in Scenario X1 and Scenario Y1

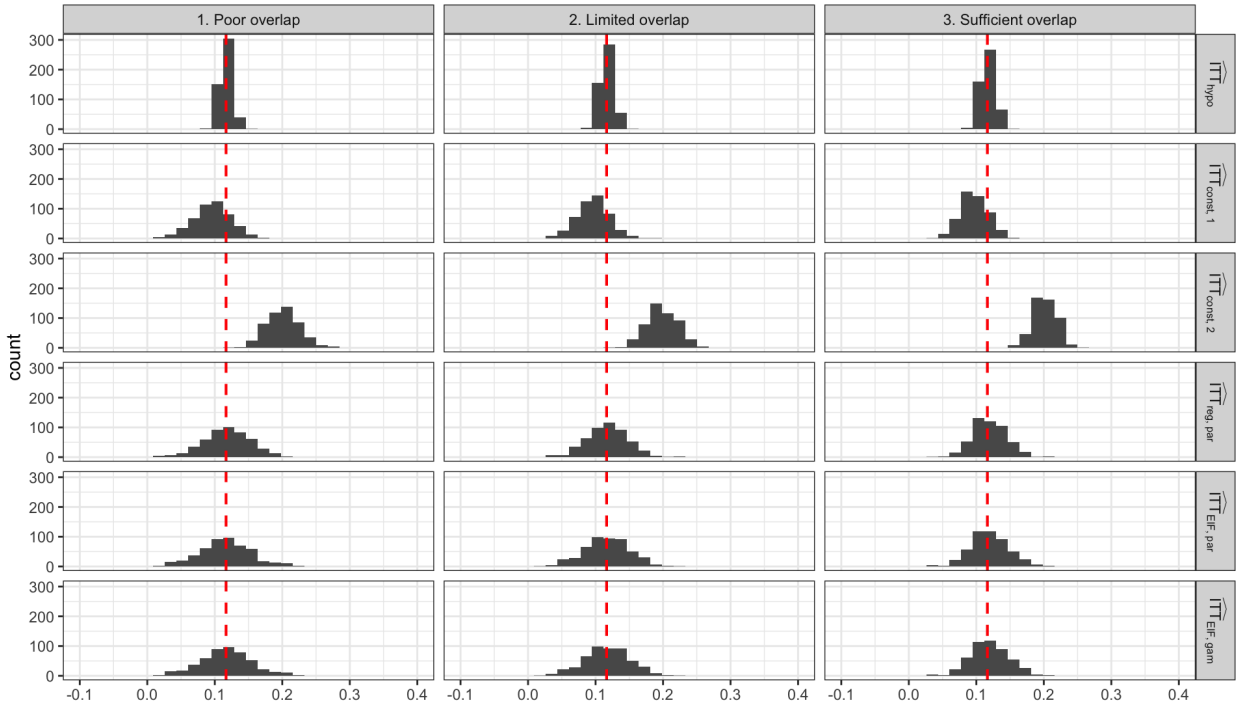

Figure S1: Sampling distributions of 6 intention-to-treat effect estimators $\widehat{ITT}_{\text{hypo}}$, $\widehat{ITT}_{\text{const},1}$, $\widehat{ITT}_{\text{const},2}$, $\widehat{ITT}_{\text{reg, par}}$, $\widehat{ITT}_{\text{EIF, par}}$, and $\widehat{ITT}_{\text{EIF, gam}}$ when $N_1 = N_2 = N = 2000$, observed covariates are generated according to Scenario X1, and outcomes are generated according to Scenario Y1. Three levels of overlap (Poor, Low, and Sufficient) are considered. Simulations are repeated 1000 times. The dashed red lines represent the ground truth intention-to-treat effects in each setting.

D.2: Additional simulation results

Figure S2 shows the sampling distributions of $\widehat{ITT}_{\text{hypo}}$, $\widehat{ITT}_{\text{const,1}}$, $\widehat{ITT}_{\text{const,2}}$, $\widehat{ITT}_{\text{reg, par}}$, and $\widehat{ITT}_{\text{EIF, gam}}$, when $N_1 = N_2 = N = 2000$, observed covariates were generated according to Scenario X2 and outcomes were generated according to Scenario Y2. Table S1 summarizes the percentage of bias and the coverage of 95% confidence intervals for different choices of sample size and the overlap parameter c. We still truncate the estimator $\widehat{ITT}_{\text{EIF, gam}}$ to be bounded in $[-1, 1]$ by letting the estimator be $\phi(\widehat{ITT}_{\text{EIF, gam}})$ where function $\phi(x) = 1,\ \forall x \geq 1$, $\phi(x) = -1,\ \forall x \leq -1$, and $\phi(x) = x$ otherwise. The ground truth intention-to-treat effects are superimposed using red dashed lines. The historical-data-driven estimator $\widehat{ITT}_{\text{EIF, gam}}$ closely resembles the ground truth ITTs but has a larger variance compared to that of $\widehat{ITT}_{\text{hypo}}$. However, $\widehat{ITT}_{\text{reg, par}}$ is biased when the overlap parameter $c \in \{0.25, 0.50\}$ because we still fit linear parametric models for $CATE(\mathbf{X})$ and $CC(\mathbf{X})$ which are not correctly specified in this data generation setting. Besides, the two estimators $\widehat{ITT}_{\text{const,1}}$ and $\widehat{ITT}_{\text{const,2}}$ based on an incorrect assumption of conditional constancy were significantly biased, and their confidence intervals did not have the nominal level of coverage.

	$\widehat{ITT}_{\text{hypo}}$		$\widehat{ITT}_{\text{const, 1}}$		$\widehat{ITT}_{\text{const, 2}}$		$\widehat{ITT}_{\text{reg, par}}$		$\widehat{ITT}_{\text{EIF, gam}}$	
Sample size	% Bias	95% CI Coverage	% Bias	95% CI Coverage	% Bias	95% CI Coverage	% Bias	95% CI Coverage	% Bias	95% CI Coverage
					$c = 0$					
1000	-0.1	95.6%	-18.7	70.9%	76.2	0.0%	0.4	95.3%	1.3	94.5%
2000	0.1	96.2%	-18.8	49.6%	76.6	0.0%	-0.1	94.2%	1.1	94.3%
5000	0.1	95.3%	-19.0	13.4%	76.0	0.0%	-0.6	94.0%	0.5	95.3%
					$c = 0.25$					
1000	0.6	96.5%	-19.8	62.6%	68.2	0.0%	7.1	93.3%	-0.5	95.9%
2000	0.2	95.2%	-19.4	40.9%	69.0	0.0%	8.0	89.3%	2.1	95.5%
5000	-0.1	96.3%	-19.7	4.5%	68.6	0.0%	7.8	82.2%	2.2	94.8%
					$c = 0.5$					
1000	0.2	95.7%	-21.6	53.7%	56.9	0.0%	18.0	77.3%	2.1	94.4%
2000	-0.4	95.5%	-22.1	21.0%	57.7	0.0%	17.6	62.6%	0.8	93.2%
5000	0.2	94.8%	-21.6	1.3%	57.4	0.0%	17.5	29.1%	1.0	94.8%

Table S1: Simulation results of 5 estimators for various sample size and choice of the overlap parameter c when the observed covariates were generated according to Scenario X2 and outcomes were generated according to Scenario Y2. The percentage of bias and coverage of 95% confidence intervals are reported. Confidence intervals of $\widehat{ITT}_{\text{EIF, gam}}$ were estimated based on the asymptotic normality and the efficient influence function. Confidence intervals of $\widehat{ITT}_{\text{hypo}}$ were based on two-sample tests. Confidence intervals of the other 3 estimators were obtained via the nonparametric bootstrap. Out of 1000 simulations, $\widehat{ITT}_{\text{EIF, gam}}$ fell outside $[-1, 1]$ a total of 13, 5 and 8 times when $c = 0$, 0.25 and 0.5, respectively, when $n = 1000$, 6, 3 and 4 times when $c = 0$, 0.25 and 0.5, respectively, when $n = 2000$, and 4, 3 and 2 times when $c = 0$, 0.25 and 0.5, respectively, when $n = 5000$.

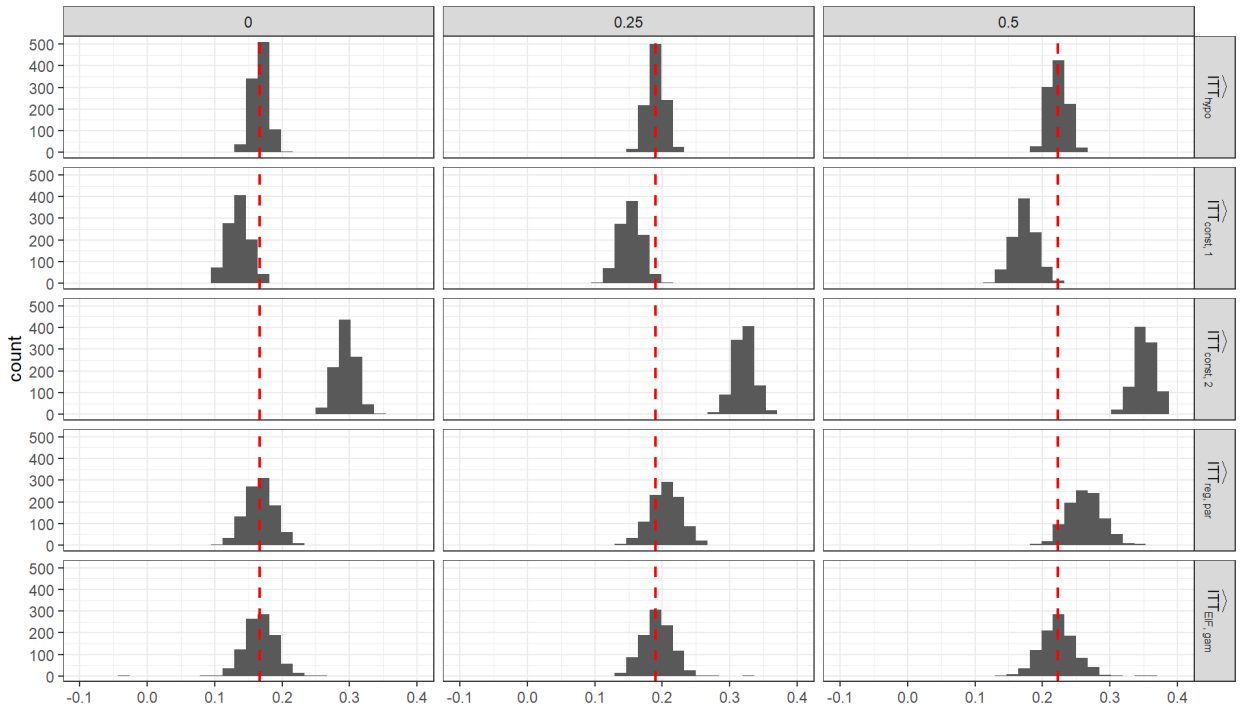

Figure S2: Sampling distributions of 5 intention-to-treat effect estimators $\widehat{ITT}_{\text{hypo}}$, $\widehat{ITT}_{\text{const},1}$, $\widehat{ITT}_{\text{const},2}$, $\widehat{ITT}_{\text{reg, par}}$, and $\widehat{ITT}_{\text{EIF, gam}}$ when $N_1 = N_2 = N = 2000$, different choices of the overlap parameter c and observed covariates were generated according to Scenario X2 and outcomes were generated according to Scenario Y2. Simulations are repeated 1000 times. The dashed red lines represent the ground truth intention-to-treat effects in each setting. $\widehat{ITT}_{\text{EIF, par}}$ is not included because it has to be biased as we estimate EIFs by fitting linear parameter models which are not correctly specified.

Figure S3 shows the sampling distributions of $\widehat{ITT}_{\text{hypo}}$, $\widehat{ITT}_{\text{const},1}$, $\widehat{ITT}_{\text{const},2}$, $\widehat{ITT}_{\text{reg, par}}$, and $\widehat{ITT}_{\text{EIF, gam}}$, when $N_1 = N_2 = N = 2000$, observed covariates were generated according to Scenario X2 and outcomes were generated according to Scenario Y1. Table S2 summarizes the percentage of bias and the coverage of 95% confidence intervals for different choices of sample size and the overlap parameter c.

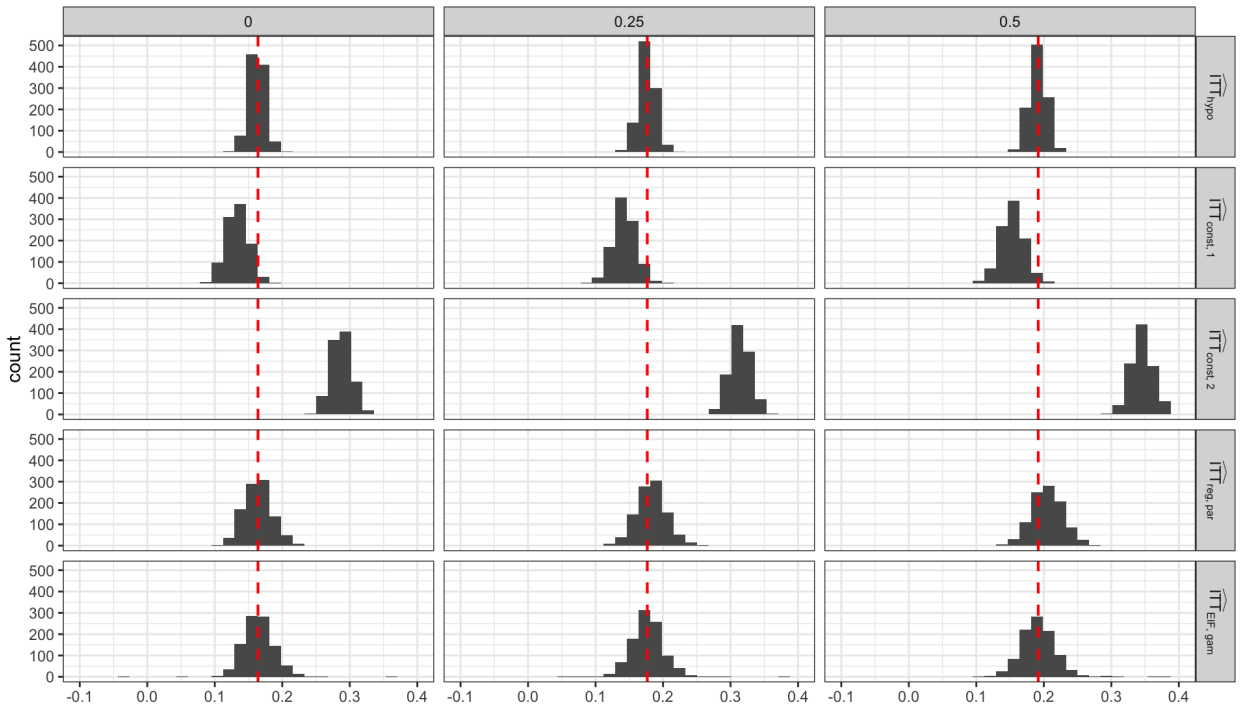

Figure S3: Sampling distributions of 6 intention-to-treat effect estimators $\widehat{ITT}_{\text{hypo}}$, $\widehat{ITT}_{\text{const},1}$, $\widehat{ITT}_{\text{const},2}$, $\widehat{ITT}_{\text{reg, par}}$, and $\widehat{ITT}_{\text{EIF, gam}}$ in Scenario X2 when $N_1 = N_2 = N = 2000$, different choices of the overlap parameter c, and observed covariates were generated according to Scenario X2 and outcomes were generated according to Scenario Y1. Simulations are repeated 1000 times. The dashed red lines represent the ground truth intention-to-treat effects in each setting.

	$\widehat{ITT}_{\text{hypo}}$		$\widehat{ITT}_{\text{const, 1}}$		$\widehat{ITT}_{\text{const, 2}}$		$\widehat{ITT}_{\text{reg, par}}$		$\widehat{ITT}_{\text{EIF, gam}}$	
Sample size	% Bias	95% CI Coverage	% Bias	95% CI Coverage	% Bias	95% CI Coverage	% Bias	95% CI Coverage	% Bias	95% CI Coverage
					$c = 0$					
1000	0.0	95.9%	-19.1	70.7%	75.9	0.0%	-0.7	94.7%	-1.4	95.6%
2000	-1.0	94.9%	-18.6	51.7%	75.2	0.0%	-0.1	95.0%	-1.3	94.8%
5000	-0.4	94.5%	-18.8	14.7%	74.9	0.0%	-0.3	95.2%	0.8	95.4%
					$c = 0.25$					
1000	-0.1	94.8%	-19.4	70.7%	77.7	0.0%	2.7	94.3%	-0.8	94.7%
2000	-0.2	94.8%	-19.1	46.8%	77.8	0.0%	3.3	94.3%	1.9	95.2%
5000	-0.2	94.8%	-19.4	9.3%	77.7	0.0%	3.0	92.7%	-0.1	94.5%
					$c = 0.5$					
1000	-0.4	95.0%	-19.8	68.3%	80.0	0.0%	7.6	92.7%	-0.4	94.9%
2000	-0.4	96.0%	-20.0	44.3%	79.9	0.0%	6.9	91.8%	1.4	95.3%
5000	-0.2	95.7%	-19.5	10.3%	80.1	0.0%	7.1	86.1%	-0.2	94.4%

Table S2: Simulation results of 5 estimators for various sample sizes and choice of the overlap parameter c when the observed covariates were generated according to Scenario X2 and outcomes were generated according to Scenario Y1. The percentage of bias and coverage of 95% confidence intervals are reported. Confidence intervals of $\widehat{ITT}_{\text{EIF, gam}}$ were estimated based on the asymptotic normality and the efficient influence function. Confidence intervals of $\widehat{ITT}_{\text{hypo}}$ were based on two-sample tests. Confidence intervals of the other 4 estimators were obtained via the nonparametric bootstrap. Out of 1000 simulations, $\widehat{ITT}_{\text{EIF, gam}}$ fell outside $[-1, 1]$ (considered outliers) a total of 14, 11 and 9 times when $c = 0$, 0.25 and 0.5, respectively, when $n = 1000$, 7, 6 and 5 times when $c = 0$, 0.25 and 0.5, respectively, when $n = 2000$, and 3, 3 and 0 times when $c = 0$, 0.25 and 0.5, respectively, when $n = 5000$.

Web Appendix E: Additional details on real data

E.1: Kaplan-Meier estimates in HPTN 084

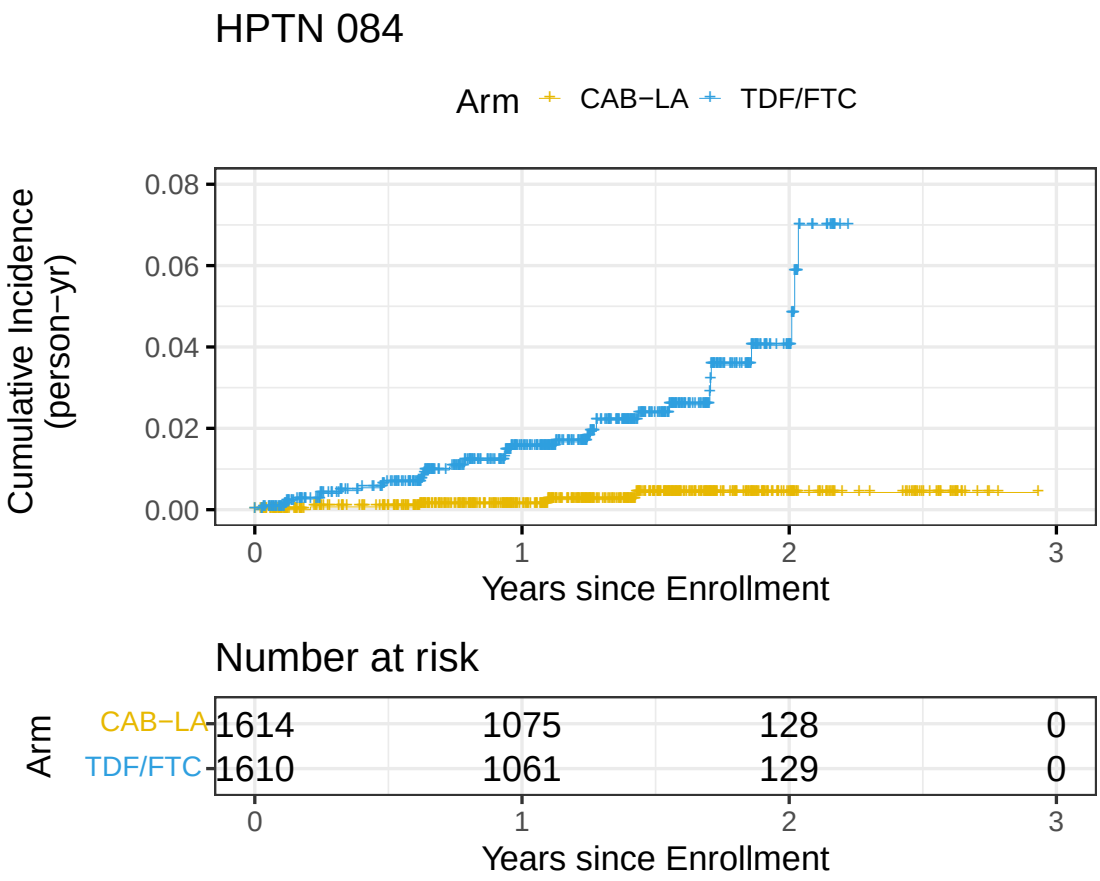

Figure S4: Kaplan-Meier estimates of incident HIV acquisition in the HPTN 084 trial population

E.2: A summary of 5 historical trials

Table S3: Year, target population, locations, and relative risk of five major placebo-controlled trials of daily oral TDF/FTC against placebo (Cohen and Baden, 2012). † The heterosexual men and women are in HIV discordant partnerships

Trial (year)	**Target population**	**Locations**	**Relative risk (placebo vs. TDF/FTC) [95% CI]**
iPrEx (2010)	Men who have sex with men	Brazil, Ecuador, Peru, South Africa, Thailand, and U.S.	1.79 [1.20, 2.67]
Partners PrEP (2012)	Heterosexual men and women†	Kenya and Uganda	4.00 [2.19, 7.32]
TDF2 (2012)	Heterosexual men and women	Botswana	2.68 [1.26, 5.73]
FEM-PrEP (2012)	Young women	Kenya, South Africa, and Tanzania	1.05 [0.66, 1.68]
VOICE (2015)	Women	South Africa, Uganda, and Zimbabwe	0.97 [0.69, 1.37]

E.3: Overlap of HPTN 084 and Partners PrEP

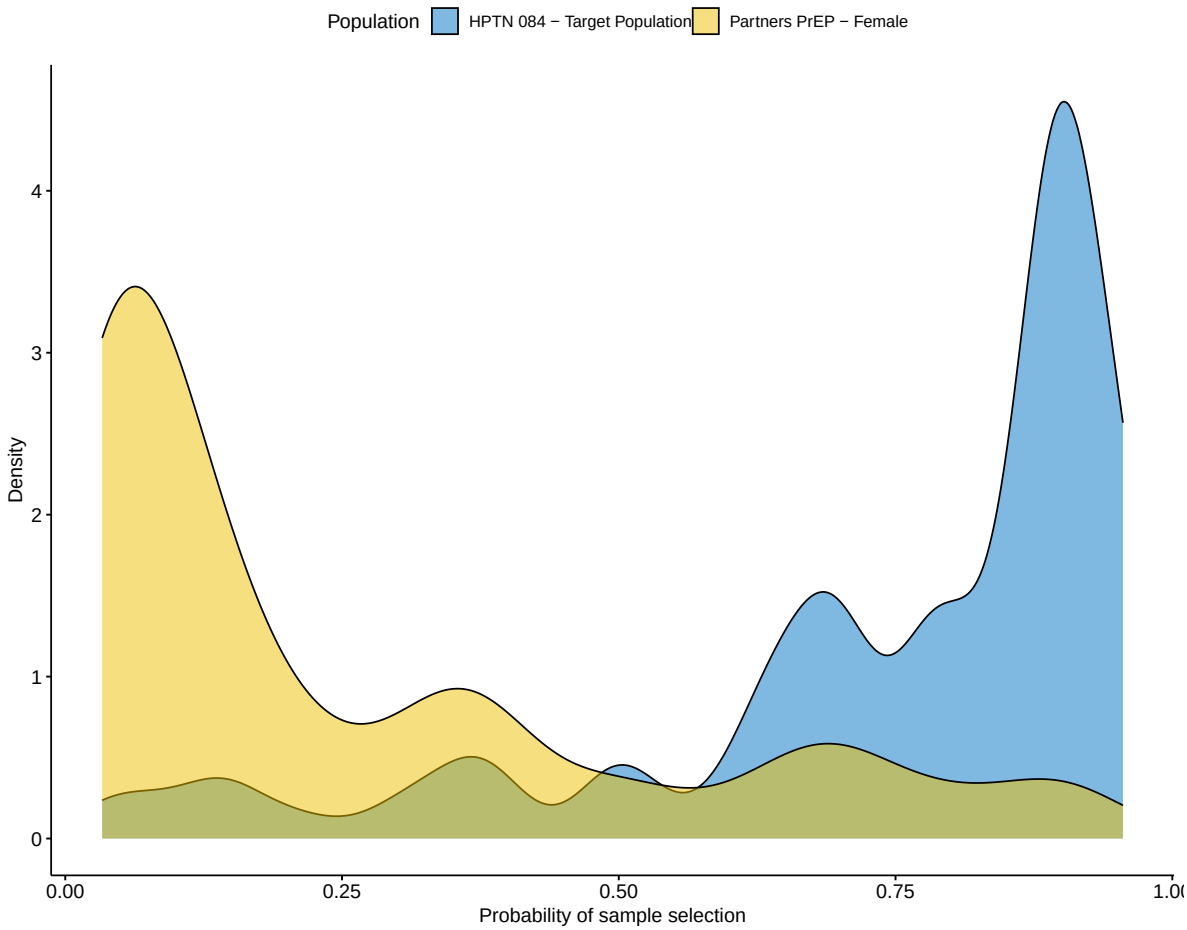

Figure S5: The probability of sample selection which is the probability of a participant to be in the HPTN 084 - Target Population given the covariates.

References

Balke, A. and Pearl, J. (1997). Bounds on treatment effects from studies with imperfect compliance. *Journal of the American Statistical Association*, 92(439):1171–1176.

Cohen, M. S. and Baden, L. R. (2012). Preexposure prophylaxis for HIV—where do we go from here? *New England Journal of Medicine*, 367(5):459–461.

Frangakis, C. E. and Rubin, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, 86(2):365–379.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58(1):21–29.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323.

Manski, C. F. and Pepper, J. V. (2000). Monotone instrumental variables: with an application to the returns to schooling. *Econometrica*, 68:997–1010.

Okui, R., Small, D. S., Tan, Z., and Robins, J. M. (2012). Doubly robust instrumental variable regression. *Statistica Sinica*, pages 173–205.

Robins, J. M. (1994). Correcting for non-compliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and Methods*, 23(8):2379–2412.

Tan, Z. (2006). Regression and weighting methods for causal inference using instrumental variables. *Journal of the American Statistical Association*, 101(476):1607–1618.

Wang, L. and Tchetgen Tchetgen, E. (2018). Bounded, efficient and multiply robust estimation of average treatment effects using instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(3):531–550.

Zhang, Z. (2009). Covariate-adjusted putative placebo analysis in active-controlled clinical trials. *Statistics in Biopharmaceutical Research*, 1(3):279–290.