Improving Inferences from Randomized Trials: Using per-protocol analyses obtain better estimated of HIV treatment effects.

by Timothy Feeney
SER June 20 2024

e: feeney@unc.edu, bsky: @tfeend.bsky.social

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

Improving Inferences from Randomized Trials: Using per-protocol analyses obtain better estimated of HIV treatment effects.

by Timothy Feeney
SER June 20 2024

e: feeney@unc.edu, bsky: @tfeend.bsky.social

UNC GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

Outline

Outline
RCTs
Per-protocol effects
Example using ACTG 5202 Trial
Population
Analysis Plan
Results
Limitations and Future

Today I will be working with other presenters to both explain how per-protocol analyses can facilitate better estimates and also convince you that this is the way forward.

Ill start with background about RCTs, define per protocol effects, and then finish up with an example.

Randomized Trials are a gold standard

- Require clear enrollment criteria
- Unabmiguous intervention protocol
- Exchangeability: $Y^a \perp\!\!\!\perp A \quad$ for $A \in \{0, 1\}$
- Consistency[1]: $Y = Y^{a=1}A + Y^{a=0}(1 - A)$
- Positivity[2]: $Pr(A = a \mid \mathbf{L}) > 0, \ \forall \ell$ where $f(a, \ell) > 0$

☞ This allows for unbiased estimation of treatment effects.

[1] also known as treatment variation irrelevance
[2] $\mathbf{L}$ is covariate vector

Randomized Trials are a gold standard
Require clear enrollment criteria
Unabmiguous intervention protocol
Exchangeability: $Y^a \perp\!\!\!\perp A$ for $A \in \{0, 1\}$
Consistency[1]: $Y = Y^{a=1}A + Y^{a=0}(1 - A)$
Positivity[2]: $Pr(A = a \mid \mathbf{L}) > 0, \ \forall \ell$ where $f(a, \ell) > 0$
This allows for unbiased estimation of treatment effects.
[1] also known as treatment variation irrelevance
[2] $\mathbf{L}$ is covariate vector

- RCTs are classically considered the gold standard in evaluating treatment effects. The status of RCTs originates from aspects of the design that help assure causal estimates are identified
- Some of these included clear enumeration of the population unders study, clearly laid out treatment protocols, and other causal identification criteria of marginal exchangeability, causal consistency and positivity being met by design.

RCT estimands

- Intention-to-treat (ITT) effect: $E[Y^{r=1} - Y^{r=0}]$
- This is the effect of treatment assignment on outcomes
- Public health focused

- Per-protocol (PP) effect: $E[Y^{r=1,\bar{a}=1} - Y^{r=0,\bar{a}=0}]$
- This is the effect of treatment assignment and adherence on outcomes
- Patient focused[1]

[1] Hernan and Robins, NEJM 2016
[2] $\bar{a}$ is the history of treatment adherence

RCT estimands
- Intention-to-treat (ITT) effect: $E[Y^{r=1} - Y^{r=0}]$
- This is the effect of treatment assignment on outcomes
- Public health focused
- Per-protocol (PP) effect: $E[Y^{r=1,\bar{a}=1} - Y^{r=0,\bar{a}=0}]$
- This is the effect of treatment assignment and adherence on outcomes
- Patient focused[1]

[1] Hernan and Robins, NEJM 2016
[2] $\bar{a}$ is the history of treatment adherence

- There are two commonly reported estimands in an RCT. The primary result reported is typically the intention-to-treat results which present the outcomes had everyone been assigned to treatment 1 versus everyone being assigned to treatment 0.
- However, while this is public health focused, it is limited to evaluating the effect of randomization and not the effect of treatment.
- Instead, decision makers (e.g. physician) and those on the receiving end of the treatment may want to know instead waht is the effect if adherent to the treatment assignment. This cannot be answered with the ITT estimand and instead requires a per-protocol analysis that takes into account deviation or loss to follow up.

There is no *one* per-protocol effect

- Accounts for adherence
- "Doc, what if I take all my doses like you tell me to?"
- There are six per-protocol parameters that can be estimated[1]
- This can also depend on how the investigator(s) define adherence.

$$E\left[Y^{r=1,\bar{a}=1}\right] - E\left[Y^{r=0,\bar{a}=1}\right]$$
$$E\left[Y^{r=1,\bar{a}=1}\right] - E\left[Y^{r=0,\bar{a}=0}\right]$$
$$E\left[Y^{r=0,\bar{a}=1}\right] - E\left[Y^{r=0,\bar{a}=0}\right]$$
$$E\left[Y^{r=1,\bar{a}=1}\right] - E\left[Y^{r=1,\bar{a}=0}\right]$$
$$E\left[Y^{r=0,\bar{a}=1}\right] - E\left[Y^{r=1,\bar{a}=0}\right]$$
$$E\left[Y^{r=1,\bar{a}=0}\right] - E\left[Y^{r=0,\bar{a}=0}\right]$$

[1] Rudolph *et al.* Epidemiology 2020

[2] r = randomization, $\bar{a}$ = treatment history

There is no *one* per-protocol effect
Accounts for adherence
"Doc, what if I take all my doses like you tell me to?"
There are six per-protocol parameters that can be estimated[1]
This can also depend on how the investigator(s) define adherence.
$E\left[Y^{r=1,\bar{a}=1}\right] - E\left[Y^{r=0,\bar{a}=1}\right]$
$E\left[Y^{r=1,\bar{a}=1}\right] - E\left[Y^{r=0,\bar{a}=0}\right]$
$E\left[Y^{r=0,\bar{a}=1}\right] - E\left[Y^{r=0,\bar{a}=0}\right]$
$E\left[Y^{r=1,\bar{a}=1}\right] - E\left[Y^{r=1,\bar{a}=0}\right]$
$E\left[Y^{r=0,\bar{a}=1}\right] - E\left[Y^{r=1,\bar{a}=0}\right]$
$E\left[Y^{r=1,\bar{a}=0}\right] - E\left[Y^{r=0,\bar{a}=0}\right]$
[1] Rudolph *et al.* Epidemiology 2020
[2] r = randomization, $\bar{a}$ = treatment history

However there is no ONE per protocol effect and there are at least 6 estimands that can be considered per protocol effect. They are llustrated here.

- For instance if you look at the second line this answers "what if I took my treatment as assigned the whole way through the trial" where as line five asks "what if I did the opposite of what I was assigned the whole way through the trial?"
- These estimands can be made even more precise to account for deviating from 1 to many doses–more on this in a bit.

Per-protocol effects can be biased

- Frequently done by exluding those not adhering[1]
- Estimation of the per-protocol effect is much like an *observational study*
- Thus, bias can arise from non-exchangeability, non-consistency, or non-positivity

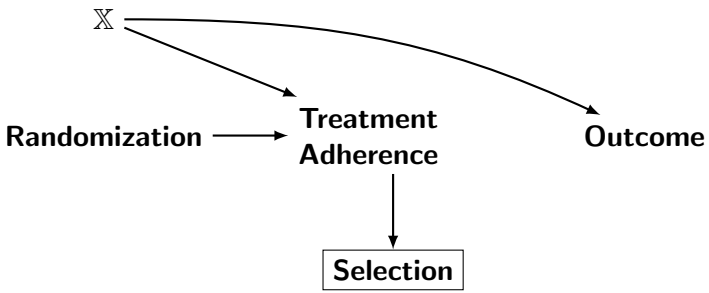
☞ This can lead to biased estimates

[1] Cole *et al.* JAMA Net. Open 2023, Dodd *et al.* Trials 2012

Per-protocol effects can be biased
Frequently done by exluding those not adhering[1]
Estimation of the per-protocol effect is much like an observational study
Thus, bias can arise from non-exchangeability, non-consistency, or non-positivity
This can lead to biased estimates
[1] Cole et al. JAMA Net. Open 2023, Dodd et al. Trials 2012

- Per protocol effects are typically estimated by excluding those that deviate from the protocol
- However, these estimates should be thought of more like analyzing an observational study where common causes of the outcome of interest and adherence should be accounted for–this has been well reported since 2001 by Robins and Finkelstein
- As a result of this per protocol estimates can be biased.

Per-protocol effects can be biased



Can be addresed: inverse probability weighting (IPW), the g-formula, TMLE, can be used to account for selection bias (see next talk)

[0] $\mathbb{X}$: vector of covariates

Per-protocol effects can be biased
$\mathbb{X}$
Randomization
Treatment Adherence
Outcome
Selection
Can be addresed: inverse probability weighting (IPW), the g-formula, TMLE, can be used to account for selection bias (see next talk)
[0] $\mathbb{X}$: vector of covariates

- Here is a simple DAG illustrating the problem. If per protocol analyses are done where only those who adhere are included you create a selection bias.
- Here this is by conditioning on a selection which is downstream of a collider, treatment adherence.
- However instead of restricting to only those that adhere there are ways around this which will be covered by the other speakers.

Per Protocol Causal Identification

- Conditional Exchangeability:
  $Y^g \perp\!\!\!\perp (A_t, C_{t+1}) \mid \left(\bar{A}_{t-1} = \bar{a}^g_{t-1}, \bar{L}_t = \bar{\ell}_t, C_t = Y_t = 0\right) \quad \forall t$
- Consistency: if $\bar{A}_t = \bar{A}^g_t$ then $\bar{Y}_t = \bar{Y}^g_t$
- Positivity: $f(a^g_t, C_t = 0 \mid \bar{a}^g_t, \bar{\ell}_t, C_t = Y_t = 0) > 0$ where $f(\bar{a}^g_t, \bar{\ell}_t, C_t = Y_t = 0) > 0 \quad \forall t$
- No interference: $\bar{A}^g_{it} \perp\!\!\!\perp \bar{Y}^g_{jt}$ where $i \neq j$
- No missclassification and correct model specification

Where, C=censoring, t =time point from $0 \ldots t$, g is a deterministic treatment strategy, overbar denotes history of values

[0]Wen *et al.* Biometrics 2019

Per Protocol Causal Identification

- Conditional Exchangeability: $Y^g \perp\!\!\!\perp (A_t, C_{t+1}) \mid \left(\bar{A}_{t-1} = \bar{a}^g_{t-1}, \bar{L}_t = \bar{\ell}_t, C_t = Y_t = 0\right) \quad \forall t$
- Consistency: if $\bar{A}_t = \bar{A}^g_t$ then $\bar{Y}_t = \bar{Y}^g_t$
- Positivity: $f(a^g_t, C_t = 0 \mid \bar{a}^g_t, \bar{\ell}_t, C_t = Y_t = 0) > 0$ where $f(\bar{a}^g_t, \bar{\ell}_t, C_t = Y_t = 0) > 0 \quad \forall t$
- No interference: $\bar{A}^g_{it} \perp\!\!\!\perp \bar{Y}^g_{jt}$ where $i \neq j$
- No misclassification and correct model specification

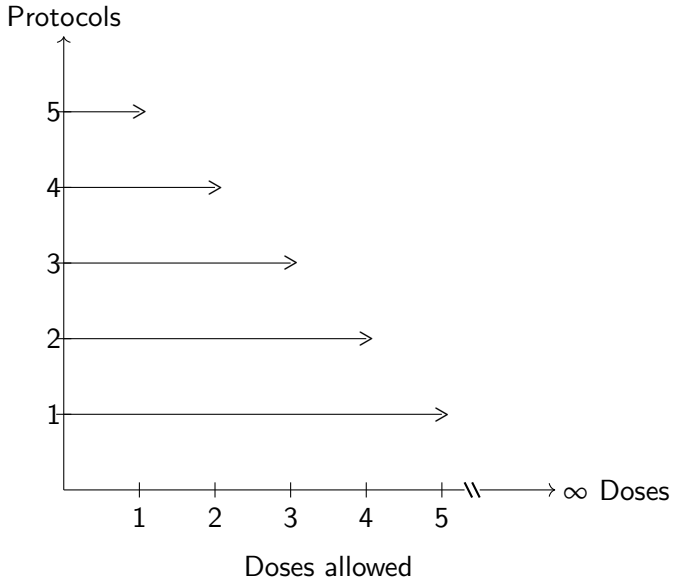
Where, C=censoring, t =time point from $0 \ldots t$, g is a deterministic treatment strategy, overbar denotes history of values

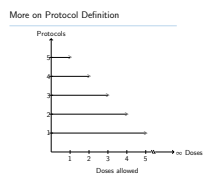
[0]Wen *et al.* Biometrics 2019

The causal idenfication criteria are slightly modified to account for the treatment regimens described by the trial protocol. This allows for us to take time into account

- now we assume that adherence and censoring is independent of an individuals counterfacutal outcome conditional on adherence under a specific treatment regimen at all prior times,and covariate values at all time points
- Consistency is now that your counterfactual outcome for a treatment regimen is the outcome observed under that treatment regimen
- Positivity requires nonzero joint probability of treatment adherence and being uncensored at each time point conditional on covariates
- No interference at each time point.
- and of course no missclassification or model misspecification

More on Protocol Definition





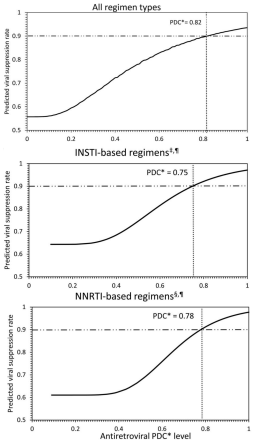
- Before I mentioned that there is both no one protocol, but also that this definition can be modulated; This figure attempts to illustrate this point.
- On the x-axis there are the number of doses allowed before the protocol has been violated and censoring occurs. The ITT analysis would be infinite number of doses, becuase the only thing that matters is randomization assignment.
- Person one is allowed to miss 4 doses and on the 5th missed dose or treatment they are censored. Other protocols are still possible, for instance protocol 2 only allows 3 missed doses and porotocl 5 allows none, and a participant is censored as soon as they miss one dose.
- We will use this approach in the example I will highlight next

An example using an HIV Trial: AIDS Clinical Trial Group (ACTG) 5202

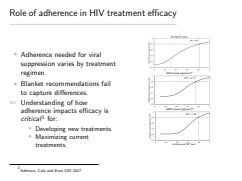
Ill now go over some results from a per protocol analysis that varies how protocol is defined in order to illustrate my point.

Role of adherence in HIV treatment efficacy

- Adherence needed for viral suppression varies by treatment regimen.
- Blanket recommendations fail to capture differences.
- ☞ Understanding of how adherence impacts efficacy is *critical*[1] for:
  - Developing new treatments.
  - Maximizing current treatments.



[1] Adimora, Cole and Eron CID 2017



The efficacy of HIV treatments depends both on the mechanism of the medication and the adherence to a dosing scheme. In general it is not possible to make a blanket statement that any number of missed doses per time-frame is ok or not. Thus evaluating adherence in in terms of doses taken should be considered.

Example: ACTG 5202 Reanalysis

Overview:

- Phase 3b RCT
- 59 sites, US and Puerto Rico

Objective:

- Per-protocol analysis modulating protocol definition.
- Target population: The population of HIV+ individuals in the United States
- Estimand: $E\left[Y^{r=1,\bar{a}=1}\right] - E\left[Y^{r=0,\bar{a}=0}\right]$ at 48 and 96 weeks

Outcomes: Virologic Failure defined as:

- plasma HIV-1 RNA level ≥1000 copies /mL between 16 weeks and 24 weeks
- or ≥200 copies/mL at or after 24 weeks

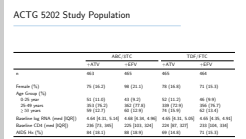
Example: ACTG 5202 Reanalysis
Overview:
Phase 3b RCT
59 sites, US and Puerto Rico
Objective:
Per-protocol analysis modulating protocol definition.
Target population: The population of HIV+ individuals in the United States
Estimand: $E\left[Y^{r=1,\bar{a}=1}\right] - E\left[Y^{r=0,\bar{a}=0}\right]$ at 48 and 96 weeks
Outcomes: Virologic Failure defined as:
plasma HIV-1 RNA level ≥1000 copies /mL between 16 weeks and 24 weeks
or ≥200 copies/mL at or after 24 weeks

The ACTG 5202 study was a phase 3 trial throughout the US and Puerto Rico. The findings were published in 2009 and then in 2011

- We reanalyzed this data with the goal to evaluate treatment effects in under multiple protocol definitions to estimate treatment efficacy in the population of HIV+ person in the United states.
- We aimed to estimate a 48 and 96 week risk difference.

ACTG 5202 Study Population

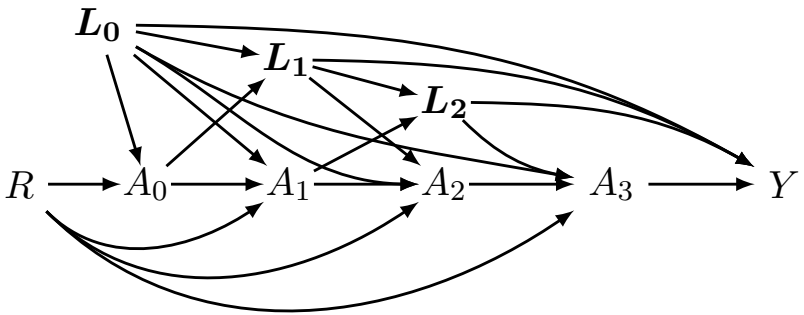
	ABC/3TC		TDF/FTC	
	+ATV	+EFV	+ATV	+EFV
n	463	465	465	464
Female (%)	75 (16.2)	98 (21.1)	78 (16.8)	71 (15.3)
Age Group (%)				
0-25 year	51 (11.0)	43 (9.2)	52 (11.2)	46 (9.9)
25-49 years	353 (76.2)	362 (77.8)	339 (72.9)	356 (76.7)
≥ 50 years	59 (12.7)	60 (12.9)	74 (15.9)	62 (13.4)
Baseline log RNA (med [IQR])	4.64 [4.31, 5.14]	4.68 [4.34, 4.96]	4.65 [4.31, 5.05]	4.65 [4.35, 4.91]
Baseline CD4 (med [IQR])	236 [73, 345]	225 [103, 324]	224 [87, 327]	233 [104, 334]
AIDS Hx (%)	84 (18.1)	88 (18.9)	69 (14.8)	71 (15.3)


ACTG 5202 Study Population

participants were Randomized 1:1:1:1 to

-
- TDF/FTC + (EFV or ATV/r) + ABC/3TC placebo
- ABC/3TC + (EFV or ATV/r) + TDF/FTC placebo
- Stratified by HIV-1 RNA screening level of
- $< 100,000$
- or $\geq 100,000$

Identification Reminder with DAG



j: follow up time, $\mathbf{L_j}$: vector of covariates at follow up time j
A_j: tx adherence at follow up time j, R: randomization, Y: viral failure or death

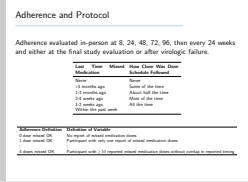
Identification Reminder with DAG
j: follow up time, $\mathbf{L_j}$: vector of covariates at follow up time j
A_j: tx adherence at follow up time j, R: randomization, Y: viral failure or death

Adherence and Protocol

Adherence evaluated in-person at 8, 24, 48, 72, 96, then every 24 weeks and either at the final study evaluation or after virologic failure.

Last Time Missed Medication	**How Close Was Dose Schedule Followed**
Never	Never
>3 months ago	Some of the time
1-3 months ago	About half the time
2-4 weeks ago	Most of the time
1-2 weeks ago	All the time
Within the past week	

Adherence Definition	**Definition of Variable**
0 dose missed OK	No report of missed medication doses
1 dose missed OK	Participant with only one report of missed medication doses
⋮	⋮
4 doses missed OK	Participant with ≥ 10 reported missed medication doses without overlap in reported timing

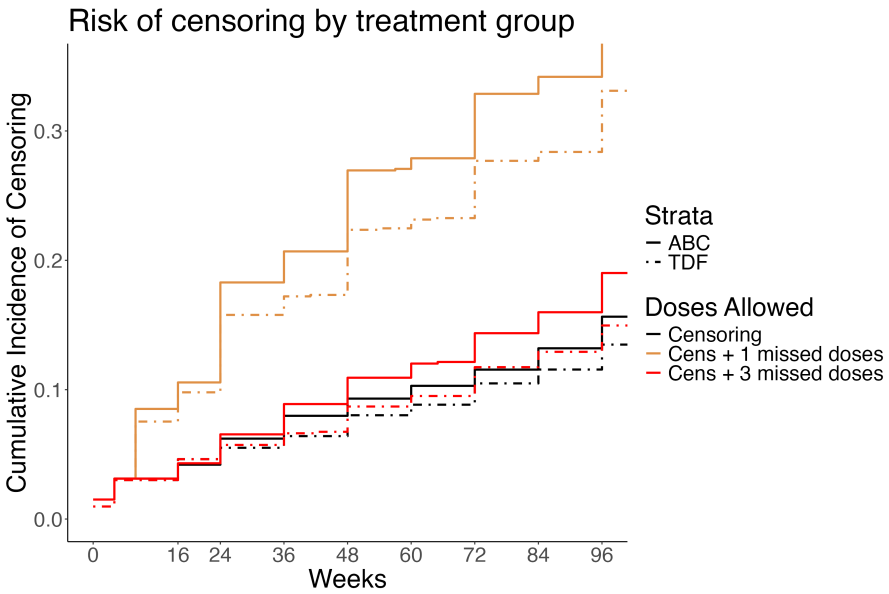


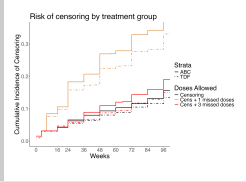
Deviation from Defined Protocols

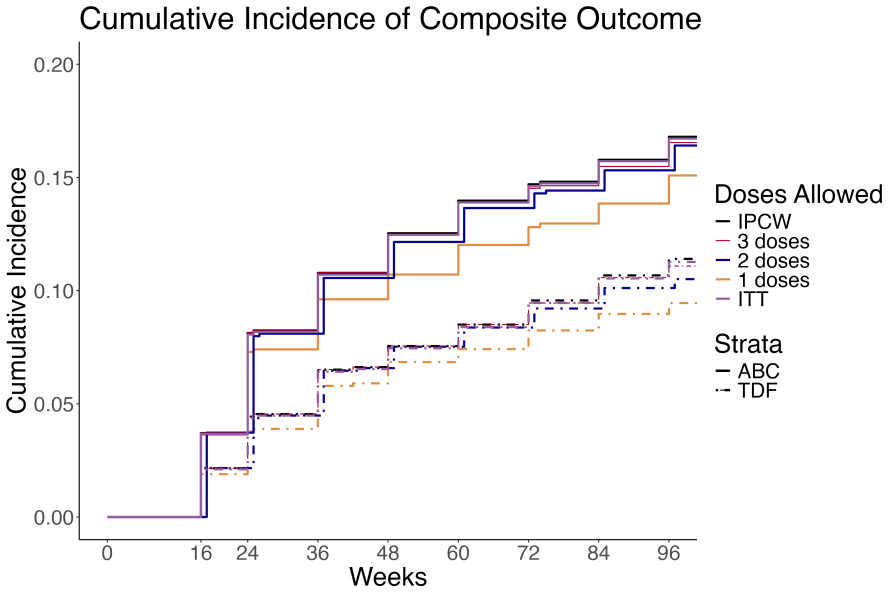
Treatment Group	Censored	1 Dose	2 Dose	3 Dose	4 Dose	5 Dose	Total
ABC/3TC	234	276	110	57	18	7	928
TDF/FTC	211	263	79	38	23	7	929

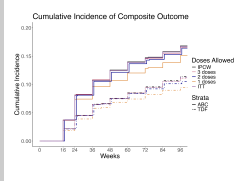
Deviation from Defined Protocols

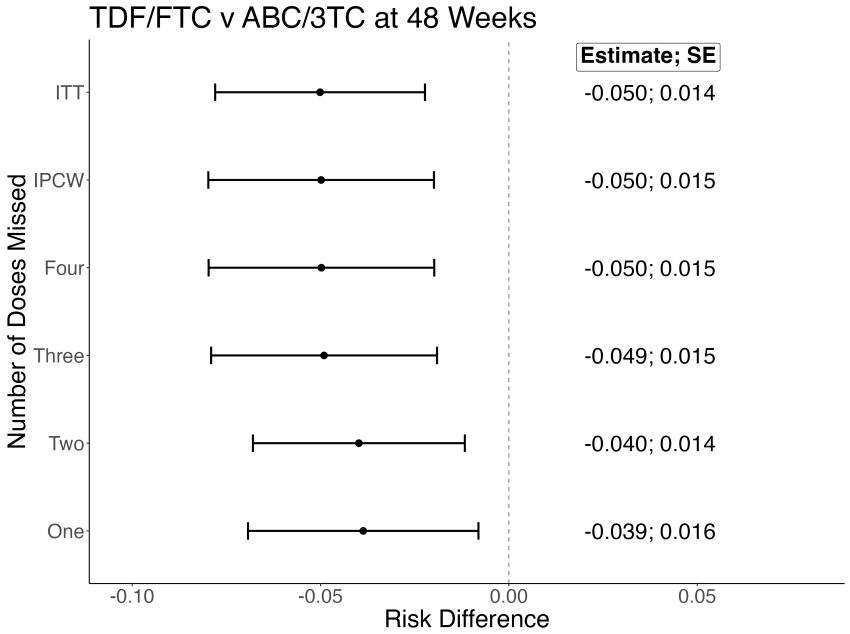
Treatment Group	Censored	1 Dose	2 Dose	3 Dose	4 Dose	5 Dose	Total
ABC/3TC	234	276	110	57	18	7	928
TDF/FTC	211	263	79	38	23	7	929

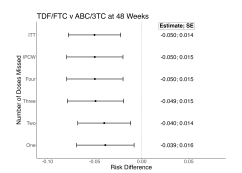

Risk of censoring by treatment group
Cumulative Incidence of Censoring
0.0
0.1
0.2
0.3
0
16
24
36
48
60
72
84
96
Weeks
Strata
ABC
TDF
Doses Allowed
Censoring
Cens + 1 missed doses
Cens + 3 missed doses

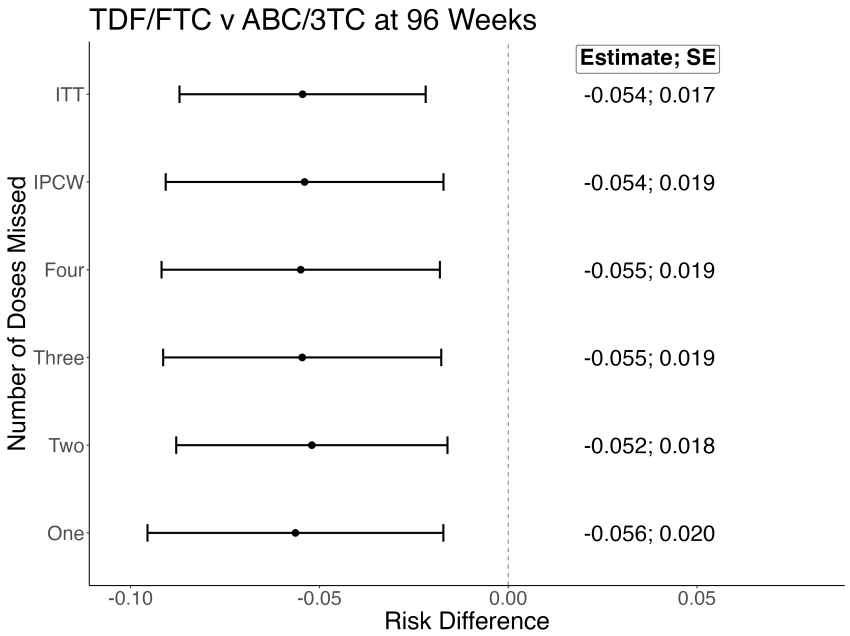


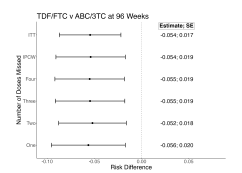

Cumulative Incidence of Composite Outcome
Cumulative Incidence
0.20
0.15
0.10
0.05
0.00
0
16
24
36
48
60
72
84
96
Weeks
Doses Allowed
IPCW
3 doses
2 doses
1 doses
ITT
Strata
ABC
TDF




TDF/FTC v ABC/3TC at 48 Weeks
Estimate; SE
ITT
IPCW
Four
Three
Two
One
-0.050; 0.014
-0.050; 0.015
-0.050; 0.015
-0.049; 0.015
-0.040; 0.014
-0.039; 0.016
Number of Doses Missed
-0.10
-0.05
0.00
0.05
Risk Difference




TDF/FTC v ABC/3TC at 96 Weeks
Estimate; SE
ITT
IPCW
Four
Three
Two
One
-0.054; 0.017
-0.054; 0.019
-0.055; 0.019
-0.055; 0.019
-0.052; 0.018
-0.056; 0.020
Number of Doses Missed
-0.10
-0.05
0.00
0.05
Risk Difference


TDF/FTC v ABC/3TC at 96 Weeks
Estimate; SE
ITT
IPCW
Four
Three
Two
One
-0.054; 0.017
-0.054; 0.019
-0.055; 0.019
-0.055; 0.019
-0.052; 0.018
-0.056; 0.020
Number of Doses Missed
-0.10
-0.05
0.00
0.05
Risk Difference

Limitations and Future Plans

1. Completed with public access data[1]
2. Reliance on coarse, self-reported medication adherence
3. Assume identification conditions met.[2]
4. Future directions include repeating analysis with g-formula, considering additional protocols.

[1] Approved for more granular data from ACTG, awaiting dataset
[2] NB: not guaranteed in per-protocol setting even though it is a trial

Limitations and Future Plans

1. Completed with public access data[1]
2. Reliance on coarse, self-reported medication adherence
3. Assume identification conditions met.[2]
4. Future directions include repeating analysis with g-formula, considering additional protocols.

[1] Approved for more granular data from ACTG, awaiting dataset
[2] NB: not guaranteed in per-protocol setting even though it is a trial

Thank you!

UNC | GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

I'd like to acknowledge:

- Steve Cole
- Paul Zivich
- Catherine Li
- ACTG 5202
- ACTG 5202 Particpants
- Cole Lab Members

My website where you can find a link to my github.

Thank you!
UNC | GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH
I'd like to acknowledge:
- Steve Cole
- Paul Zivich
- Catherine Li
- ACTG 5202
- ACTG 5202 Particpants
- Cole Lab Members
My website where you can find a link to my github.