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

GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

2024-06-14

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

GILLINGS SCHOOL OF GLOBAL PUBLIC HEALTH

Outline

Outline

RCTs

Per-protocol effects

Example using ACTG 5202 Trial
Population
Analysis Plan
Results
Limitations and Future

Appendix

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: $Y = Y^{a=1}A + Y^{a=0}(1 - A)$
- Positivity[1]: $Pr(A = a) > 0$, $\forall \ell$ where $f(a) > 0$

☞ This allows for unbiased estimation of treatment effects.

[1] $\mathbf{L}$ is covariate vector

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: $Y = Y^{a=1}A + Y^{a=0}(1 - A)$
- Positivity[1]: $Pr(A = a) > 0$, $\forall \ell$ where $f(a) > 0$
☞ This allows for unbiased estimation of treatment effects.
[1] $\mathbf{L}$ 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: $E[Y^{r=1}] - E[Y^{r=0}]$
- This is the effect of treatment assignment on outcomes
- Public health focused

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

r = randomization; $\bar{a}$ = history of treatment adherence
[1]Hernan and Robins, NEJM 2016

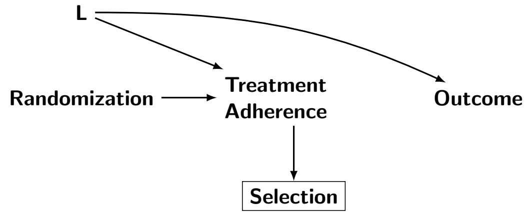
RCT estimands
- Intention-to-treat (ITT) effect: $E[Y^{r=1}] - E[Y^{r=0}]$
- This is the effect of treatment assignment on outcomes
- Public health focused
- *Typical* Per-protocol (PP) effect: $E[Y^{r=1,\bar{a}=1}] - E[Y^{r=0,\bar{a}=0}]$
- This is the effect of treatment assignment and adherence on outcomes
- Patient focused[1]
r = randomization; $\bar{a}$ = history of treatment adherence
[1]Hernan and Robins, NEJM 2016

- 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.

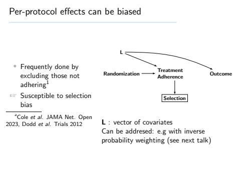
Per-protocol effects can be biased

- Frequently done by excluding those not adhering[1]
- ☞ Susceptible to selection bias

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



L : vector of covariates
Can be addresed: e.g with inverse probability weighting (see next talk)



- 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.
- 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.

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 *at least* 6 per-protocol parameters that can be estimated[1]
- There are also $k \in \{1, \ldots, \infty\}$ protocols 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 *at least* 6 per-protocol parameters that can be estimated[1]
- There are also $k \in \{1, \ldots, \infty\}$ protocols depend on how the investigator(s) define adherence.

[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 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 $0 \dots t$, 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 missclassification and correct model specification
Where, C=censoring, t =time point from $0 \dots t$, 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

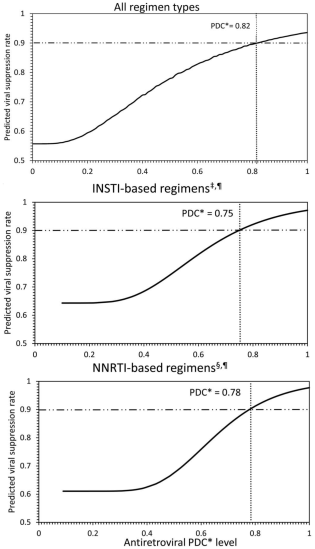
An example using an HIV Trial:
AIDS Clinical Trial Group (ACTG) 5202

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

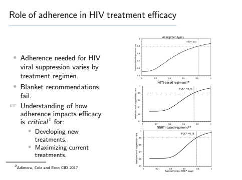
Ill now go over some results from a per protocol analysis that varies how protocol is defined in order to illustrate my point.

Role of adherence in HIV treatment efficacy

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



[a] 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.

ACTG 5202 Study Population

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

	ABC/3TC	TDF/FTC
N	928	929
Male at birth %	81.4	84.0
Age Group %		
≤ 25	10.1	10.5
26-49	77.0	74.8
≥ 50	12.8	14.6
Baseline $\log_{10}$ RNA copies/mL(med [IQR])	4.66 [4.31, 5.06]	4.65 [4.34, 4.96]
Baseline CD4 count/mL (med [IQR])	229 [84, 338]	230 [97, 330]

ACTG 5202 Study Population

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

	ABC/3TC	TDF/FTC
N	928	929
Male at birth %	81.4	84.0
Age Group %		
≤ 25	10.1	10.5
26-49	77.0	74.8
≥ 50	12.8	14.6
Baseline $\log_{10}$ RNA copies/mL(med [IQR])	4.66 [4.31, 5.06]	4.65 [4.34, 4.96]
Baseline CD4 count/mL (med [IQR])	229 [84, 338]	230 [97, 330]

Phase 3 RCT in 59 sites in the US and Puerto Rico participants Randomized 1:1 TDF/FTC or ABC/3TC

- 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 $\geq 100,000$

Example: ACTG 5202 Reanalysis

Objective:

- Per-protocol analyses modulating protocol definition.
- Estimand: $E\left[Y^{r=1,\bar{a}=1}\right] - E\left[Y^{r=0,\bar{a}=0}\right]$ at 48 weeks

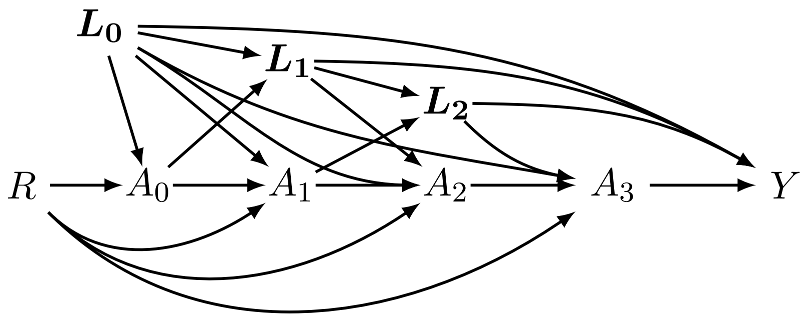
Outcomes: Composite Virologic Failure and all-cause mortality:

Example: ACTG 5202 Reanalysis
Objective:
Per-protocol analyses modulating protocol definition.
Estimand: $E\left[Y^{r=1,\bar{a}=1}\right] - E\left[Y^{r=0,\bar{a}=0}\right]$ at 48 weeks
Outcomes: Composite Virologic Failure and all-cause mortality:

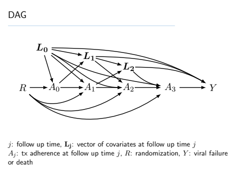
The ACTG 5202 study was a phase 3 trial throughout the US and Puerto Rico. The findings were published in 2009 and then in 2011

- 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 the risk difference at 48 weeks

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



A simplified DAG illustrates our assumptions. We are assuming that randomization to treatment impacts adherence, and that baseline covariate values and covariate values at the preceeding time point also impacts adherence and the outcome.

we will use IPW in the analysis here to account for these factors that leads to differences in adherence and the outcome

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

Adhrence in ACTG 5202 was collecetd by self report and whether or not there were missed doses within a previous time frame. There was also information about how closesly a dosing regimen was followed but we did not incorporate that here.

We defined protocol deviation, illustrated in the bottom table, by how many missed doses were acceptable. For instance if 0 doses missed were OK, then as soon as a person reported missing any doses they were censored. For those where missing 4 doses was ok, as soon as they missed the fifth dose they were censored.

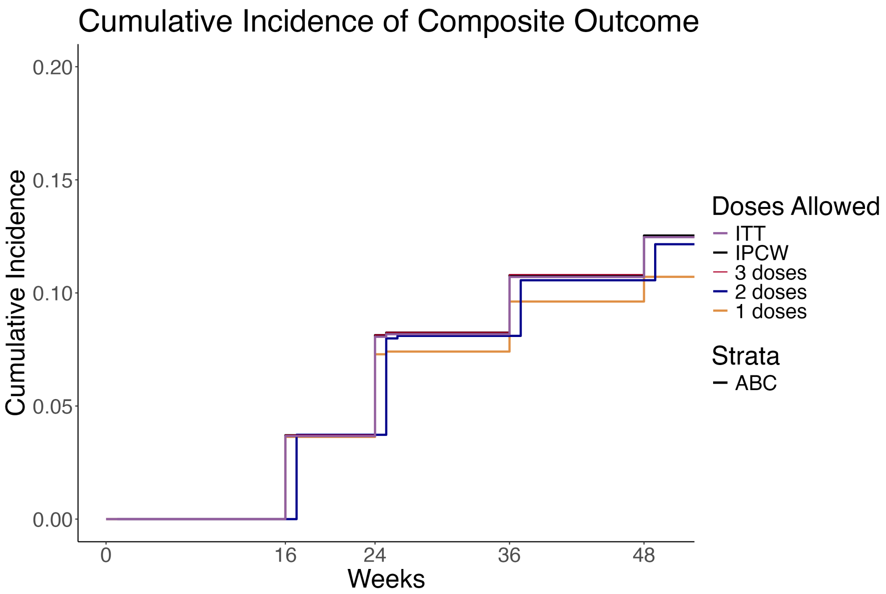
Deviation from Defined Protocols

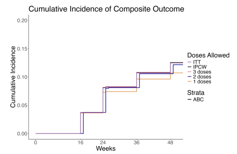
Treatment Group	Censored	1 Dose	2 Dose	3 Dose	4 Dose	5 Dose	Total
ABC/3TC	234	276	110	57	18	7	928
TDF/FTC	211	263	79	38	23	7	929

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

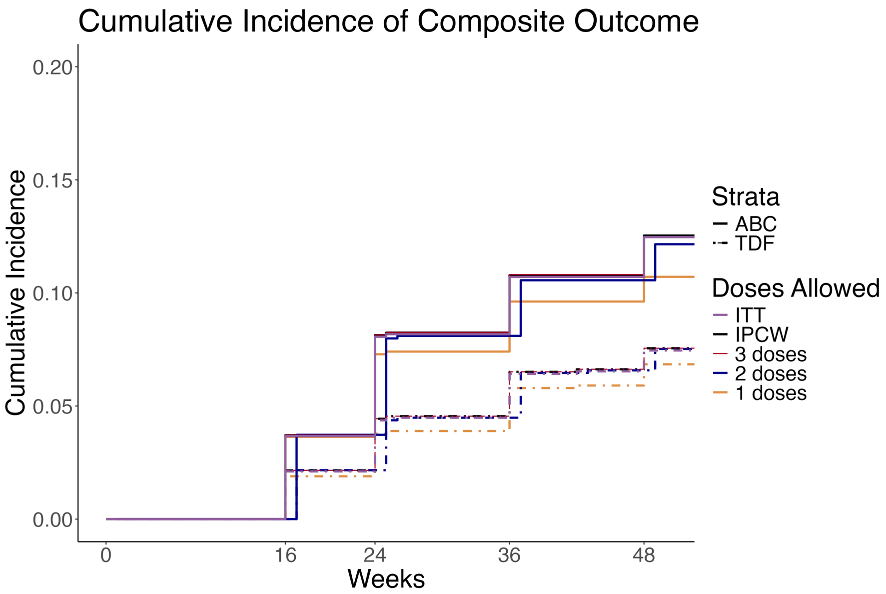
As might be expected as the definition of protocol deviation becomes more strict, requiring more missed doses to be censored, there is a decrease in the number of participants deviating from the protocol. This goes from 276 in the ABC arm and 263 in the TDF arm for the 1 dose protocol down to 7 in each arm under the 5 dose protocol definition.

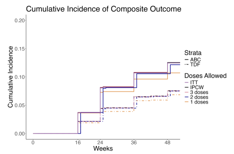




now we turn our attention to the risk of viral failure and death. There are two notable findings here.

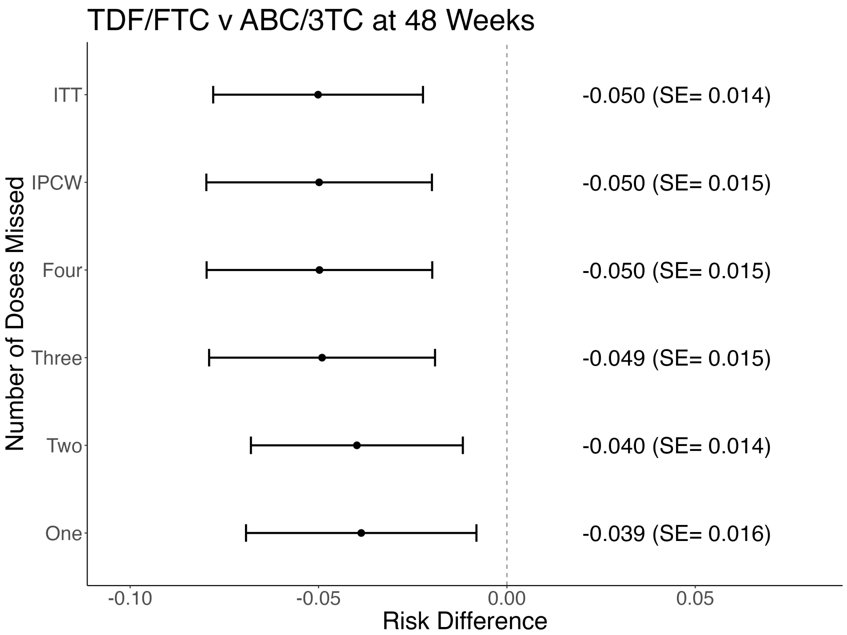
1. the absolute risk of an outcome in both arms under the 1 dose missed protocol is lower suggesting higher risk people are missing 1 dose of medication.
2. There is a smaller difference in the risk when accounting for 1 dose protocol deviation. This is a bit harder to see here, but will become a little more apparent on the next slide.

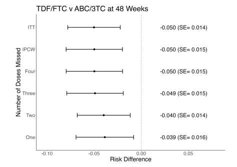




now we turn our attention to the risk of viral failure and death. There are two notable findings here.

1. the absolute risk of an outcome in both arms under the 1 dose missed protocol is lower suggesting higher risk people are missing 1 dose of medication.
2. There is a smaller difference in the risk when accounting for 1 dose protocol deviation. This is a bit harder to see here, but will become a little more apparent on the next slide.





here we can see the risk differences and 95% confidence intervals based on robust standard errors. As expected as we censored individuals the standard error increased. The changes in the risk difference become more apparent with smaller risk differences when accounting for 1 and 2 dose deviations. This suggest the efficacy of the TDF versus ABC is not quite as large as you might expect if you relied only on a naive analysis.

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

This is a work in progress and There were some limitations.

1. this analysis was completed with public data. We have approval from ACTG to obtain more granular data and we will update our results with that
2. currently this analysis relied on coarse data on adherence and I am currently working to improved estimates of the number of doses that people actually missed.
3. we assume that we have met all the required identification condigions based on covariates in the IPW models.

Our future directions are to also complete this analysis using the g-formula, and also consider additional granular data once the ACTG provides us with an updated data set

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.

Thank you for your time. Id like to acknowledge the Steve Cole, the cole lab, Paul Zivich, Catherine Li and all the participants of the ACTG 5202 study whom without their participation this work would not be possible.

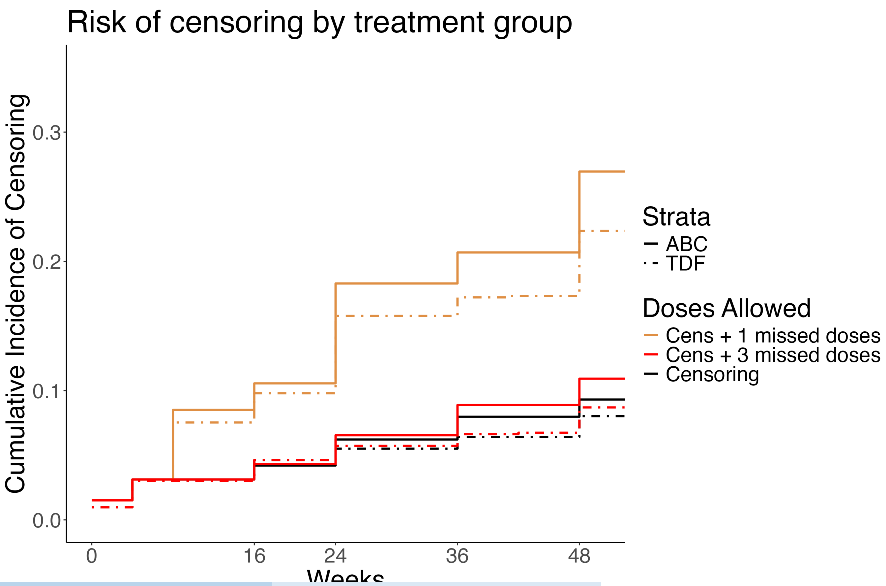
Outcome Definition

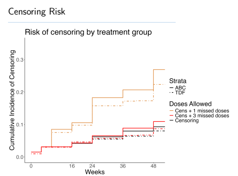
- plasma HIV-1 RNA level ≥1000 copies /mL between 16 weeks and 24 weeks
- or ≥200 copies/mL at or after 24 weeks
- all cause mortality

Outcome Definition

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

Censoring Risk





here are risk curves for censoring and censoring +protocol deviation. The black line shows only the risk for loss to follow up and there is little difference between the two arms. The red line shows a similar pattern that overlaps closely with the censoring arm. We would expect the more doses required to be missed to approximate the censoring arm. However, if you look at the yellow line the overall risk of being censorid is higher and there is a slightly higher risk in the ABC arm versus the TDF arm in the later time points.