Fusion to address systematic errors across data sources

Paul Zivich

Department of Epidemiology
Causal Inference Research Laboratory
UNC Gillings School of Global Public Health

May 23, 2022

Acknowledgements

Supported by NIH T32-AI007001.

Thanks to Stephen Cole, Jessie Edwards, Bonnie Shook-Sa, Alex Breskin, Michael Hudgens, and others at the UNC Causal Inference Research Lab (causal.unc.edu).

Works in-progress, so errors are mine.[1]

✉ pzivich@unc.edu pzivich @PausalZ

[1] Footnotes are reserved for asides or references

Background

Motivation

Recent developments in quantitative methods

- All promise to fundamentally improve how we learn
- Examples: causal inference, machine learning, big data

However, study design remains the foundation

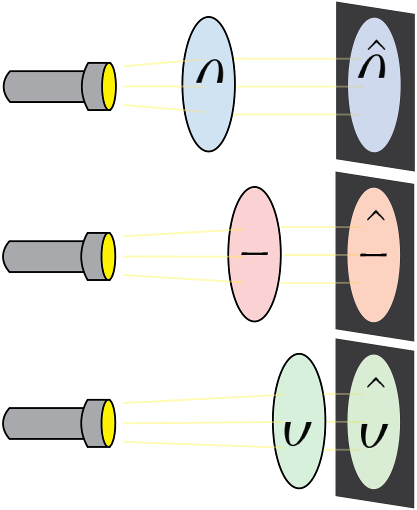
Definitions of Fusion Study Designs

Combine heterogeneous data sources to answer a question that could not be answered (as well) by any subset[2]

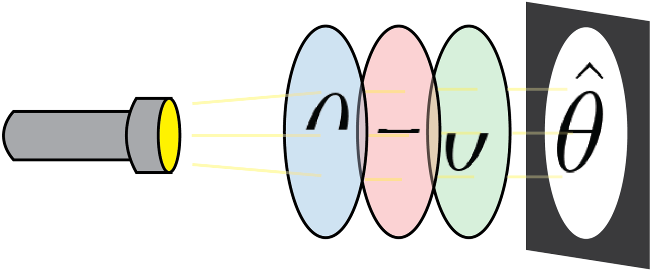
- Meta-analysis: combine 'similar' studies to reduce random error
- Fusion: combine (possibly dissimilar) studies to reduce systematic and random error

[2] Cole et al. *Am J Epi* (In-press)

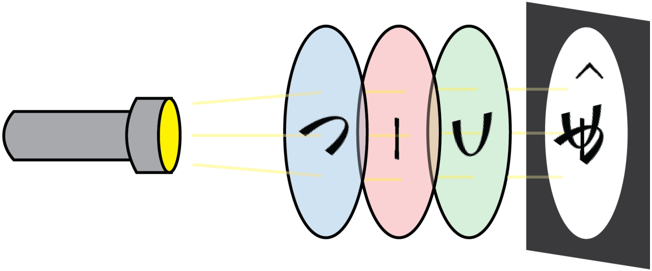
A Visual Analogy



A Visual Analogy



A Visual Analogy



A Didactic Example

Motivating Question

A collaborator asks us to help them estimate the mean of some variable (Y) for a defined population ($S = 1$). However, Y was not measured in the target population.[3]

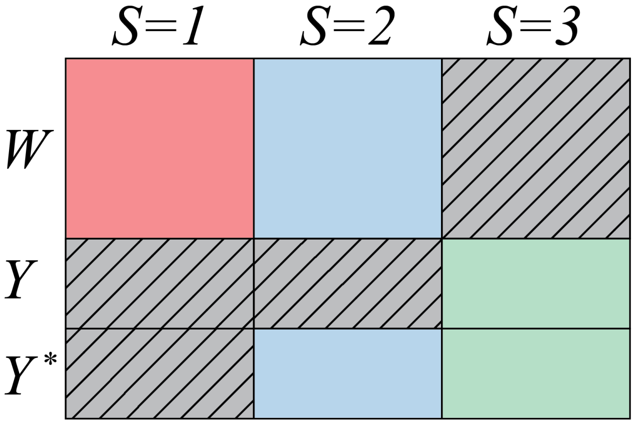
However, several sources of partially overlapping information are available

So, under what assumptions could $\mu = E[Y | S = 1]$ be estimated?[4]

[3] This problem is a simplified version of transportability

[4] A variation of this is presented in Cole et al. (in-press) *Am J Epidemiol*

Available Information



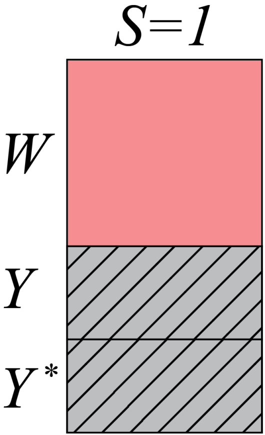
	$S=1$	$S=2$	$S=3$
W	🟥	🟦	▨
Y	▨	▨	🟩
Y^*	▨	🟦	🟩

Approach 1

Single data source: sample of $S = 1$

- Y not measured
- So can make no further progress

$$\hat{\mu}_1 = \varnothing$$



Approach 2

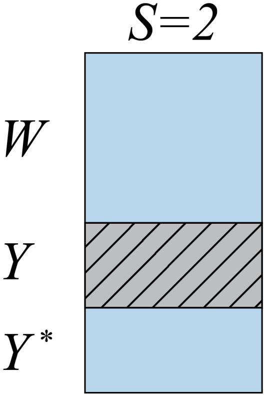
Single data source: sample of $S = 2$

- Y not measured
- Mismeasured Y, Y^*, is available

Assumptions[5]

- $Y = Y^*$
- $E[Y|S = 1] = E[Y|S = 2]$

$$\hat{\mu}_2 = n_2^{-1} \sum_i I(S_i = 2) Y_i^*$$



[5] Webster-Clark & Breskin (2021) *Am J Epidemiol*

Approach 3

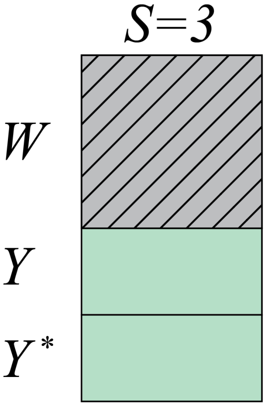
Single data source: sample of $S = 3$

- Y and Y^* were measured

Assumptions

- $E[Y|S=1] = E[Y|S=3]$

$$\hat{\mu}_3 = n_3^{-1} \sum_i I(S_i = 3) Y_i$$



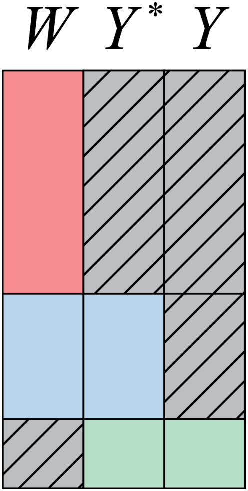
Fusion (Approach 4)

All data sources

- Sample of $S = 1$
  - Contribute W
- Sample of $S = 2$
  - Contribute W, Y^*
  - Measure of Y conditional on W
- Sample of $S = 3$
  - Contribute Y, Y^*
  - Account for measurement error

Not enough

- Need to combine data sources correctly



Fusion: Identification

Link $S = 1$ and $S = 2$:

- Conditional transportability assumptions[6]

$$E[Y|W, S = 1] = E[Y|W, S = 2]$$

$$\Pr(S = 2|W = w) > 0 \text{ where } \Pr(S = 1|W = w) > 0$$

Link $S = 2$ and $S = 3$:

- Non-differential measurement error

$$\Pr(Y^* = y|Y = y) = \Pr(Y^* = y|Y = y, W = w)$$

[6] Westreich et al. (2017) *Am J Epidemiol*

Fusion: Estimation

M-estimator:[7]

$$\sum_{i=1}^{n} \psi(O_i; \hat{\theta}) = 0$$

where $O_i = \{S_i, W_i, Y_i, Y_i^*\}$ and $\theta = (\mu, \eta)$

[7] See Stefanski & Boos (2002) *Am Stat* for an introduction

Fusion: Estimation

Stacked estimating equation[8]

$$\psi(O_i;\theta) = \begin{bmatrix} I(S_i = 3)\, Y_i\, (Y_i^* - \eta_1) \\ I(S_i = 3)(1 - Y_i)\,((1 - Y_i^*) - \eta_2) \\ I(S_i \neq 3)\,(I(S_i = 1) - \text{expit}(W_i \beta))\, W_i \\ I(S_i = 2)(Y_i^* - \eta_3)\frac{1 - \text{expit}(W_i\beta)}{\text{expit}(W_i\beta)} \\ \mu(\eta_1 + \eta_2 - 1) - (\eta_3 + \eta_2 - 1) \end{bmatrix}$$

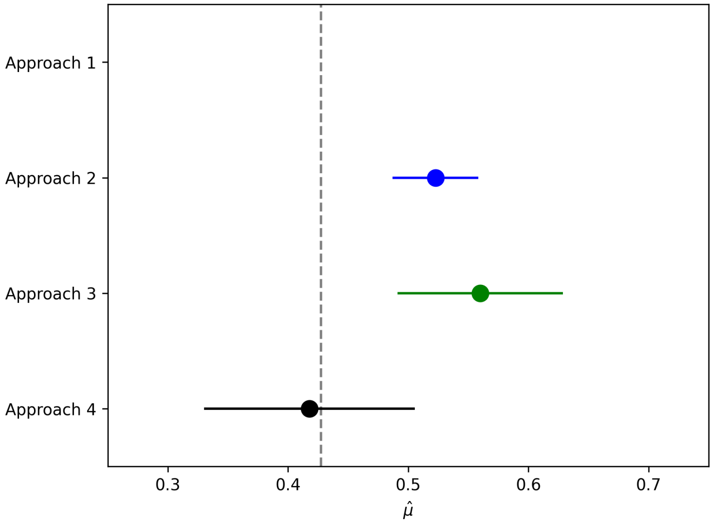
Sandwich variance estimator to estimate the variance.

Automated computation[9]

[8] Measurement correction from Rogan & Gladen (1978) *Am J Epidemiol*

[9] Python: *delicatessen* (Zivich et al. (2022) *arXiv*), R: *geex* (Saul & Hudgens (2020) *J Stat Softw*), SAS: PROC IML.

Example


Approach 1
Approach 2
Approach 3
Approach 4
0.3
0.4
0.5
0.6
0.7
$\hat{\mu}$

Bridged Comparisons

A Case Study

What is the risk difference at one-year of follow-up for AIDS, death, or more than a 50% decline in CD4 count if everyone had been assigned triple ART versus mono ART?

Data sources:

- ACTG 320
  - Randomized trial comparing triple to dual ART
- ACTG 175
  - Randomized trial comparing dual to mono ART

Target population: ACTG 320

Default Approach

Transitive Comparison

$$\text{Triple} > \text{Dual}$$
$$\text{Dual} > \text{Mono}$$
$$\therefore \text{Triple} > \text{Mono}$$

- Appealing argument
  - Fundamentally underlies many comparisons
- Often left implicit
  - Example: FDA approval following non-inferiority trial
- Formalization
  - Network meta-analysis, counterfactual placebos

Problem with Transitive Comparisons

Assumes "similar" target populations

- Marginal exchangeability between populations

Highly suspect assumption

- Sample from different populations
- Define endpoints differently
- Different rates of loss to follow-up or adherence

How can we relax this assumption?

Notation

$A_i = \{1, 2, 3\}$: ART regimens

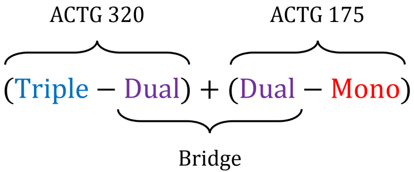
T_i^a: potential time of the event under treatment a
T_i: time of the event under assigned A_i
C_i: time of censoring
$T_i^* = \min(T_i, C_i), \quad \delta_i = I(T_i = T_i^*)$

W_i: set of baseline covariates
V_i: distinct set of baseline covariates
S_i: population membership, $\{0, 1\}$

$F_s^a(t) = \Pr(T^a < t | S = s)$

Bridged Treatment Comparisons

Make the indirect comparisons explicit via fusion[10]



ACTG 320 ACTG 175

(Triple − Dual) + (Dual − Mono)

Bridge

Estimand

$$\begin{aligned}
\psi(t) &= F_1^{3}(t) - F_1^{1}(t) \\
&= F_1^{3}(t) - F_1^{1}(t) + \big(F_1^{2}(t) - F_1^{2}(t)\big) \\
&= \big(F_1^{3}(t) - F_1^{2}(t)\big) + \big(F_1^{2}(t) - F_1^{1}(t)\big)
\end{aligned}$$

[10] See Breskin et al. (2021) *SIM* for details

Identification: Triple vs Dual

$$F_1^3(t) - F_1^2(t)$$

Treatment

$$T_i = T_i^a \text{ for } a = A_i$$

$$\Pr(T^a < t | S = 1) = \Pr(T^a < t | A = a, S = 1) \text{ for } a \in \{2, 3\}$$

$$\Pr(A = a | S = 1) > 0 \text{ for } a \in \{2, 3\}$$

Censoring

$$\Pr(T < t | A, W, S = 1) = \Pr(T < t | C > t, A, W, S = 1)$$

$$\Pr(C > T | A = a, W = w, S = 1) > 0 \; \forall \; \Pr(A = a, W = w | S = 1) > 0$$

Estimation: Triple vs Dual

Inverse probability weighting estimator[11]

$$\hat{F}^a_{320}(t) = n_{320}^{-1} \sum_{i=1}^{n} \frac{I(A_i = a) I(S_i = 1) I(T_i^* \leq t) \delta_i}{\Pr(A_i = a | S_i = 1) \pi_C(W_i, A_i, S_i; \hat{\alpha})}$$

for $a \in \{2, 3\}$, where

$$n_{320} = \sum_{i=1}^{n} I(S_i = 1)$$

$$\pi_C(W_i, A_i, S_i; \hat{\alpha}) = \Pr(C_i > t | W_i, A_i, S_i; \hat{\alpha})$$

[11] Identification implies estimation hereafter following stability from parametric or semiparametric restrictions. Estimation also requires correct model specification

Identification: Dual vs Mono

$$F_1^{2}(t) - F_1^{1}(t)$$

Similar identification assumption for treatment and censoring

- Unwilling to assume trials are random samples of same population

Transport[12]

$$\Pr(T^a < t | V, S = 1) = \Pr(T^a < t | V, S = 0)$$

$$\Pr(S = 0 | V = v) > 0 \text{ for all } v \text{ where } \Pr(S = 1 | V = v) > 0$$

[12] Simple transitivity arguments are the special case where $V = \emptyset$

Estimation: Dual vs Mono

$$\hat{F}^{a}_{175}(t) = \hat{n}_{175}^{-1} \sum_{i=1}^{n} \frac{I(A_i = a) I(S_i = 1) I(T_i^* \leq t)\delta_i}{\Pr(A_i = a | S_i = 1)\pi_C(W_i, A_i, S_i; \hat{\alpha})} \times \frac{1 - \pi_S(V_i; \hat{\beta})}{\pi_S(V_i; \hat{\beta})}$$

for $a \in \{1, 2\}$, where

$$\hat{n}_{175} = \sum_{i=1}^{n} I(S_i = 0) \frac{1 - \pi_S(V_i; \hat{\beta})}{\pi_S(V_i; \hat{\beta})}$$

$$\pi_S(V_i; \hat{\beta}) = \Pr(S_i = 1 | V_i; \hat{\beta})$$

Application to ACTG

Implemented in Python 3.6+[13]

$\pi_C(W_i, A_i, S_i; \hat{\alpha})$

- Stratified Cox PH model & Breslow estimator
- Stratified by trial and ART
- W: age, gender, race, injection drug use, Karnofsky score

$\pi_S(V_i; \hat{\beta})$

- Logistic regression
- $V = W$

[13] Using `NumPy`, `SciPy`, `statsmodels`

A Testable Implication

Identification strategy for ψ required

$$F_1^2(t) - F_1^2(t) = 0$$

which implies

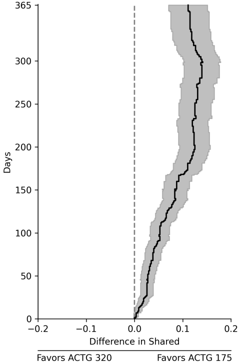
$$E\left[\hat{F}_{320}^2(t)\right] - E\left[\hat{F}_{175}^2(t)\right] = 0$$

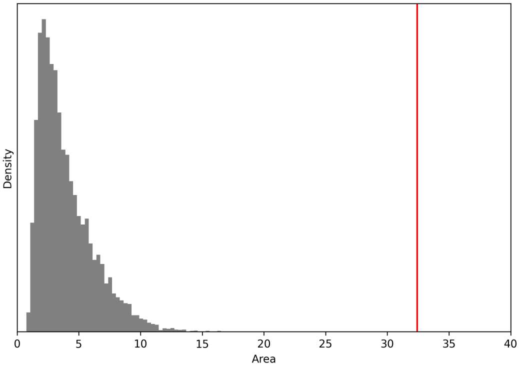
Therefore, can compare the shared arms

- If non-zero then at least one assumption is wrong
- Ways to assess
  - Graphically[14]
  - Numerically via a permutation test based on area between the risk functions

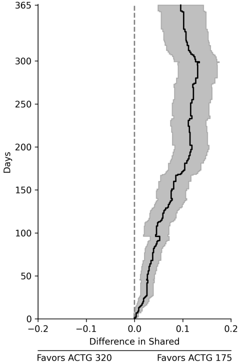
[14] Twister plot as described in Zivich et al. (2021) *Am J Epidemiol*

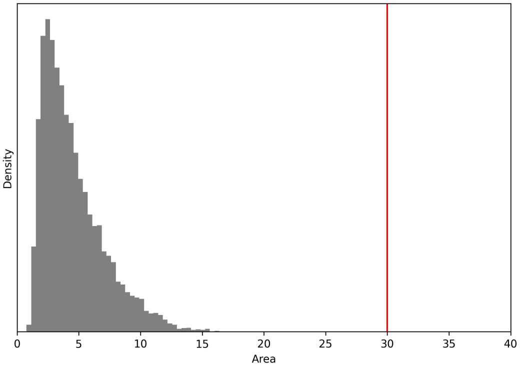
Testable Implication: Naive





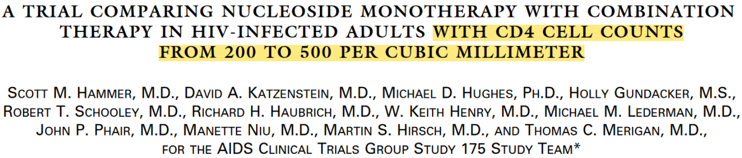
Testable Implication: transported





The Distinction

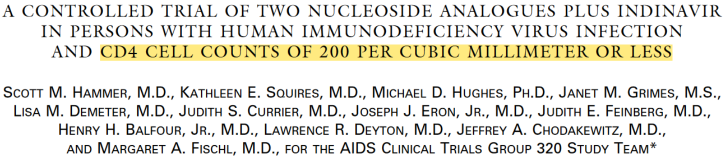
ACTG 175



A TRIAL COMPARING NUCLEOSIDE MONOTHERAPY WITH COMBINATION THERAPY IN HIV-INFECTED ADULTS WITH CD4 CELL COUNTS FROM 200 TO 500 PER CUBIC MILLIMETER

Scott M. Hammer, M.D., David A. Katzenstein, M.D., Michael D. Hughes, Ph.D., Holly Gundacker, M.S.,
Robert T. Schooley, M.D., Richard H. Haubrich, M.D., W. Keith Henry, M.D., Michael M. Lederman, M.D.,
John P. Phair, M.D., Manette Niu, M.D., Martin S. Hirsch, M.D., and Thomas C. Merigan, M.D.,
for the AIDS Clinical Trials Group Study 175 Study Team*

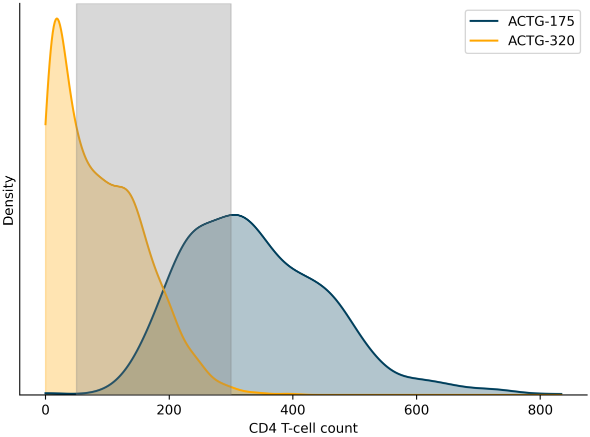
ACTG 320



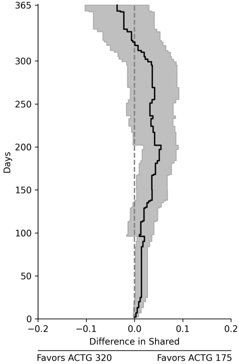
A CONTROLLED TRIAL OF TWO NUCLEOSIDE ANALOGUES PLUS INDINAVIR IN PERSONS WITH HUMAN IMMUNODEFICIENCY VIRUS INFECTION AND CD4 CELL COUNTS OF 200 PER CUBIC MILLIMETER OR LESS

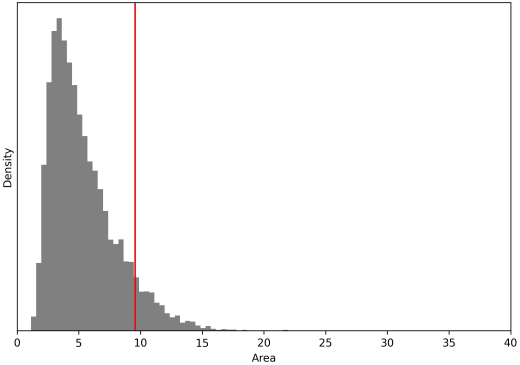
Scott M. Hammer, M.D., Kathleen E. Squires, M.D., Michael D. Hughes, Ph.D., Janet M. Grimes, M.S.,
Lisa M. Demeter, M.D., Judith S. Currier, M.D., Joseph J. Eron, Jr., M.D., Judith E. Feinberg, M.D.,
Henry H. Balfour, Jr., M.D., Lawrence R. Deyton, M.D., Jeffrey A. Chodakewitz, M.D.,
and Margaret A. Fischl, M.D., for the AIDS Clinical Trials Group 320 Study Team*

Baseline CD4 counts



Testable Implication: transported and CD4 restricted





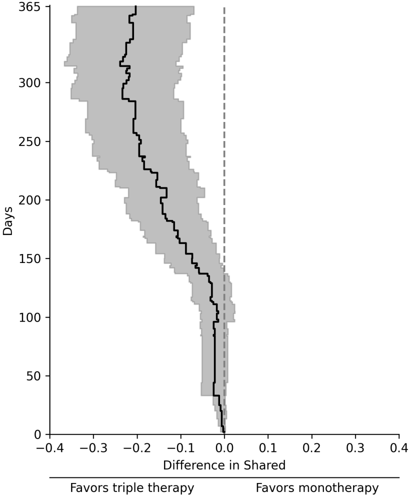
Estimation: Triple vs Mono

Estimator for parameter of interest[15]

$$\hat{\psi}(t) = \left(\hat{F}^{3}_{320}(t) - \hat{F}^{2}_{320}(t)\right) + \left(\hat{F}^{2}_{175}(t) - \hat{F}^{1}_{175}(t)\right)$$

[15] Variance estimator proposed in Breskin et al. (2021) *SIM*

Comparison of interest



Summary

Bridged comparisons offer

- Comparisons across trials
    - Analytical corrections for differences
- A testable condition

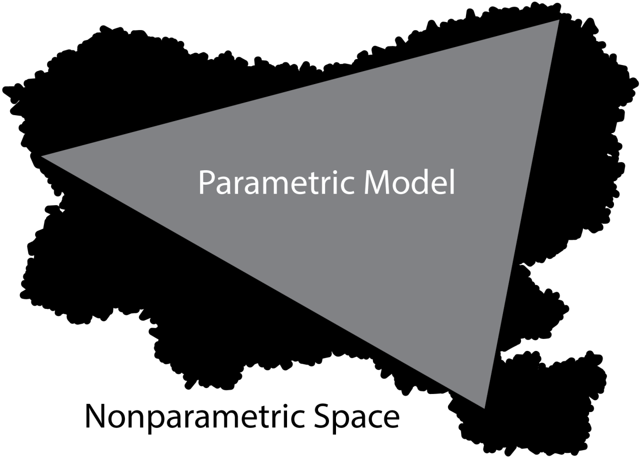
Future Work and Extensions

Ongoing Applications

Pre-exposure prophylaxis for the prevention of HIV

- TAF/FTC vs Placebo
  - Alternative tenofovir pro-drug
  - Comparison
    - DISCOVER (TAF/FTC vs. TDF/FTC)
    - iPrEx (TDF/FTC vs. Placebo)
- LA-CAB vs Placebo
  - Long-acting injectible
  - Comparison
    - HPTN-083 (LA-CAB vs. TDF/FTC)
    - iPrEx (TDF/FTC vs. Placebo)

Statistical Models



Other Extensions

Nested studies

Measurement error

- Other corrective approaches

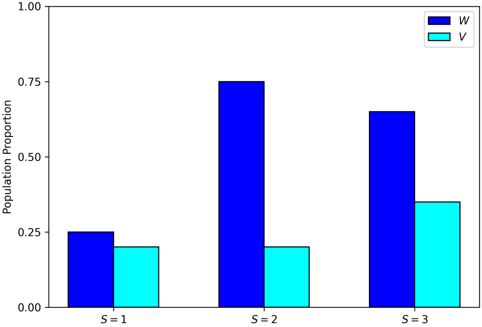
Diverse data sources

- Subject-matter knowledge
  - Semi-Bayes
- Pharmacokinetic data

Questions

Supplement

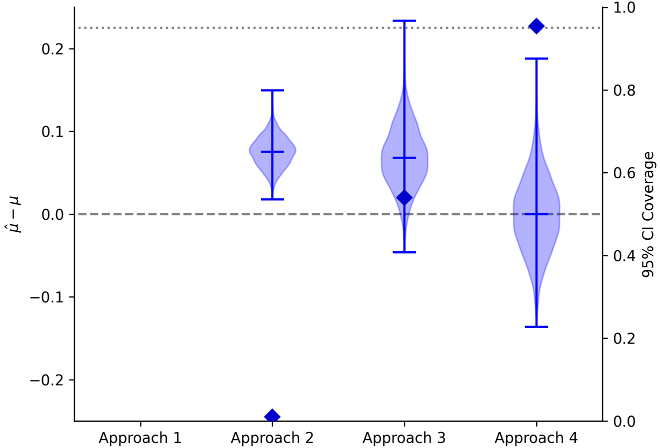
Didactic Simulation: Setup



$$\Pr(Y|W,V) = \text{logit}(-0.5 + 2W - V - 2WV + \epsilon)$$

$$\Pr(Y^*|Y) = 0.80X + (1 - 0.95)(1 - X)$$

Didactic Simulation: Results


$\hat{\mu} - \mu$
95% CI Coverage
Approach 1
Approach 2
Approach 3
Approach 4