Statistical and Practical Issues in Implementation Trials: Connecting the Estimand Framework to Implementation Science Trials

Erinn M. Hade
Associate Director & Associate Professor
Division of Biostatistics
Department of Population Health

Joint work with: Andrea Troxel, Anushree Iyengar, Benjamin Stockton, Natasha Williams, Dike Ojji, Mary Jo Messito, Gbenga Ogedegbe

May 20, 2026

NYU Langone Health

Disclosures

- Funding from NIH, PCORI, and the Educational Foundation of America.

- The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

NYU Langone Health

Today we will…..

- Review implementation science goals for randomized trials of implementation strategies
- Connect implementation science randomized trial to the estimands framework
- Illustrate these through two recent or ongoing implementation science cluster randomized trials.

Implementation Science & Goals

- Study of ***methods*** to increase **uptake** of evidence-based interventions
  - Context and environment

- Goal to ***improve delivery*** of an intervention and the **effectiveness** in key populations

- How do we help **people, institutions, places** utilize effective practices/treatments

Improve health care quality

NYU Langone Health

Implementation Science & Definitions

- Practice or Intervention: **what** we provide to improve health
  - *Shown to be efficacious in our discussion*

- Implementation strategy: **how** we provide/offer the intervention/practice
  - Components of how we deliver … to help people, providers, institutions use use the practice/intervention.

NYU Langone Health

Implementation Randomized Trial Examples

Managing Hypertension Among People Living with HIV: An Integrated Model (MAP-IT)

- ***HLB-SIMPLe Alliance***
- ***University of Abuja and NYUGSoM***

- To evaluate practice facilitation (PF) on the integration of a task strengthening strategy for HTN control (TASSH) into the HIV care platform.

Just Mothers Program

- ***NYUGSoM***
- ***NYC Health and Hospitals***
- ***NYU FHCs***

- To compare the adoption and clinical outcomes of an evidence based behavioral lifestyle intervention among clinics that receive the Digital health/low touch strategy compared with the Perinatal CHW-led/high touch strategy.

NYU Langone Health

Implementation Randomized Trial Examples

Managing Hypertension Among People Living with HIV: An Integrated Model (MAP-IT)

- Hybrid type III design
- **Intervention**: Task shifting protocol
- **Implementation strategy**:
  - Practice facilitation
  - No practice facilitation

Just Mothers Program

- Hybrid type III design
- **Intervention**: Prenatal and nutrition education program
- **Implementation strategies:**
  - Digital health/low touch strategy
  - CHW-led/high touch strategy.

NYU Langone Health

Implementation Science & Estimands

- ICH-E9 addendum (Estimands and Sensitivity Analysis) – R1 (2021)
- Aims to clarify research questions and the planned analysis framework

  – **Estimands** are a structured description of treatment effects that the study plans to measure.

  – **One or more estimands** is defined for each outcome

    - *Population*
    - *Treatment*
    - *Endpoint*
    - *Summary measure*
    - *Intercurrent events*

NYU Langone Health

Implementation Science & Estimands

How may these elements be unique randomized implementation science trials?

- *Population – cluster and individual level for some outcomes*
- *Treatment – implementation strategy*
- *Endpoint – adoption/uptake; reach; clinical measures*
- *Summary measure – difference in proportions, means, risk ratio, OR, HR*
- *Intercurrent events – refinements of the strategy, system wide implementations of another strategy, patient deaths, discontinuation of program, missed visits*

NYU Langone Health

MAP-IT: Adoption

- **Population** – Persons living with HIV with hypertension from 30 primary healthcare clinics in Akwa Ibom State of Nigeria.
- **Implementation strategy** – *Practice facilitation of the TASSH protocol by trained nurses as compared to self directed care.*
- ***Endpoint*** *– Adoption of TASSH protocol*
  - *When? Following 12 months of PF as compared to the 3 months prior to transition*
  - *How defined? Multi-component – all or some*
  - *Measured monthly at the clinic level – interest in cluster average effect to understand uptake*

NYU Langone Health

MAP-IT: Adoption

- ***Endpoint*** *– Adoption of TASSH*
  - *When? Following 12 months of PF as compared to the 3 months prior to transition*
  - *How defined to measure adoption? Multi-component – all or some*
  - *Measured monthly at the clinic level – interest in cluster average effect to understand uptake*

Block	2020				2021				2022				2023				2024				2025			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1	UG3 Phase and Pilot											SD	Tr	PF	PF	PF	PF	FU	FU	FU	FU	FU	FU	
2												SD	SD	Tr	PF	PF	PF	PF	FU	FU	FU	FU	FU	
3												SD	SD	SD	Tr	PF	PF	PF	PF	FU	FU	FU	FU	
4												SD	SD	SD	SD	Tr	PF	PF	PF	PF	FU	FU	FU	
5												SD	SD	SD	SD	SD	Tr	PF	PF	PF	PF	FU	FU	
												T_0	T_3	T_6	T_9	T_{12}	T_{15}	T_{18}	T_{21}	T_{24}	T_{27}	T_{30}	T_{33}	T_{36}

TASSH introduced to all 30 facilities

NYU Langone Health

MAP-IT: Adoption

- ***Summary measure**: difference in proportion of clinic adoption*
- **Intercurrent events**: repeated cross-sectional assessment of adoption across each month (treatment policy).

Analysis strategy

- Estimation of difference in proportion adoption of TASSH under PF as compared to self directed care.
- Cluster average effect estimate – within cluster pre and post implementation
- Site size not substantially influential or variable

NYU Langone Health

Just Mothers: Adoption

- **Population** – 12 prenatal healthcare clinics in safety net hospitals across NYC serving pregnant women before 20w0d gestation speaking English or Spanish

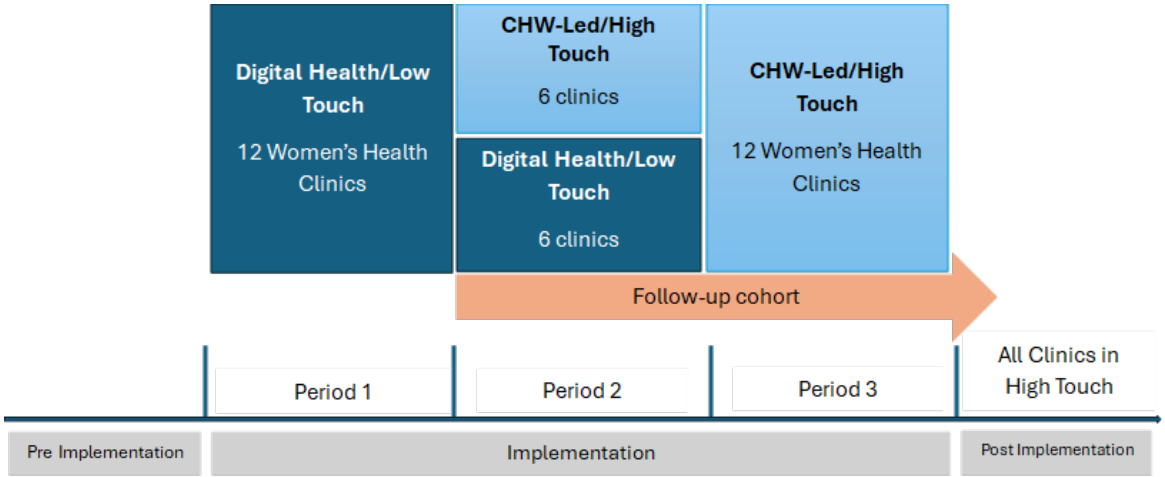
- **Implementation strategy** – Digital health/low touch strategy as compared to CHW-led/high touch strategy

- ***Endpoint*** *– Referral rate to the program measures adoption*
  - *Pregnancy cohorts in three periods*
  - *Number of individuals correctly referred relative to the number eligible for referral at a site*
  - *Interest in cluster average effect to understand uptake*

NYU Langone Health

Just Mothers: Adoption

- ***Endpoint*** *– Referral rate to the program measures adoption*
  - *Pregnancy cohorts in three periods*
  - *Number of individuals correctly referred relative to the number eligible for referral at a site*
  - *Interest in cluster average effect to understand uptake*



NYU Langone Health

Just Mothers: Adoption

- ***Summary measure**: difference in referral rate between strategies*
- **Intercurrent events**: repeated assessment of adoption across each period (treatment policy); low engagement with trial team (treatment policy)

Analysis strategy

- Planned cluster average effect estimate – within cluster pre and post implementation
  - GEEs with working independence correlation and robust standard errors and small cluster correction……

NYU Langone Health

Just Mothers- Pragmatic hiccups

- Recruitment during period 1 extended 2 months longer than expected – recruiting more in period 1
- Implications for period 3 recruitment – budget/funders
- Implications for overall estimand (CRT versus SWT) if limited or eliminated period 3

Analysis strategy

– Amend strategy to add sensitivity analysis to utilize only period 1 and 2 data to make direct comparisons between strategies.

NYU Langone Health

MAP-IT and Just Mothers: clinical outcomes

- Intercurrent events may typically have more impact on clinical and patient level outcomes
- Focus on **participant average** treatment effect

MAP-IT

- SBP and BP control
- Intercurrent events
  - Incomplete visits (treatment policy)
  - Death and loss to follow up (composite)

Just Mothers

- PNC adequacy, preterm delivery, maternal weight gain, gest diabetes
- Intercurrent events
  - Missed content (treatment policy)
  - Stillbirth (composite)
  - Nutrition education from providers (treatment policy)

NYU Langone Health

Implementation Science & Estimand Framework

- Once we clarify the **estimand components** we can more clearly **define estimation and missing** data strategy for these estimands – to be consistent with our defined quantities of interest

    - *But what happens when…..*
        - *Clinics do not initiate the strategy when we expect*
        - *Clinics close*
        - *Clinics consider adopting a competing strategy*
        - *Clinics change how a strategy is provided*

NYU Langone Health

Funders & References

Just Mothers: NIH OTHL158287
MAP-IT UG3 HL154498

- US Dept HHS ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. 2021.
- Wolfenden L, Foy R, Presseau J, Grimshaw JM, Ivers NM, Powell BJ, Taljaard M, Wiggers J, Sutherland R, Nathan N, Williams CM, Kingsland M, Milat A, Hodder RK, Yoong SL. Designing and undertaking randomised implementation trials: guide for researchers. BMJ. 2021 Jan 18;372:m3721. doi: 10.1136/bmj.m3721. PMID: 33461967; PMCID: PMC7812444.
- Kahan BC, Li F, Copas AJ, Harhay MO. Estimands in cluster-randomized trials: choosing analyses that answer the right question. Int J Epidemiol. 2023 Feb 8;52(1):107-118. doi: 10.1093/ije/dyac131. PMID: 35834775; PMCID: PMC9908044.
- Aifah, A. A., Hade, E. M., Colvin, C., Henry, D., Mishra, S., Rakhra, A., ... & Ojji, D. (2023). Study design and protocol of a stepped wedge cluster randomized trial using a practical implementation strategy as a model for hypertension-HIV integration—the MAP-IT trial. *Implementation Science*, *18*(1), 1-12.
- Kahan BC, Blette BS, Harhay MO, Halpern SD, Jairath V, Copas A, Li F. Demystifying estimands in cluster-randomised trials. Stat Methods Med Res. 2024 Jul;33(7):1211-1232. doi: 10.1177/09622802241254197. Epub 2024 May 23. PMID: 38780480; PMCID: PMC11348634

Learn-As-you-GO (LAGO) to adapt the intervention in an ongoing trial to prevent trial failure

Judith J. Lok
Professor of Mathematics and Statistics
Boston University, jjlok@bu.edu

Society for Clinical Trials Annual Meeting
Phoenix, Arizona, 5/20/2026

Joint work with **Ante Bing**, Boston University, **Daniel Nevo**, Tel Aviv University, and **Donna Spiegelman**, Yale

Supported by NIH 1R01HL167936-01A1 and NIH/NIEHS DP1 ES025459 and (but not the view of the NIH)

Background: Adaptive Designs

Popular in clinical trials and other randomized studies.

Goal: allow researchers to modify or stop a study intermittently to improve the chances of reaching desired conclusions, without jeopardizing the legitimacy and soundness of the trial.

FDA: "...a clinical study design that allows for prospectively planned modifications based on accumulating study data without undermining the study's integrity and validity" (FDA, 2016).

Examples: sample size adjustment, removing treatment arms, changing target population.

Our focus: Learn-As-you-GO (LAGO) studies

Large-scale intervention studies with a "Learn-As-you-GO" (LAGO) element (Nevo et al. (2021), Bing et al. (2025)).

Intervention:

- Treatment / device / new way to organize care / implementation of an intervention or intervention package.
- Can be composed of various components, possibly some continuous.
- While a subject-matter expert has some knowledge with regards to the preferred **intervention package**, its exact configuration is an inherent part of the study goals.

Example of LAGO: the "BetterBirth" study

Intervention: encourage use of WHO checklist, to ultimately improve safety of mother and newborn child (Hirschhorn et al. (2015)).

Details on how to encourage use of the checklist, so, the precise **intervention component composition**, changed over time.

Location: Uttar Pradesh, India.

Outcomes: Semrau et al. (2017): Outcomes of a Coaching-Based WHO Safe Childbirth Checklist Program in India, NEJM.

LAGO example: the "BetterBirth" study: continued

Package components: Intervention launch duration, Number of coaching visits, Data feedback form.....

BetterBirth began with a series of pilot studies to **find the optimal intervention package**. Pilots: collected process outcomes before and after intervention (check list use, completion of essential birth practices such as wash hands), not health outcomes (mother and baby alive).

RCT: more collected, including health outcomes 7 days after hospital discharge in 3600 participants.

LAGO: Goal of the analysis of a LAGO study

Know how the probability of "success" (for example, the percentage of births for which the BetterBirth Checklist is used) depends on the intervention package X, let's say X consists of both the launch duration and the number of coaching visits.

Or, e.g., finding optimal intervention package $\boldsymbol{x}^{opt}$, e.g., the $\boldsymbol{x}$ that solves (with i indexing participants and j indexing centers):

$$P_\beta(Y_{ij} = 1|\boldsymbol{x}) \geq 0.85,$$

while minimizing the cost of $\boldsymbol{x}$, and subject to (natural) constraints on $\boldsymbol{x}$.

In words: $\boldsymbol{x}^{opt}$ **minimizes the cost** of the intervention package while achieving a **success probability of at least** 0.85.

In the presence of center effects α_j: e.g., finding optimal intervention package $\boldsymbol{x}^{opt}$, e.g., the $\boldsymbol{x}$ that solves

$$\frac{1}{J}\sum_{j=1}^{J} P_\beta(Y_{ij} = 1|\boldsymbol{x}) \geq 0.85,$$

while minimizing the cost of $\boldsymbol{x}$, and subject to (natural) constraints on $\boldsymbol{x}$.

In words: $\boldsymbol{x}^{opt}$ **minimizes the cost** of the intervention package while achieving a **success probability of at least** 0.85.

LAGO: Learning-As-you-GO aspect

First: 2 stages.

We consider the situation where the stage 2 intervention package can be adapted/learnt using stage 1 data.

Typical reason for Learn-As-you-GO (LAGO): study team decides: if it is broken, try to fix it.

LAGO: Main statistical analysis complication

We cannot just condition on the observed interventions:

Later-stage observed interventions are a function of prior data, so conditioning on them would mean conditioning on outcomes.

LAGO: Notation

- 2 stages, no early stopping. Each center in 1 stage only.
- k: stage. Can be 1 or 2.
- Same n, number of participants, in both stages (for now).
- $\boldsymbol{x}^{(1)}$: recommended (multivariate) intervention for stage 1. $\boldsymbol{X}^{(2,n)}$: recommended (multivariate) intervention for stage 2 (assume: bounded random variable).
- $\boldsymbol{A}_j^{(k,n)}$: actual intervention in center/site j of stage k. Assume: depends on $X^{(k,n)}$ but not further on prior data.
- $\boldsymbol{Y}^{(k,n)} = (Y_1^{(k,n)}, Y_2^{(k,n)}, ..., Y_n^{(k,n)})$: binary outcomes in stage k, with $Y_i^{(k,n)}$ the outcome of participant i in stage k.
- (($\Rightarrow$ $Y_i^{(1,n)}$ and $Y_i^{(2,n)}$ denote outcomes of two different participants (one in stage 1, one in stage 2))).
- The n superscript is needed in stages after stage 1: the intervention in stage 2 depends on n stage 1 participants.

LAGO: Statistical considerations

The main statistical complication with "Learn-As-you-GO" (LAGO):

Recommended intervention $\boldsymbol{X}^{(2,n)}$ is random and determined by some function of the prior data:

Recommended intervention $\boldsymbol{X}^{(2,n)}$ depends on what researchers decide upon, looking at the stage 1 data.

In terms of random variables: $\boldsymbol{X}^{(2,n)}$ is not independent of the data of stage 1 participants, *and thus outcomes in the different stages are also dependent*.

LAGO: Learning-As-you-GO in stage 2

We make sure that $\boldsymbol{X}^{(2,n)}$ depends deterministically on the stage 1 data; a pre-specified rule.

Example 1: the decision rule is based on means/percentages based on the entire stage 1 sample in a smooth, say differentiable, way.

Example 2: the decision rule is based on an mle calculated from the stage 1 data, in a smooth, say differentiable, way.

Focus on asymptotic inference

What happens as the total number of observations gets large?

Seems reasonable approach in the BetterBirth study.

Will assume that the number of observations in both stages gets large at the same rate, to obtain realistic approximations of what happens in finite samples.

How to Learn-As-you-GO (LAGO) in stage 2? Assumptions

Recommended stage 2 intervention: $\boldsymbol{X}^{(2,n)}$, learnt from stage 1.

Main assumption: given $\boldsymbol{X}^{(2,n)}$, the stage 2 data are independent of the stage 1 data: conditionally on $\boldsymbol{X}^{(2,n)}$, $(\boldsymbol{A}^{(2,n)}, \boldsymbol{Y}^{(2,n)})$ is independent of $(\boldsymbol{A}^{(1,n)}, \boldsymbol{Y}^{(1,n)})$.

Assumption: $\boldsymbol{X}^{(2,n)}$ converges in probability to a constant $\boldsymbol{x}^{(2)}$ as the number of observations in stage 1 goes to infinity.

((Will happen in the previous examples because of the Weak Law of Large Numbers, the Continuous Mapping Theorem, and theory on MLE's))

LAGO: Actual interventions

Assumption: In each stage k and center j, the **actual intervention** $A_j^{(k,n)}$ is a function $h_j^{(k)}$ of the recommended intervention $X^{(k,n)}$, with $h_j^{(k)}$ a deterministic, continuous, bounded, center-specific function.

Don't need to know the $h_j^{(k)}$, do need to observe the $A_j^{(k,n)}$.

LAGO: logistic regression and an outcome goal

Assumption: In each stage k, for all $i = 1, \ldots, n$, success probability (logistic regression):

$$\text{logit} P(Y_i^{(k,n)} = 1 | \boldsymbol{X}^{(k,n)}, \boldsymbol{A}_{j(i)}^{(k,n)} = \boldsymbol{a}) = \beta_0 + \beta_1 a_1 + \ldots + \beta_p a_p,$$

$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)$, with $\boldsymbol{\beta}$: parameter of interest.

Outcome goal: a natural adaption of the intervention is to estimate $\boldsymbol{\beta}$ from the stage 1 data ($\hat{\boldsymbol{\beta}}^{(1)}$), and to obtain the recommended $\boldsymbol{X}^{(2,n)}$ by solving

$$P_{\hat{\boldsymbol{\beta}}^{(1)}}(Y_i = 1 | \boldsymbol{x}) \geq 0.85$$

for $\boldsymbol{x}$, while minimizing cost($\boldsymbol{x}$), and subject to constraints on $\boldsymbol{x}$.

LAGO: Summary of set-up:

Stage 1: Recommended intervention $\boldsymbol{x}^{(1)}$.
Collect outcomes.

Stage 2: Recommended intervention $\boldsymbol{X}^{(2,n)}$, learnt from stage 1.
Collect outcomes.

Assumption 1: Given $\boldsymbol{X}^{(2,n)}$, the stage 1 and stage 2 data are independent.

Assumption 2: $\boldsymbol{X}^{(2,n)} \rightarrow^{P} \boldsymbol{x}^{(2)}$.

LAGO likelihood, under Assumption 1:

$$
\begin{aligned}
& P(\boldsymbol{Y}^{(1,n)}, \boldsymbol{Y}^{(2,n)}, \boldsymbol{A}^{(2,n)}, \boldsymbol{A}^{(1,n)} | \boldsymbol{x}^{(1)}; \beta) \\
&= P(\boldsymbol{A}^{(1,n)} | \boldsymbol{x}^{(1)}, \beta) P(\boldsymbol{Y}^{(1,n)} | \boldsymbol{A}^{(1,n)}; \beta) P(\boldsymbol{A}^{(2,n)} | \boldsymbol{Y}^{(1,n)}, \boldsymbol{A}^{(1,n)}) \\
&\quad P(\boldsymbol{Y}^{(2,n)} | \boldsymbol{A}^{(2,n)}, \boldsymbol{Y}^{(1,n)}, \boldsymbol{A}^{(1,n)}; \beta) \\
&= P(\boldsymbol{Y}^{(1,n)} | \boldsymbol{A}^{(1,n)}; \beta) P(\boldsymbol{Y}^{(2,n)} | \boldsymbol{A}^{(2,n)}; \beta) \\
&\quad P(\boldsymbol{A}^{(2,n)} | \boldsymbol{Y}^{(1,n)}, \boldsymbol{A}^{(1,n)}) P(\boldsymbol{A}^{(1,n)} | \boldsymbol{x}^{(1)}).
\end{aligned}
$$

The mle for β solves the score equations. Score

$$
U(\beta) = \sum_{i=1}^{n} \boldsymbol{A}^{(1)}_{j(i)} \left(Y_i^{(1)} - p_{\boldsymbol{A}^{(1)}_{j'(i)}}(\beta) \right) + \boldsymbol{A}^{(2,n)}_{j'(i)} \left(Y_i^{(2,n)} - p_{\boldsymbol{A}^{(2,n)}_{j'(i))}}(\beta) \right).
$$

Same as estimating equations had the $\boldsymbol{A}^{(2,n)}$ not been dependent on stage 1. $\Rightarrow$ **Solving LAGO score equations can be done with standard software.**

LAGO: Unbiased estimating equations, Consistency, Asymptotic normality

Let β^* the true values of β. Note:

$$E_{\beta^*}(U(\beta^*)) = 0 + \sum_{i=1}^{n} E\left(E\left[\boldsymbol{A}_{j'(i)}^{(2,n)}(Y_i^{(2,n)} - p_{\boldsymbol{A}_{j'(i)}^{(2,n)}}(\beta^*)) \mid \boldsymbol{A}_{j'(i)}^{(2,n)}\right]\right) = 0.$$

Consistency and asymptotic normality do not follow from this immediately by applying theory from e.g. Van der Vaart (1998): we are **not dealing with iid sums:** stage 2 interventions depend on stage 1 data.

Consistency and asymptotic normality of $\hat{\beta}$: proven through non-standard means (tools from survival analysis, coupling).

LAGO: test for H_0: no effect of any of the package components

Because of the asymptotic normality result, the Wald test for $H_0 : \beta_1 = \beta_1^*$ is asymptotically valid for any constant β_1^*.

LAGO: alternative test for H_0: no effect of any of the package components

In a LAGO trial, the dependence between the stage 2 and stage 1 data is solely due to the stage 1 data determining the stage 2 recommended intervention, which, in turn, affects the actual stage 2 intervention, and thus the stage 2 outcomes.

However, under the null, there is no effect of the actual interventions on the stage 2 outcomes.

$\Rightarrow$ Under the null, regardless of the way the intervention was adapted, the stage 1 and stage 2 outcomes are independent!

$\Rightarrow$ **A standard test for equal probabilities in the control and the intervention arms is valid.**

Learn-As-you-GO (LAGO): Adding a power goal

Power goal: to add a power goal, two options, neither of them changing the test at the end of the LAGO trial. First estimate $\hat{\beta}^{(1)}$ from the stage 1 data. Then:

1. Conditional power goal: at the end of stage 1, estimate how the power of the end-of-study test depends on the recommended intervention conditional on the stage 1 data. Add power $\geq$ power goal to the LAGO optimization.
2. Unconditional power goal: at the end of stage 1, estimate how the unconditional power of the end-of-study test depends on the recommended intervention conditional on the stage 1 data. Add power $\geq$ power goal to the LAGO optimization.

Unconditional power can be shown to lead to higher power. Better if power hard to achieve.

Bing, Spiegelman, and Lok 2026;
`https://arxiv.org/abs/2509.11479v1`.

LAGO: Simulation study

Aim: verify the above results.

Setting: 2 stages, and a single normally distributed center-specific covariate.

LAGO: Simulation study

For all stages in each intervention center: We simulated $\boldsymbol{A}_j =$ recommended intervention + iid normal noise. ((If some component of $\boldsymbol{A}_j$ was out of the permitted intervention space, we took the boundary.))

After stage 1, $\boldsymbol{X}^{(2,n)}$ was chosen as follows.

First, we computed an estimator $\hat{\boldsymbol{\beta}}^{(1)}$ from stage 1 data only:

$$\hat{\boldsymbol{\beta}}^{(1)}(\boldsymbol{Y}^{(1,n)}, \boldsymbol{A}^{(1,n)}).$$

Then, $\boldsymbol{X}^{(2,n)}$ was chosen by finding $\boldsymbol{x}$ that solves the equation

$$p_{\boldsymbol{x}}(\hat{\boldsymbol{\beta}}^{(1)}) \geq 0.85$$

while minimizing cost, under the constraints on $\boldsymbol{x}$.

Results of simulation study: parameter estimates

Summary of the results obtained from 2,000 simulated datasets for each design:

Simulation study results for individual package component effects under linear cost function with $c_1 = 1$ and $c_2 = 8$. Results include for each individual component estimator relative bias, ratio between mean estimated standard error and empirical standard deviation (times 100) and coverage rate of 95% confidence interval.

Simulation study: results for individual package component effects. Unit costs were $c_1 = 1$ and $c_2 = 8$

$e^{\beta^\star}$	n_{1k}	$n_{2k'}$	K	%RelBias	$\frac{SE}{EMP.SD}$ ($\times 100$)	CP95	%RelBias	$\frac{SE}{EMP.SD}$ ($\times 100$)	CP95
				$\hat{\beta}_{11}$			$\hat{\beta}_{12}$		
(1.2, 1.5)	50	100	6	-2.3	96.5	95.1	-1.9	84.1	94.0
			10	-2.7	98.8	94.9	-1.2	92.2	95.2
			20	-1.4	101.3	95.2	-0.3	102.7	95.6
		200	6	-1.8	95.0	94.9	-2.6	81.0	95.4
			10	-4.4	92.7	94.2	-1.0	91.9	95.2
			20	-2.1	102.2	95.5	-0.2	99.7	95.2
	100	100	6	-1.7	92.9	94.7	-1.5	86.2	95.5
			10	2.8	101.9	95.7	-1.4	100.9	95.4
			20	2.1	101.1	95.5	-0.5	101.6	95.0
		200	6	-3.2	91.4	94.6	-0.8	83.6	95.5
			10	-1.6	99.5	95.4	-0.6	94.9	95.3
			20	-0.4	98.4	95.0	-0.3	97.5	94.5
(1.2, 2)	50	100	6	-16.0	91.6	95.4	0.7	86.0	96.0
			10	-7.4	101.4	95.8	0.2	102.2	96.0
			20	-3.6	99.6	95.2	-0.1	101.4	94.8
		200	6	-11.8	89.9	95.1	0.7	89.7	95.1
			10	-9.2	94.9	95.5	0.1	97.6	96.0
			20	-2.7	100.0	95.0	-0.2	101.4	96.2
	100	100	6	-7.6	94.5	95.8	-0.1	94.1	95.2
			10	-2.1	98.2	94.8	-0.0	102.7	95.2
			20	-3.7	100.3	95.2	0.2	102.7	95.5
		200	6	-7.1	84.6	95.2	0.3	95.8	95.9
			10	-4.6	96.4	94.7	0.0	99.6	95.5
			20	-3.5	98.0	94.6	0.1	104.8	95.9

%RelBias, percent relative bias $100(\hat{\beta} - \beta^\star)/\beta^\star$; SE, mean estimated standard error; EMP.SD, empirical standard deviation; CP95, empirical coverage rate of 95% confidence intervals.

Confidence bands for probabilities

We can use the resulting estimates and CIs of the β's to construct confidence bands for the probability of success under the various possible intervention packages:

((using Scheffé's method, which is based on the asymptotic normality of $\hat{\beta}$))

Estimates and confidence regions for the optimal intervention package $\boldsymbol{x}^{opt}$

We can use the resulting estimates and CIs of the $\boldsymbol{\beta}$'s to construct estimates and confidence regions for the optimal intervention $\boldsymbol{x}^{opt}$:

Keep only package configurations for which the confidence interval for the success probability contains 0.85.

The confidence bands for the probabilities of success can be used to obtain (conservative) confidence intervals for the probability of success under the estimated optimal intervention.

Results of simulation study: optimal intervention package and confidence bands for the success probabilities

Simulation study results for estimated optimal package and coverage probability of confidence bands for success probability, under linear cost function with $c_1 = 1$ and $c_2 = 8$.

Reported:

- bias for optimal individual components
- empirical root mean squared error of the estimated optimal package
- coverage rate of 95% confidence set for optimal intervention package and its relative size
- coverage rate of 95% confidence bands for outcome probability under all values on a grid of $\mathcal{X}$.

Results presented for $K = 10$ centers per stage.

Results of simulation study: optimal intervention package

$\exp(\beta^\star)$	$\boldsymbol{x}^{opt}$	n_{1k}	$n_{2k'}$	Bias(x_1^{opt}) times 100	Bias(x_2^{opt}) times 100	root MSE($\boldsymbol{x}^{opt}$) times 100	SetCP95	SetPerc%	BandsCP95
$(1, 2)$	$(0, 3.2)$	50	100	34.5	-4.7	85.0	94.0	7.6	97.0
			500	16.5	-2.1	58.5	95.0	4.0	97.2
		100	100	24.0	-2.5	71.0	94.8	6.3	96.5
			500	10.6	-0.9	47.0	95.3	3.7	97.4
$(1.2, 1.5)$	$(2, 4.5)$	50	100	-9.5	2.7	51.6	94.8	13.3	96.0
			500	-2.7	2.1	27.8	95.1	7.6	95.9
		100	100	-3.6	1.2	35.9	94.8	12.3	95.4
			500	-0.7	1.7	18.1	95.3	7.1	95.4
$(1.2, 2)$	$(2, 2.6)$	50	100	-33.1	4.5	84.0	94.6	14.3	95.5
			500	-14.9	3.3	56.6	94.4	8.1	95.3
		100	100	-23.2	3.3	70.3	95.6	12.4	96.0
			500	-8.8	2.3	43.6	95.1	7.5	95.8

Learn-As-you-GO (LAGO): The BetterBirth Study

Intervention: encourage use of WHO checklist, to ultimately improve safety of mother and newborn child (Semrau et al. (2017)).

Details on how to encourage use of the checklist, so, the precise **intervention component composition**, changed over time, in 3 stages.

Outcome considered:

- Oxytocin administration immediately after delivery, as recommended by the WHO (World Health Organization (2016)).

Learn-As-you-GO (LAGO): The BetterBirth Study

Outcome considered:

- Oxytocin administration immediately after delivery, as recommended by the WHO (World Health Organization (2016)).

Outcome goal considered:

- $\geq 85\%$ uptake.

Intervention components considered: LAGO table:

- Number of coaching visits. In $\{1, 2, \ldots, 40\}$. \$170 per coaching visit.
- Launch duration (days). In $\{1, 2, 3, 4, 5\}$. \$800 per launch day.

Center characteristic considered:

- Monthly birth volume.

Learn-As-you-GO (LAGO): The BetterBirth Study

Package component effect estimates and confidence intervals, calculated after each stage.

	Stage 1 $n_1 = 73$ OR (CI-OR)	Stages 1-2 $(n_1 + n_2 = 1780)$ OR (CI-OR)	Stages 1-3 $(n_1 + n_2 + n_3 = 6124)$ OR (CI-OR)
Intercept	1.07 (0.00, 280.80)	0.10 (0.07,0.15)	0.10 (0.09,0.11)
Launch Duration (days)	1.41 (0.76,2.64)	2.65 (1.95,3.77)	2.79 (2.41,3.23)
Coaching Visits (per 3 visits)	7.95 (1.77,73.95)	1.11 (0.96,1.28)	1.08 (1.04,1.12)
Birth Volume (monthly, per 100)	0.37 (0.00,32.33)	2.11 (1.93,2.33)	1.94 (1.84,2.06)
	$\hat{\boldsymbol{x}}^{opt,(2,n_1)} = (1,5)$	$\hat{\boldsymbol{x}}^{opt,(3,(n_1,n_2))} = (3,1)$	$\hat{\boldsymbol{x}}^{opt} = (3,1)$

OR, estimated odds ratio $\exp(\hat{\boldsymbol{\beta}})$;
CI-OR, 95% Confidence interval for the odds ratio.
In the estimated optimal interventions:
the first component is the launch duration (in days)
the second component is the number of coaching visits.

Learn-As-you-GO (LAGO): The BetterBirth Study: estimated optimal intervention

95% confidence set for the optimal intervention $CS(\boldsymbol{x}^{opt})$ over the grid of $\mathcal{X}$, taking all possible numbers of coaching visits, $1, ..., 40$, and $1, 1.5., 2, 2.5, ..., 5$ for intervention launch duration.

Result: Out of 360 potential intervention packages, 38 (10.5%) were included in the 95% confidence set.

Result: included in the confidence set: 1.5 days launch duration and 40 coaching visits; 2 days launch durations and 27 or more coaching visits; 2.5 days launch duration and less than 20 coaching visits; and 3 days launch duration and less than 5 coaching visits.

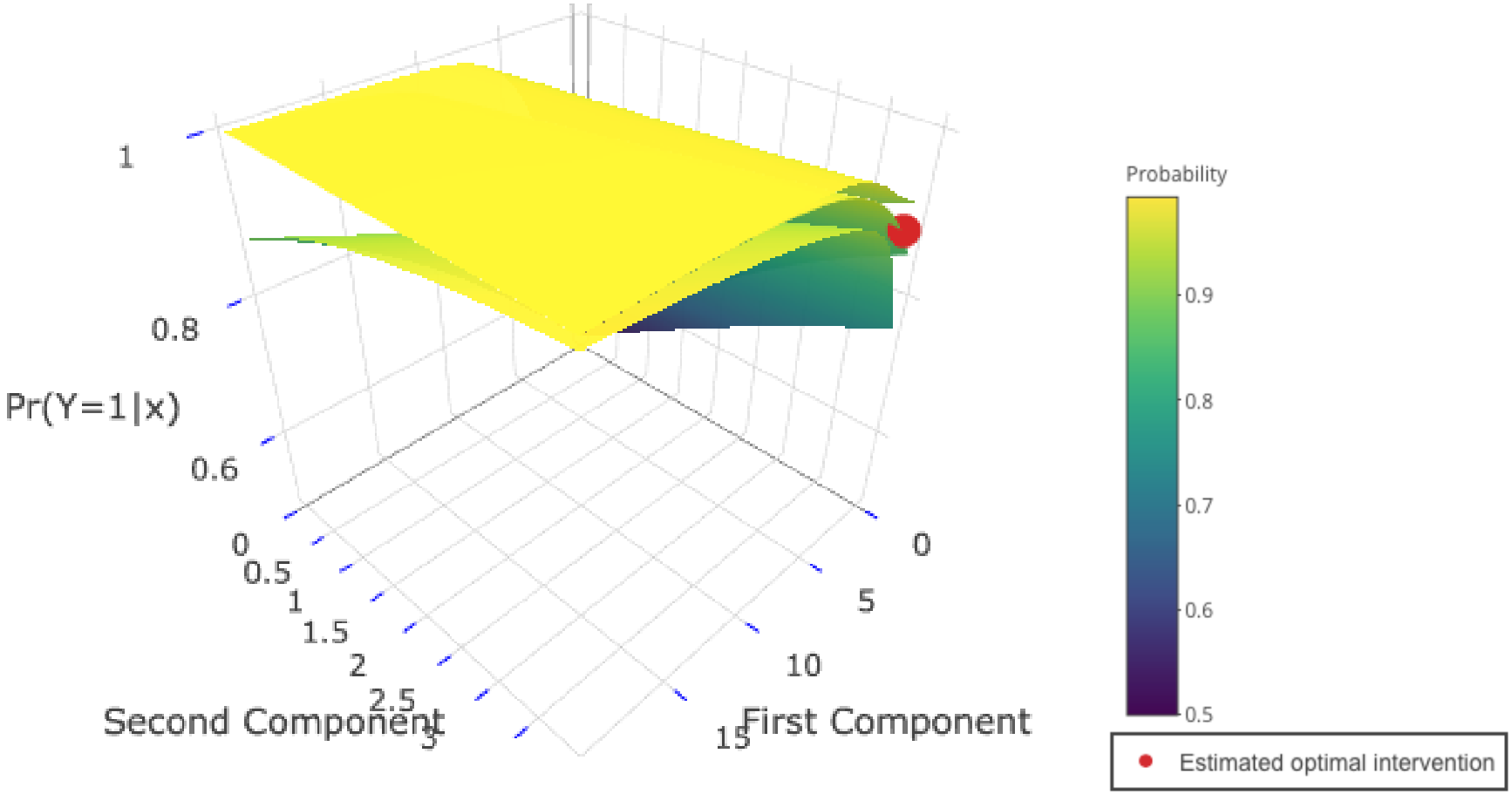
Result: cost: The first, second and third quartiles of the cost distribution within $CS(\boldsymbol{x}^{opt})$ were $Q1$=\$2462, $Q2$=\$4035, and $Q3$=\$6797.

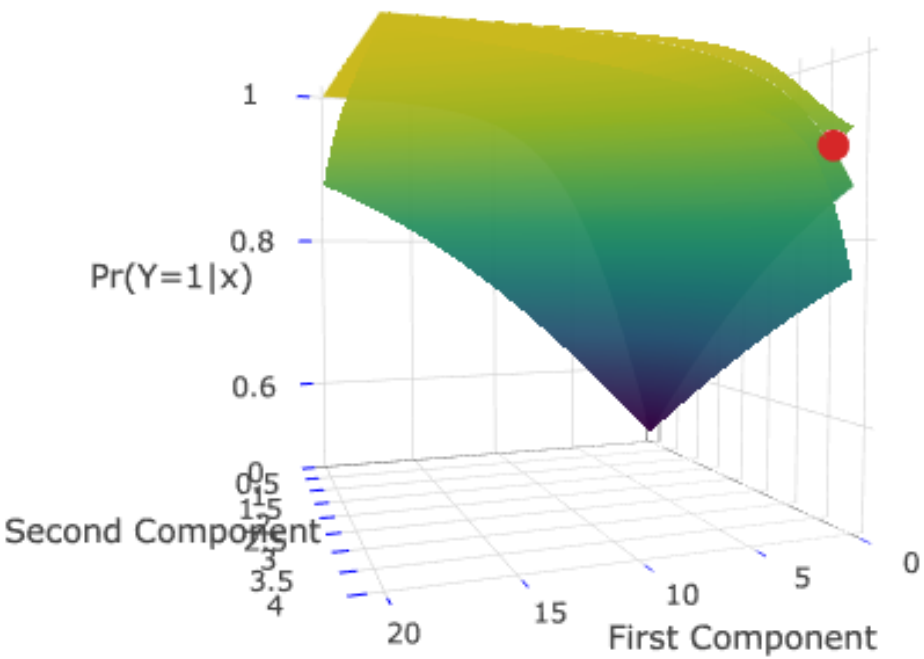
Learn-As-you-GO (LAGO): The BetterBirth Study: confidence band for the success probability

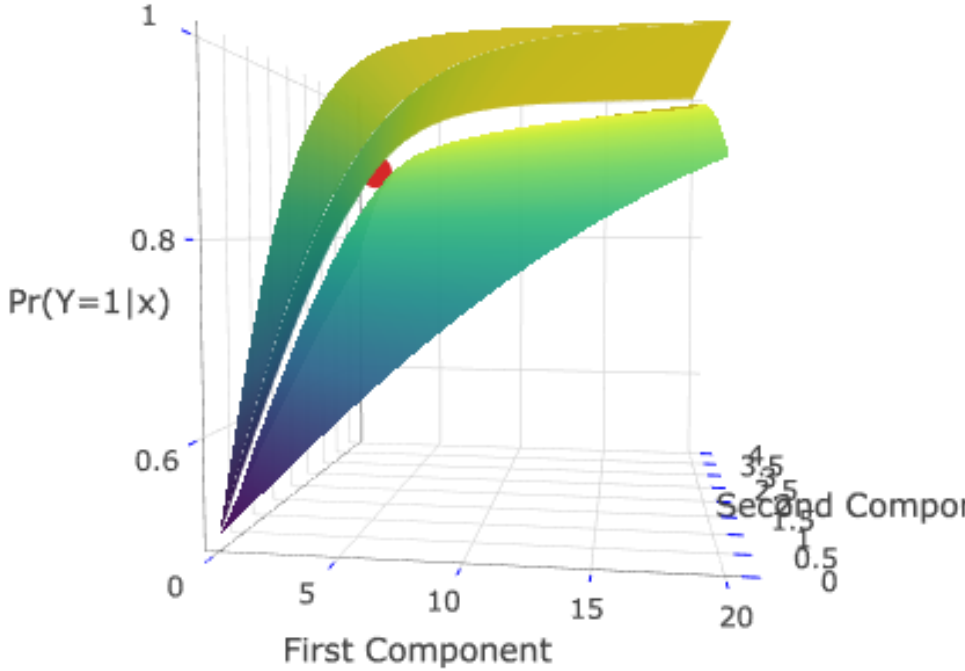
95% simultaneous confidence band for the probability of success under all 360 intervention compositions.

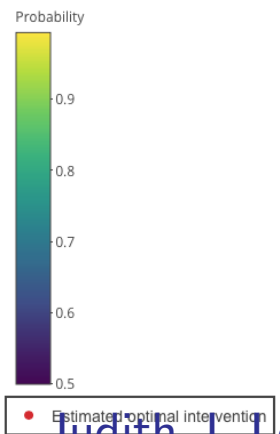
For the estimated optimal intervention $\hat{\boldsymbol{x}}^{opt} = (1, 3)$, the obtained confidence interval (within the bands) for the probability of oxytocin administration was $(0.79, 0.93)$.

The mean difference between the top and bottom of the confidence band over all 360 intervention compositions was 0.07.


1
0.8
Pr(Y=1|x)
0.6
0
0.5
1
1.5
2
2.5
3
Second Component
0
5
10
15
First Component
Probability
0.9
0.8
0.7
0.6
0.5
Estimated optimal intervention


1
0.8
Pr(Y=1|x)
0.6
Second Component
20
15
10
5
0
First Component


1
0.8
Pr(Y=1|x)
0.6
0
5
10
15
20
First Component


Probability
0.9
0.8
0.7
0.6
0.5
Estimated optimal intervention

Summary: Learn-As-you-GO (LAGO)

- Adaptive "Learn-As-you-GO" (LAGO) designs aim to find the optimal intervention package composition, and its effect.
- Dependence between stage 1 and stage 2 data.
- We used a coupling argument to provide theoretical justification for using existing software for estimation and inference.
- We provided a confidence set for the optimal intervention.
- We provided a confidence band for the probabilities under every package composition, including under the estimated optimal intervention.
- LAGO is currently used in PULESA-Uganda, MAPIT-Nigeria, and TASKPEN-Zambia. Large-scale multi-component implementation trials in people living with HIV.

Extensions (solved, not shown):

1. Replace $n = n_1 = n_2$ with $n_1/n_2 \to \alpha$, $0 < \alpha < \infty$.
2. Include center-specific covariates Z in the logistic regression model to predict success (or "outcome=1"), and also extended to center-specific recommended interventions.
3. > 2 stages, assuming recommended intervention in each stage converges in probability to a constant.
4. LAGO for continuous outcomes with iid errors, based on a generalized linear model for the outcomes. Instead of coupling, we replaced the errors under the LAGO design by errors under the limiting deterministic design, similar to $\tilde{Y}$ above, plus empirical process theory (led by Ante Bing, recent PhD graduate at Boston University; *Biometrics* 2025).

Future research on Learn-As-you-GO (LAGO)

1. Publish an R package for LAGO optimizations and analyses. Joint work with recent Boston University PhD graduate Ante Bing and Donna Spiegelman. `https://github.com/correspondMerchant/LAGO-R-Package`
2. Consider (un)measured confounding by indication: what if center intervention compositions correlate with their potential outcomes? Solution: fixed center effects. Joint work with Minh Bui, PhD student Boston University, Chris Longenecker, University of Washington in Seattle and Donna Spiegelman. Also allows center-dependent error distribution.
3. Allow for random center effects. Joint work with Yale postdoc Jingyu Cui and Donna Spiegelman.
4. Study design: explicitly consider other choices of $\boldsymbol{X}^{(2,n)}$ (Minh Bui), strategizing number of participants per stage (Ruyi Liyu, Yale PhD student).
5. LAGO for longitudinal data. Joint work with Minh Bui, Chris Longenecker, and Donna Spiegelman.

Bing, A., D. Spiegelman, and J. J. Lok (2025). Learn-as-you-go (lago) trials: Optimizing trials for effectiveness and power to prevent failed trials. *arXiv preprint arXiv:2509.11479*.

Bing, A., D. Spiegelman, D. Nevo, and J. J. Lok (2025). Learn-As-you-GO (LAGO) trials: optimizing treatments and preventing trial failure through ongoing learning. *Biometrics 81*(2), ujaf061. arXiv preprint arXiv:2307.06552.

FDA (2016). Adaptive designs for medical device clinical studies: Guidance for industry and food and drug administration staff.

Hirschhorn, L. R., K. Semrau, B. Kodkany, R. Churchill, A. Kapoor, J. Spector, S. Ringer, R. Firestone, V. Kumar, and A. Gawande (2015). Learning before leaping: integration of an adaptive study design process prior to initiation of betterbirth, a large-scale randomized controlled trial in uttar pradesh, india. *Implementation Science 10*(1), 117.

Lindvall, T. (2002). *Lectures on the coupling method*. Courier Corporation.

Nevo, D., J. J. Lok, and D. Spiegelman (2021). Analysis of "Learn-As-you-GO" (LAGO) studies. *Annals Of Statistics 49*(2), 793.

Semrau, K. E. A., L. R. Hirschhorn, M. M. Delaney, V. P. Singh, R. Saurastri, N. Sharma, D. E. Tuller, R. Firestone, S. Lipsitz, N. Dhingra-Kumar, and B. S. Kodkany (2017). Outcomes of a Coaching-Based WHO Safe Childbirth Checklist Program in India. *New England Journal of Medicine 377*(24), 2313–2324.

Van der Vaart, A. W. (1998). *Asymptotic statistics*. Cambridge series in statistical and probabilistic mathematics. Cambridge: Cambridge University Press.

World Health Organization (2016). *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection*. http://www.who.int/hiv/pub/arv/arv-2016/en/.

Thanks for listening!

LAGO: Consistency

$$\frac{U(\beta^*)}{n} =$$

$$\sum_{i=1}^{n} \frac{\boldsymbol{A}_{j(i)}^{(1,n)}}{n} \left(Y_i^{(1,n)} - p_{\boldsymbol{A}_{j(i)}^{(1,n)}}(\beta^*) \right) + \sum_{i=1}^{n} \frac{\boldsymbol{A}_{j'(i)}^{(2,n)}}{n} \left(Y_i^{(2,n)} - p_{\boldsymbol{A}_{j'(i)}^{(2,n)}}(\beta^*) \right)$$

1. The covariance between the two terms is 0, since, conditionally on $\boldsymbol{X}^{(2,n)}$, $\left(\boldsymbol{Y}^{(2,n)}, \boldsymbol{A}^{(2,n)}\right)$ is independent of $\left(\boldsymbol{Y}^{(1)}, \boldsymbol{A}^{(1)}\right)$.

((2. First term: variance $\to 0$ by standard theory: no learning.

3. Second term: Assume for now: finitely many centers j. Then, can be shown: variance also $\to 0$, from our assumption that $\boldsymbol{X}^{(2,n)} \xrightarrow{P} \boldsymbol{x}^{(2)}$.

$\Rightarrow U(\beta^*)/n \xrightarrow{P} 0$ by Chebyshev's inequality. Consistency: then follows as in Van der Vaart (1998).))

A note on consistency

For consistency we use: since $\boldsymbol{X}^{(2,n)} \rightarrow^{P} \boldsymbol{x}^{(2)}$, $\boldsymbol{A}^{(2,n)}_{j'(i)}$ converge in probability to constants. They are also bounded. Thus, we can use Lebesgue's Dominated Convergence Theorem.

Consistency, continued

Now, let $\psi_n(\beta) = U(\beta)/n$ and recall that $\hat{\beta}$ solves the equation $\psi_n(\beta) = 0$. In order to prove consistency, we follow reasoning similar to Van der Vaart (1998):

$$\begin{aligned}\psi_n(\beta) &= \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}^{(1)}_{j(i)} \left(Y_i^{(1,n)} - p_{\boldsymbol{A}^{(1)}_{j(i)}}(\beta) \right) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}^{(2,n)}_{j'(i)} \left(Y_i^{(2,n)} - p_{\boldsymbol{A}^{(2,n)}_{j'(i)}}(\beta) \right) \\ &= \psi_n(\beta^*) + \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}^{(1)}_{j(i)} (p_{\boldsymbol{A}^{(1)}_{j(i)}}(\beta^*) - p_{\boldsymbol{A}^{(1)}_{j(i)}}(\beta)) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{A}^{(2,n)}_{j'(i)} (p_{\boldsymbol{A}^{(2,n)}_{j'(i)}}(\beta^*) - p_{\boldsymbol{A}^{(2,n)}_{j'(i)}}(\beta)).\end{aligned}$$

Before: $\psi_n(\beta^*) \to^P 0$. Then, the Continuous Mapping Theorem. Then, under mild regularity conditions on the limiting function $\psi(\beta)$, we have **consistency**.

LAGO: Asymptotic normality

Taylor expansion:

$$0 = U(\hat{\beta}) = U(\beta^*) + \frac{\partial}{\partial \beta} U(\tilde{\beta})(\hat{\beta} - \beta^*)$$

where $\tilde{\beta}$ is between $\hat{\beta}$ and β^*.

$\Rightarrow$ as usual:

$$\sqrt{n}(\hat{\beta} - \beta^*) = \left(\frac{1}{n} \frac{\partial}{\partial \beta} U(\tilde{\beta}) \right)^{-1} U(\beta^*)/\sqrt{n}.$$

Numerator $U(\beta^*)/\sqrt{n}$ not an iid sum: not standard!

LAGO: Asymptotic normality: numerator

$$\sqrt{n}\frac{1}{n}\left(\sum_{i=1}^{n} \boldsymbol{A}^{(1)}_{j(i)}\left(Y_i^{(1)} - p_{\boldsymbol{A}^{(1)}_{j(i)}}(\boldsymbol{\beta}^*)\right) + \sum_{i=1}^{n} \boldsymbol{A}^{(2,n)}_{j'(i)}\left(Y_i^{(2,n)} - p_{\boldsymbol{A}^{(2,n)}_{j'(i)}}(\boldsymbol{\beta}^*)\right)\right)$$

We saw: terms uncorrelated, and limiting variance we calculated when showing consistency.

Asymptotic normality not straightforward: not an iid sum, and cannot condition on the $\boldsymbol{A}^{(2,n)}$.

$\Rightarrow$ A coupling argument for the second term!

LAGO: Asymptotic normality: coupling

Because of our assumptions, $\boldsymbol{A}_j^{(2,n)}$ converges in probability to a constant for each j.

Let

$$\boldsymbol{a}_j^{(2)} = h_j^{(2)}(\boldsymbol{x}^{(2)})$$

be the limit in probability of

$$\boldsymbol{A}_j^{(2,n)} = h_j^{(2)}(\boldsymbol{X}^{(2,n)}).$$

LAGO: Asymptotic normality: coupling

Construct $\boldsymbol{Y}^{(\boldsymbol{x}^{(2)})} = \left(Y_i^{(\boldsymbol{x}^{(2)})}, i = 1, \ldots, n \right)$: n independent Bernoulli variables with $P(Y_i^{(\boldsymbol{x}^{(2)})} = 1) = p_{h_{j'(i)}(\boldsymbol{x}^{(2)})}(\beta^*) = p_{\boldsymbol{a}^{(2)}_{j'(i)}}(\beta^*)$. These are iid random variables with success probability under $\boldsymbol{x}^{(2)}$, independent of all the quantities we had so-far.

Notice that we are after a limiting distribution, so as long as we do not change the distribution we can replace any of the variables in the numerator by something else.

We will replace the terms with $\boldsymbol{A}^{(2,n)}_{j'(i)}$ and $Y_i^{(2,n)}$ by terms depending on $\boldsymbol{a}^{(2)}_{j'(i)}$, $Y_i^{(\boldsymbol{x}^{(2)})}$, and $\boldsymbol{X}^{(2,n)}$ without changing the distribution of the numerator.

LAGO: Asymptotic normality: coupling

Construct variables $\tilde{Y}_i^{(2,n)}$ which, given the stage 1 data (including $\boldsymbol{X}^{(2,n)}$), have the same distribution as the original $Y_i^{(2,n)}$, but are coupled (see e.g. Lindvall (2002)) with the $Y_i^{(\boldsymbol{x}^{(2)})}$ as follows. Let U_i be a uniform$[0,1]$ random variable independent of all other variables that we introduced so-far. If $p_{\boldsymbol{a}_{j'(i)}^{(2)}}(\boldsymbol{\beta}^*) > p_{\boldsymbol{A}_{j'(i)}^{(2,n)}}(\boldsymbol{\beta}^*)$, then:

$$\tilde{Y}_i^{(2,n)} = \begin{cases} 0 & \text{if } Y_i^{(\boldsymbol{x}^{(2)})} = 0 \\ 0 & \text{if } Y_i^{(\boldsymbol{x}^{(2)})} = 1 \text{ and } U_i < \dfrac{p_{\boldsymbol{a}_{j'(i)}^{(2)}}(\boldsymbol{\beta}^*) - p_{\boldsymbol{A}_{j'(i)}^{(2,n)}}(\boldsymbol{\beta}^*)}{p_{\boldsymbol{a}_{j'(i)}^{(2)}}(\boldsymbol{\beta}^*)} \\ 1 & \text{if } Y_i^{(\boldsymbol{x}^{(2)})} = 1 \text{ and } U_i \geq \dfrac{p_{\boldsymbol{a}_{j'(i)}^{(2)}}(\boldsymbol{\beta}^*) - p_{\boldsymbol{A}_{j'(i)}^{(2,n)}}(\boldsymbol{\beta}^*)}{p_{\boldsymbol{a}_{j'(i)}^{(2)}}(\boldsymbol{\beta}^*)}. \end{cases}$$

LAGO: Asymptotic normality: coupling

Similarly, if $p_{\boldsymbol{a}^{(2)}_{j'(i)}}(\boldsymbol{\beta}^*) \leq p_{\boldsymbol{A}^{(2,n)}_{j'(i)}}(\boldsymbol{\beta}^*)$, then:

$$\tilde{Y}_i^{(2,n)} = \begin{cases} 1 & \text{if } Y_i^{(\boldsymbol{x}^{(2)})} = 1 \\ 1 & \text{if } Y_i^{(\boldsymbol{x}^{(2)})} = 0 \\ & \quad \text{and } U_i < \frac{p_{\boldsymbol{A}^{(2,n)}_{j'(i)}}(\boldsymbol{\beta}^*) - p_{\boldsymbol{a}^{(2)}_{j'(i)}}(\boldsymbol{\beta}^*)}{1 - p_{\boldsymbol{a}^{(2)}_{j'(i)}}(\boldsymbol{\beta}^*)} \\ 0 & \text{if } Y_i^{(\boldsymbol{x}^{(2)})} = 0 \\ & \quad \text{and } U_i \geq \frac{p_{\boldsymbol{A}^{(2,n)}_{j'(i)}}(\boldsymbol{\beta}^*) - p_{\boldsymbol{a}^{(2)}_{j'(i)}}(\boldsymbol{\beta}^*)}{1 - p_{\boldsymbol{a}^{(2)}_{j'(i)}}(\boldsymbol{\beta}^*)}. \end{cases}$$

LAGO: Asymptotic normality: coupling

$\tilde{Y}^{(2,n)}$ depend on the stage 1 data and $\boldsymbol{X}^{(2,n)}$ the same way as the original $Y^{(2,n)} \Rightarrow \tilde{Y}_i^{(2,n)}$ can replace $Y_i^{(2,n)}$ in numerator without changing the distribution. Now,

$$
\begin{aligned}
& n^{-1/2} \sum_{i=1}^{n} \boldsymbol{A}_{j'(i)}^{(2,n)} \left(\tilde{Y}_i^{(2,n)} - p_{\boldsymbol{A}_{j'(i)}^{(2,n)}}(\beta^*) \right) \\
&= \quad n^{-1/2} \sum_{i=1}^{n} \boldsymbol{a}_{j'(i)}^{(2,n)} \left(Y_i^{(\boldsymbol{x}^{(2)})} - p_{\boldsymbol{a}_{j'(i)}^{(2,n)}}(\beta^*) \right) + B_n.
\end{aligned}
$$

Using the coupling, we have shown that $B_n \xrightarrow{P} 0 \Rightarrow$ Reduced to standard logistic regression with $\boldsymbol{x}^{(2)}$ fixed!

$\Rightarrow$ Asymptotic nomality.

Taylor expansion for asymptotic normality: denominator

$$\frac{\partial}{\partial \boldsymbol{\beta}} U(\boldsymbol{\beta}) = -\sum_{i=1}^{n} \boldsymbol{A}^{(1)}_{j(i)} \frac{\partial}{\partial \boldsymbol{\beta}} p_{\boldsymbol{A}^{(1)}_{j(i)}}(\boldsymbol{\beta}) - \sum_{i=1}^{n} \boldsymbol{A}^{(2,n)}_{j'(i)} \frac{\partial}{\partial \boldsymbol{\beta}} p_{\boldsymbol{A}^{(2,n)}_{j'(i)}}(\boldsymbol{\beta}).$$

This converges using Continuous Mapping Theorem and Lebesgue's Dominated Convergence Theorem. Consistent estimator $\hat{\tau}$: substitute $\boldsymbol{\beta}^0$ with $\hat{\boldsymbol{\beta}}$ and $\boldsymbol{a}$'s with $\boldsymbol{A}$'s.

Yale SCHOOL OF PUBLIC HEALTH

Center for Methods in Implementation and Prevention Science

Statistical Design Methodologies for Hybrid Type 2 Effectiveness-Implementation Studies

Donna Spiegelman, Sc.D.

Susan Dwight Bliss Professor of Biostatistics and

Director, Center for Methods in Implementation and Prevention Science

Yale School of Public Health

https://ysph.yale.edu/cmips/

Donna.Spiegelman@yale.edu

Work made possible by: ***Training in Implementation Science Research Methods*** *(T32HL155000)*

Yale SCHOOL OF PUBLIC HEALTH

Center for Methods in Implementation and Prevention Science

Owen MA, Curran GM, Smith JD, Tedla Y, Cheng C, Spiegelman D.

Power and Sample Size Calculations for Cluster Randomized Hybrid Type 2 Effectiveness-Implementation Studies

Stat Med. 2025 Feb 28;44(5):e70015. doi: 10.1002/sim.70015. PMID: 39930740.

Owen, M., Li, F., Liu, R., & Spiegelman, D. (2025).

A comparison of methods for designing hybrid type 2 cluster-randomized trials with continuous effectiveness and implementation endpoints.

arXiv preprint arXiv:2510.20741.

https://github.com/melodyaowen/crt2power.

https://cran.rproject.org/web/packages/crt2power/index.html

Work made possible by: ***Training in Implementation Science Research Methods*** *(T32HL155000)*

Outline

- **Background**
  - What are Implementation Outcomes?
  - Defining Hybrid 2 Studies
  - Cluster Randomized Trials
  - Literature Review
- **Study Design Methods**
- **Illustrative Example: The CIRCL Study**
- **Summary**

Background

What are Implementation Outcomes?

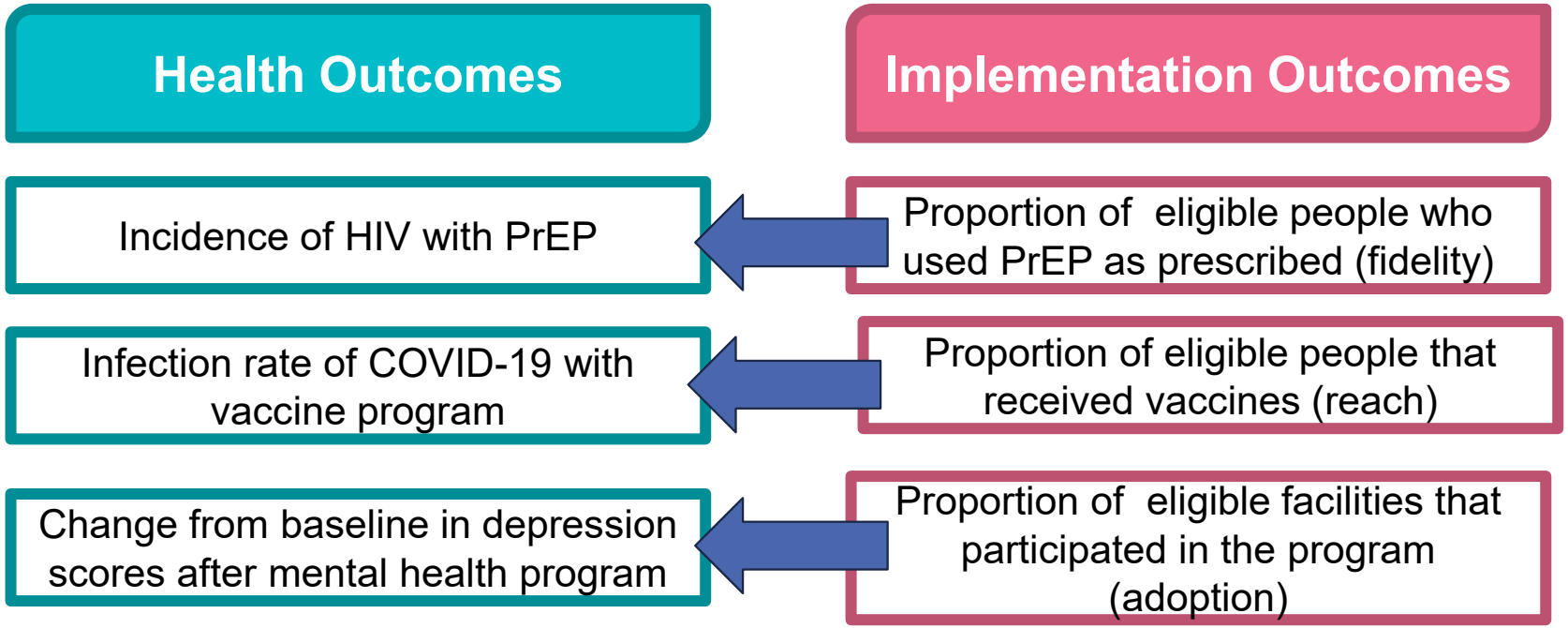
Various implementation outcome frameworks exist that help us **define implementation outcomes**, including the RE-AIM framework [1], and the Proctor's taxonomy of implementation outcomes [2].

RE-AIM framework:

Reach	• Proportion of target individuals who receive the intervention or program
Effectiveness	• The impact an intervention has on the individual outcomes of interest
Adoption	• the proportion of settings (e.g., clinics, schools) and staff willing to initiate the program
Implementation	• The degree to which the intervention or program is delivered as intended (fidelity)
Maintenance	• The extent to which the intervention or program can be sustained long-term

Health vs. Implementation Outcomes

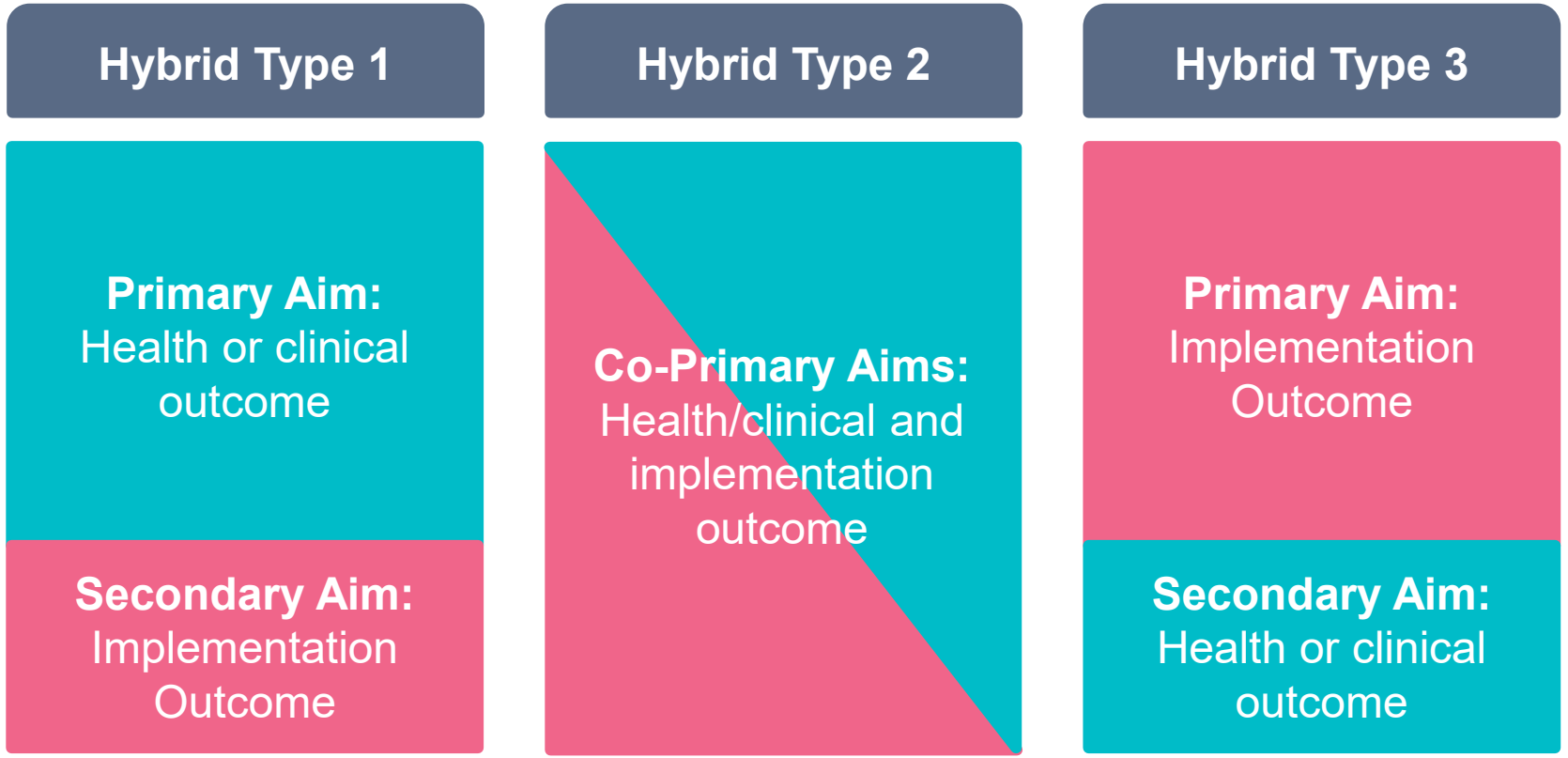
Hybrid studies focus on both **health (clinical or effectiveness)** and **implementation** outcomes [3]. To address public health issues, both types of outcomes are necessary to adequately address and improve the issue.



There are **three types of hybrid studies** that differ based on the primary goal of the study [3].

What are hybrid studies?

Hybrid Study Typology



Motivation and Statistical Issues for Hybrid Type 2 Studies

- **Motivation** for Utilizing Hybrid Type 2 Studies
  - Allows investigators to explore the implementation of an intervention along with the health outcomes of the intervention
    - **Prevents delays** that can occur when we study implementation outcomes afterwards
  - Ideal and efficient when we have **preliminary evidence of clinical effectiveness**, but there is a need for more data in **other settings** and information on the **implementation**
- **Statistical Issues** for Hybrid Type 2 Studies
  - Introduce the topic of **multiple testing**
  - There may be differences between the **nature** and **measurement** of the implementation and intervention outcomes
  - Complexity of **clustered data**, which is often utilized in hybrid studies

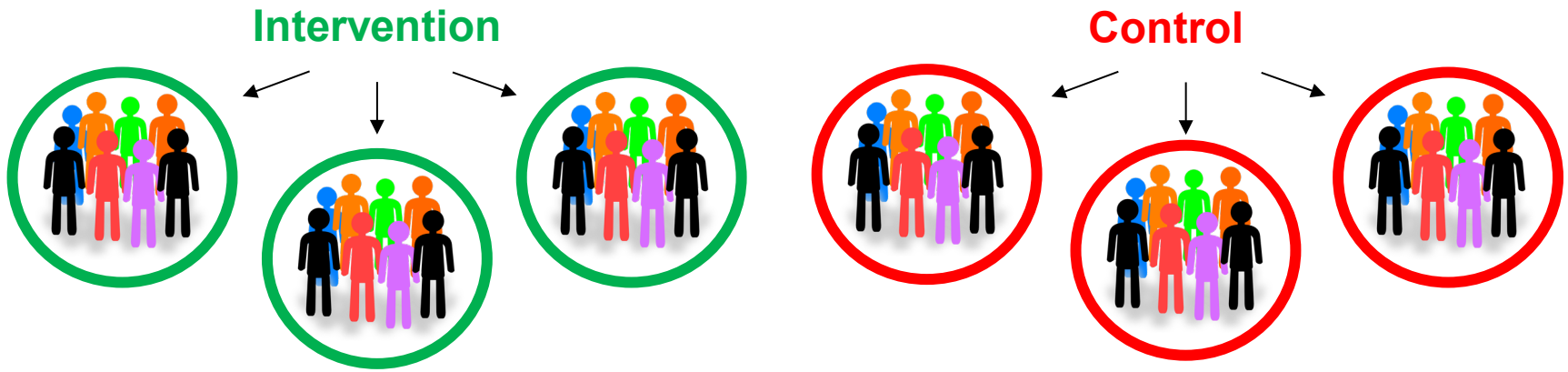
Cluster Randomized Trials Overview

What are CRT's?

- Experiments in which social units, or **clusters**, of individuals are **randomized to intervention groups** rather than individuals [4]
- Example: hospitals randomized to using a new vaccine program

Why use CRT's?

- **Logistical** convenience and increased compliance
- **Minimizes** treatment group contamination
- Many interventions **naturally applied at the cluster level**, specifically programs intended to improve the implementation of an intervention, which is why Hybrid studies are usually CRTs



Results from Literature Search

Literature search in **PubMed** and **Google Scholar** was conducted to find relevant study design methods publications.

18 relevant publications were identified.

Five key study design methods were identified that could be used in the context of hybrid type 2 designs

1. P-Value Adjustments for Multiple Testing
2. Creating Combined Outcomes
3. Single 1-DF Combined Test for Two Outcomes
4. Disjunctive 2-DF Test for Two Outcomes
5. Conjunctive Intersection-Union Test for Two Outcomes

These tests differ in the **hypothesis setup** and overall research questions that can be answered through the study.

In our work, Methods 2 and 3 have been *extended to the case of CRTs*.

Study Design Methods

Notation and Setup for Hybrid 2 CRT

Notation

- $Q = 2$ is the number of primary outcomes, with index $q = 1,2$
- α is the family-wise false positive rate
- β_1^* and β_2^* are the treatment effects of the first and second primary outcomes, respectively
- π as the statistical power, which is the probability of detecting a true effect under the alternative hypothesis
- $i = 1,2$ is the index of the two treatment groups
- $k = 1, \ldots, 2K$ is the index of the clusters; K clusters in each treatment group, thus $2K$ clusters total
- $j = 1, \ldots, m$ is the index of the subjects in each cluster, where there are m subjects in each of the $2K$ total clusters
- λ is the non-centrality parameter, and is used to calculate the power of a test given K, m, α-level

Then, in a **cluster randomized setting for a continuous endpoint**, m, K, and λ are obtained by the following equations [4]:

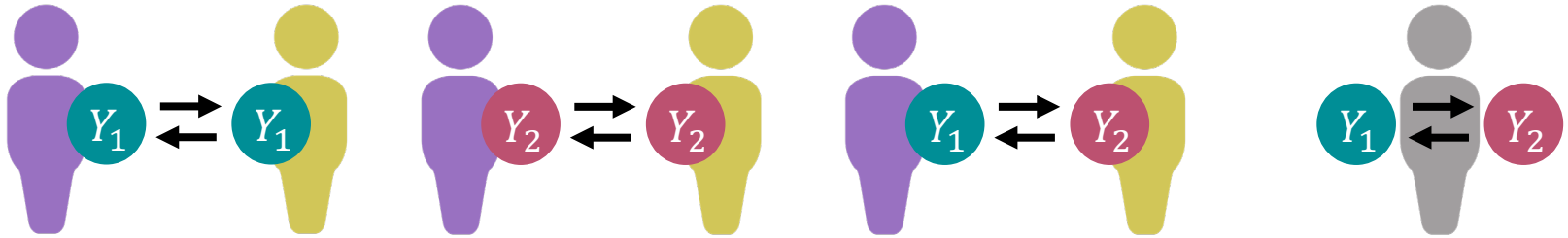
$$m = \frac{2\,\lambda\,\sigma_q^2(1-\rho)}{(\beta_q^*)^2 K - 2\,\lambda\,\sigma_q^2\rho}\;;\quad K = \frac{2\,\lambda\,\sigma_q^2\,[1+(m-1)\rho]}{m(\beta_q^*)^2}\;;\quad \lambda = \frac{(\beta_q^*)^2}{2\,\frac{\sigma_q^2}{Km}\,[1+(m-1)\rho_0^{(q)}]}$$

Correlations in CRTs

Correlations Needed for the Design of Cluster Randomized Trials

Four correlations are used throughout the five study design methods.

$\rho_0^{(1)}$	$\rho_0^{(2)}$	$\rho_1^{(1,2)}$	$\rho_2^{(1,2)}$
Endpoint-specific ICC for Y_1	Endpoint-specific ICC for Y_2	Inter-subject between-endpoint ICC	Intra-subject between-endpoint ICC
Correlation for Y_1 for two different individuals in the same cluster	Correlation for Y_2 for two different individuals in the same cluster	Correlation between Y_1 and Y_2 for two different individuals in the same cluster	Correlation between Y_1 and Y_2 for the same individual



Hypothesis Setup

- For two primary outcomes, there are **four scenarios**:
  1. $\beta_1^* = 0$ and $\beta_2^* = 0$
  2. $\beta_1^* = 0$ and $\beta_2^* \neq 0$
  3. $\beta_1^* \neq 0$ and $\beta_2^* = 0$
  4. $\beta_1^* \neq 0$ and $\beta_2^* \neq 0$
- Depending on the research question, these scenarios can be grouped in different ways to create the desired hypothesis test
- **Disjunctive Test**: $H_0: \beta_1^* = 0$ and $\beta_2^* = 0$ vs. $H_A: \beta_1^* \neq 0$ or $\beta_2^* \neq 0$
  - Groups the **last three scenarios together** under H_A
  - Intervention must have an effect on **at least one outcome**, but not necessarily both, in order to reject H_0
- **Conjunctive Test**: $H_0: \beta_1^* = 0$ or $\beta_2^* = 0$ vs. $H_A: \beta_1^* \neq 0$ and $\beta_2^* \neq 0$
  - Groups the **first three scenarios together** under H_0
  - Alternative hypothesis requires that the intervention have an effect on exactly **both outcomes** [6]
  - Since the intervention must have an effect on both the outcomes, we argue that this framework is most in line with the investigators research goals when conducting a hybrid type 2 design.

Method 1: P-Value Adjustments

Hypothesis Test: $H_0: \beta_1^* = 0 \text{ and } \beta_2^* = 0$ vs. $H_A: \beta_1^* \neq 0 \text{ or } \beta_2^* \neq 0$

Rationale: Tests the disjunctive hypothesis, and we reject H_0 when the intervention has a significant effect on **one or both of the outcomes**. Allocates the overall family-wise Type I error rate across the outcomes. **Three ways of adjusting the α-level** [8, 9, 10]:

1. Bonferroni Correction: $\alpha_{Bonf} = \alpha/Q$
2. Sidak Method: $\alpha_{Sdk} = 1 - (1-\alpha)^{\frac{1}{Q}}$
3. D/AP Approach: $\alpha_{DAP,2} = 1 - (1-\alpha)^{1/M_q}$

   where $M_q = Q^{1-\rho_2^{(1,2)}}$

Method		α-level
Bonferonni		$\alpha_{Bonf} = 0.025$
Sidak		$\alpha_{Sdk} = 0.0253$
D/AP	$\rho_2^{(1,2)} = 0.01$	$\alpha_{DAP} = 0.0255$
	$\rho_2^{(1,2)} = 0.05$	$\alpha_{DAP} = 0.0262$
	$\rho_2^{(1,2)} = 0.1$	$\alpha_{DAP} = 0.0271$

Strengths and Limitations:

- ✓ Common approach and easy to enact
- ✓ Choices of how to adjust for α
- X Tends to result in lower power and higher sample size requirements
- X H_A doesn't require the treatment to be efficacious on both outcomes

Required Parameters

Parameter	Description
β_1^* and β_2^*	The treatment effects on Y_1 and Y_2
$\rho_0^{(1)}$ and $\rho_0^{(2)}$	The endpoint specific ICC's for Y_1 and Y_2
$\rho_2^{(1,2)}$	Correlation between Y_1 and Y_2 for the same individual
σ_1^2 and σ_2^2	The outcome variances, $\mathrm{Var}(Y_1)$ and $\mathrm{Var}(Y_2)$
m, K, π	Desired number of clusters, cluster size, and power
α_{adj}	Adjusted type I error rate from Bonferroni, Sidak, or D/AP

Parameter	Description
β_1^* and β_2^*	The treatment effects on Y_1 and Y_2
$\rho_0^{(1)}$ and $\rho_0^{(2)}$	The endpoint specific ICC's for Y_1 and Y_2
$\rho_1^{(1,2)}$	Correlation between Y_1 and Y_2 for two different individuals in the same cluster
$\rho_2^{(1,2)}$	Correlation between Y_1 and Y_2 for the same individual
σ_1^2 and σ_2^2	The outcome variances, $\mathrm{Var}(Y_1)$ and $\mathrm{Var}(Y_2)$
m, K, π	Desired number of clusters, cluster size, and power

Method 1: P-Value Adjustments

Required Parameters:

Parameter	Description
β_1^* and β_2^*	The treatment effects on Y_1 and Y_2
$\rho_0^{(1)}$ and $\rho_0^{(2)}$	The endpoint specific ICC's for Y_1 and Y_2
$\rho_2^{(1,2)}$	Correlation between Y_1 and Y_2 for the same individual
σ_1^2 and σ_2^2	The outcome variances, $\mathrm{Var}(Y_1)$ and $\mathrm{Var}(Y_2)$
m, K, π	Desired number of clusters, cluster size, and power
α_{adj}	Adjusted type I error rate from Bonferroni, Sidak, or D/AP

Design Equations:

$$\lambda^{(q)} = \frac{\left(\beta_q^*\right)^2}{2\,\frac{\sigma_q^2}{Km}\left[1+(m-1)\rho_0^{(q)}\right]};\quad K^{(q)} = \frac{2\lambda\sigma_q^2\,[1+(m-1)\rho_0^{(q)}]}{m\left(\beta_q^*\right)^2};\quad m^{(q)} = \frac{2\lambda\sigma_q^2\,(1-\rho_0^{(q)})}{\left(\beta_q^*\right)^2 K - 2\,\lambda\sigma_q^2\,\rho_0^{(q)}}$$

Final sample size and power is found as:

$$\pi = \min\left(\pi^{(1)}, \pi^{(2).}\right);\ K = \max\left(K^{(1)}, K^{(2).}\right);\ m = \max\left(m^{(1)}, m^{(2).}\right)$$

Method 2: Creating Combined Outcomes

Hypothesis Test: $H_0: \beta_c^* = 0$ vs. $H_A: \beta_c^* \neq 0$

Rationale: β_c^* represents the treatment effect on the combined outcome. We can define Y_c from Y_1 and Y_2 in **three main ways** [7]:

1. Total Score: Sum or average of scores from the same scale
2. Event Rate: Number of events that occur from pre-defined list
3. Time to Event: Time to first event from pre-defined list of events

When we reject H_0, we are concluding that the **treatment is efficacious on the combined outcome, Y_c.**

Strengths and Limitations:

- ✓ Easy to transform outcomes to be in the same direction
- ✓ Can be efficient and results in lower sample size requirements
- X Relies on the outcomes being substantially similar
- X Not ideal for hybrid studies when implementation outcomes are not similar to the health outcomes

Method 2: Creating Combined Outcomes

Required Parameters:

Parameter	Description
β_c^*	The treatment effect on Y_c
$\rho_0^{(c)}$	The endpoint specific ICC, calculated from $\rho_0^{(1)}$, $\rho_0^{(2)}$, $\rho_1^{(1,2)}$, $\rho_2^{(1,2)}$, σ_1^2, and σ_2^2
σ_c^2	The outcome variance of the combined outcome, $\mathrm{Var}(Y_c)$. Can be calculated from $\rho_2^{(1,2)}$, σ_1^2 and σ_2^2
m, K, π	Desired number of clusters, cluster size, and power

Design Equations:

$$\lambda = \frac{(\beta_c^*)^2}{2\,\frac{\sigma_c^2}{Km}\left[1 + (m-1)\rho_0^{(c)}\right]}; K = \frac{2\,\lambda\,\sigma_c^2\left[1 + (m-1)\rho_0^{(c)}\right]}{m\,(\beta_c^*)^2}; m = \frac{2\,\lambda\,\sigma_c^2\left(1-\rho_0^{(c)}\right)}{(\beta_c^*)^2\,K - 2\,\lambda\,\sigma_c^2\,\rho_0^{(c)}}$$

$$\beta_c^* = \beta_1^* + \beta_2^*;\ \rho_0^{(c)} = \frac{\rho_0^{(1)}\sigma_1^2 + \rho_0^{(2)}\sigma_2^2 + 2\rho_1^{(1,2)}\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + 2\rho_2^{(1,2)}\sigma_1\sigma_2}; \sigma_c^2 = \sigma_1^2 + \sigma_2^2 + 2\rho_2^{(1,2)}\sigma_1\sigma_2$$

Method 3: Single 1-DF Combined Test

Hypothesis Test: $H_0: \beta_1^* = \beta_2^* = 0$ vs. $H_A: \beta_1^* \neq 0$ or $\beta_2^* \neq 0$

Rationale: The two outcomes are weighted into a single test statistic. Similar to combined outcomes approach but accounts for all correlations and both outcome variances.

Can also write hypotheses as $H_0: \beta_1^* + \beta_2^* = 0$ vs. $H_A: \beta_1^* + \beta_2^* \neq 0$.

$$Z_c^2 = \left[\frac{Z_1+Z_2}{\sqrt{2\left(1+\text{Corr}(Z_1,Z_2)\right)}}\right]^2;\ Z_q^2 = \frac{\left(\beta_q^*\right)^2}{\frac{2\sigma_q^2}{Km}\left[1+(m-1)\rho_0^{(q)}\right]};$$

$$\text{Corr}(Z_1,Z_2) = \frac{\left(\rho_2^{(1,2)}+(m-1)\rho_1^{(1,2)}\right)}{\sqrt{\left(1+(m-1)\rho_0^{(1)}\right)\left(1+(m-1)\rho_0^{(2)}\right)}}$$

General setup for RCTs was proposed by O'Brien et al. and Pocock et al., and we extend to the case of CRTs [11, 12].

Strengths and Limitations:

- ✓ Uses a weighted test statistic that utilizes variances and correlations
- X H_A doesn't require treatment be efficacious on both outcomes

Method 3: Single 1-DF Combined Test

Required Parameters:

Parameter	Description
β_1^* and β_2^*	The treatment effects on Y_1 and Y_2
$\rho_0^{(1)}$ and $\rho_0^{(2)}$	The endpoint specific ICC's for Y_1 and Y_2
$\rho_1^{(1,2)}$	Correlation between Y_1 and Y_2 for two different individuals in the same cluster
$\rho_2^{(1,2)}$	Correlation between Y_1 and Y_2 for the same individual
σ_1^2 and σ_2^2	The outcome variances, $\mathrm{Var}(Y_1)$ and $\mathrm{Var}(Y_2)$
m, K, π	Desired number of clusters, cluster size, and power

Design Equations:

$$\lambda = \left[\frac{\sqrt{\frac{(\beta_1^*)^2}{\frac{2\sigma_1^2}{Km}[1+(m-1)\rho_0^{(1)}]}} + \sqrt{\frac{(\beta_2^*)^2}{\frac{2\sigma_2^2}{Km}[1+(m-1)\rho_0^{(2)}]}}}{\sqrt{2\left(1+\frac{\left(\rho_2^{(1,2)}+(m-1)\rho_1^{(1,2)}\right)}{\sqrt{\left(1+(m-1)\rho_0^{(1)}\right)\left(1+(m-1)\rho_0^{(2)}\right)}}\right)}}\right]^2 ; \; K = \frac{2\lambda\left(1+\frac{\left(\rho_2^{(1,2)}+(m-1)\rho_1^{(1,2)}\right)}{\sqrt{\left(1+(m-1)\rho_0^{(1)}\right)\left(1+(m-1)\rho_0^{(2)}\right)}}\right)}{\left[\sqrt{\frac{(\beta_1^*)^2}{\frac{2\sigma_1^2}{m}\left[1+(m-1)\rho_0^{(1)}\right]}} + \sqrt{\frac{(\beta_2^*)^2}{\frac{2\sigma_2^2}{m}\left[1+(m-1)\rho_0^{(2)}\right]}}\right]^2}$$

m calculated in R with an equation solver

Method 4: Disjunctive 2-DF Test for Two Outcomes

Hypothesis Test: $H_0: \boldsymbol{L\beta^*} = \mathbf{0}$ vs. $H_A: \boldsymbol{L\beta^*} \neq \mathbf{0}$

Rationale: Simultaneously test both outcomes for any departure from H_0 which says there's no treatment effect across any outcome [13]. $\boldsymbol{L}$ is a $S \times Q$ contrast matrix, where S is the number of hypotheses being specified, and Q is the dimension of the outcome variable ($S \times Q = 2 \times 2$)

$$\boldsymbol{L} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \; H_0: \boldsymbol{L\beta^*} = \mathbf{0} \implies H_0: \begin{bmatrix} \beta_1^* = 0 \\ \beta_2^* = 0 \end{bmatrix}$$

Overall test statistic: $F^* = K(\boldsymbol{L}\beta^*)^T \left(\boldsymbol{L}\,\Omega_{\beta^*}\boldsymbol{L}^T\right)^{-1}\boldsymbol{L}\beta^*$

Where Ω_{β^*} is a 2×2 variance-covariance matrix for the treatment effect estimate β^* and has elements

$$\omega_1^2 = \frac{4\,\sigma_1^2\,[1+(m-1)\rho_0^{(1)}]}{m}; \; \omega_2^2 = \frac{4\,\sigma_2^2\,[1+(m-1)\rho_0^{(2)}]}{m}; \; \omega_{1,2} = \frac{4\sigma_1\sigma_2\,[\rho_2^{(1,2)}+(m-1)\rho_1^{(1,2)}]}{m}$$

assuming equal treatment allocation.

Strengths and Limitations:

- ✓ Setup and contrast matrix generalizable to ≥ 2 outcomes
- X H_A doesn't require treatment be efficacious on both outcomes

Method 4: Disjunctive 2-DF Test for Two Outcomes

Required Parameters:

Parameter	Description
β_1^* and β_2^*	The treatment effects on Y_1 and Y_2
$\rho_0^{(1)}$ and $\rho_0^{(2)}$	The endpoint specific ICC's for Y_1 and Y_2
$\rho_1^{(1,2)}$	Correlation between Y_1 and Y_2 for two different individuals in the same cluster
$\rho_2^{(1,2)}$	Correlation between Y_1 and Y_2 for the same individual
σ_1^2 and σ_2^2	The outcome variances, $\mathrm{Var}(Y_1)$ and $\mathrm{Var}(Y_2)$
m, K, π	Desired number of clusters, cluster size, and power

Design Equations:

$$\lambda = \left[\frac{Km[(\beta_1^*)^2\,\sigma_2^2\,VIF_2 \;-\; 2\,\beta_1^*\,\beta_2^*\,\sigma_1\,\sigma_2\,VIF_{12} \;+\; (\beta_2^*)^2\,\sigma_1^2\,VIF_1]}{2\,\sigma_1^2\,\sigma_2^2\,[VIF_1\,VIF_2 \;-\; VIF_{12}^2]}\right]$$

F-Distribution: $\pi = \int_{F_{1-\alpha}(S,2K-S-Q)}^{\infty} f(x;\lambda,2,2K-2-Q)dx$

χ^2 -Distribution: $\pi = \int_{\chi^2_{1-\alpha}(2)}^{\infty} \chi^2(x;\,2,\lambda)dx$

m can be calculated using an equation solver in R. If using χ^2, we can obtain equation for K, since in this case the degrees of freedom do not depend on the number of clusters. Note that $VIF_1 = 1 + (m-1)\rho_0^{(1)}$, $VIF_2 = 1 + (m-1)\rho_0^{(2)}$, and $VIF_{12} = \rho_2^{(1,2)} + (m-1)\rho_1^{(1,2)}$

Method 5: Conjunctive Intersection-Union Test

Hypothesis Test: $H_0: \beta_1^* = 0 \text{ or } \beta_2^* = 0$ vs. $H_A: \beta_1^* \neq 0 \text{ and } \beta_2^* \neq 0$

Rationale: This setup is powered for the alternative hypothesis that the intervention have an effect on both outcomes.

Yang et al. consider the vector of Wald test statistics; for two outcomes, this is $\boldsymbol{\zeta} = (\zeta_1, \zeta_2)^T$, where $\zeta_q = \sqrt{K_1 + K_2}\beta_q^*/\omega_1$ where ω_q is the estimated standard error of the treatment effect estimator and $K_1 + K_2$ are the total clusters in the study. Under equal treatment allocation, and 2 outcomes, we have:

$$\boldsymbol{\zeta} = [\zeta_1, \zeta_2]^T = \left[\frac{\beta_1^*\sqrt{2K}}{\sqrt{\frac{4\sigma_1^2 VIF_1}{m}}}, \quad \frac{\beta_2^*\sqrt{2K}}{\sqrt{\frac{4\sigma_2^2 VIF_2}{m}}}\right]^T,$$

where again $VIF_1 = 1 + (m-1)\rho_0^{(1)}$, $VIF_2 = 1 + (m-1)\rho_0^{(2)}$.

Strengths and Limitations:

- ✓ Only test that uses a conjunctive hypothesis, which is more in-line with the goal of Hybrid Type 2 studies
- X Has higher sample size requirements compared to other methods (but not as high as P-Value adjustment methods)

Method 5: Conjunctive Intersection-Union Test

Required Parameters:

Parameter	Description
β_1^* and β_2^*	The treatment effects on Y_1 and Y_2
$\rho_0^{(1)}$ and $\rho_0^{(2)}$	The endpoint specific ICC's for Y_1 and Y_2
$\rho_1^{(1,2)}$	Correlation between Y_1 and Y_2 for two different individuals in the same cluster
$\rho_2^{(1,2)}$	Correlation between Y_1 and Y_2 for the same individual
σ_1^2 and σ_2^2	The outcome variances, $\mathrm{Var}(Y_1)$ and $\mathrm{Var}(Y_2)$
m, K, π	Desired number of clusters, cluster size, and power

Design Equations:

$$\pi = \Pr\left(R = \bigcap_{q=1}^{2}\{\zeta_q > c_q\} \mid H_A\right) = \int_{c_1}^{\infty}\int_{c_2}^{\infty} f_{\boldsymbol{W}}(w_1, w_2)\, dw_1 dw_2$$

m and K can be calculated using software. *R* denotes the pre-specified rejection region, and c_1 and c_2 are the corresponding endpoint specific critical values for rejection for outcomes 1 and 2 respectively. $f_{\boldsymbol{W}\,(w_1,w_2)}$ can either be a multivariate normal density, or a multivariate t-distribution. When using t, it is denoted as $f_{\boldsymbol{T}(\omega_1,\omega_2)}$ with $2K - 2Q = 2K - 4$ degrees of freedom. Then, $c_1 = c_2 = t_\alpha(2K - 2Q)$,

Illustrative Example

The CIRCL Study

To illustrate the procedures and implications of the design methodologies for Hybrid 2 studies, we use **design input parameters** motivated by a study being conducted by the **Chicago Implementation Research Center** (CIRCL) [14, 15].

The **study specifications** are as follows:

Study Goal	• Evaluate the use of practice facilitation (PF) in order to help community health centers better utilize the community-adapted Kaiser bundle
Study Type	• Parallel cluster randomized design • Participating clinics (clusters) are randomized to 2 treatment arms
Treatment Arms	• **Experimental Group**: Kaiser bundle with PF • **Control Group**: Kaiser bundle without PF
Outcomes	• **Health Outcome**: Proportion of patients with controlled blood pressure (Yes/No) • **Implementation Outcome**: Reach, defined as the proportion of patients, among those who are eligible, who received the Kaiser bundle (Yes/No)

Input Parameters from CIRCL Study

Input Parameter	Description	Statistical Notation	Value
Number of clusters	Number of clinics in each treatment arm	K	15
Cluster Size	Number of patients in each clinic	m	300
Statistical Power	Probability of detecting a true effect under the alternative hypothesis	π	80%
Family-wise False Positive Rate	Probability of one or more Type I error(s)	α	0.05
Effect for Y_1	Estimated intervention effect on hypertension control, in percentage point increase	β_1^*	10%
Effect for Y_2	Estimated Intervention effect on reach, in percentage point increase	β_2^*	10%
Endpoint-specific ICC for Y_1	Correlation for hypertension control (Y_1) for two different individuals in the same clinic	$ICC(Y_1) = Corr(Y_{1,ikj}, Y_{1,ikj'}) = \rho_0^{(1)}$	0.025
Endpoint-specific ICC for Y_2	Correlation for reach (Y_2) for two different individuals in the same clinic	$ICC(Y_2) = Corr(Y_{2,ikj}, Y_{2,ikj'}) = \rho_0^{(2)}$	0.025
Inter-subject between-endpoint ICC	Correlation between hypertension control (Y_1) and reach (Y_2) for two different individuals in the same cluster	$Corr(Y_{1,ikj}, Y_{2,ikj'}) = \rho_1^{(1,2)}$	0.01
Intra-subject between-endpoint ICC	Correlation between hypertension control (Y_1) and reach (Y_2) for the same individual	$Corr(Y_{1,ikj}, Y_{2,ikj}) = \rho_2^{(1,2)}$	0.05
Total Variance of Y_1	Total variance of the hypertension control outcome	$\mathrm{Var}(Y_1) = \sigma_1^2$	0.23
Total Variance of Y_2	Total variance of the reach outcome	$\mathrm{Var}(Y_2) = \sigma_2^2$	0.25

Output from Five Design Methods

Results for the five study design approaches for CIRCL Hybrid 2 Study

Method	Power	K	m
1. P-Value Adjustments			
a. Bonferroni	84.55%	14	149
b. Sidak	84.67%	14	147
c. D/AP	84.98%	14	141
2. Combined Outcomes	**98.18%**	**8**	**23**
3. Single 1-DF Combined Test	98.11%	**8**	**23**
4. Disjunctive 2-DF Test			
a. Using χ^2-distribution	96.01%	9	34
b. Using F-distribution	93.63%	11	45
5. Conjunctive IU Test			
a. Using MVN-distribution	91.43%	11	74
b. Using t-distribution	89.92%	12	86

Output from Five Design Methods

Key Findings from CIRCL Study

Method 1: P-Value Adjustments

- Lowest power and higher sample size requirements
- We prove in the manuscript that $\alpha^{Bonferroni} < \alpha^{Sidak} < \alpha^{DAP}$, meaning that the Bonferroni will yield less power than the Sidak, which will yield less power than the D/AP approach
- Difference between the p-value adjustment methods will increase as the number of outcomes increases, but for $Q = 2$, the difference is very small

Method 2: Combined Outcomes and **Method 3**: Single 1-DF Test

- Yield the highest power and lowest sample size requirements
- This is a consequence of the hypothesis framework essentially combining the treatment effects of the two outcomes, and testing for them simultaneously
- When the two outcome ICC's and variances are the same (i.e. $\rho_0^{(1)} = \rho_0^{(2)}$ and $\sigma_1^2 = \sigma_2^2$), these two methods are mathematically equivalent

Method 4: Disjunctive 2-DF Test

- Uses linear hypothesis, less conservative than P-Value adjustments but larger rejection region than Method 5 since the alternative hypothesis does not require the intervention be effective on both outcomes

Method 5: Conjunctive IU Test

- Less power, but performed better than the p-value adjustment methods

Summary of Theoretical Findings

Description
P-Value Adjustment Methods are always less powerful than the **Combined Outcomes Approach**
P-Value Adjustment Methods are **always less powerful** than the **Single Weighted 1-DF Test**.
P-Value Adjustment Methods always have a smaller non-centrality parameter than the **Disjunctive 2-DF Test**. However, due to the differing degrees of freedom, there are cases where the P-Value Adjustment Methods can result in higher power than the Disjunctive 2-DF Test (though this is not typically observed in practice).
Combined Outcomes Approach is theoretically equivalent to the **Single Weighted 1-DF Test** when the outcome specific ICCs and variances between the two outcomes are the same, resulting in the same statistical power.
Single Weighted 1-DF Test has the same non-centrality parameter as the **Disjunctive 2-DF Test** when the cluster-corrected standardized effect sizes of the first and second outcomes are equal.

Summary

Summary

Summary of Findings

- Methods 2 and 3 yielded highest power, and were found to be mathematically equivalent under certain conditions
- Methods 5 uses conjunctive power, which requires treatment to significantly impact both outcomes in order to reject the null hypothesis; **performed better than the p-value adjustment methods**
- Paper detailing these results is under review and on arxiv that will serve as a resource for researchers to learn about design options for Hybrid Type 2 studies
  - Provides steps for how to calculate each study design parameter
  - Offers flexibility of unequal treatment assignment
- Software to enact these methods in R has been added to CRAN and github, and is called `crt2power`
- Paper includes closer examination of the performance of each study design method under different circumstances shown here is in progress, including a numerical study

Try our software – it seems to work really well

- https://cran.rproject.org/web/packages/crt2power/index.html
- https://github.com/melodyaowen/crt2power.

Thank You!

Yale University School of Public Health *and the* ***Center for Methods in Implementation and Prevention Science***

Co-Authors *Melody Owen, Geoffrey M. Curran, Justin D. Smith, Chao Cheng, Donna Spiegelman, Yacob Tedla, Ruyi Liu, and Fan Li*

Training in Implementation Science Research Methods *- T32HL155000*

References

[1] Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. American Journal of Public Health. 1999; 89:1322-7. DOI: https://doi.org/10.2105/AJPH.89.9.1322

[2] Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health. 2011; 38(2):65-76. DOI: 10.1007/s10488-010-0319-7

[3] Curran GM, Bauer M, Mittman B, Pyne JM, Stetler C. Effectiveness-implementation hybrid designs: combining elements of clinical effectiveness and implementation research to enhance public health impact. Med Care. 2012; 50(3):217-26. DOI: 10.1097/MLR.0b013e3182408812

[4] Donner A, Klar N. Design and Analysis of Cluster Randomization Trials in Health Research: Wiley; 2000 October 10, 2000. 192 p.

[5] Senn S, Bretz F. Power and sample size when multiple endpoints are considered. Pharm Stat. 2007; 6(3):161-70. DOI: 10.1002/pst.301

[6] Dmitrienko A, Bretz F, Westfall P, Troendle J, Wiens B, Tamhane A, et al. Multiple Testing Methodology. Taylor and Francis Group, LLC; 2010. p. 35-97.

[7] Chi GY. Some issues with composite endpoints in clinical trials. Fundam Clin Pharmacol. 2005; 19(6):609-19. DOI: 10.1111/j.1472-8206.2005.00370.x

[8] Rupert M. Simultaneous statistical inference: Springer Science; 2012.

[9] Sidak Z. Rectangular Confidence Regions for the Means of Multivariate Normal Distributions. Journal of the American Statistical Association. 1967; 62(318):626-33. DOI: https://doi.org/10.2307/2283989

[10] Sankoh AJ, Hueque MF, Dubey SD. Some Comments on Frequently Used Multiple Endpoint Adjustment Methods in Clinical Trials. Statistics in Medicine. 1997; 16:2529-42. DOI: 10.1002/(sici)1097-0258(19971130)16:22<2529::aid-sim692>3.0.co;2-j

[11] Pocock SJ, Geller NL, Tsiatis AA. The Analysis of Multiple Endpoints in Clinical Trials. Biometrics. 1987; 43(3):487-98. DOI: https://doi.org/10.2307/2531989

[12] O'Brien PC. Procedures for Comparing Samples with Multiple Endpoints. Biometrics. 1984; 40(4):1079-87. DOI: https://doi.org/10.2307/2531158

[13] Yang S, Moerbeek M, Taljaard M, Li F. Power analysis for cluster randomized trials with continuous coprimary endpoints. Biometrics. 2022. DOI: 10.1111/biom.13692

[14] Smith JD, Davis P, Kho AN. Community-Driven Health Solutions on Chicago's South Side. Stanf Soc Innov Rev. 2021; 19(3):A27-A9. DOI: 10.48558/85p7-3113

[15] Kho A, Daumit GL, Truesdale KP, Brown A, Kilbourne AM, Ladapo J, et al. The National Heart Lung and Blood Institute Disparities Elimination through Coordinated Interventions to Prevent and Control Heart and Lung Disease Alliance. Health Serv Res. 2022; 57 Suppl 1(Suppl 1):20-31. DOI: 10.1111/1475-6773.13983