ADVANCING CAUSAL INFERENCE METHODS FOR CLUSTER RANDOMIZED TRIALS

- Dr. Brennan Kahan
  - Senior Research Fellow, MRC Clinical Trials Unit at University College London, Institute of Clinical Trials and Methodology, London, UK
- Dr. Laura Balzer
  - Associate Professor, Division of Biostatistics, School of Public Health, University of California, Berkeley, USA
- Dr. Bingkai Wang
  - Postdoctoral Researcher, Department of Statistics and Data Science, The Wharton School, The University of Pennsylvania, USA
- Dr. Fan Li
  - Assistant Professor, Department of Biostatistics, Center for Methods in Implementation and Prevention Science, Yale University School of Public Health, USA
- Discussant: Dr. Karla Hemming
  - Professor, Institute of Applied Health Research, and College of Medical and Dental Sciences, University of Birmingham, UK

MRC Clinical Trials Unit

UCL

Estimands in Cluster-Randomized Trials: Choosing Analyses That Answer the Right Question

Brennan Kahan

Smarter Studies
Global Impact
Better Health

Outline

- Motivation
- Estimands for CRTs
- Estimation
- Conclusions

Motivation

Analyses of CRTs require must account for clustering

However, different methods of doing so lead to different results

Example: the TRIGGER trial[1]

Method	Odds ratio[a]
GEEs with independence correlation	3.69
GEEs with exchangeable correlation	3.83
Mixed-effects model	4.21
Cluster-level summaries	4.85

[a]For outcome “overall adherence”

MRC Clinical Trials Unit

UCL

[1]Jairath V, Kahan BC, Gray A, *et al*. Restrictive versus liberal blood transfusion for acute upper gastrointestinal bleeding (TRIGGER): a pragmatic, open-label, cluster randomised feasibility trial. Lancet. 2015 Jul 11;386(9989):137-44.

Example: the TRIGGER trial

Method	Odds ratio[a]
GEEs with independence correlation	3.69
GEEs with exchangeable correlation	3.83
Mixed-effects model	4.21
Cluster-level summaries	4.85

[a]For outcome “overall adherence”

- Each method of analysis targets a different **estimand**

MRC Clinical Trials Unit

UCL

[1]Jairath V, Kahan BC, Gray A, *et al*. Restrictive versus liberal blood transfusion for acute upper gastrointestinal bleeding (TRIGGER): a pragmatic, open-label, cluster randomised feasibility trial. Lancet. 2015 Jul 11;386(9989):137-44.

Estimands

They describe what happens to the **same** set of patients under **different** treatments[1]

Structured approach to describe treatment effect

Aim is to:

- Provide **clarity** around what's being estimated[2]
- Allow investigators to align **statistical analyses** to their **research question**

MRC Clinical Trials Unit

UCL

[1]ICH E9 (R1) addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials

[2]Kahan BC, Cro S, Li F, et al. Eliminating ambiguous treatment effects using estimands. *American Journal of Epidemiology* 2023

Estimands for CRTs

These have **additional** considerations compared to individually randomised trials[1, 2]

- Marginal vs. cluster-specific effect
- Participant- vs. cluster-average effect

MRC Clinical Trials Unit

UCL

[1]Brennan C Kahan, Fan Li, Andrew J Copas, Michael O Harhay, Estimands in cluster-randomized trials: choosing analyses that answer the right question, *International Journal of Epidemiology*, 2023.
[2]Brennan C Kahan, Michael Harhay, Scott Halpern, *et al*. Demystifying estimands in cluster-randomised trials. *ArXiv* 2023 https://arxiv.org/abs/2303.13960

Marginal vs. cluster-specific effects

Well studied; difference is in how summaries are applied

	Intervention	Control
Cluster 1	50/100	25/100
Cluster 2	75/100	60/100

Marginal vs. cluster-specific effects

Marginal: summarise potential outcomes under each treatment condition and contrast

	Intervention	Control
Cluster 1	50/100	25/100
Cluster 2	75/100	60/100
Overall	0.625	0.425

$$\frac{0.625/(1-0.625)}{0.425/(1-0.425)} = \mathbf{2.25}$$

Marginal vs. cluster-specific effects

Cluster-specific: summarise and contrast potential outcomes within each cluster, then take an average across clusters

	Intervention	Control	OR
Cluster 1	50/100	25/100	3.0
Cluster 2	75/100	60/100	2.0

$$\frac{(100)(3.0) + (100)(2.0)}{200} = \mathbf{2.50}$$

Marginal vs. cluster-specific effects

Marginal = cluster-specific for **collapsible** summary measures (e.g. differences)
Marginal ≠ cluster-specific for **non-collapsible** measures (e.g. odds ratio)

(magnitude of difference depends on ICC)

Participant- vs. cluster-average effects

Less well studied; difference is in how data is weighted

	Cluster size	Difference in POs (intervention vs. control)
Cluster 1	10	2
Cluster 2	10	2
Cluster 3	100	5
Cluster 4	100	5

Participant- vs. cluster-average effects

Participant-average: each participant is given equal weight

	Cluster size	Difference in POs (intervention vs. control)
Cluster 1	10	2
Cluster 2	10	2
Cluster 3	100	5
Cluster 4	100	5

$$\frac{(2)(10)(2) + (2)(100)(5)}{10 + 10 + 100 + 100} = \mathbf{4.7}$$

Participant- vs. cluster-average effects

Cluster-average: each cluster is given equal weight

	Cluster size	Difference in POs (intervention vs. control)
Cluster 1	10	2
Cluster 2	10	2
Cluster 3	100	5
Cluster 4	100	5

$$\frac{(2)(2) + (2)(5)}{4} = \mathbf{3.5}$$

Participant- vs. cluster-average effects

Participant-average = cluster-average when **there is no informative cluster size**

Informative cluster size (ICS)

- When outcome and/or treatment effect depends on size of cluster
- E.g. patients experience better outcomes when they present to larger hospitals as compared to smaller hospitals

Choice of estimand in CRTs

Need to define both marginal vs. cluster-specific **and** participant- vs. cluster-average aspects

- E.g. marginal cluster-average effect

Choice of estimand in CRTs

Need to define both marginal vs. cluster-specific **and** participant- vs. cluster-average aspects

- E.g. marginal cluster-average effect

Choice depends on study objectives

- Pragmatic trial aiming to assess impact of intervention as used in practice -> **marginal participant-average**
- How much an intervention is likely to improve test scores in a typical school -> **cluster-average**

Estimation

Each estimand requires **different** estimation methods

[1]Sullivan Pepe M, Anderson GL.. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Commun Stat Simul Comput* 1994.
[2]Seaman S, Pavlou M, Copas A.. Review of methods for handling confounding by cluster and informative cluster size in clustered data. *Stat Med* 2014.
[3]Williamson JM, Datta S, Satten GA.. Marginal analyses of clustered data when cluster size is informative. *Biometrics* 2003.

Estimation

Each estimand requires **different** estimation methods

Informative cluster size poses a **problem** for standard estimators

- Mixed-effects models/GEEs with exchangeable correlation may be **biased** for both participant- and cluster-average effects[1-3]
- This is because they use inverse-variance weighting

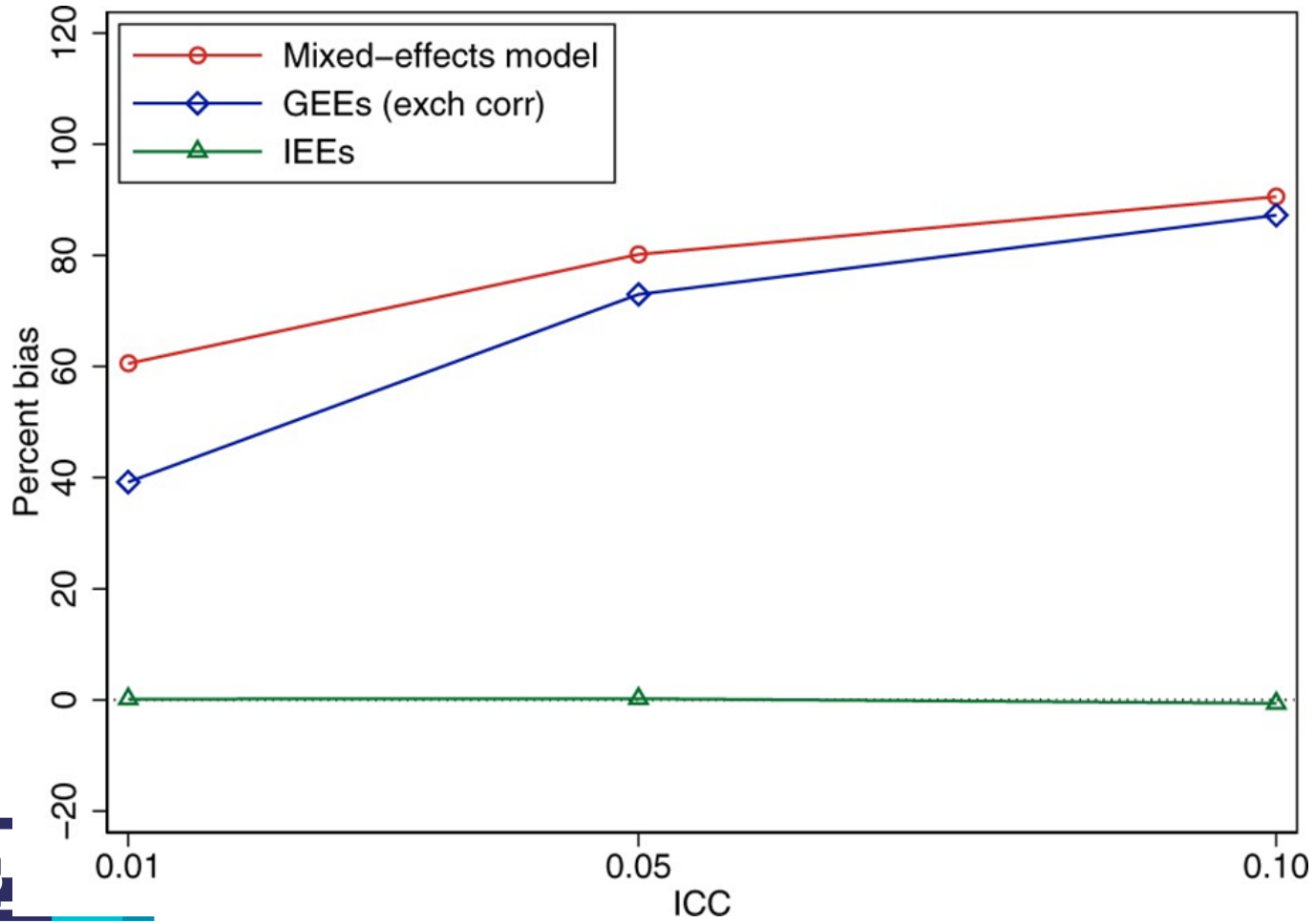
Degree of bias **increases** with ICC

MRC Clinical Trials Unit

UCL

[1]Sullivan Pepe M, Anderson GL.. A cautionary note on inference for marginal regression models with longitudinal data and general correlated response data. *Commun Stat Simul Comput* 1994.
[2]Seaman S, Pavlou M, Copas A.. Review of methods for handling confounding by cluster and informative cluster size in clustered data. *Stat Med* 2014.
[3]Williamson JM, Datta S, Satten GA.. Marginal analyses of clustered data when cluster size is informative. *Biometrics* 2003.

Example of bias[1]



Simulated example based on:

- 30 clusters with N=100 (treatment effect = 1)
- 30 with N=10 (treatment effect = 5)

Bias is for participant-average effect

[1]Brennan C Kahan, Fan Li, Andrew J Copas, Michael O Harhay, Estimands in cluster-randomized trials: choosing analyses that answer the right question, *International Journal of Epidemiology*, 2023.

Unbiased alternatives[1,2]

Independence estimating equations (IEEs)

- Uses working independence correlation structure in conjunction with cluster-robust SEs

Analysis of cluster-level summaries

- Calculate mean in each cluster and apply regression model to cluster-level summaries

MRC Clinical Trials Unit

UCL

[1]Brennan C Kahan, Fan Li, Andrew J Copas, Michael O Harhay, Estimands in cluster-randomized trials: choosing analyses that answer the right question, *International Journal of Epidemiology*, 2023.
[2]Brennan C Kahan, Michael Harhay, Scott Halpern, *et al*. Demystifying estimands in cluster-randomised trials. *ArXiv* 2023 https://arxiv.org/abs/2303.13960

What do they estimate?

IEEs estimate **marginal** effects[1]

- Can be used to estimate **either** participant- or cluster-average effects
- Depends on how data is weighted

Cluster-level summaries can estimate **either** marginal or cluster-specific effects

- **And** either participant- or cluster-average effects
- Depends on implementation[1]

MRC Clinical Trials Unit

UCL

[2]Brennan C Kahan, Michael Harhay, Scott Halpern, *et al*. Demystifying estimands in cluster-randomised trials. *ArXiv* 2023 https://arxiv.org/abs/2303.13960

Revisiting TRIGGER

Estimand	Estimator	Odds ratio (95% CI)
Marginal participant-average		
	IEEs (unweighted)	3.69 (1.83 to 7.43)
	Cluster-level summaries (weighted)	3.69 (1.83 to 7.43)
Cluster-specific participant-average		
	Cluster-level summaries (weighted)	4.28 (1.11 to 16.48)
Marginal cluster-average		
	IEEs (weighted)	3.92 (1.59 to 9.64)
	Cluster-level summaries (unweighted)	3.92 (1.51 to 10.19)
Cluster-specific cluster-average		
	Cluster-level summaries (unweighted)	4.85 (0.85 to 27.53)

Summary

Start with **estimand** -> **estimator**

If ICS is possible, standard estimators may be **biased**

Can use simple alternatives

- IEEs or cluster-level summaries

References

Brennan C Kahan, Fan Li, Andrew J Copas, Michael O Harhay, Estimands in cluster-randomized trials: choosing analyses that answer the right question, *International Journal of Epidemiology*, Volume 52, Issue 1, February 2023, Pages 107–118, https://doi.org/10.1093/ije/dyac131

Brennan C Kahan, Michael Harhay, Scott Halpern, Vipul Jairath, Andrew Copas, Fan Li. Demystifying estimands in cluster-randomised trials. *ArXiv* 2023 https://arxiv.org/abs/2303.13960

Acknowledgements

Fan Li
Michael Harhay
Andrew Copas
Bryan Blette
Vipul Jairath

B.C.K. and A.J.C. are funded by the UK MRC, grants MC_UU_00004/07 and MC_UU_00004/09. M.O.H. was supported by the National Heart, Lung, and Blood Institute of the United States National Institutes of Health (NIH) under award R00HL141678. M.O.H. was also supported by the Patient-Centered Outcomes Research Institute (PCORI) Award (ME-2020C1-19220). F.L. was supported by PCORI Awards (ME-2020C3-21072 and ME-2020C1-19220). All statements in this report are solely those of the authors and do not necessarily represent the views of the NIH, PCORI, its Board of Governors or Methodology Committee.

MRC Clinical Trials Unit

UCL

Thank you!

Two-Stage TMLE to reduce bias & improve efficiency in cluster randomized trials

Laura B. Balzer, PhD MPhil

Division of Biostatistics

School of Public Health

University of California, Berkeley

Berkeley Public Health

SEARCH

SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

@LauraBBalzer (laura.balzer@berkeley.edu)

No relevant disclosures

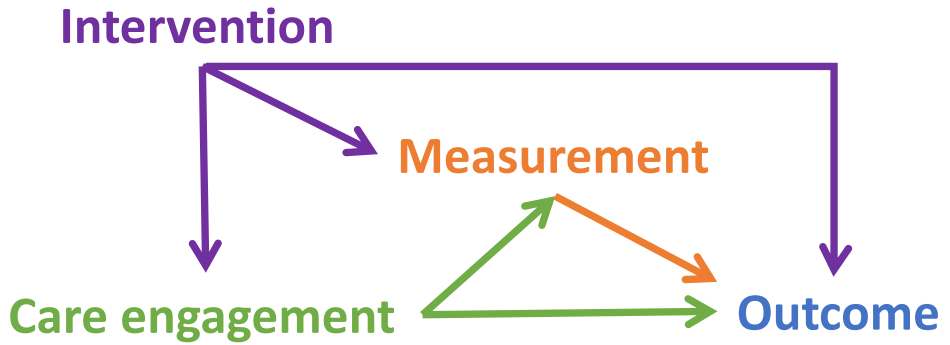
Berkeley Public Health

SEARCH
SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

@LauraBBalzer (laura.balzer@berkeley.edu)

Missing data happen!

- >90% of CRTs have missing participant outcomes (Fiero, *Trials,* 2016)
- Ignoring missingness can result in bias & misleading inference
  - e.g., through a complete-case analysis
- Exacerbated with post-baseline causes of missingness
- e.g., suppose the intervention increases care engagement, which improves participant outcomes and their chances of being measured
  - Care status mediates the treatment-outcome relationship & confounds the measurement-outcome relationship
  - Standard analytic approaches fail

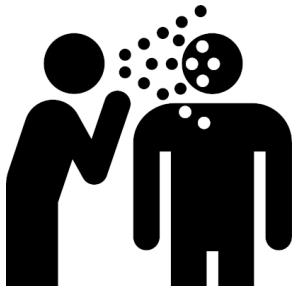


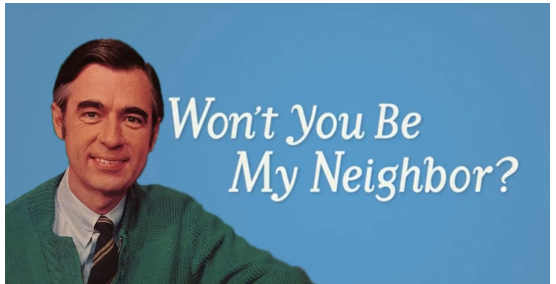
(laura.balzer@berkeley.edu)

Few clusters!

- By design, participants are dependent
  - e.g., clustered or grouped within households, clinics, schools, communities…
- # of independent units <<< # participants
  - Ignoring this dependence can result in bias & misleading inference
- Median of 30 clusters randomized (Selvaraj, *JAMA Int Med,* 2013)

Social network from Uganda-East with nodes as residents, edges as connections, & colors as village within the community. (Chen, Epi, 2021)





(laura.balzer@berkeley.edu)

Solution: Two-Stage TMLE

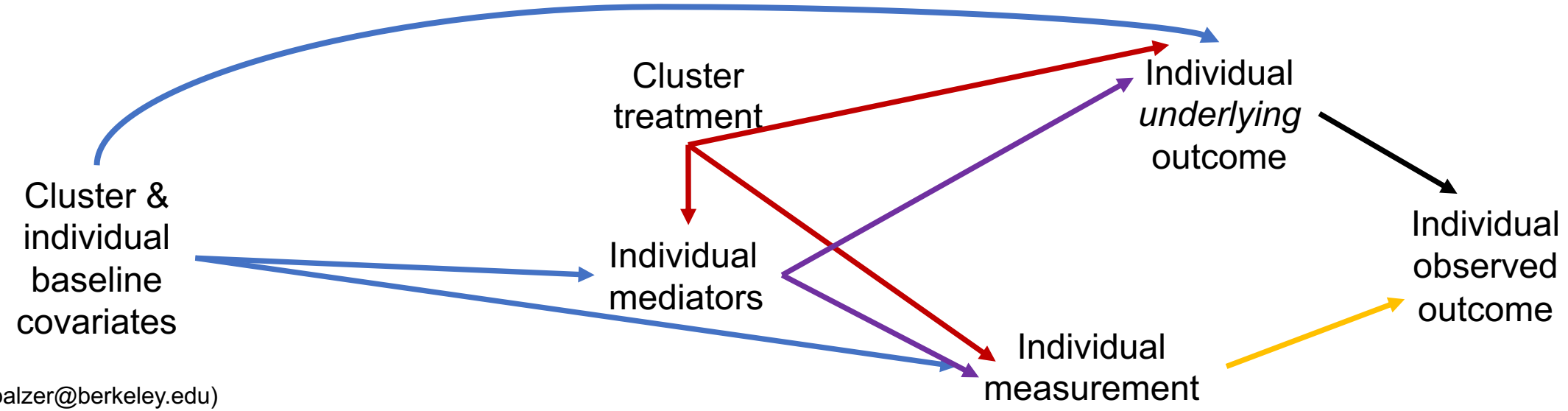
- Goal 1: Reduce bias due to missing outcomes
- Goal 2: Improve efficiency when estimating effects with few units

- Proposed, evaluated, and applied Two-Stage TMLE (targeted minimum loss-based estimation)
  - Separates control for differential missingness/censoring at the individual-level from evaluation of the intervention effect
- Stage 1: in each cluster separately, identify and estimate a cluster-level endpoint that appropriately accounts for missingness
- Stage 2: use the estimates from Stage 1 to evaluate the intervention effect with maximum precision

(Balzer, *Biostats*, 2021)

Stage 1: Identify & estimate in each cluster

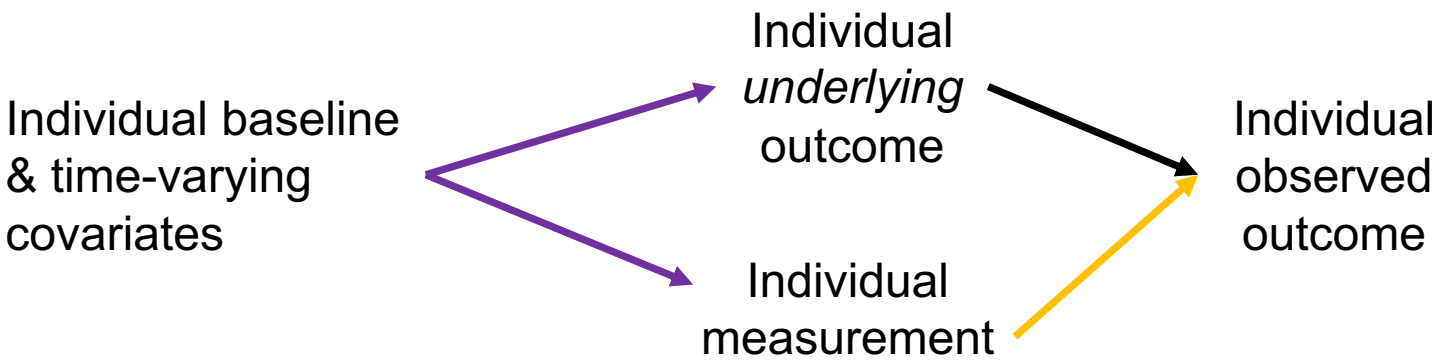
- **Identify:** Causal parameter, corresponding to hypothetical interventions to ensure complete outcome measurement, to a statistical estimand
- By stratifying on each cluster, we move from a complex problem
  - Involving multi-level data structure and the missing data equivalent of time-dependent confounding



(laura.balzer@berkeley.edu)

Stage 1: Identify & estimate in each cluster

- To a simpler identification problem
  - Since the cluster-level covariates and treatment are constant, we can effectively collapse the causal graph to



  - Lets the missingess mechanism vary by cluster
  - Applies to more complex settings, such as time-to-event outcomes

(laura.balzer@berkeley.edu)

Stage 1: Identify & estimate in each cluster

- We could implement a parametric approach:
  - G-computation (Robins, *Mathematical modeling,* 1986)
  - Inverse probability weighting (Horvitz, *JASA,* 1952)
  - But misspecification → bias and misleading inference

Like Brainy Smurf, parametric regressions pretend to know more than we actually do

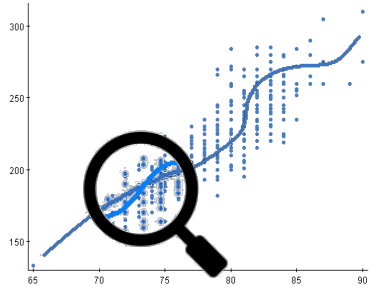
- We want to harness recent advances in machine learning
  - e.g., Super Learner: an ensemble method to combine predictions from multiple algorithms (van der Laan, *Stat App Gen*, 2007)
  - But simple application of machine learning results in the wrong bias-variance tradeoff + challenges for statistical inference

By our algorithms combined…

(laura.balzer@berkeley.edu)

Solution: Targeted Learning

- TMLE (Targeted minimum loss-based estimation): a general framework for the construction of doubly robust, semiparametric efficient, substitution estimators
- Updates of initial estimates of the expected outcome using information in treatment mechanism
  - Focus estimation where it matters most for our question
- Accurate quantification of uncertainty
  - Integrates machine learning with formal statistical theory



Like Robin Hood, we target to hit the bullseye!

(laura.balzer@berkeley.edu)

(van der Laan & Rubin, *IJB,* 2006; van der Laan & Rose, Targeted Learning, 2011)

Solution: Two-Stage TMLE

- Stage 1: within each cluster separately, identify and estimate an endpoint
  - Adjusting for differential missingness/censoring at the individual-level

- Stage 2: given those cluster-specific endpoint estimates, evaluate the intervention effect with maximum precision
  - Observed data are cluster-level
    - *W*: baseline covariates
    - *A*: indicator that randomized to the intervention
    - *Y*: cluster-specific endpoint, estimated from Stage 1

(laura.balzer@berkeley.edu)

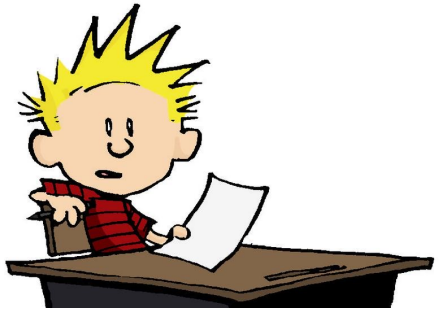
Stage 2: Unadjusted effect estimator

- At cluster-level, no confounding and complete follow-up
  - Identification is trivial
- Estimate the intervention effect by contrasting the average outcomes between treatment arms
  - e.g., $\widehat{\mathbb{E}}(Y \mid A = 1) - \widehat{\mathbb{E}}(Y \mid A = 0)$ or $\widehat{\mathbb{E}}(Y \mid A = 1) \div \widehat{\mathbb{E}}(Y \mid A = 0)$
- Unadjusted estimator
  - Unbiased - assuming Stage 1 appropriately accounts for missingness
  - But inefficient
  - Adjustment for baseline covariates increases precision in randomized trials (e.g., Gail, *Stat Med,* 1996; Hayes & Moulton, CRTs, 2009; Moore, *Stat Med,* 2009; Rosenblum, *IJB,* 2010; Turner, *Am J Pub Health*, 2017; Murray, *Annu Rev Public Health*, 2020; Benkeser, *Biometrics,* 2020)

(laura.balzer@berkeley.edu)

Stage 2 challenge: which covariates and what form?

- The analysis plan must be pre-specified
- Unclear *a priori* which covariates to include in the adjustment set *W* when
  - Predicting the outcome: $\mathbb{E}(Y|A,W)$
  - Estimating the treatment mechanism: $\mathbb{P}(A=1|W)$
  - Also what form?
- Breadth of adjustment is restricted by number of independent units
  - Must improve precision without sacrificing Type-I error control



(laura.balzer@berkeley.edu)

Solution: Adaptive Pre-specification

- Procedure to data-adaptively select the adjustment approach to maximize precision, while preserving Type-I error
- Pre-specify
  1. Candidate estimators of the conditional mean outcome $\mathbb{E}(Y|A,W)$
     - e.g., "working" GLMs (Rosenblum, *IJB,* 2010)
  2. Candidate estimators of the treatment mechanism $\mathbb{P}(A=1|W)=0.5$
     - Each candidate yields a different update to an initial estimator $\widehat{\mathbb{E}}(Y|A,W)$
  3. Measure of performance
     - Loss function: squared influence curve for the TMLE
     - Risk: variance of the TMLE
  4. Cross-validation (CV) scheme to select the candidate with the smallest CV-variance estimate

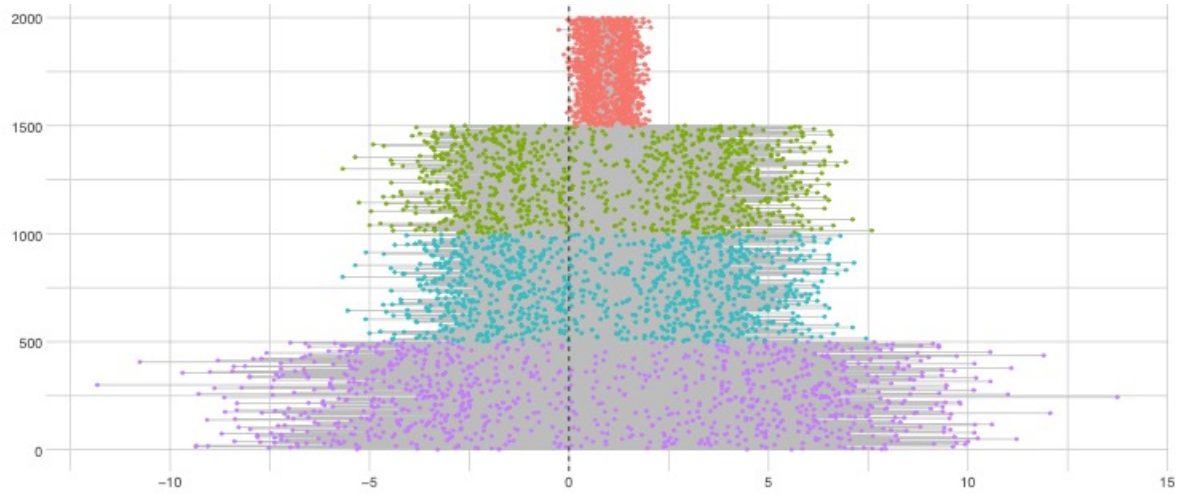
(Rubin, *IJB,* 2008; Balzer, *StatMed*, 2016)

Solution: Adaptive Pre-specification

- With these ingredients, we now have a fully pre-specified and automated procedure to data-adaptively select the adjustment approach that maximizes empirical efficiency

- Exciting side note: recently extended this procedure for larger randomized trials (N>40) with substantial gains in precision

- Figure: 95% confidence intervals with
  - unadjusted estimator (purple)
  - fixed adjustment (blue)
  - small-trial adaptive pre-specification (green)
  - large-trial implementation (red)
- Tech report: https://arxiv.org/abs/2210.17453

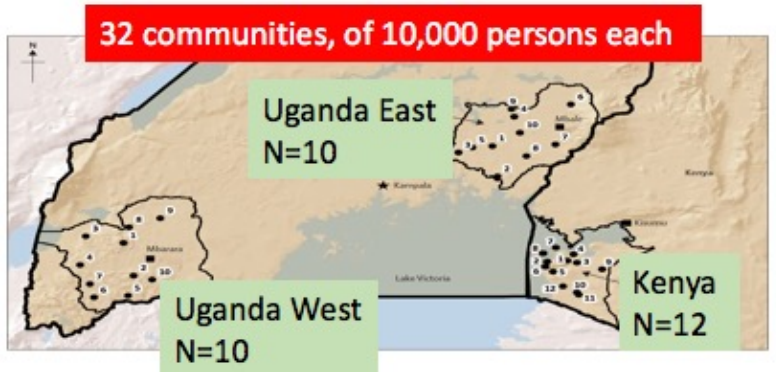


(laura.balzer@berkeley.edu)

Application to SEARCH

SEARCH
SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

- Universal HIV Test-and-Treat Trial
  - Test all for HIV and immediate treatment for persons with HIV
  - SEARCH used a community-based, multi-disease, patient-centered care approach
- Design: Cluster randomized trial of >320,000 participants in 32 communities in rural Uganda and Kenya
- www.searchendaids.com (NCT:01864603)



(laura.balzer@berkeley.edu)

Application to SEARCH

SEARCH
SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

- Stage 1 demonstration: impact of missing data on estimates of population-level HIV viral suppression in the intervention arm
  - Proportion of persons with HIV who are suppressing viral replication (<500c/mL)

$$\mathbb{P}(Suppressed = 1 \,|HIV = 1)$$

- Statistical analysis plan: https://arxiv.org/abs/1808.03231
- Code: https://github.com/LauraBalzer/SEARCH_Analysis_Adults
- Results: Havlir, *NEJM*, 2019; Balzer, *Epi,* 2020

(laura.balzer@berkeley.edu)

Application to SEARCH

SEARCH
SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

- High levels of testing achieved through community-based, multi-disease health fairs with follow-up for non-participants
  - e.g., at baseline, 90% of 150,395 adults (15+yrs) tested for HIV
  - Same mechanism to measure HIV RNA (viral loads)
- Tempting to estimate population-level suppression with the raw proportion

$$\frac{\#\ suppressed\ viral\ load}{\#\ measured\ viral\ load}$$

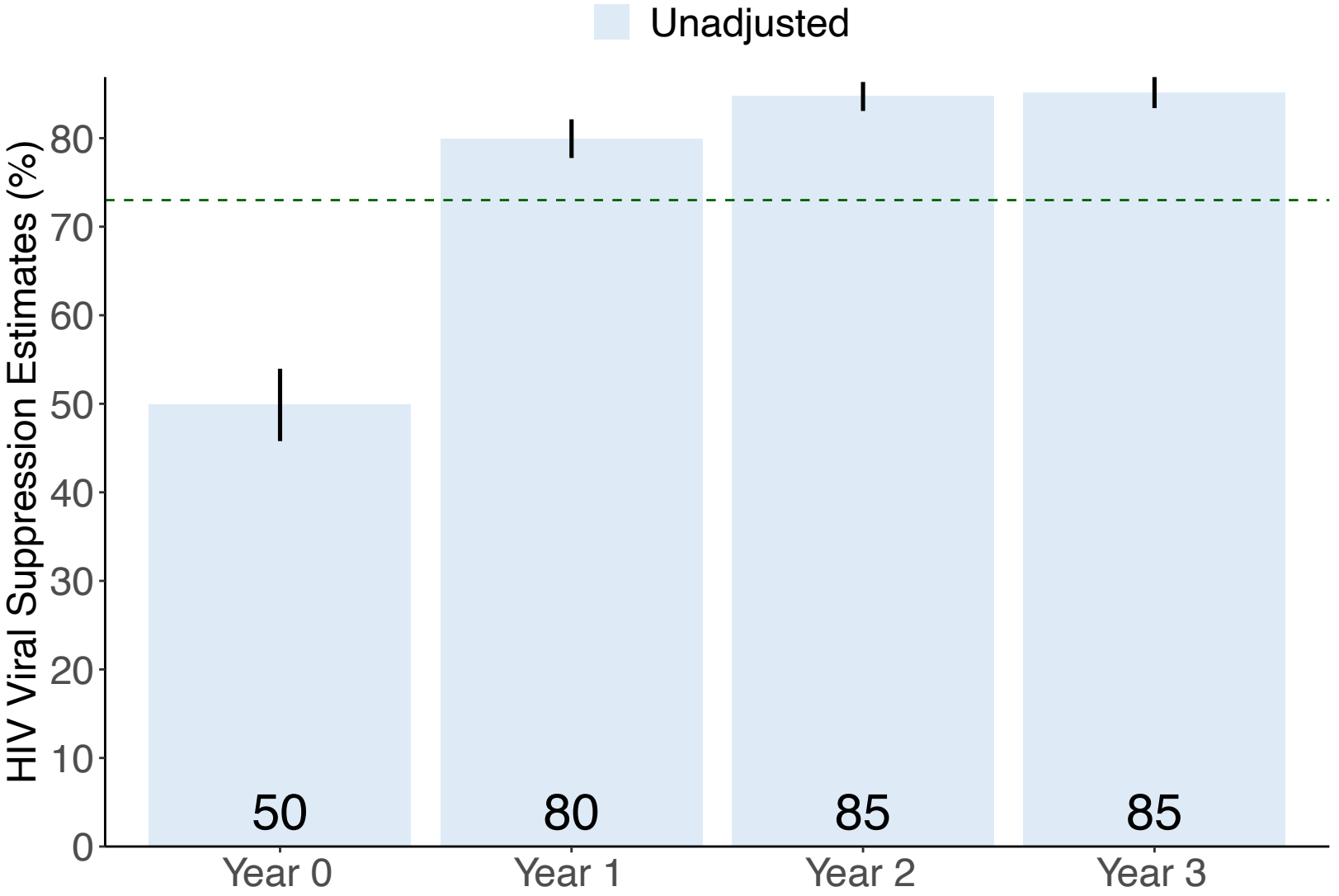
  - Rely on the missing-completely-at-random assumption

(laura.balzer@berkeley.edu)

(Chamie, *Lancet HIV*, 2016; Havlir, *NEJM*, 2019; Balzer, *Epi*, 2020)

Results:

- Unadjusted: Raw proportion among those measured

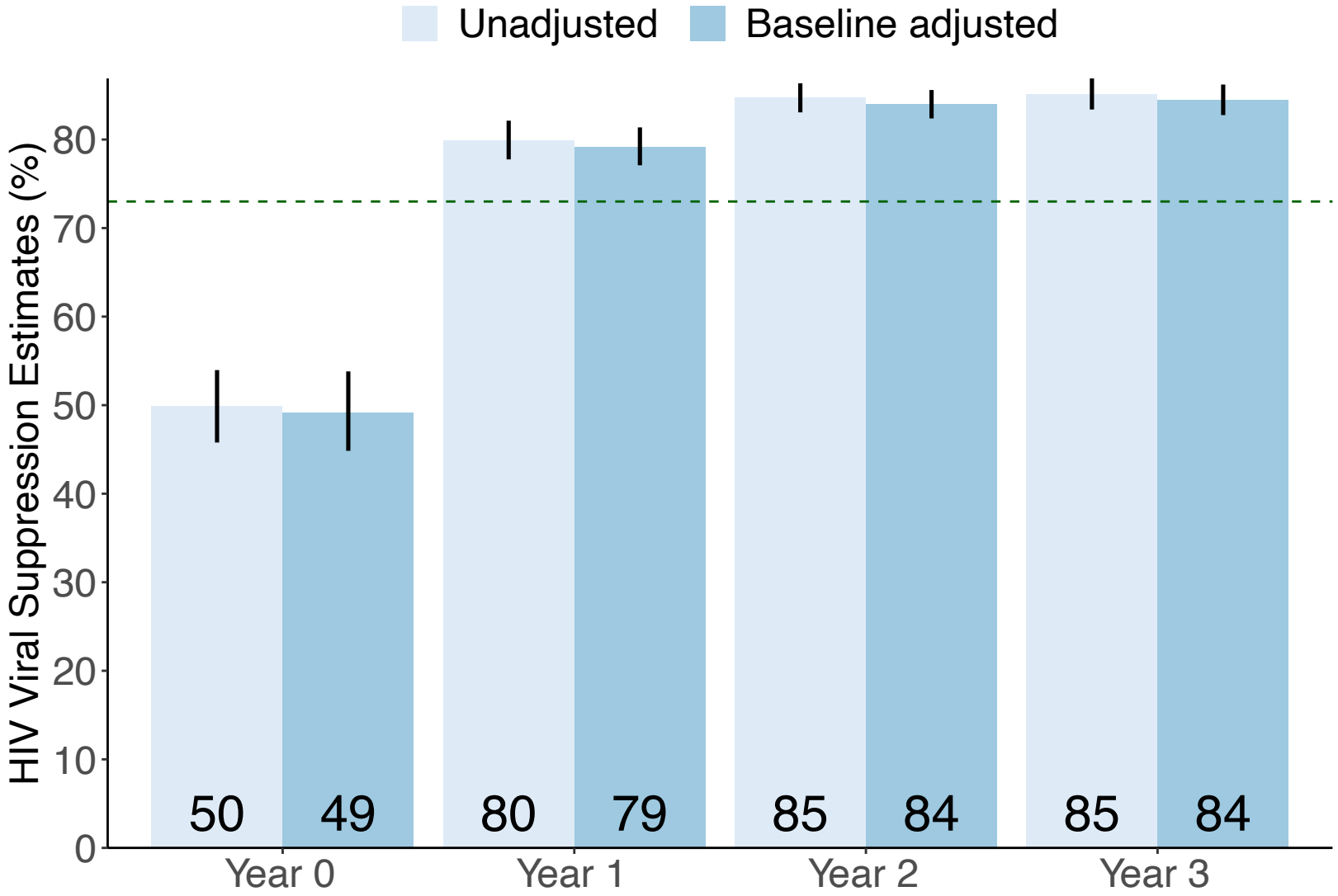


(laura.balzer@berkeley.edu)

(Balzer, *Epi*, 2020)

Results:

- Unadjusted (light blue)
- Baseline adjusted (medium blue): stratifying on sex, age-group, & community
  - Equivalent to saturated regressions in Gcomp. or inverse-weighting



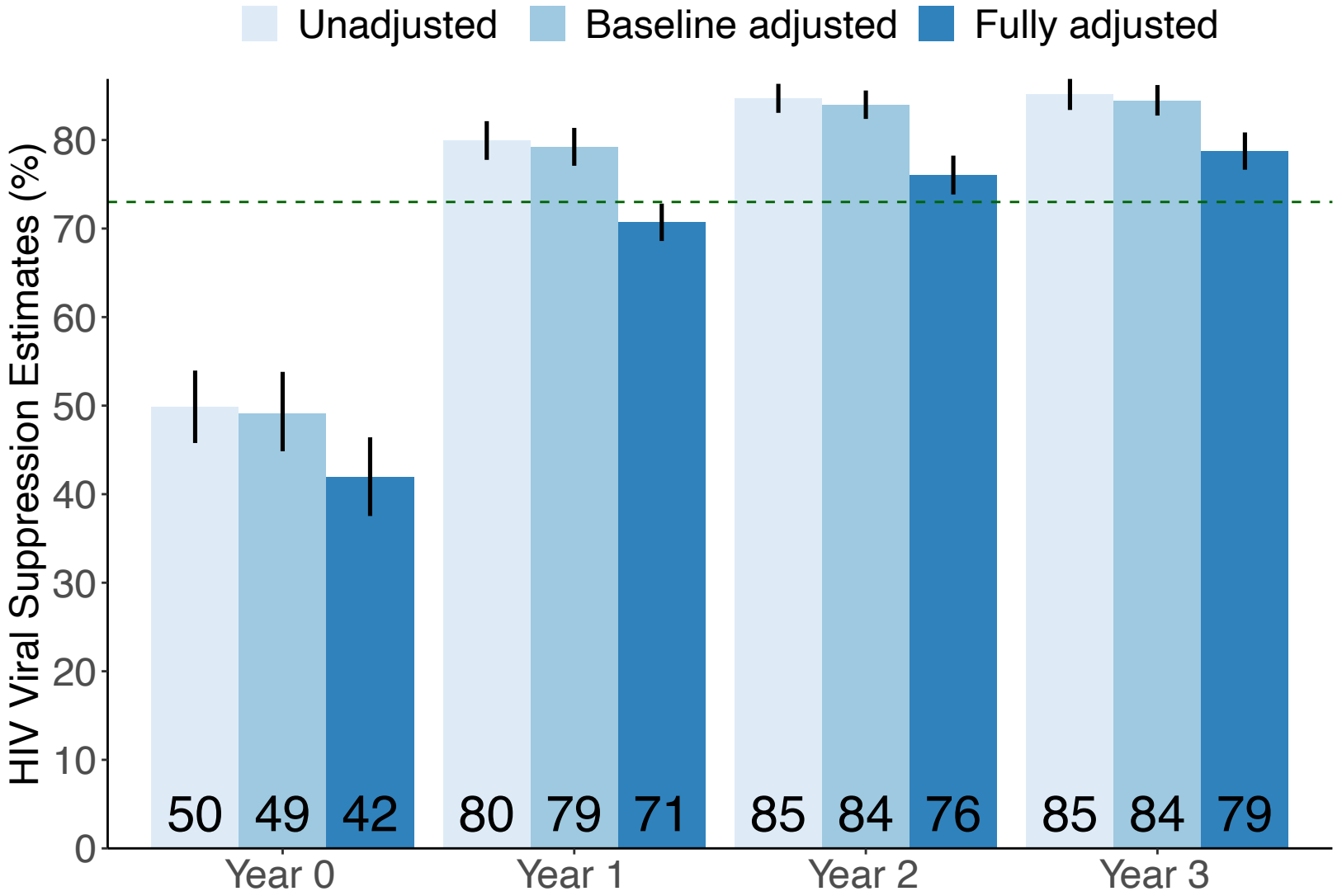
- Stratification on baseline predictors yielded slightly lower estimates
  - But not meaningfully different

(laura.balzer@berkeley.edu)

(Balzer, *Epi*, 2020)

Results:

- Unadjusted (light blue)
- Baseline adjusted (medium blue)
- Fully adjusted (dark blue): control for baseline & time-varying covariates with TMLE + Super Learner



- Full adjustment yielded substantially & meaningfully lower estimates
  - Sub-population with measured viral loads was enriched for persons with prior diagnoses, ART use, & suppression→ major drivers of current suppression

(Balzer, *Epi*, 2020)

Application to SEARCH

SEARCH
SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

- Stage 2 demonstration: impact of adjustment when estimating the intervention effect
  - Three-year cumulative incidence of HIV (primary study outcome)
    - Assessed through longitudinal follow-up of a closed cohort
  - Incidence of HIV-associated active tuberculosis (TB) or death to due illness
    - Time-to-event outcome subject to right-censoring
  - Population-level HIV viral suppression
    - Assessed cross-sectionally

- Focus on efficiency gains when using TMLE with Adaptive Pre-specification in Stage 2 after adjusting for missingness/censoring in Stage 1
  - Compare with an unadjusted effect estimator in Stage 2

(laura.balzer@berkeley.edu)

Focusing on Stage 2	Effect	95%CI	Efficiency*
Three-year cumulative HIV Incidence			
Unadjusted	0.98	(0.66, 1.45)	-
TMLE with Adaptive Prespec.	0.96	(0.80, 1.17)	4.6
Incidence of HIV-associated TB or death			
Unadjusted	0.79	(0.64, 0.98)	-
TMLE with Adaptive Prespec.	0.80	(0.69, 0.91)	2.6
Population-level Viral Suppression			
Unadjusted	1.15	(1.11, 1.20)	-
TMLE with Adaptive Prespec.	1.15	(1.11, 1.20)	1.0

*Precision gain: $\frac{variance\ of\ unadjusted}{variance\ of\ TMLE}$

(laura.balzer@berkeley.edu)

Summary

SEARCH
SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

- Two-Stage TMLE simultaneously addresses bias due to missing individual-level outcomes and imprecision due to few randomized units (i.e., clusters)
- Applicable to wide range of
  - Measurement schemes (e.g., cross-sectional sampling, longitudinal follow-up)
  - Endpoint types (e.g., binary, continuous, rate, time-to-event outcomes)
  - Causal parameters (e.g., population, conditional, sample effects)
- Simulations [not shown]: potential to overcome the shortcomings of existing methods, especially with post-baseline causes of missingness
- Application to SEARCH: real-life improvements

(laura.balzer@berkeley.edu)

Other ongoing work in CRTs

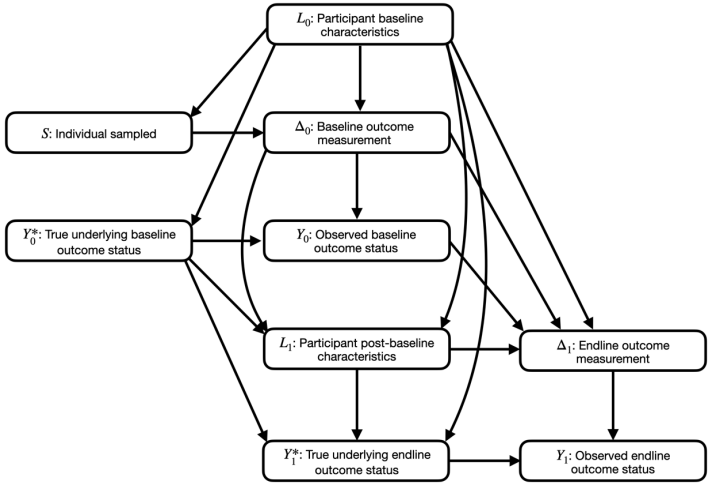
- Importance of specifying the research question
  - Traditionally, analyses of CRTs have allowed the method to determine the effect estimated
    - e.g., GEE with a logit-link yields odds ratios
  - With TMLE, we can learn effects
    - on any scale
    - at any level
  - e.g., in the PTBi study, the intervention reduced the **individual-level risk** of preterm infant mortality by 35% and the **cluster-level incidence** of preterm infant mortality by only 19%
  - Led by Alejandra Benitez

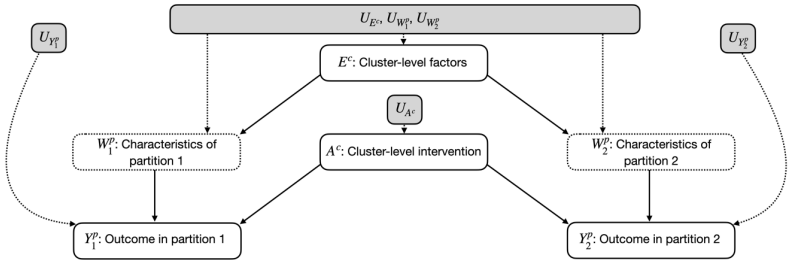
Effect	Estimate (95%CI)
Individual-level	0.65 (0.43, 0.98)
Cluster-level	0.81 (0.59, 1.11)

(Benitez, https://arxiv.org/abs/2110.09633, 2021)

Other ongoing work in CRTs

- More missingnesss
  - Sub-sampling, on the baseline outcome status among those sampled, and final outcome among those at risk
- Far fewer clusters
  - Assumptions increase our effective sample size, but blur the lines between CRTs and observational studies
- Application to SEARCH-TB:
  - Under *unrealistic* assumptions: SEARCH increased incident TB infection by 18%
  - After accounting for the sampling scheme, missingness, and within community dependence: SEARCH decreased the incident TB infection by 27%
- Led by Josh Nugent





(Nugent, https://arxiv.org/abs/2208.09508, 2022)

Acknowledgements: Study Participants and Communities
Republic of Kenya Ministry of Health, Republic of Uganda Ministry of Health

Funding: NIH/NIAID/NHLBI/NIMH/NIAAA (U01AI099959, UM1AI068636, U01AI150510)

Diane Havlir, Moses Kamya, Maya Petersen – Co-PIs

Makerere University & Infectious Diseases Research Collaboration
Jane Kabami
Elijah Kakande
Hellen Nakato
Asiphas Owaraganise
Edith Biira
Helen Sunday

University of California Berkeley
Maya Petersen
Laura Balzer
Josh Schwab

University of Pittsburg
Urvi Parikh

University College London
Andrew Phillips

Kenya Medical Research Institute
James Ayieko
Colette Aoko
George Agengo
Janice Lithunya
Marilyn Nyabuti
Norton Sang
Erick Wafula Mugoma
Elizabeth Bukusi

University of California San Francisco
Diane Havlir, Gabe Chamie, Cait Koss
Carol Camlin, Judy Hahn, Starley Shade,
Edwin Charlebois, Craig Cohen, Eric Goosby
Ted Ruel, Carina Marquez, Priscilla Hsue,
John Schrom, Jenny Temple, Monica Gandhi
Doug Black, Tamara Clark, Colie Sutter

NIH
Melanie Bacon
Joana Roe
Carl Dieffenbach
Carolyn Williams
Kendall Bryant
Xinzhi Zhang
Makeda Williams
Chris Gordon
Dianne Rausch

SEARCH
SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

PEPFAR
Eric Goosby
Deborah Birx
John Nkengesong

Gilead and ViiV for drug donation

NIH National Institute of Allergy and Infectious Diseases
NIH National Heart, Lung, and Blood Institute
NIH National Institute of Mental Health
NIH National Institute on Alcohol Abuse and Alcoholism

REPUBLIC OF KENYA
MINISTRY OF HEALTH
KENYA MEDICAL RESEARCH INSTITUTE KEMRI
THE REPUBLIC OF UGANDA MINISTRY OF HEALTH
IDRC
UCSF University of California San Francisco

Thanks to many others on the SEARCH team and to those who serve in advisory capacity!

Backup/bonus slides

SEARCH Study: Arms

SEARCH
SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

- Intervention:
  - Testing: Health fairs* at baseline and annually
  - Treatment: Universal eligibility
  - Care: Patient-centered, streamlined**
- Active control:
  - Testing: Health fairs* at baseline
  - Treatment: Country guidelines, changed over time
  - Care: Country standard

*Multi-disease (HIV, hypertension, diabetes, malaria, . . .) screening/linkage with follow-up for non-participants (Chamie, *LancetHIV*, 2016)

**Chronic care model, rapid ART start, welcoming environment, flexible clinic hours, mobile phone triage and reminders (Kwarisiimia, *JIAS*, 2017)

(laura.balzer@berkeley.edu)

SEARCH Study: Summary

SEARCH
SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

- A community health approach with a patient-centered, multi-disease model rapidly increased population-level HIV viral suppression from 42% to 79% (intervention) – compared to control (68%)

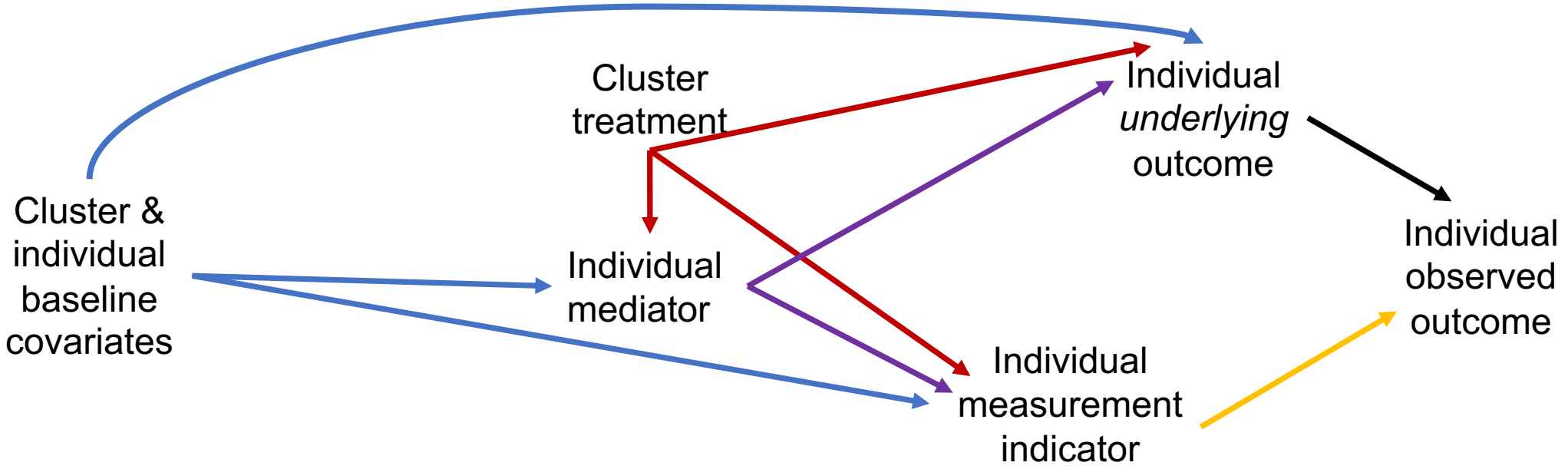
HIV incidence	Improved Community Health
↓ 32% Annual HIV incidence within arm	↓ 21% HIV mortality
↔ Cumulative HIV incidence between arms*	↓ 59% HIV/TB year 3 annual incidence
	↑ 26% HT control
*Explanation: Active control	Explanation: SEARCH intervention

(laura.balzer@berkeley.edu)

(Havlir, NEJM, 2019)

Simulation study: Design

- *N=30* clusters with ~150 participants/cluster
- Randomized the cluster-level intervention
- Included an individual-level mediator (i.e., post-baseline covariate)
- Included differential outcome measurement by arm



Simulation study: Estimators compared

1. Student's t-test, the only unadjusted estimator
2. Covariate adjusted residuals estimator (CARE) (Hayes & Moulton, CRTs, 2009)
3. Mixed models (a.k.a., random effects models)
4. Generalized estimating equations (GEE)
5. DR-GEE, an extension of GEE to incorporate weights to adjust for baseline drivers of missingness (Prague, *Biometrics*, 2016)
6. Two-Stage TMLE (Balzer, *Biostats*, 2021)

- Comparators selected, because they are most commonly implemented in CRTs (Turner, *AJPH*, 2017; Murray, *Prev Med,* 2018)

(laura.balzer@berkeley.edu)

Simulation study: Results for the risk difference (RD = -9.2%)

	$\widehat{pt}$	bias	σ	$\hat{\sigma}$	coverage
t-test	-32.0	-22.9	0.048	0.050	0.8
CARE	-21.8	-12.7	0.037	0.037	7.8
Two-Stage TMLE	-9.8	-0.7	0.038	0.046	98.8

$\widehat{pt}$: average point estimate (%)
bias: average deviation between estimates and effect (%)
σ: standard deviation of the point estimates
$\hat{\sigma}$: average standard error estimate
coverage: proportion of 95% confidence intervals containing the true effect (%)

(laura.balzer@berkeley.edu)

Simulation study: Results for the risk ratio (RR = 0.88)

	$\widehat{pt}$	bias	σ	$\hat{\sigma}$	coverage
Mixed models	0.72	-0.16	0.049	0.069	7.0
GEE	0.72	-0.16	0.049	0.056	4.8
DR-GEE	0.68	-0.20	0.049	0.054	0.2
Two-Stage TMLE	0.88	-0.01	0.051	0.062	98.4

$\widehat{pt}$: average point estimate
bias: average deviation between estimates and effect
σ: standard deviation of the point estimates (on log-scale)
$\hat{\sigma}$: average standard error estimate (on log-scale)
coverage: proportion of 95% confidence intervals containing the true effect (%)

(laura.balzer@berkeley.edu)

Stage 1: TMLE for missing individual-level outcomes

1. Among those measured, use Super Learner to flexibly model the relationship between the outcome and adjustment factors
2. Use the output from #1 to predict the outcome for all observations
3. Target these machine learning-based predictions with information in the propensity score (i.e., the probability of measurement, given the adjustment set)
   - Also fit with Super Learner
4. Average the targeted predictions

(laura.balzer@berkeley.edu)

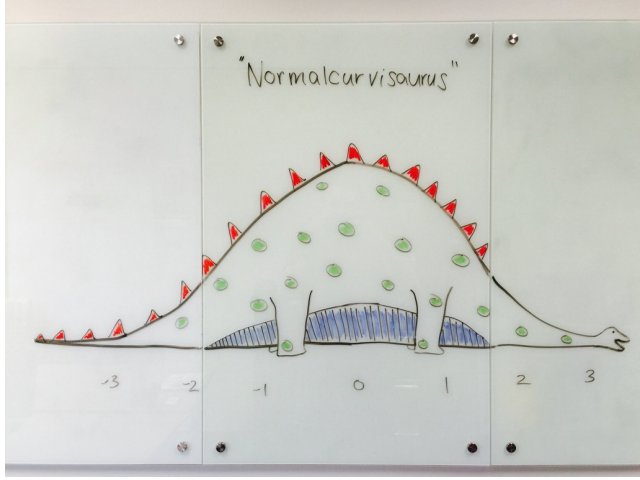
Stage: TMLE for the intervention effect in (without adaptive pre-specification)

1. Regress the outcome *Y* on the treatment *A* & adjustment variables *W:* $\widehat{\mathbb{E}}(Y|A,W)$
   - e.g., working logistic regression of estimated viral suppression on the intervention assignment and baseline HIV prevalence
2. Use the coefficients from #1 to predict the outcome under the intervention $\widehat{\mathbb{E}}(Y|A=1,W)$ and under the control $\widehat{\mathbb{E}}(Y|A=0,W)$
3. Target these predictions with information in the estimated treatment mechanism: $\widehat{\mathbb{P}}(A=1|W)$
   - Second opportunity to adjust for predictive covariates
4. Average the targeted predictions and contrast

(laura.balzer@berkeley.edu)

(Rosenblum, *IJB,* 2010; Balzer, SATE, *StatMed*, 2016)

A note on statistical inference



- Two-Stage TMLE is an asymptotically linear estimator

$$\hat{\psi} - \psi = \frac{1}{N}\sum_{i=1}^{N} IC_i + R_N$$

where $\hat{\psi}$ is the estimator, ψ the estimand, IC_i the influence curve for cluster *i,* and $R_N = o_P\left(N^{-\frac{1}{2}}\right)$ the remainder term, going to zero in probability

- Implies, by the Central Limit Theorem, normally distributed in the limit
  - Construction of 95% confidence intervals & hypothesis testing
- Briefly, requires a consistent estimator of the cluster-level endpoint in Stage 1
  - Missingness matters!
    - Having sufficiently sized clusters is important
  - Biased estimators of cluster-specific endpoints → biased estimates and misleading inference for the intervention effect

(Balzer, *Biostatistics*, 2021)

Summary

SEARCH
SUSTAINABLE EAST AFRICA RESEARCH IN COMMUNITY HEALTH

- Remaining challenge: Stage 1 adjustment for missingness/censoring occurs within each cluster separately
  - Limits breadth and flexibility of adjustment when the outcome is rare or the cluster-specific sample size is small

- Future work:
  - Address above challenge through a single stage approach
    - Extend Hierarchical TMLE for missing individual-level outcomes (Balzer, *SMMR*, 2019)
  - Extend to observational studies, where clustering is also common
    - Stage 2: Adjustment for the purposes of confounding control (instead of efficiency gains)

(laura.balzer@berkeley.edu)

Model-robust and Efficient Covariate Adjustment for Cluster-randomized Trials

Bingkai Wang
Department of Statistics and Data Science

Disclosure

- Research in this presentation was supported by a Patient-Centered Outcomes Research Institute Award. (PCORI. Award ME-2020C3-21072). The statements presented in this presentation are solely the responsibility of the authors and do not necessarily represent the official views of PCORI., its Board of Governors or Methodology Committee.

- Bingkai Wang also receives support from NIAID R01-AI148127.

Wharton
University of Pennsylvania

Motivating example: PPACT

Cognitive behavioral therapy for chronic pain

1:1 to 106 doctors

Model-based inference:
Interpret β_A as “treatment effect”

Linear mixed models (LMM) &
Generalized estimating equations (GEE)

Wharton
University of Pennsylvania

Model-based inference

- GEE and LMM were used by 68% of CRTs. *(Fiero et al. 2015, Trials)*
- **Issue 1:** The marginal estimand is unclear.
- **Issue 2:** Model misspecification.

Goal 1: robust inference with GEE or LMM

- robust : valid inference in large samples when models are arbitrarily wrong.

Our contribution 1: We adapted GEE and LMM to achieve **robust** inference for cluster- and individual-average estimands.

Wharton
University of Pennsylvania

Model-based inference

Issue 3: Parametric models are too restrictive to fully capture the power gain from covariate adjustment.

Goal 2: efficient inference

- efficient: theoretically maximized power in large samples.

Our contribution 2: We proposed efficient estimators using machine learning.

- We extended *Balzer (2019), Benitez (2021)* to handle cluster size variation dependent on treatment and covariates.
- Omitted from this talk.

Wharton
University of Pennsylvania

Outline

1. Setup

2. Robust inference with GEE and LMM

3. Data application

Wharton
University of Pennsylvania

Estimands

- The cluster-average estimand answers

  "What is the expected change in outcome associated with treatment for a typical cluster?"

- The individual-average estimand answers

  "What is the expected change in outcome associated with treatment for a typical patient?"

- We focus on the difference, ratio, and other scales of treatment effect.

Wharton
University of Pennsylvania

Setup

- Each cluster $i = 1, \dots, m$ has
  - population size N_i,
  - observed cluster size $M_i \leq N_i$,
  - a random treatment A_i.

- Each individual $j = 1, \dots, N_i$ has
  - an outcome Y_{ij},
  - covariates X_{ij}.

Assumptions

- Inter-cluster independence
- Arbitrary intra-cluster correlation
- (For simplicity) Independent observed cluster size with random sampling.

Wharton
University of Pennsylvania

Outline

1. Setup

2. Robust inference with GEE and LMM

3. Data application

Wharton
University of Pennsylvania

GEE for CRTs

- Mean model:

$$E\left[Y_{ij}\middle|A_i, L_{ij}\right] = g^{-1}\left(\beta_0 + \beta_A A_i + \beta_L L_{ij}\right)$$

Canonical link

User-specified covariates

- Estimation :

$$\sum_{i=1}^{m} \mathrm{D}_i^T \mathrm{V}_i^{-1}(Y_i^o - \mu_i^o) = 0.$$

Commonly use exchangeable correlation

- $\hat{\beta}_A$ targets neither estimand.

Wharton
University of Pennsylvania

Weighted g-computation estimator

- Step 1: estimate marginal means: $a = 0,1$

$$\hat{\mu}^{GEE}(a) = \frac{1}{m}\sum_{i=1}^{m}\frac{1}{M_i}\sum_{j} g^{-1}(\hat{\beta}_0 + \hat{\beta}_A a + \hat{\beta}_L L_{ij})$$

weighting g-computation

- Step 2: estimator $\hat{\Delta}^{GEE} = \hat{\mu}^{GEE}(1) - \hat{\mu}^{GEE}(0)$ for estimand Δ.
- Step 3: Sandwich variance estimator $\hat{V}^{GEE}$.

Wharton
University of Pennsylvania

Results for GEE

Theorem 1

$$\left(\hat{V}^{GEE}\right)^{-1/2}\left(\hat{\Delta}^{GEE} - \Delta\right) \xrightarrow{d} \mathrm{N}(0,1)$$

if GEE uses

(a) a correct mean model,

or (b) Gaussian working model,

or (c) an independence working correlation,

or (d) no individual-level covariate.

Wharton
University of Pennsylvania

Implications

- Many linear-adjusted estimators in the literature are special cases of GEE.
  (Imai et al, 2009, *Stat Sci*; Su and Ding 2021, *JRSSB*; Bugni et al. 2022, *arXiv*)
- Okay to not estimate within-cluster correlation.
- For non-independent cluster size, Theorem 1 still holds with independence working correlation and cluster weights $1/M_i$.

Wharton
University of Pennsylvania

LMM for CRTs

$$Y_{ij} = \alpha_0 + \alpha_A A_i + \alpha_L L_{ij} + \gamma_i + \epsilon_{ij}$$

random effect residual error

- LMM is a special case of GEE.
- Like GEE, we construct the weighted g-computation estimator $\hat{\Delta}^{LMM}$ with sandwich variance estimator $\hat{V}^{LMM}$.

Wharton
University of Pennsylvania

Results for LMM

Theorem 2

$$\left(\hat{V}^{LMM}\right)^{-1/2}\left(\hat{\Delta}^{LMM}-\Delta\right)\xrightarrow{d}\mathrm{N}(0,1).$$

Special case

Given 1:1 randomization, model-based inference based on $\hat{\alpha}_A$ is valid for Δ.

Wharton
University of Pennsylvania

Outline

1. Setup

2. Robust inference with GEE and LMM

3. Data application

Wharton
University of Pennsylvania

Results of data analyses with GEE

Study	Number of clusters	Cluster sizes	Estimates for Δ	RE
PPACT	106	1-10	-0.53 (-0.83, -0.22)	2.50
Work, Family, and Health Study	56	3-50	0.21 (0.11, 0.31)	2.86
Improving rational use of artemisinin	32	39-129	0.07 (-0.01, 0.15)	1.10

RE: relative efficiency compared to unadjusted analyses

We highly recommend using robust covariate adjustment in analyses of CRTs.

Wharton
University of Pennsylvania

More about this paper

Full paper
https://arxiv.org/abs/2210.07324

Personal webpage
https://bingkaiwang.com/

Co-authors
Chan Park, Department of Statistics and Data Science, University of Pennsylvania
Dylan Small, Department of Statistics and Data Science, University of Pennsylvania
Fan Li, Department of Biostatistics and Center for Methods in Implementation and Prevention Science, Yale University

Thank you!

Wharton
University of Pennsylvania

References

1. DeBar, L., Mayhew, M., Benes, L., Bonifay, A., Deyo, R. A., Elder, C. R., Keefe, F. J., Leo, M. C., McMullen, C., Owen-Smith, A., et al. (2022). A primary care–based cognitive behavioral therapy intervention for long-term opioid users with chronic pain: a randomized pragmatic trial. *Annals of Internal Medicine*, 175(1):46–55.
2. Fiero, M. H., Huang, S., Oren, E., & Bell, M. L. (2016). Statistical analysis and handling of missing data in cluster randomized trials: a systematic review. *Trials*, *17*(1), 1-10.
3. Wang, X., Turner, E. L., Li, F., Wang, R., Moyer, J., Cook, A. J., ... & Heagerty, P. J. (2022). Two weights make a wrong: Cluster randomized trials with variable cluster sizes and heterogeneous treatment effects. *Contemporary Clinical Trials*, *114*, 106702.
4. Small, Dylan S., Thomas R. Ten Have, and Paul R. Rosenbaum. "Randomization inference in a group–randomized trial of treatments for depression: covariate adjustment, noncompliance, and quantile effects." Journal of the American Statistical Association 103, no. 481 (2008): 271-279.
5. Imai, K., King, G., and Nall, C. (2009). The essential role of pair matching in cluster-randomized experiments, with application to the Mexican universal health insurance evaluation. Statistical Science, 24(1):29–53.
6. Stephens, A. J., Tchetgen Tchetgen, E. J., and Gruttola, V. D. (2012). Augmented generalized estimating equations for improving efficiency and validity of estimation in cluster randomized trials by leveraging cluster-level and individual-level covariates. Statistics in Medicine, 31(10):915–930.
7. Middleton, J. A. and Aronow, P. M. (2015). Unbiased estimation of the average treatment effect in cluster-randomized experiments. Statistics, Politics and Policy, 6(1-2):39–75.
8. Schochet, P. Z., Pashley, N. E., Miratrix, L. W., and Kautz, T. (2021). Design-based ratio estimators and central limit theorems for clustered, blocked RCTs. Journal of the American Statistical Association, pages 1–22.
9. Balzer, L. B., van der Laan, M., Ayieko, J., Kamya, M., Chamie, G., Schwab, J., Havlir, D. V., and Petersen, M. L. (2021). Two-Stage TMLE to reduce bias and improve efficiency in cluster randomized trials. Biostatistics. kxab043.
10. Su, F. and Ding, P. (2021). Model-assisted analyses of cluster-randomized experiments. Journal of the Royal Statistical Society, Series B, 83(5):994–1015.
11. Bugni, F., Canay, I., Shaikh, A., and Tabord-Meehan, M. (2022). Inference for cluster randomized experiments with non-ignorable cluster sizes. arXiv preprint arXiv:2204.08356.
12. Work, Family, and Health Study (WFHS) (2018). Work, family and health network. *Inter-university Consortium for Political and Social Research [distributor].*
13. Prudhomme O'Meara, W., Menya, D., Laktabai, J., Platt, A., Saran, I., Maffioli, E., Kipkoech, J., Mohanan, M., and Turner, E. L. (2018). Improving rational use of acts through diagnosisdependent subsidies: Evidence from a cluster-randomized controlled trial in western kenya. *PLoS medicine*, 15(7):e1002607.

A framework for causal inference with stepped-wedge cluster randomized trials

Fan Li

Department of Biostatistics
Center for Methods in Implementation and Prevention Science (CMIPS)
Yale Center for Analytical Sciences (YCAS)
Yale School of Public Health

https://lifan90.com/

Joint work with Bingkai Wang (Penn) and Xueqi Wang (Yale)

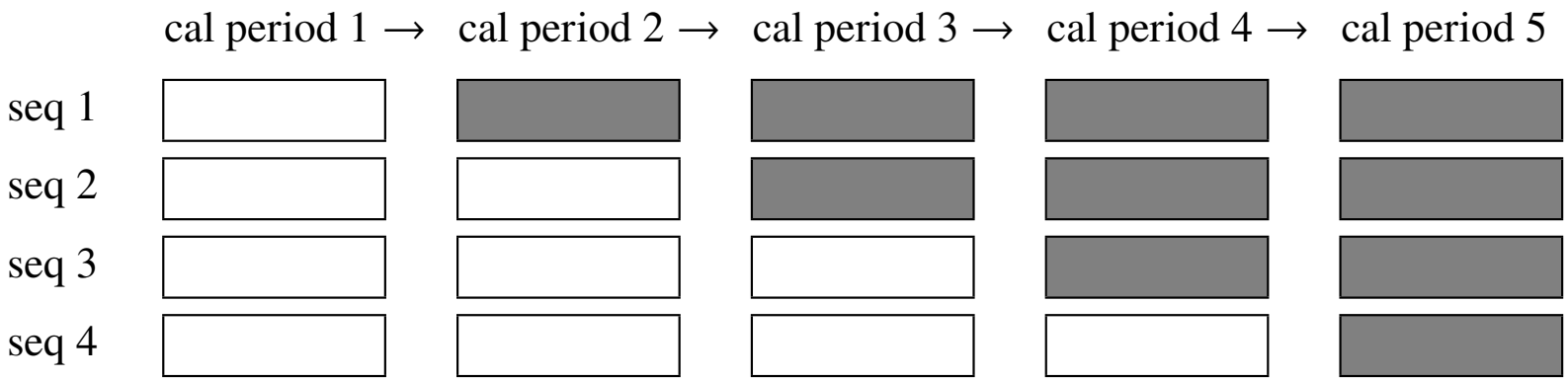
The Society for Clinical Trials (SCT) 44th Annual Meeting
May, 2023

Disclosure

- Research in this presentation was partially supported by a Patient-Centered Outcomes Research Institute Award® (PCORI® Award ME-2020C3-21072).
  - *New Methods for Planning Cluster Randomized Trials to Detect Treatment Effect Heterogeneity*
- F.L. acknowledges an incoming PCORI® award (PCORI® Award ME-2022C2-27676) that will support the future development of this ongoing work.
  - *Toward Improved Design and Analysis of Stepped Wedge Trials: An Estimand-aligned and Efficiency-focused Framework*
- The statements presented are solely the responsibility of the authors/study team and do not necessarily represent the official views of PCORI®, its Board of Governors or Methodology Committee.
- F.L. also receives support from the NIA IMPACT Collaboratory and NIH Pragmatic Trials Collaboratory

Stepped wedge cluster randomized trials

- Stepped wedge cluster randomized trials (SW-CRTs) sequentially transition experiment units (clusters) from control to the intervention conditions



  - each sequence (seq) can include multiple clusters
  - calendar period (cal period) often defined evenly

- SW-CRTs are increasingly popular in pragmatic trials

  - ensure full roll-out of an intervention during the study period
  - feasible to allocate finite resources at multiple calendar periods

Elements for causal inference

- Estimands in SW-CRTs
  - informative cluster size or cluster-period size
  - two time scales: calendar time and exposure time
- Baseline covariate adjustment are common[1]
  - covariate adjustment methods that preserve the target estimand under misspecification
- A clear conceptual framework is necessary
  - **Goal**: discuss the set of nonparametric model space for which these developments can be supported

[1]Nevins, Pascale, et al. A Scoping Review described diversity in methods of randomization and reporting of baseline balance in Stepped-Wedge Cluster Randomized Trials. *JCE* (2023).

Elements for causal inference

- Estimands in SW-CRTs
  - informative cluster size or cluster-period size (✗)
  - two time scales: calendar time and exposure time (✓)
- Baseline covariate adjustment are common[2]
  - covariate adjustment methods that preserve the target estimand under misspecification (✓)
- A clear conceptual framework is necessary
  - **Goal**: clarify the set of nonparametric model space for which these developments can be supported

[1]Nevins, Pascale, et al. A Scoping Review described diversity in methods of randomization and reporting of baseline balance in Stepped-Wedge Cluster Randomized Trials. *JCE* (2023).

Notation

- I: total number of clusters
- M_i: finite, source population size in cluster $i = 1, \ldots, I$
- $J + 2$: total number of periods, indexed by $j = 0, \ldots, J + 1$
  - in period 0, all clusters are in the control condition
  - as time proceeds, each cluster will start receiving treatment in a period randomly chosen among $\{1, \ldots, J + 1\}$
  - in period $J + 1$, all clusters are in the treatment condition
- Y_{ijk}: observed outcome of individual k in cluster i in period j
- $\boldsymbol{X}_{ik}$: baseline covariates for individual k in cluster i (time invariant)
- S_{ijk}: binary sampling indicator
- $N_{ij} = \sum_{k=1}^{M_i} S_{ijk}$: number of observed individuals in period j in cluster i

Conceptualization - cross-sectional sampling

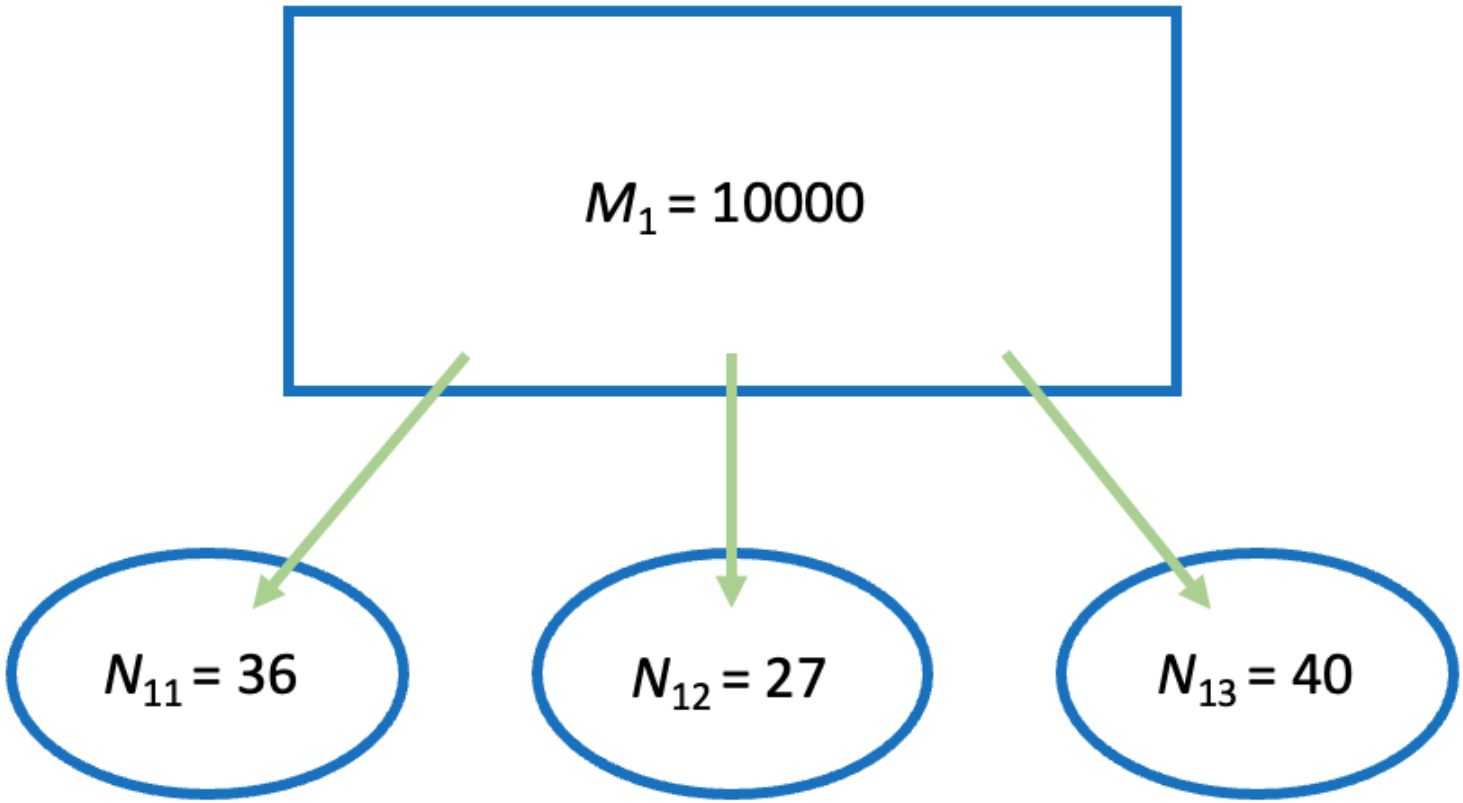


Figure 1: An illustrative example of the sampling for a cluster ($i = 1$) in a SW-CRT with $J = 3$.

Notation - cont'd

- $Z_i = j$: if cluster i starts receiving treatment in the beginning of period j, $j \in \{1, \ldots, J+1\}$
- Observed data for each cluster:
  $\boldsymbol{O}_i = \{Y_{ijk}, \boldsymbol{X}_{ik}, Z_i : S_{ijk} = 1, j = 0, 1, \ldots, J+1, k = 1, \ldots, M_i\}$
- Pursue the potential outcomes framework
  - $Y_{ijk}(d)$: potential outcome of individual k in cluster i at period j if the cluster has been treated for exactly d periods, i.e., $Z_i = j - d + 1$
  - a version of Stable Unit Treatment Value Assumption

$$Y_{ijk} = \sum_{d=1}^{j} I\{Z_i = j - d + 1\} Y_{ijk}(d) + I\{Z_i > j\} Y_{ijk}(0).$$

  - complete, but not fully observed, data vector for each cluster i:
    $\boldsymbol{W}_i = \{(Y_{ijk}(d), S_{ijk}, \boldsymbol{X}_{ik}, Z_i, M_i) : k = 1, \ldots, M_i, 1 \leq d \leq j \leq J\}$
  - omit data from period 0 and $J+1$ for now (positivity/overlap)

Nonparametric assumptions

Formalize a super-population framework to enable causal inference

Assumption 1 (Super-population): $\boldsymbol{W}_i$ for $i = 1, \ldots, I$ are i.i.d. samples from an unknown distribution with finite second moments

Assumption 2 (Staggered randomization): $Z_i \perp \boldsymbol{W}_i$. In addition, $P(Z_i = j) = \pi_j > 0$ for $j = 1, \ldots, J+1$ so that $\sum_{j=1}^{J+1} \pi_j = 1$

Assumption 3 (Non-informative cluster size): Within each cluster i, the vectors $\{Y_{i1k}(0), \ldots, Y_{iJk}(J), \boldsymbol{X}_{ik}\}$ for $k = 1, \ldots, M_i$ are identically distributed given M_i

Assumption 4 (Non-informative enrollment): $\{S_{ijk} : j = 1, \ldots, J, k = 1, \ldots, M_i\}$ is independent of the other random variables in $\boldsymbol{W}_i$ given M_i, and $\{N_{i1}, \ldots, N_{iJ}\} \perp M_i$

Estimands

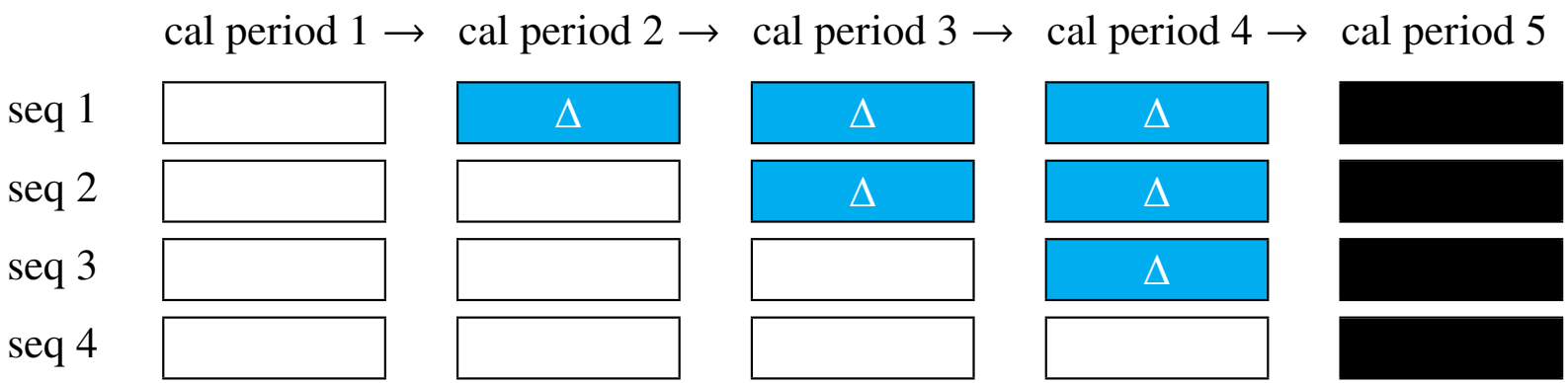
- Interested in inferring the class of average treatment effect estimands

$$\Delta_j(d) = E\{Y_{ijk}(d)\} - E\{Y_{ijk}(0)\}, \quad \text{for } 1 \leq d \leq j \leq J.$$

  - model-free definition!
  - $J(J+1)/2$ estimands in total
  - as a starting point, we do not consider estimands defined for the last period $j = J+1$ (partially because $Y_{i,J+1,k}(0)$ is not well-defined, or truncated by design)
  - likewise, $Y_{i1k}(d)$ for $d \geq 1$ is also truncated by design
  - a few examples to follow

Example 1: Constant treatment effect Δ

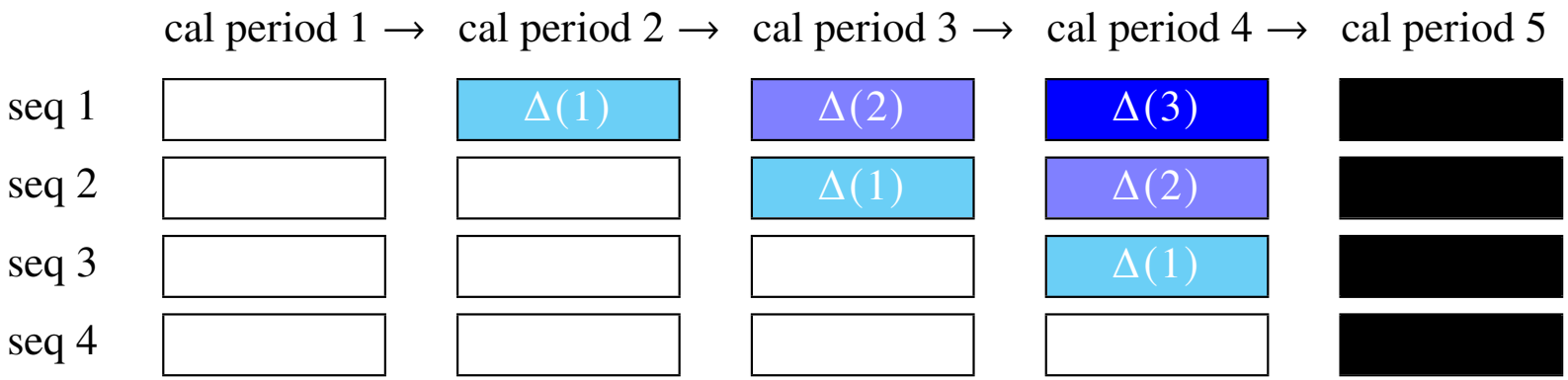
- Assume $\Delta_j(d)$ is constant for all j and d, i.e., $\Delta_j(d) \equiv \Delta$



  - The majority of the existing literature is built on this assumption (Hussey and Hughes, 2007)

Example 2: Duration-specific treatment effect $\boldsymbol{\Delta}^D = (\Delta(1), \ldots, \Delta(J))^\top$

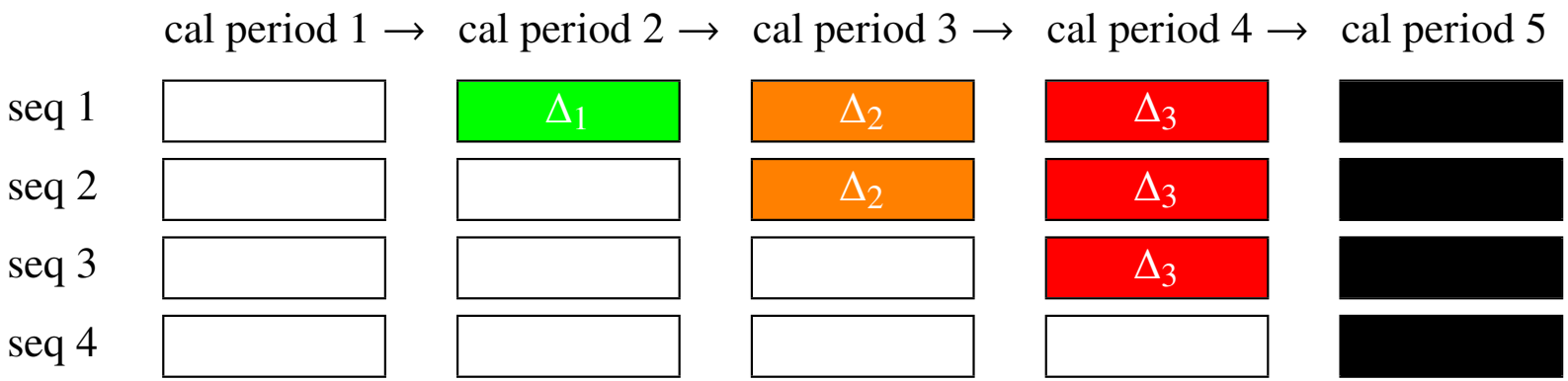
- Consider $\Delta_j(d)$ to be constant across j, but vary by d, i.e., $\Delta_j(d) = \Delta(d)$



  - ignoring duration effects (exposure-time heterogeneity) can lead to erroneous conclusions (Kenny et al, 2022; Mayeleff et al. 2022)
  - dynamic causal effects in the difference-in-differences literature (e.g. Sun and Abraham, 2021)

Example 3: Period-specific treatment effect $\boldsymbol{\Delta}^P = (\Delta_1, \ldots, \Delta_J)^\top$

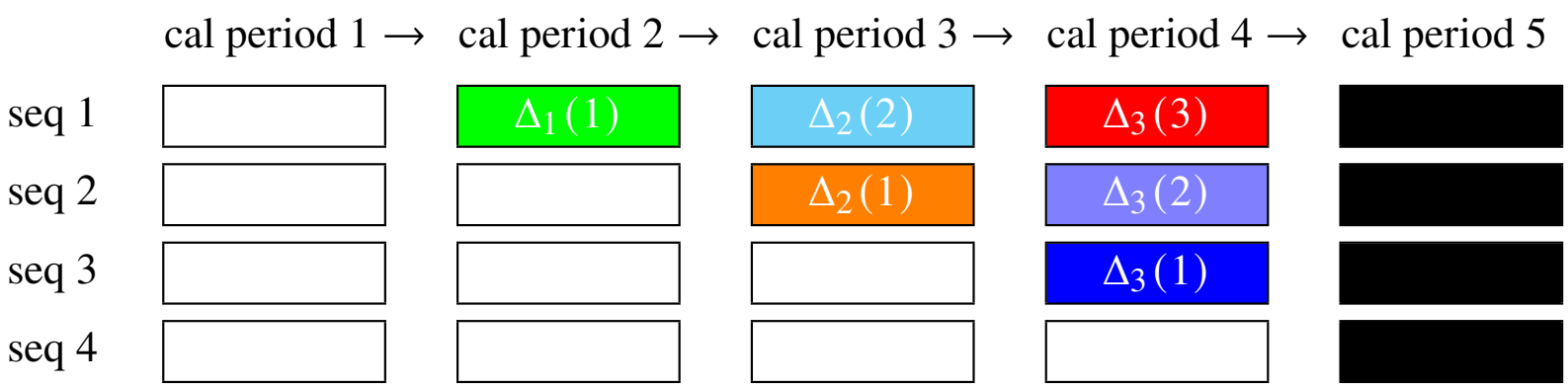
- Allow the treatment effect to vary by the period of measurement, but not the duration of treatment, i.e.,$\Delta_j(d) = \Delta_j$



  - account for unmeasured, exogenous time-varying factors (such as external shock) that contribute to variations of potential outcomes
  - not sufficiently explored in prior literature

Example 4: Saturated treatment effect $\boldsymbol{\Delta}^S$ $= (\Delta_1(1), \Delta_2(1), \Delta_2(2), \ldots, \Delta_J(J))^\top$

- The most granular set of estimands with fine resolution



  - make the least assumptions but include the most parameters to estimate

Mixed-model analysis of covariance (ANCOVA)

- Consider the following working linear mixed model

$$Y_{ijk} = \beta_{0j} + TE_{ij} + \boldsymbol{\beta}_{\boldsymbol{X}}^{\top} \boldsymbol{X}_{ik} + RE_{ij} + \varepsilon_{ijk}, \tag{1}$$

  - β_{0j}: intercept parameter for period j
  - TE_{ij}: term encoding treatment effects
  - $\boldsymbol{\beta}_{\boldsymbol{X}}$: coefficient for baseline variables
  - RE_{ij}: term for random effects to account for intracluster correlation
  - $\varepsilon_{ijk} \sim N(0, \sigma^2)$: residual random error
  - $\boldsymbol{\beta}_{\boldsymbol{X}}$ can be replaced by $\boldsymbol{\beta}_{\boldsymbol{X}j}$ to reflect a period-varying correlation, and all of our asymptotic results will still hold under this setting
  - mixed ANCOVA 1 model (Wang et al., 2021+)

Parameterization of treatment effect

- Specification of the treatment effect term TE_{ij}
  - constant treatment effect Δ:

    $TE_{ij} = \beta_Z I\{Z_i \leq j\}$ with a scalar treatment effect coefficient β_Z
  - duration-specific treatment effect $\boldsymbol{\Delta}^D$:

    $TE_{ij} = \sum_{d=1}^{j} \beta_{Zd} I\{Z_i = j - d + 1\}$, where β_{Zd} targets the treatment effect for duration d
  - period-specific treatment effect $\boldsymbol{\Delta}^P$:

    $TE_{ij} = \beta_{jZ} I\{Z_i \leq j\}$, where β_{jZ} targets the treatment effect in period j
  - saturated treatment effect $\boldsymbol{\Delta}^S$:

    $TE_{ij} = \sum_{d=1}^{j} \beta_{jZd} I\{Z_i = j - d + 1\}$ with coefficient β_{jZd} targeting $\Delta_j(d)$

Random effects specification

- Specification of the random effect term RE_{ij} (Li et al, 2021, SMMR)
  - $RE_{ij} = 0$: an independence working correlation structure
  - $RE_{ij} = \alpha_i$ with $\alpha_i \sim N(0, \tau^2)$: an exchangeable working correlation structure
  - $RE_{ij} = \alpha_i + \gamma_{ij}$ with $\alpha_i \sim N(0, \tau^2)$ and $\gamma_{ij} \sim N(0, \kappa^2)$: a nested exchangeable working correlation structure
  - only a working assumption!
- These specifications allow us to analytically investigate model-robustness of the average treatment effect estimators

Model-robustness

- **Main Result**. *Under Assumptions 1–4,*
  (a) assuming a constant treatment effect,
  $\widehat{V}^{-1/2}(\widehat{\beta}_Z - \Delta) \xrightarrow{d} N(0,1)$;
  (b) assuming duration-specific treatment effects,
  $\left(\widehat{\mathbf{V}}^D\right)^{-1/2}\left(\widehat{\boldsymbol{\beta}}_Z^D - \boldsymbol{\Delta}^D\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_J)$;
  (c) assuming period-specific treatment effects,
  $\left(\widehat{\mathbf{V}}^P\right)^{-1/2}\left(\widehat{\boldsymbol{\beta}}_Z^P - \boldsymbol{\Delta}^P\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_J)$;
  (d) assuming saturated treatment effects,
  $\left(\widehat{\mathbf{V}}^S\right)^{-1/2}\left(\widehat{\boldsymbol{\beta}}_Z^S - \boldsymbol{\Delta}^S\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{J(J+1)/2})$,

  where $\mathbf{I}_q$ *is the* $q \times q$ *identity matrix for a positive integer* q.

- **Punchline**: mixed ANCOVA offers consistent causal estimates (a) as long as the treatment effect structure (w.r.t. 2 time scales) is correct; (b) even if all other aspects of the working model are incorrect

Generalization to other effect measures

- When the outcome is binary, treatment effects on a ratio, odds ratio may be of interest
- Define the saturated model-free estimands

$$\Phi_j(d) = f\left[E\{Y_{ijk}(d)\}, E\{Y_{ijk}(0)\}\right]$$

for a user-defined function f and $1 \leq d \leq j \leq J$.

  - $f(x, y) = x/y$ for treatment effects on the ratio scale
  - $f(x, y) = \log\{x/(1-x)\} - \log\{y/(1-y)\}$ for treatment effects on the log-odds-ratio scale
  - vector of estimands: $\mathbf{\Phi}^S = (\Phi_1(1), \ldots, \Phi_J(J))$

- **Solution**: linear mixed ANCOVA working model + g-computation formula $\Rightarrow$ Main results on model-robustness still hold

Discussion

- Mixed-effects models dominate applications to cluster trials
- This is another attempt to demystify the question: *when can linear mixed model provide robust causal inference in cluster trials*?
  - this work is sequel to [*Wang, B., Harhay, M. O., Small, D. S., Morris, T. P., Li, F. (2021). On the mixed-model analysis of covariance in cluster-randomized trials. arXiv e-prints, arXiv-2112.*]
- Work in progress not mentioned here
  - initial simulation results
  - parallel set of results for GEE (working model can be based on canonical link but with other constraints on working correlation)
- Do not address complications due to informative cluster size (ICS) & informative enrollment (Kahan et al, 2022; Wang et al. 2022; Wang et al. 2022+)
  - extensions to meet such new challenges is an immediate next step

References - cont'd

Nevins, P., Davis-Plourde, K., Macedo, J. P., Ouyang, Y., Ryan, M., Tong, G., ... Taljaard, M. (2023). A Scoping Review described diversity in methods of randomization and reporting of baseline balance in Stepped-Wedge Cluster Randomized Trials. Journal of Clinical Epidemiology.

Hussey, M. A., Hughes, J. P. (2007). Design and analysis of stepped wedge cluster randomized trials. Contemporary clinical trials, 28(2), 182-191.

Kenny, A., Voldal, E. C., Xia, F., Heagerty, P. J., Hughes, J. P. (2022). Analysis of stepped wedge cluster randomized trials in the presence of a timeâvarying treatment effect. Statistics in Medicine, 41(22), 4311-4339.

Maleyeff, L., Li, F., Haneuse, S., Wang, R. (2022). Assessing exposure-time treatment effect heterogeneity in stepped wedge cluster randomized trials. Biometrics

Wang, B., Harhay, M. O., Small, D. S., Morris, T. P., Li, F. (2021+). On the mixed-model analysis of covariance in cluster-randomized trials. arXiv e-prints, arXiv-2112.

Li, F., Hughes, J. P., Hemming, K., Taljaard, M., Melnick, E. R., Heagerty, P. J. (2021). Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: an overview. *Statistical Methods in Medical Research, 30*(2), 612-639.

Kahan, B. C., Li, F., Copas, A. J., Harhay, M. O. (2023). Estimands in cluster-randomized trials: choosing analyses that answer the right question. International Journal of Epidemiology, 52(1), 107-118.

Wang, X., et al. (2022). Two weights make a wrong: cluster randomized trials with variable cluster sizes and heterogeneous treatment effects. Contemporary Clinical Trials, 114, 106702.

Wang, B., Park, C., Small, D. S., Li, F. (2022+). Model-robust and efficient inference for cluster-randomized experiments. arXiv preprint arXiv:2210.07324.

Causal inference for cluster randomized trials—Where we are at?

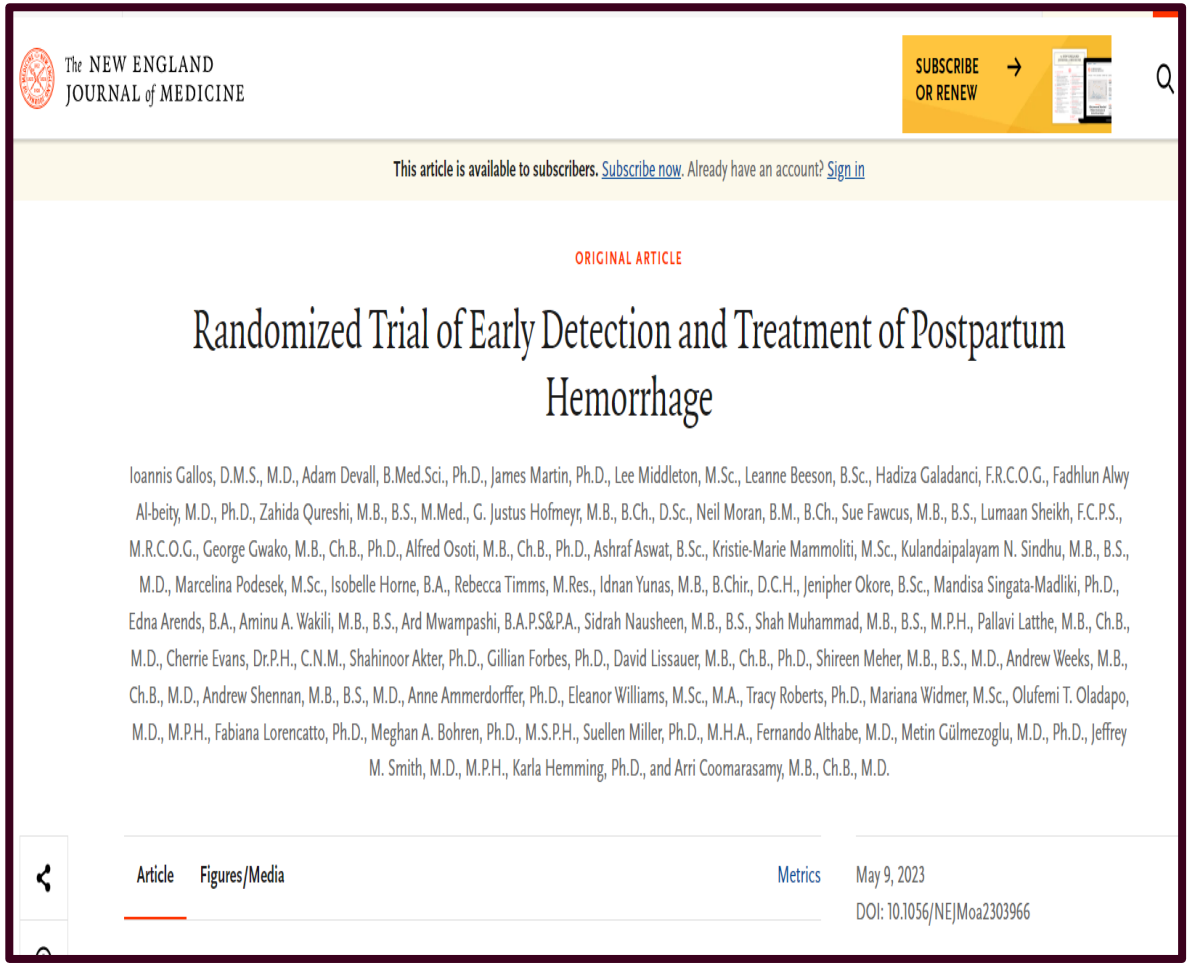
SCT May 2023

Karla Hemming, PhD
University of Birmingham

Objectives

- **How much does this matter in practice?**
  - Large case study (remember TRIGGER has only 6 clusters)
  - Bias vs decision making?
  - Evidence of informative cluster sizes more generally?
- **How do we translate this information?**
  - How well can statisticians and clinical academics understand this?
  - Evidence from a qualitative study on the interactions between statisticians and investigators when planning CRTs

EMOTIVE CRT: Bundle of evidence-based interventions



The NEW ENGLAND JOURNAL of MEDICINE

SUBSCRIBE OR RENEW

This article is available to subscribers. Subscribe now. Already have an account? Sign in

ORIGINAL ARTICLE

Randomized Trial of Early Detection and Treatment of Postpartum Hemorrhage

Ioannis Gallos, D.M.S., M.D., Adam Devall, B.Med.Sci., Ph.D., James Martin, Ph.D., Lee Middleton, M.Sc., Leanne Beeson, B.Sc., Hadiza Galadanci, F.R.C.O.G., Fadhlun Alwy Al-beity, M.D., Ph.D., Zahida Qureshi, M.B., B.S., M.Med., G. Justus Hofmeyr, M.B., B.Ch., D.Sc., Neil Moran, B.M., B.Ch., Sue Fawcus, M.B., B.S., Lumaan Sheikh, F.C.P.S., M.R.C.O.G., George Gwako, M.B., Ch.B., Ph.D., Alfred Osoti, M.B., Ch.B., Ph.D., Ashraf Aswat, B.Sc., Kristie-Marie Mammoliti, M.Sc., Kulandaipalayam N. Sindhu, M.B., B.S., M.D., Marcelina Podesek, M.Sc., Isobelle Horne, B.A., Rebecca Timms, M.Res., Idnan Yunas, M.B., B.Chir., D.C.H., Jenipher Okore, B.Sc., Mandisa Singata-Madliki, Ph.D., Edna Arends, B.A., Aminu A. Wakili, M.B., B.S., Ard Mwampashi, B.A.P.S&P.A., Sidrah Nausheen, M.B., B.S., Shah Muhammad, M.B., B.S., M.P.H., Pallavi Latthe, M.B., Ch.B., M.D., Cherrie Evans, Dr.P.H., C.N.M., Shahinoor Akter, Ph.D., Gillian Forbes, Ph.D., David Lissauer, M.B., Ch.B., Ph.D., Shireen Meher, M.B., B.S., M.D., Andrew Weeks, M.B., Ch.B., M.D., Andrew Shennan, M.B., B.S., M.D., Anne Ammerdorffer, Ph.D., Eleanor Williams, M.Sc., M.A., Tracy Roberts, Ph.D., Mariana Widmer, M.Sc., Olufemi T. Oladapo, M.D., M.P.H., Fabiana Lorencatto, Ph.D., Meghan A. Bohren, Ph.D., M.S.P.H., Suellen Miller, Ph.D., M.H.A., Fernando Althabe, M.D., Metin Gülmezoglu, M.D., Ph.D., Jeffrey M. Smith, M.D., M.P.H., Karla Hemming, Ph.D., and Arri Coomarasamy, M.B., Ch.B., M.D.

Article Figures/Media Metrics

May 9, 2023
DOI: 10.1056/NEJMoa2303966

Cluster (unit of randomization)	Health facility (n=80)
Intervention	Bundle (treatments, blood collection drape)
Primary outcome	PPH (binary) (n=~200,000)
Implementation outcome	Bundle compliance

Gallos I, Devall A, Martin J et al. Randomized Trial of Early Detection and Treatment of Postpartum Hemorrhage. N Engl J Med. 2023 May 9. doi: 10.1056/NEJMoa2303966. Epub ahead of print. PMID: 37158447.

Key findings

Outcome	Intervention	Control	RR (95% CI)
PPH	**(C=39, N=49,101)**	**(C=39, N=50,558)**	
	794 (1.6%)	2,139 (4.3%)	0.40 (0.32 to 0.50)
Bundle Compliance	**(C=39, N=4,158)**	**(C=39, N=8,351)**	
	3,791 (91.2%)	1,623 (19.4%)	4.94 (3.88 to 6.28)

*Estimated using marginal standardisation, following mixed effects logistic regression (additionally included baseline period)
ICC for PPH: 0.01; ICC for Compliance: 0.15
Very limited missing outcome data; 80 clusters and so small sample corrections probably not necessary;
adjusted for variables used in restricted randomisation

Sensitivity analyses using Independent Estimating Equations (IEE)

Outcome	Intervention	Control	RR (95% CI)
PPH	**(C=39, N=49,101)**	**(C=39, N=50,558)**	
	794 (1.6%)	2,139 (4.3%)	0.40 (0.32 to 0.50)
			IEE: 0.39 (0.30 to 0.50)
Bundle Compliance	**(C=39, N=4,158)**	**(C=39, N=8,351)**	
	3,791 (91.2%)	1,623 (19.4%)	4.94 (3.88 to 6.28)
			IEE: 4.69 (3.11 to 7.07)

*IEE: Estimated using GEE log link robust standard errors and working independent correlation structure

Sensitivity analyses using Independent Estimating Equations (IEE)

Outcome	Intervention	Control	RR (95% CI)
PPH	**(C=39, N=49,101)**	**(C=39, N=50,558)**	
	794 (1.6%)	2,139 (4.3%)	0.40 (0.32 to 0.50)
			IEE:0.39 (0.30 to 0.50)
Bundle Compliance	**(C=39, N=4,158)**	**(C=39, N=8,351)**	
	3,791 (91.2%)	1,623 (19.4%)	4.94 (3.88 to 6.28)
			IEE: 4.69 (3.11 to 7.07)

*IEE: Estimated using GEE log link robust standard errors and working independent correlation structure

Evidence of informative cluster size?

Bundle Compliance	Intervention (C=39, N=4,158)	Control (C=39, N=8,351)	RR (95% CI)
All clusters	3,791 (91.2)	1,623 (19.4)	4.94 (3.88 to 6.28)
Sub-group			
Small clusters	1,578 (89.8%)	619 (22.1%)	4.78 (3.61 to 6.33)
Large clusters	2,213 (92.2%)	1,004 (18.1%)	5.16 (3.73 to 7.14)
Ratio of Ratios			1.08 (0.75 to 1.54)

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.

Evidence of informative cluster size?

Bundle Compliance	Intervention (C=39, N=4,158)	Control (C=39, N=8,351)	RR (95% CI)
All clusters	3,791 (91.2)	1,623 (19.4)	4.94 (3.88 to 6.28)
Sub-group			
Small			61 to 6.33)
Large			73 to 7.14)
Ratio of Ratios			1.08 (0.75 to 1.54)

No evidence of "informative cluster sizes"
Whilst there are some apparent differences in the point estimates these are of insignificant magnitude to be important in interpretation

Correlation between cluster size and compliance was -0.47 (treatment arm only)

Evidence of informative cluster sizes more generally?

- Recall: Informative cluster sizes exists when impact varies by cluster size (~treatment by sub-group interaction) or prevalence varies by size
- **Patient randomized trials:**
  1. Sub-group claims common
  2. Mostly do not stand up to scrutiny
- **Cluster randomized trials**:
  1. Behavior change type interventions
  2. Perhaps more grounds to hypothesize interactions
  3. More heterogenous populations – prevalence might vary

Impact of most interventions relatively stable across sub-groups?

Unclear?

Sun X, Briel M, Busse JW, You JJ, Akl EA, Mejza F, Bala MM, Bassler D, Mertz D, Diaz-Granados N, Vandvik PO, Malaga G, Srinathan SK, Dahm P, Johnston BC, Alonso-Coello P, Hassouneh B, Walter SD, Heels-Ansdell D, Bhatnagar N, Altman DG, Guyatt GH. Credibility of claims of subgroup effects in randomised controlled trials: systematic review. BMJ. 2012 Mar

Sensitivity analyses: cluster-average effect

Outcome	Intervention	Control	RR (95% CI)
PPH	**(C=39, N=49,101)**	**(C=39, N=50,558)**	
	794 (1.6%)	2,139 (4.3%)	0.40 (0.32 to 0.50)
			CA: 0.35 (0.25 to 0.49)
Bundle Compliance	**(C=39, N=4,158)**	**(C=39, N=8,351)**	
	3,791 (91.2%)	1,623 (19.4%)	4.94 (3.88 to 6.28)
			CA: 5.81 (3.89 to 8.74)

*CA: Unweighted cluster-average using cluter-level proportions, linear regression of logit p

Sensitivity analyses: cluster-average effect

Outcome	Intervention	Control	RR (95% CI)
PPH	**(C=39, N=49,101)**	**(C=39, N=50,558)**	
	794 (1.6%)	2,139 (4.3%)	0.40 (0.32 to 0.50)
			CA: 0.35 (0.25 to 0.49):
Bundle Compliance	**(C=39, N=4,158)**	**(C=39, N=8,351)**	
	3,791 (91.2%)	1,623 (19.4%)	4.94 (3.88 to 6.28)
			CA: 5.81 (3.89 to 8.74)

*CA: Unweighted cluster-average using cluter-level proportions, linear regression of logit p

What was target of inference in EMOTIVE?

Impact on typical cluster or typical individual?

- **Clinical outcome PPH:**
  - Broad agreement that interest is in effect on average individual
  - Thus, weight by cluster size
- **Implementation outcome (Bundle compliance):**
  - Clinicians adamant interest is in effect on average individual
  - In **my opinion** they were interested in effect on average cluster

  (I found it hard to communicate the difference)

What was target of inference in EMOTIVE?

Marginal or cluster-specific effect?

- **Sample representativeness:**
  - All **individuals** in a given cluster included – thus representative of individuals
  - **Clusters** unlikely to be totally representative of wider community
- **Question of interest:**
  - Question of interest around whether the intervention works at the population level (i.e., is it **cost effective**)
- **Treatment by sub-group effects:**
  - Possible that **cluster-level covariates** that will impact on how well intervention is delivered (i.e., treatment by covariate interactions)

What was target of inference in EMOTIVE?

Marginal or cluster-specific effect?

- **Sample representativeness:**
  - All **individuals** in a given cluster included – thus representative of individuals
  - **Clusters** unlikely to be totally representative of wider community
- **Question of interest:**
  - Question of interest around whether the intervention works at the population level (i.e., is it **cost effective**)
- **Treatment by sub-group effects:**
  - Possible that **cluster-level covariates** that will impact on how well intervention is delivered (i.e., treatment by covariate interactions)

Is the interest in the marginal or cluster-specific effect?

How do we communicate this?

- **Applied statisticians:**
  - Understanding of what matters and when
  - Communication key
  - Acknowledge uncertainty
- **Academic partners:**
  - Guidance that hinges on tangible concepts, e.g.:
    - How representative the sample is?
    - Is the intent to inform policy decisions?

How do statisticians and investigators work together when planning a CRT?

- **Qualitative interview study with 25 statisticians / investigators / trialists who work in CRTs:**
  - Understand how they interact
  - How they communicate
  - How decisions are made

Easter C, Kristunas C, Hemming K, Greenfield S. Risk of Bias in Cluster Randomised Controlled Trials of Individual-Level Interventions: Protocol for a Semi-Structured Interview Study. International Journal of Qualitative Methods. 2022 Jun 30;21:16094069221113112.

What we found (statisticians):

Confident statistician

“Junior” statistician

The “I’m not a people person”

Easter C, Kristunas C, Hemming K, Greenfield S. Risk of Bias in Cluster Randomised Controlled Trials of Individual-Level Interventions: Protocol for a Semi-Structured Interview Study. International Journal of Qualitative Methods. 2022 Jun 30;21:16094069221113112.

What we found (clinicians):

Enabler (“I’ll get advice”)

The misinformed

Dominant

Easter C, Kristunas C, Hemming K, Greenfield S. Risk of Bias in Cluster Randomised Controlled Trials of Individual-Level Interventions: Protocol for a Semi-Structured Interview Study. International Journal of Qualitative Methods. 2022 Jun 30;21:16094069221113112.

Summary

- To date, we have largely been driven by choosing an analysis model that has best statistical properties
  - This needs to change
- Estimands certainly helping with clarification:
  - But, it is also creating questions
  - Need for wide level understanding of what matters when
- Academic partners:
  - Guidance that hinges on tangible concepts?

Thank you …

Simple toy example (taken from Brennan Kahan)

Marginal OR

	Intervention	Control
Cluster 1	50/100	25/100
Cluster 2	75/100	60/100
Overall*	0.625	0.425

$$\frac{0.625/(1-0.625)}{0.425/(1-0.425)} = \mathbf{2.25}$$

Average Cluster ORs

	Intervention	Control	OR
Cluster 1	50/100	25/100	3.0
Cluster 2	75/100	60/100	2.0

$$\frac{(100)(3.0)+(100)(2.0)}{200} = \mathbf{2.50}$$

*risk

Simple toy example (taken from Brennan Kahan)

Marginal OR

	Intervention	Control
Cluster 1	50/100	25/100
Cluster 2	75/100	60/100
Overall		

$$\frac{0.625/(1-0.625)}{0.425/(1-0.425)} = \mathbf{2.25}$$

Notice in CRTs treatment is crossed not nested with arm
Thus this example is slightly artificial
But, same issues arise in patient RCTs with sub-groups

	Intervention	Control	OR
Cluster 1	50/100	25/100	3.0
Cluster 2	75/100	60/100	2.0

$$\frac{(100)(3.0)+(100)(2.0)}{200} = \mathbf{2.50}$$

Simple toy example with unequal cluster sizes

	Intervention	Control
Cluster 1	5/10	2/10
Cluster 2	75/100	60/100
Overall	0.727	0.564

Marginal OR

$$\frac{0.727/(1-0.727)}{0.564/(1-0.564)} = \mathbf{2.06}$$

Notice: marginal OR changes when change sample

	Intervention	Control	OR
Cluster 1	5/10	2/10	4.0
Cluster 2	75/100	60/100	2.0

Weighted average Cluster ORs

$$\frac{(10)(4.0)+(100)(2.0)}{110} = \mathbf{2.18}$$

	Intervention	Control	OR
Cluster 1	5/10	2/10	4.0
Cluster 2	75/100	60/100	2.0

Unweighted average Cluster ORs

$$\frac{(4.0)+(2.0)}{2} = \mathbf{3.00}$$

Are there others?

Get 2.16 from logistic regression adjusted for cluster!!

	Intervention Odds	Control Odds
Cluster 1	1=5/(10-5)	0.25=2/(10-2)
Cluster 2	3=75/(100-75)	1.5=60/(100-60)
Average (odds)	2=(1+3)/2	0.88=(0.25+1.5)/2

Ratio of arm specific odds?

$$\frac{2}{0.88} = \mathbf{2.29}$$

	Intervention Odds	Control Odds
Cluster 1	1=5/(10-5)	0.25=2/(10-2)
Cluster 2	3=75/(100-75)	1.5=60/(100-60)
Average (odds)	2.81=((10*1)+(100*3))/110	1.39=((10*0.25)+(100*1.5))/110

Weighted Ratio of arm specific odds?

$$\frac{2.81}{1.39} = \mathbf{2.03}$$

How many different ways of constructing OR?

	Intervention	Control
Cluster 1	5/10	2/10
Cluster 2	75/100	60/100
Overall	0.727	0.564

Marginal OR

$$\frac{0.727/(1-0.727)}{0.564/(1-0.564)} = \mathbf{2.06}$$

	Intervention	Control	OR
Cluster 1	5/10	2/10	4.0
Cluster 2	75/100	60/100	2.0

Weighted average Cluster ORs

$$\frac{(10)(4.0)+(100)(2.0)}{110} = \mathbf{2.18}$$

	Intervention	Control	OR
Cluster 1	5/10	2/10	4.0
Cluster 2	75/100	60/100	2.0

Unweighted average Cluster ORs

$$\frac{(4.0)+(2.0)}{2} = \mathbf{3.00}$$

	Intervention	Control
Cluster 1	1=5/(10-5)	0.25=2/(10-2)
Cluster 2	3=75/(100-75)	1.5=60/(100-60)
Average (odds)	2=(1+3)/2	0.88=(0.25+1.5)/2

Ratio of arm specific odds?

$$\frac{2}{0.88} = \mathbf{2.29}$$

(Weighted): **2.03**

Simple toy example RR (taken from Brennan Kahan)

Marginal RR

	Intervention	Control
Cluster 1	50/100	25/100
Cluster 2	75/100	60/100
Overall*	0.625	0.425

$$\frac{0.625}{0.425} = \mathbf{1.47}$$

Average Cluster RRs

	Intervention	Control	RR
Cluster 1	50/100	25/100	2.0
Cluster 2	75/100	60/100	1.25

$$\frac{(100)(2.0) + (100)(1.25)}{200} = \mathbf{1.63}$$

*risk

Simple toy example with unequal cluster sizes RR

	Intervention	Control
Cluster 1	5/10	2/10
Cluster 2	75/100	60/100
Overall	0.727	0.564

Marginal RR

$$\frac{0.727}{0.564} = \mathbf{1.29}$$

	Intervention	Control	RR
Cluster 1	5/10	2/10	2.5
Cluster 2	75/100	60/100	1.25

Weighted average Cluster RRs

$$\frac{(10)(2.5) + (100)(1.25)}{110} = \mathbf{1.36}$$

	Intervention	Control	RR
Cluster 1	5/10	2/10	2.5
Cluster 2	75/100	60/100	1.25

Unweighted average Cluster RRs

$$\frac{(2.5) + (1.25)}{2} = \mathbf{1.88}$$

Are there others (RR)?

	Intervention	Control
Cluster 1	0.5=5/10	0.25=2/10
Cluster 2	0.75=75/100	0.6=60/100
Average (risk)	0.625=(0.5+0.75)/2	0.40=(0.25+0.6)/2

Ratio of arm specific risks?

$$\frac{0.625}{0.40} = \mathbf{1.56}$$

	Intervention	Control
Cluster 1	0.5=5/10	0.25=2/10
Cluster 2	0.75=75/100	0.6=60/100
Average (risk)	0.727=((10*0.5)+(100*0.75))/110	0.56=((10*0.25)+(100*0.6))/110

Weighted Ratio of arm specific risks?

$$\frac{0.727}{0.56} = \mathbf{1.29}$$

Get 1.29 from binomial regression adjusted for cluster

How many different ways of constructing RR?

	Intervention	Control
Cluster 1	5/10	2/10
Cluster 2	75/100	60/100
Overall	0.727	0.564

Marginal RR

$$\frac{0.727}{0.564} = \mathbf{1.29}$$

	Intervention	Control	RR
Cluster 1	5/10	2/10	2.5
Cluster 2	75/100	60/100	1.25

Weighted average Cluster RRs

$$\frac{(10)(2.5) + (100)(1.25)}{110} = \mathbf{1.36}$$

	Intervention	Control	RR
Cluster 1	5/10	2/10	2.5
Cluster 2	75/100	60/100	1.25

Unweighted average Cluster RRs

$$\frac{(2.5) + (1.25)}{2} = \mathbf{1.88}$$

	Intervention	Control
Cluster 1	0.5=5/10	0.25=2/10
Cluster 2	0.75=75/100	0.6=60/100
Average (risk)	0.625=(0.5+0.75)/2	0.40=(0.25+0.6)/2

Ratio of arm specific risks?

$$\frac{0.635}{0.40} = \mathbf{1.56}$$

(Weighted): 1.29