ePCT Experimental Design and Analysis

David M. Murray, PhD
Associate Director for Prevention
Director, Office of Disease Prevention
National Institutes of Health (NIH)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Learning goals

- Recognize the analytical challenges and trade-offs of pragmatic study designs, focusing on what PIs need to know—highlighting design and analysis considerations and key decision points.

Analysis Considerations

Embedded Pragmatic Clinical Trials

NIH PRAGMATIC TRIALS COLLABORATORY

Rethinking Clinical Trials®

Learning goals

- Recognize the analytical challenges and trade-offs of pragmatic study designs, focusing on what PIs need to know -- highlighting design and analysis considerations and key decision points.

Important things to know

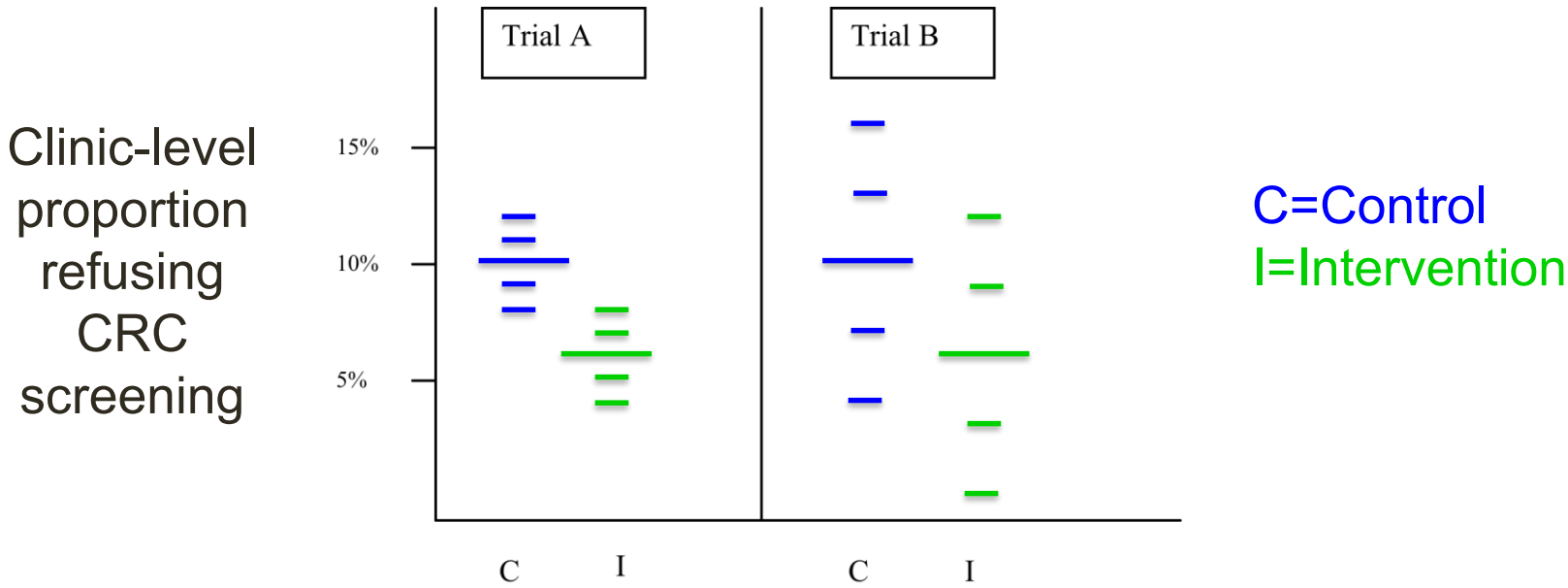
- Studies that randomize groups or deliver interventions to groups face special analytic challenges not found in traditional individually randomized trials
- Failure to address these challenges will result in an underpowered study and/or an inflated type 1 error rate
- We won't advance the science by using inappropriate methods

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Two example CRTs inspired by STOP CRC

- 10 clinics/CRT
  - 5 intervention (I) clinics & 5 control (C) clinics
  - 100 patients/clinic
- 1000 patients per trial
  - 500 intervention vs. 500 control
- Binary outcome: “No screening within year of enrollment”

Clustering in CRTs: Implications for analysis



- 5 clinics each randomized to control and intervention
- 100 eligible participants per clinic measured

Overall screening refusal proportion in both trials: 10% vs 6%

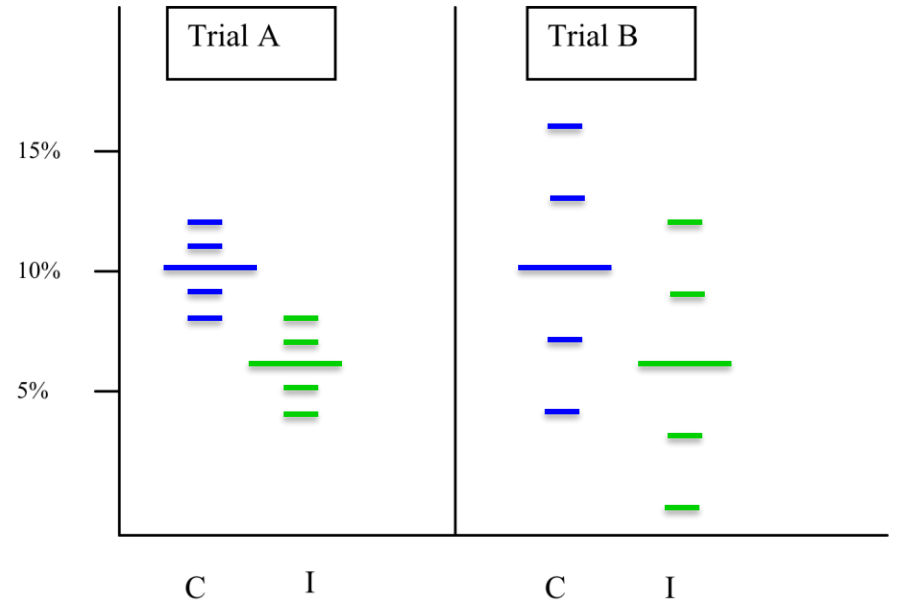
Question: is intervention effective?

Adapted from Hayes & Moulton (2009)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering in CRTs: Implications for analysis

Clinic-level proportion refusing CRC screening

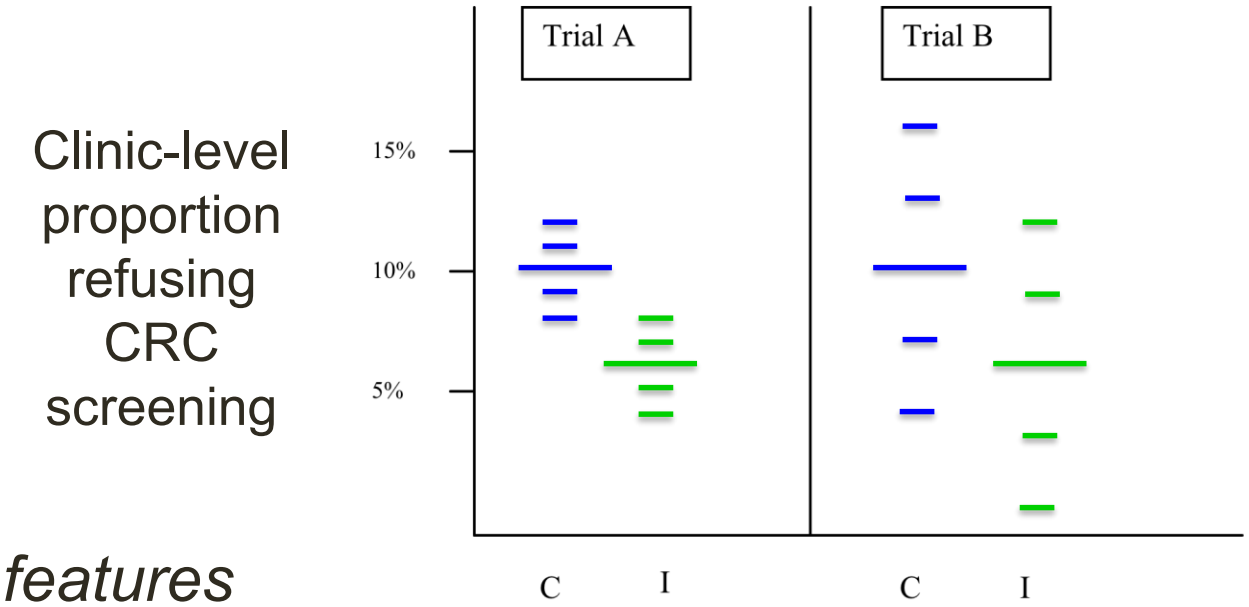


C=Control
I=Intervention

Which trial shows more evidence of benefit?

Adapted from Hayes & Moulton (2009)

Clustering in CRTs: Implications for analysis



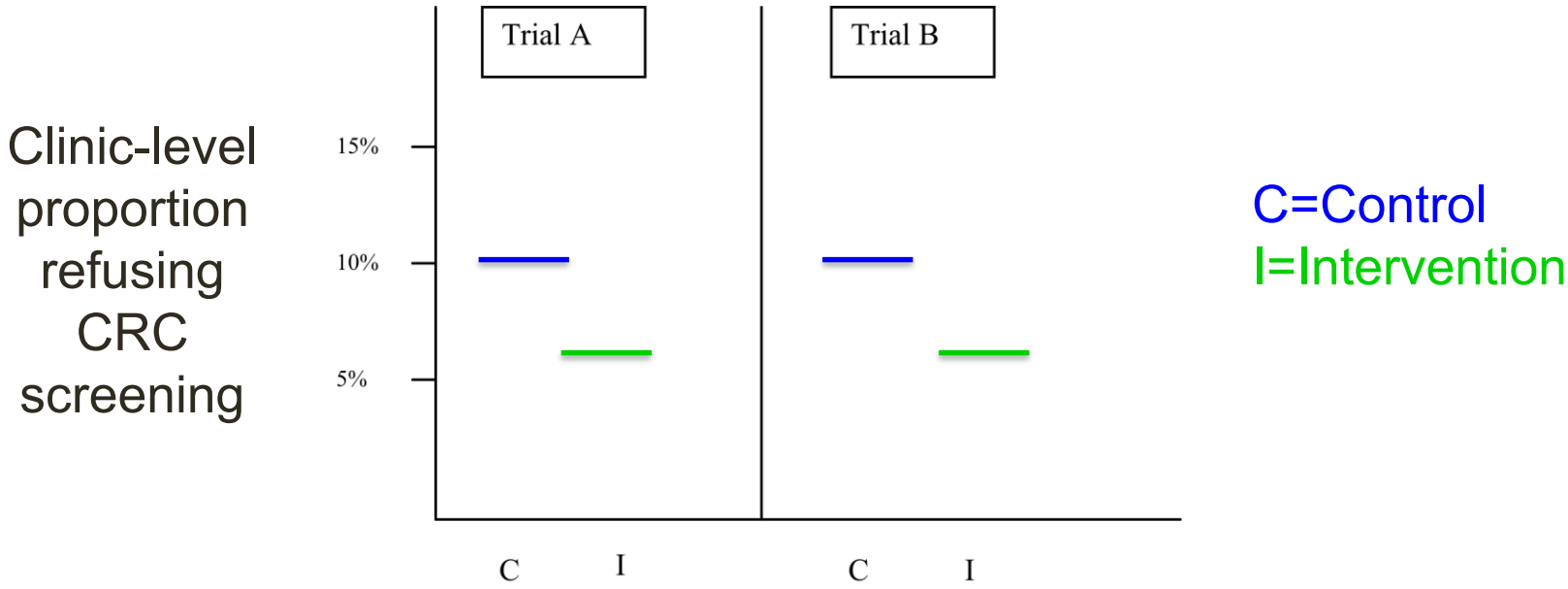
C=Control
I=Intervention

Study features

- Trial A:
  - Lower between-clinic variability (ie, less clustering)
  - Little overlap of I & C clinic-level proportions
- Trial B: overlap of intervention (I) & control (C) clinic-level proportions

Adapted from Hayes & Moulton (2009)

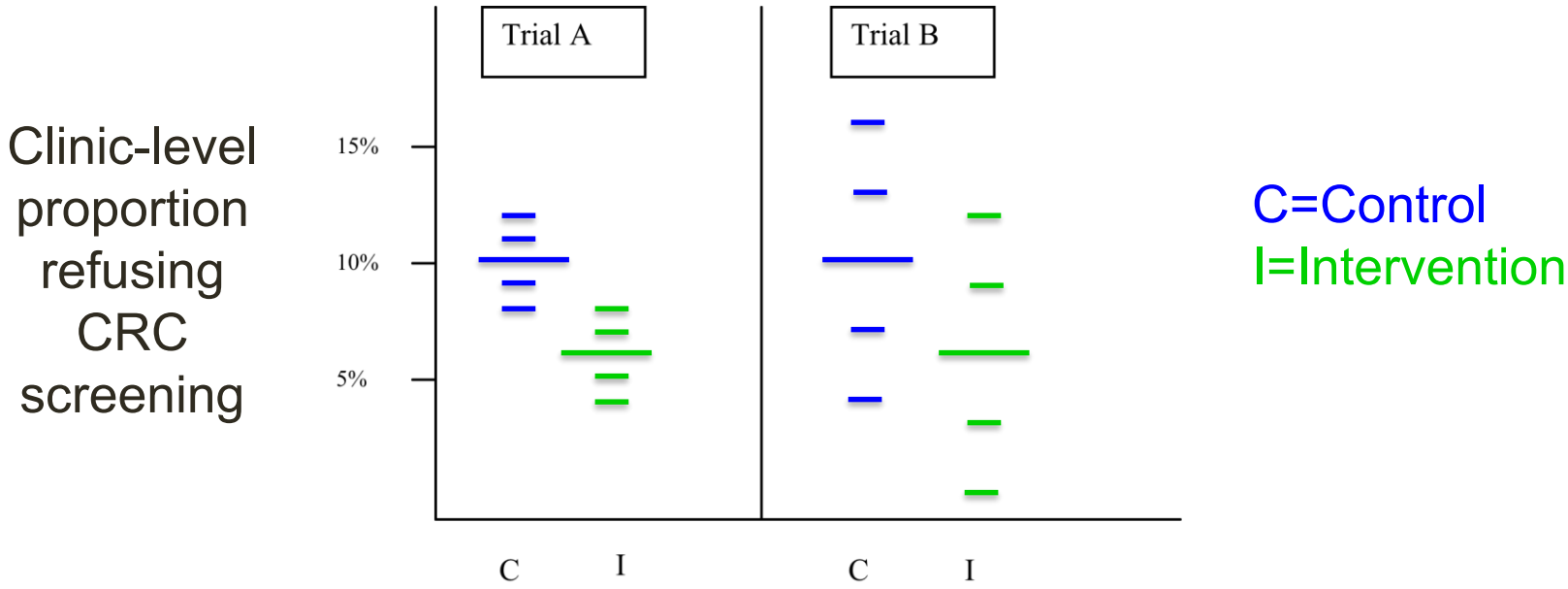
Clustering in CRTs: Implications for analysis



- If ignore clustering: p-value = **0.02** for both trials
- Comparison of 10% (50/500) vs 6% (30/500) by chi-sq. test

Adapted from Hayes & Moulton (2009)

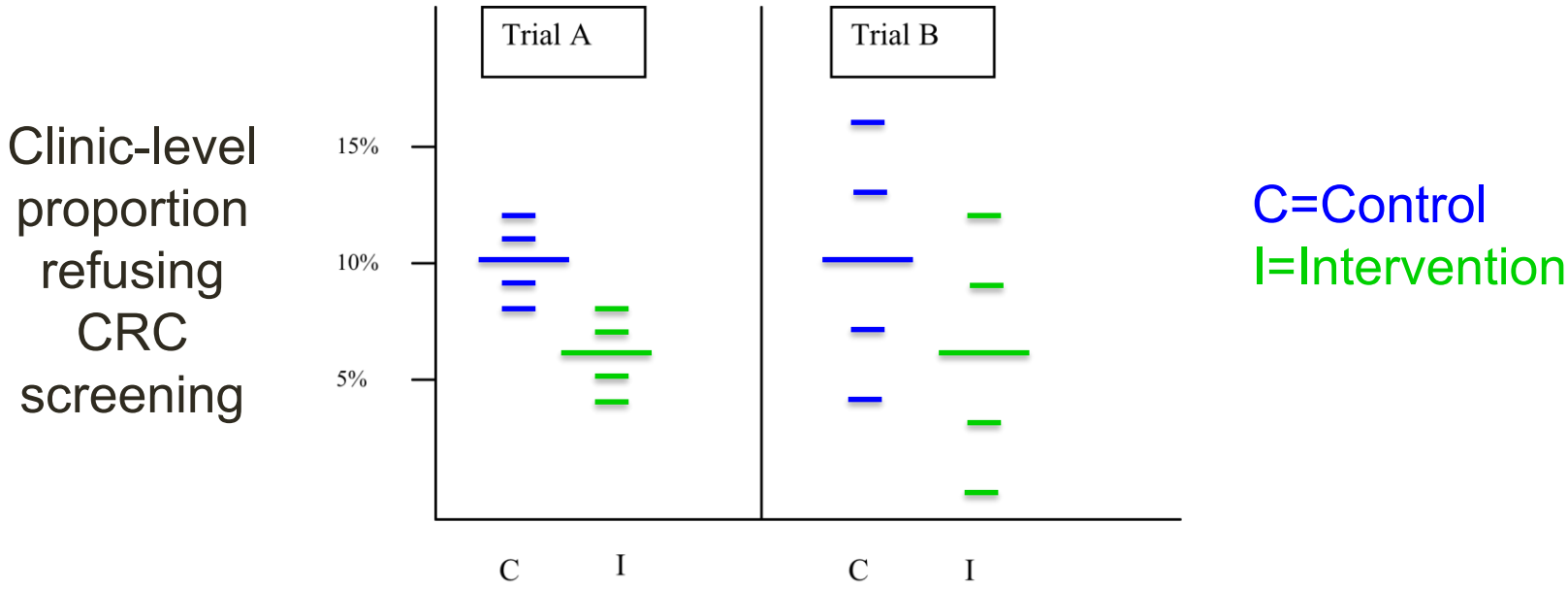
Clustering in CRTs: Implications for analysis



- Trial B p-value accounting for clustered design = **?**
- If ignore clustering: p-value = **0.02**

Adapted from Hayes & Moulton (2009)

Clustering in CRTs: Implications for analysis

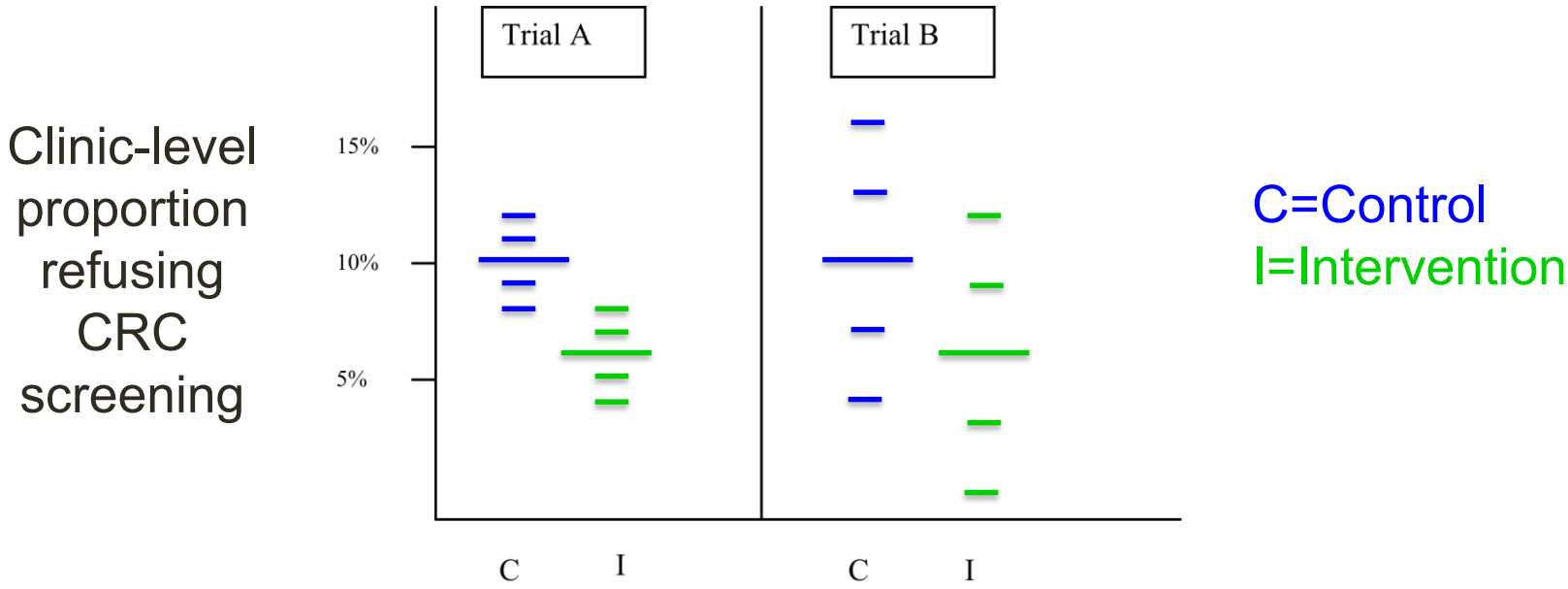


- Trial B p-value accounting for clustered design = **0.17**
- If ignore clustering: p-value = **0.02**

Adapted from Hayes & Moulton (2009)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

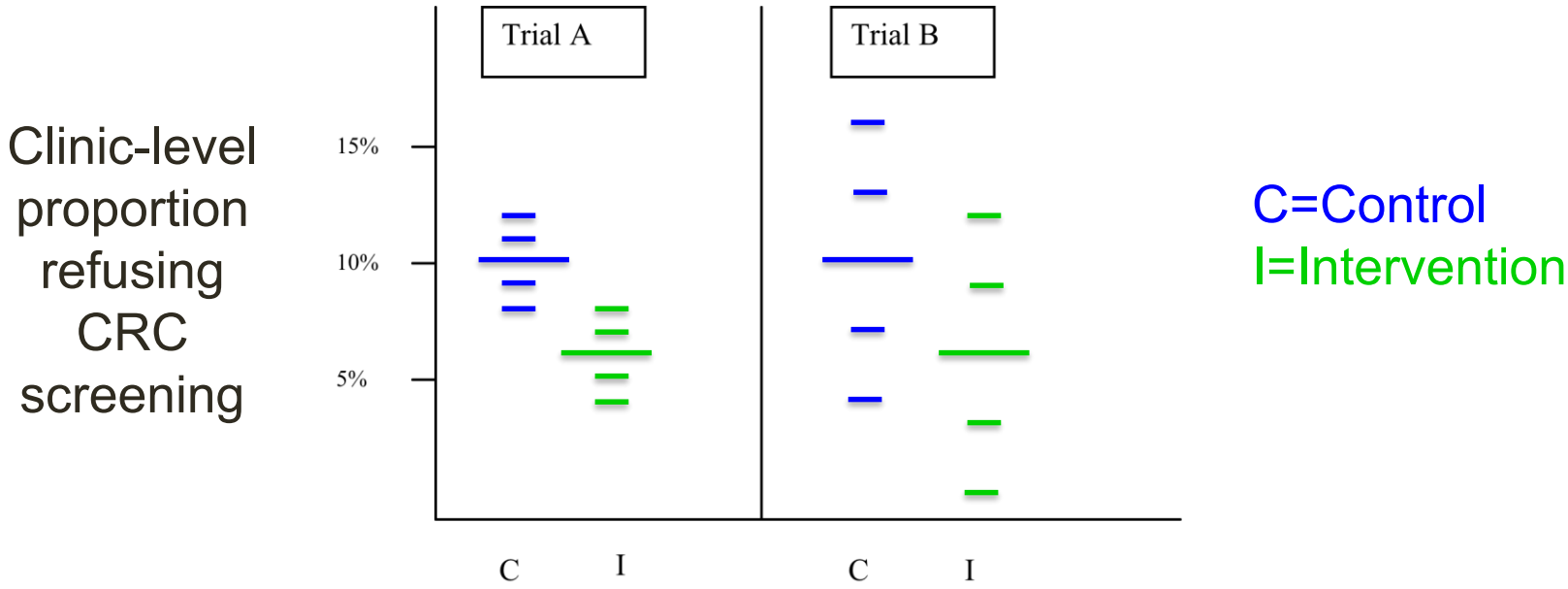
Clustering in CRTs: Implications for analysis



- Trial A p-value accounting for clustered design = **?**
- Trial B p-value accounting for clustered design = **0.17**
- If ignore clustering: p-value = **0.02**

Adapted from Hayes & Moulton (2009)

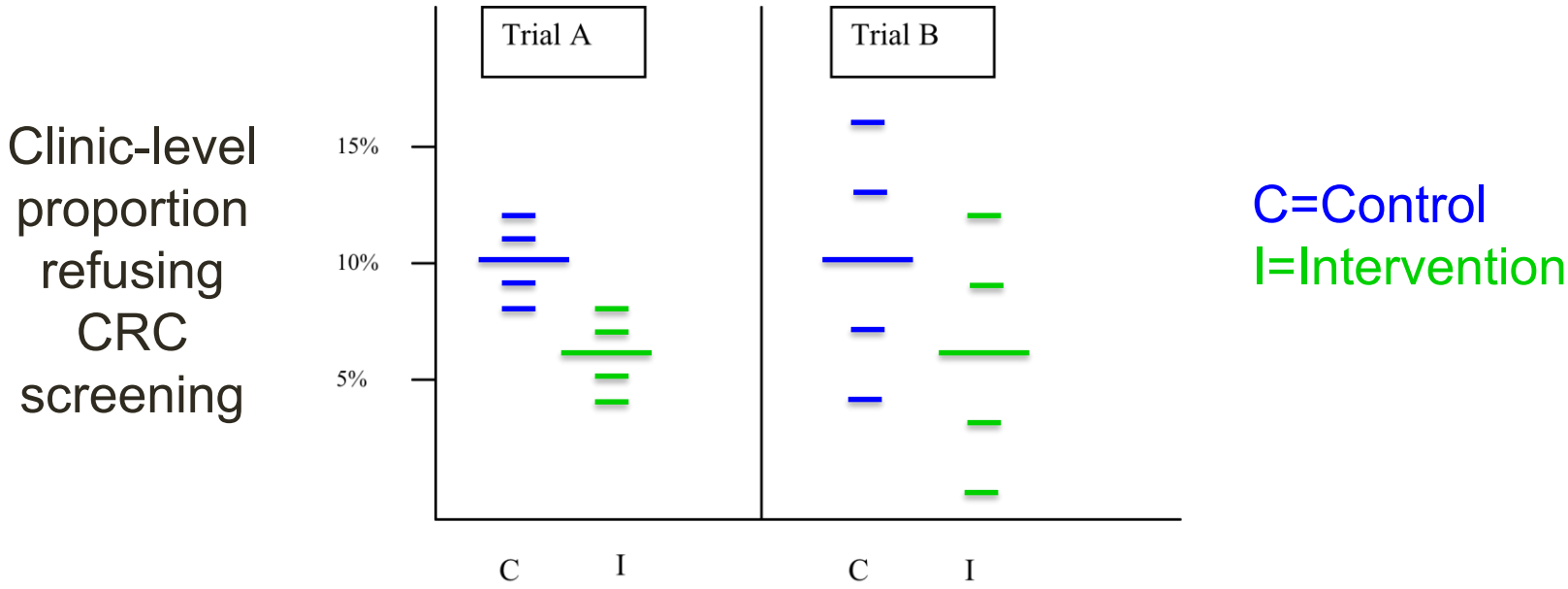
Clustering in CRTs: Implications for analysis



- Trial A p-value accounting for clustered design = **0.01**
- Trial B p-value accounting for clustered design = **0.17**
- If ignore clustering: p-value = **0.02**

Adapted from Hayes & Moulton (2009)

Clustering in CRTs: Implications for analysis



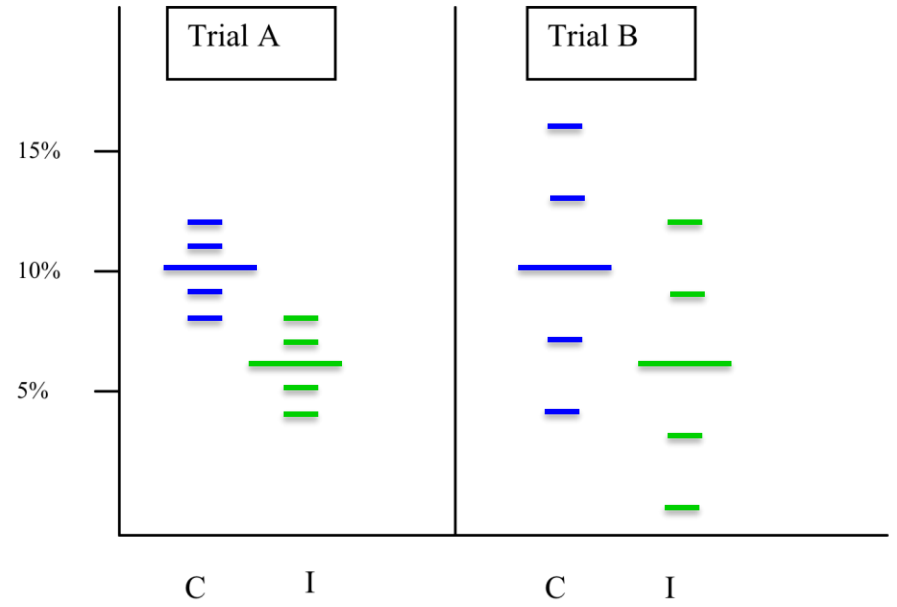
- Trial A p-value accounting for clustered design* = **0.01**
- Trial B p-value accounting for clustered design* = **0.17**

*By using a cluster-level analysis where the 10 cluster-level proportions (5 per arm) are treated as continuous variables and analyzed with Wilcoxon rank sum test

Adapted from Hayes & Moulton (2009)

Clustering in CRTs: Implications for analysis

Clinic-level proportion refusing CRC screening



C=Control
I=Intervention

- Trial A p-value accounting for clustered design* = **0.004**
- Trial B p-value accounting for clustered design* = **0.22**

*Alternative cluster-level analysis using t-test, which has stronger assumptions (ie, normality of cluster-specific prevalence) than the Wilcoxon rank sum test

Adapted from Hayes & Moulton (2009)

Summary: Analysis of two example CRTs

- Two example trials
  - Analyzed with cluster-level analysis
  - Overall sample size (# clinics/trial) =10
  - Both trials had same signal (10% vs 6%)
  - Totally different conclusions from each trial
  - Between-cluster variability (& clustering) in Trial A < Trial B
  - P-value Trial A < P-value Trial B
  - Important: if incorrectly ignore clustered design, could claim ‘significant’ when not (eg, Trial B)

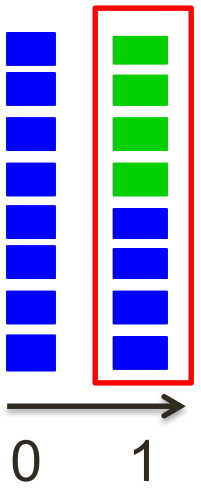
Analysis of CRTs, including SW-CRTs

- Regression analysis more common than cluster-level analysis
- Analyze individual-level data
  - eg, data from 1000 participants/trial not only one proportion/clinic
- Methods to account for clustering
  - Random effects / mixed effects models
  - Generalized estimating equations (GEE)
- If SW-CRT, **must** account for time
- Work with statistician to ensure properly account for clustering

Analysis of CRTs, including SW-CRTs

Parallel design

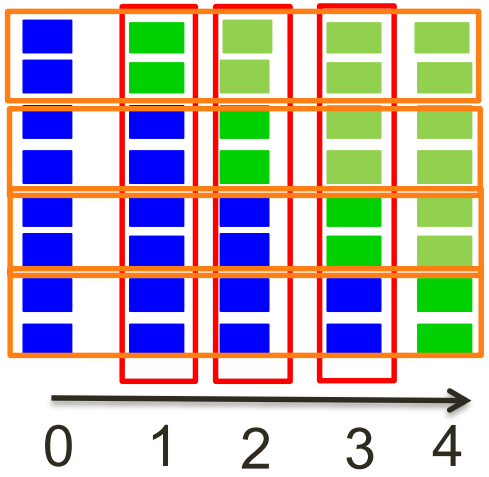
Estimated (primarily) using between-cluster ie, **vertical** information



Time since baseline

Complete SW design

Estimated using both **vertical** & **horizontal** (ie, within-cluster) information

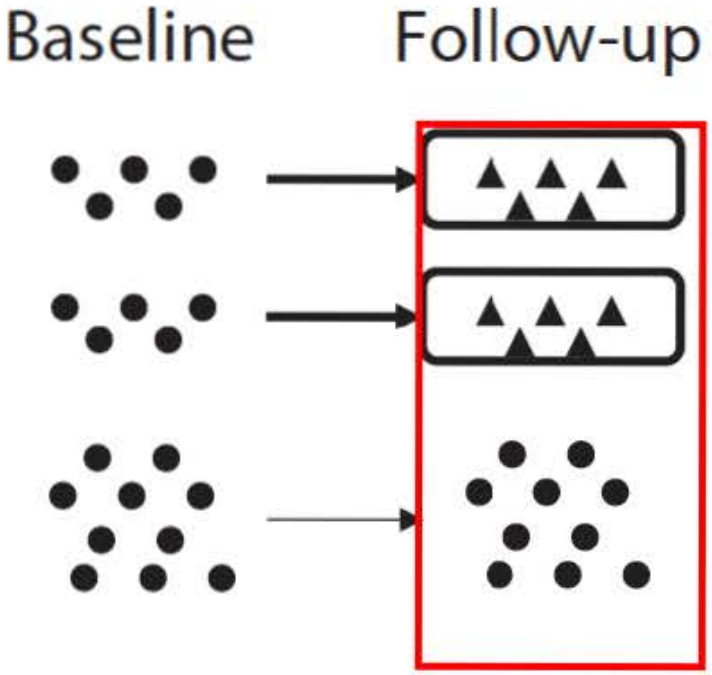


Control period Intervention period

Based on: Hemming K et al. 2015. *Stat Med*. 34:181-196.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Analysis of IRGT trials



▲ Individual measured under intervention
● Individual measured under no intervention

Parallel design

Estimated (primarily) using between-individual ie, **vertical** information

Extracted from Figure 1 in Turner et al. *Am J Public Health*. 2017;107(6).

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Analysis of IRGT trials

- Analyze individual-level data accounting for clustering
  - Random effects / mixed effects models
  - Generalized estimating equations (GEE)
- Considerations on clustering
  - Clustering in both arms: if both conditions group-based & may need different degree of clustering in two arms
  - Clustering in intervention arm only: if intervention group-based but control condition not
- Work with statistician to ensure properly account for clustering

Analysis of CRTs, SW-CRTs, and IRGTTs

- Clustering must be accounted for in analysis
- Challenges in “small” trials (# clusters < 50)
  - Limited degrees of freedom (df) for testing intervention as df driven by # clusters (i.e. groups)
  - Use t-test not Z-test & calculate correct df
  - Intervention effect SE may be under-estimated
    - Can correct e.g. finite-sample bias corrections for GEE
  - Ignore either penalty (df & SEs) leads to inflated Type I error
    - Type I error rate may be 30-50% in a CRT, even with small ICC
    - Type I error rate may be 15-25% in an IRGTT, even with small ICC
- Work with statistician to ensure properly account for clustering

Analysis of CRTs, SW-CRTs, and IRGTTs

- May need to account for complex clustering structures
  - Different clustering (ICC) in two conditions
  - Repeated measures on same individuals, if cohort
  - Decay/change in pairwise correlations over time (eg, SW-CRT)
- Other considerations
  - May need non-constant intervention effect if multiple follow-up time points (eg, like in SW-CRT)

Strategies to protect the analysis

Avoid model misspecification

- Plan analysis
  - To reflect the study design
  - Around the primary endpoints
- Anticipate
  - All sources of random variation
  - Patterns of over-time correlation
  - Pattern of the intervention effect over time
    - Important with repeated measures designs, e.g. SW-CRTs
  - Potential confounding & effect modification

Strategies to protect the analysis

Avoid low power

- Use strong interventions with good reach
- Maintain reliability of intervention implementation
- Use more & smaller groups not few large groups
- For SW-CRTs, use more steps
- Use regression adjustment
  - For covariates to reduce variance & intraclass correlation
  - In SW-CRTs, to adjust for calendar time

Challenges of pragmatic study design

- Trade-offs in flexibility, adherence, and generalizability are inevitable
- Implementation by healthcare system staff, not research staff
- New staff workflow and responsibility acknowledged
- Triage or case selection by healthcare system staff using existing structures with some modification

NIH Collaboratory: examples of analytic challenges and trade-offs

- Stepped wedge designs “roll out” over time and are more susceptible to disruption!
- Parallel cluster randomized designs are simple and powerful, but still need to address “clustering” for design and analysis.
- Individually randomized group treatment trial designs have benefits of individual-level randomization, but still need to address “clustering” for design and analysis.

It all starts with a clear research question…

- Population
- Intervention
- Comparison
- Outcome(s)

From: European Medicines Agency
ICH E9 (R1)

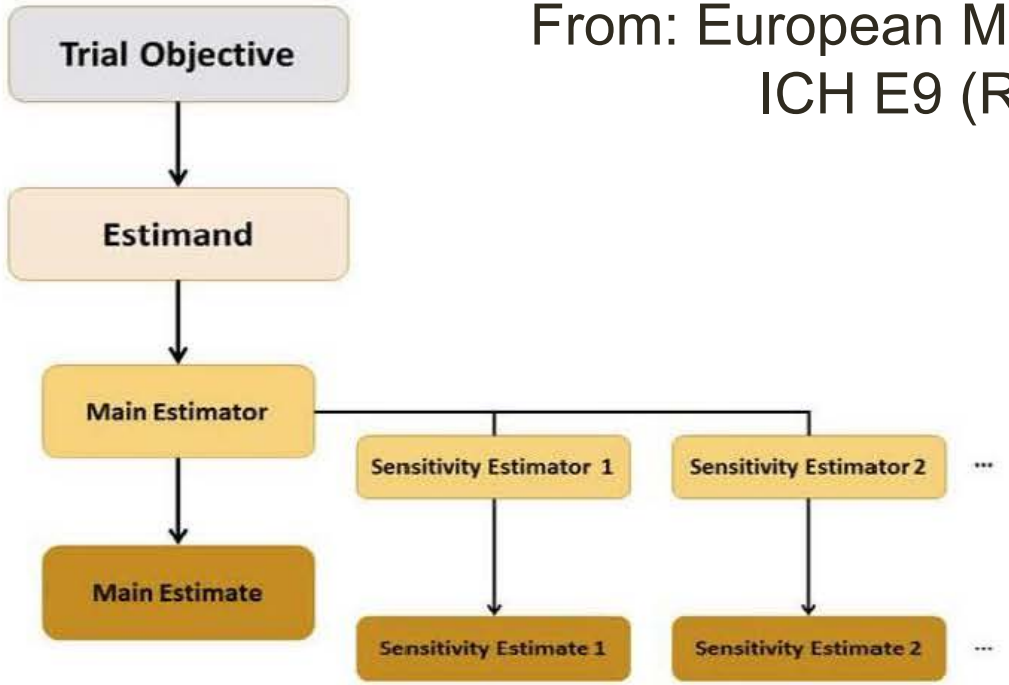


Figure 1: Aligning target of estimation, method of estimation, and sensitivity analysis, for a given trial objective

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

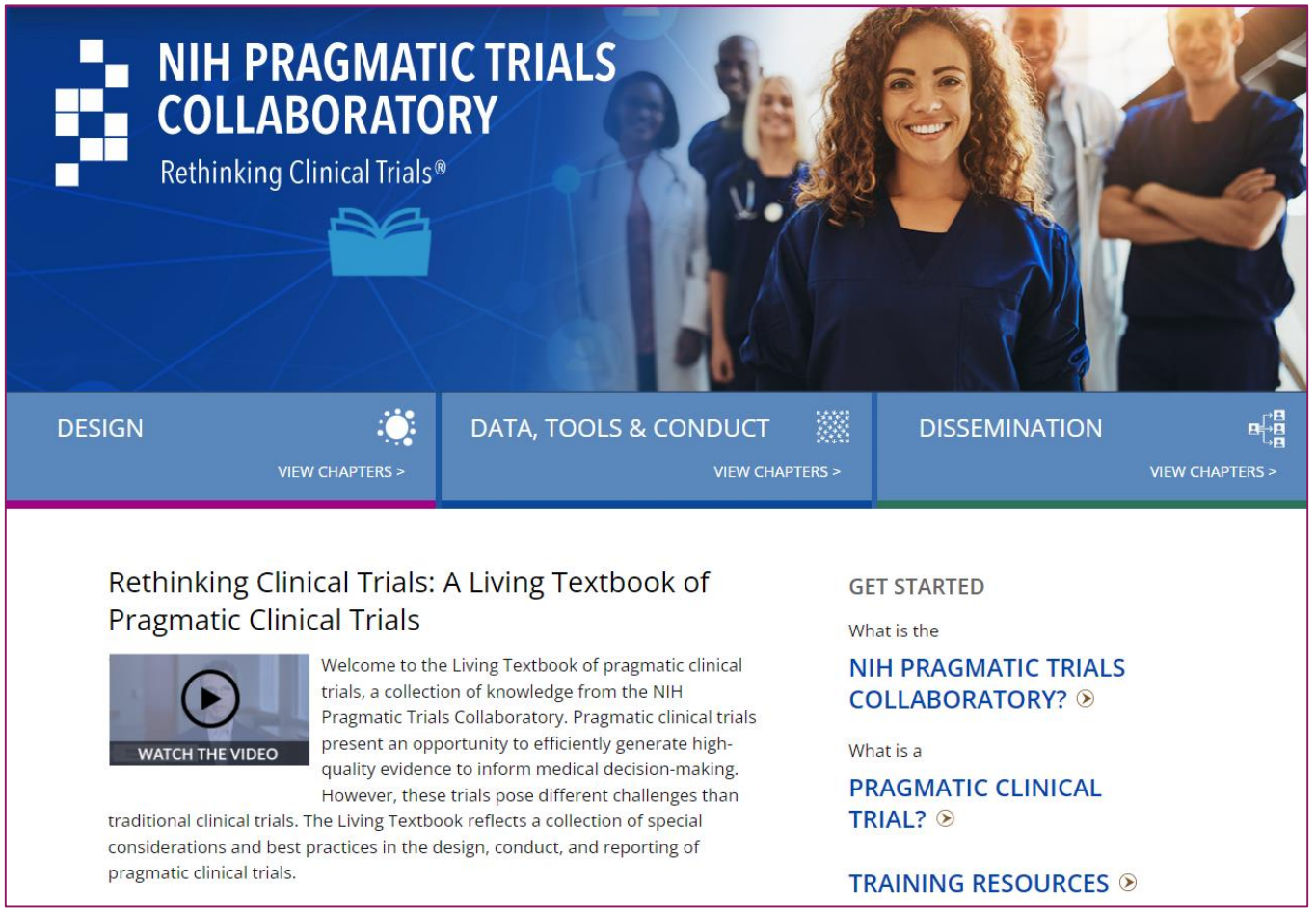
Summary: Important things to know

- Studies that randomize groups or deliver interventions to groups face special analytic challenges not found in traditional individually randomized trials
- Failure to address these challenges will result in an underpowered study and/or an inflated type 1 error rate
- We won't advance the science by using inappropriate methods

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Resource: The Living Textbook

Visit the Living Textbook of Pragmatic Clinical Trials at
www.rethinkingclinicaltrials.org



NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

NIH resources

- Pragmatic and Group-Randomized Trials in Public Health and Medicine
  - https://prevention.nih.gov/grt
  - 7-part online course on GRTs and IRGTs
- Mind the Gap Webinars
  - https://prevention.nih.gov/education-training/methods-mind-gap
    - Toward Causal Inference in Cluster Randomized Trials: Estimands and Reflection on Current Practice (Fan Li, November 3, 2022)
    - An Introduction to Cross-classified, Multiple Membership, and Dynamic Group Multilevel Models (Don Hedeker, October 20, 2022)
    - Robust Inference for Stepped Wedge Designs (Jim Hughes, May 17, 2022)
- Research Methods Resources Website
  - https://researchmethodsresources.nih.gov/
  - Material on GRTs, IRGTs, SWGRTs and a sample size calculator for each

Recommended reading

- Murray DM et al. Essential ingredients and innovations in the design and analysis of group-randomized trials. Ann Rev Public Health. 2020;41:1-19
- Kenny A et al. Analysis of stepped wedge cluster randomized trials in the presence of a time-varying treatment effect. Stat Med. 2022. PMID: 35774016.
- Kahan BC et al. Estimands in cluster-randomized trials: choosing analyses that answer the right question. Int J Epidemiol. 2022. PMID: 35834775.
- Maleyeff L et al. Assessing exposure-time treatment effect heterogeneity in stepped-wedge cluster randomized trials. Biometrics. 2022. Epub 2022/11/24. PMID: 36416302.
- Brown CH et al. Accounting for Context in Randomized Trials after Assignment. Prevention science : the official journal of the Society for Prevention Research. 2022. PMID: 36083435.