ePCT Experimental Design and Analysis

Jonathan Moyer, PhD
Statistician, National Institutes of Health
Office of Disease Prevention

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Learning goals

- Learn about cluster randomized and stepped-wedge study designs
- Recognize the analytical challenges and trade-offs of pragmatic study designs, focusing on what PIs need to know

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Important things to know

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

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Design Considerations

Embedded Pragmatic Clinical Trials

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

It all starts with a clear research question…

- Population
- Intervention
- Comparison
- Outcome(s)

From: European Medicines Agency
ICH E9 (R1)

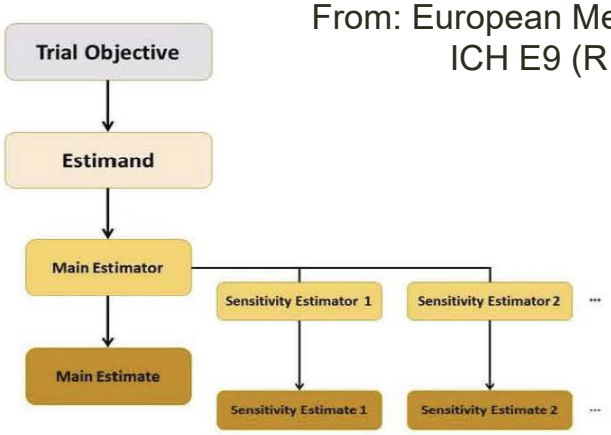


Figure 1: Aligning target of estimation, method of estimation, and sensitivity analysis, for a given trial objective

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Methods for pragmatic trials

- Pragmatic trials do not require a completely different set of research designs, measures, analytic methods, etc.
- During study design:
  - State hypotheses
  - Pre-specify analyses
  - Calculate sample size needed for desire power
  - Consider restricted randomization (e.g., stratified randomization)
  - Determine data on participant characteristics to be collected
  - Anticipate sources of heterogeneity
- Randomized trials will provide the strongest evidence.
  - What kind of randomized trial depends on the research question and how the intervention will be delivered

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

NIH Collaboratory ePCT: STOP CRC

- Strategies and Opportunities to Stop Colorectal Cancer in Priority Populations (STOP CRC)
- 40,000+ patients across 26 clinical sites
- Intervention
  - Health system–based program to improve CRC screening
  - Applied to clinical site → cluster randomization
- Unit of randomization: clinical site
- Two-arm cluster randomized trial (CRT)
  - Also referred to as a group-randomized trial

stop screen to prevent colon cancer

Coronado GD et al. *Contemp Clin Trials*. 2014;38(2):344-349.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

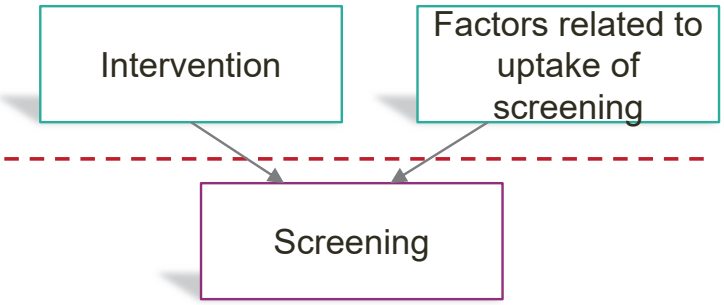
Reasons to randomize clusters instead of individuals

- Intervention targets health care units rather than individuals
  - STOP CRC: clinic-based intervention to improve screening
- Intervention targeted at individual risks "contamination"
  - Intervention spills over to members of control arm
  - For example, physicians randomized to new educational program may share knowledge with control-arm physicians in their practice
  - Contamination reduces the observed treatment effect
- Logistically easier to implement intervention by cluster

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

STOP CRC cluster randomization

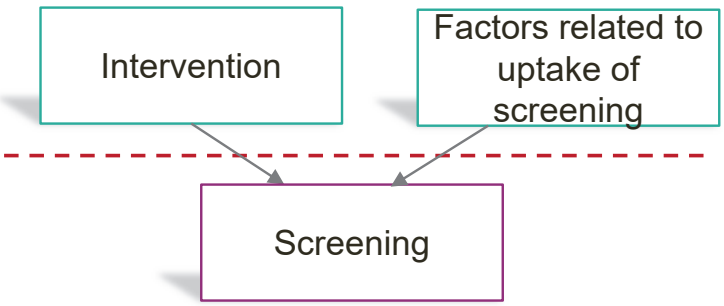
Level 2: Randomization at the level of the clinic (ie, cluster)



Level 1: Individual-level outcomes nested within clinics

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

STOP CRC cluster randomization



Level 1: Individual-level outcomes nested within clinics

- Individual-level outcomes within same clinic expected to be correlated (i.e., to *cluster*)
- Reduces power to detect treatment effect if same sample size used as under individual randomization

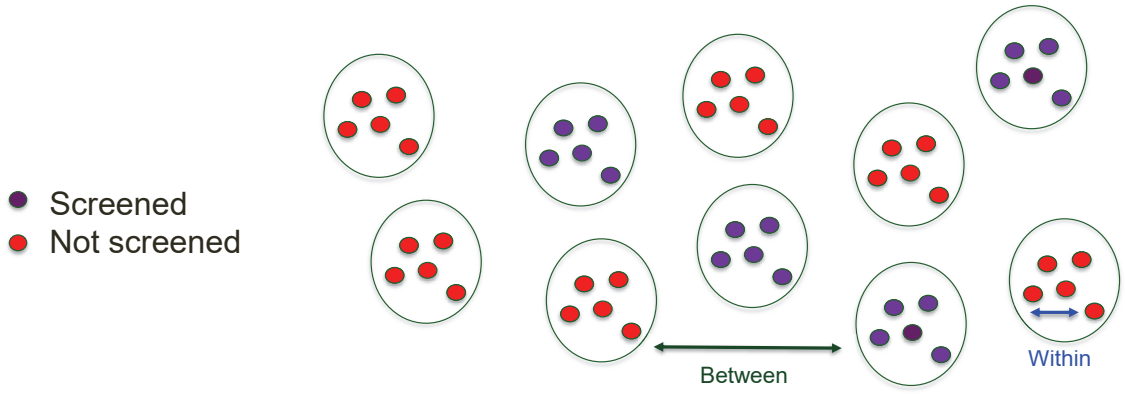
NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Understanding outcome clustering

- Consider 10 control-arm clinics (i.e., clusters)
- Each with 5 age-eligible patients: ie, who are not up to date with colorectal cancer (CRC) screening
- Binary outcome: not screened (Y/N)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Understanding outcome clustering: complete clustering (ICC =1)



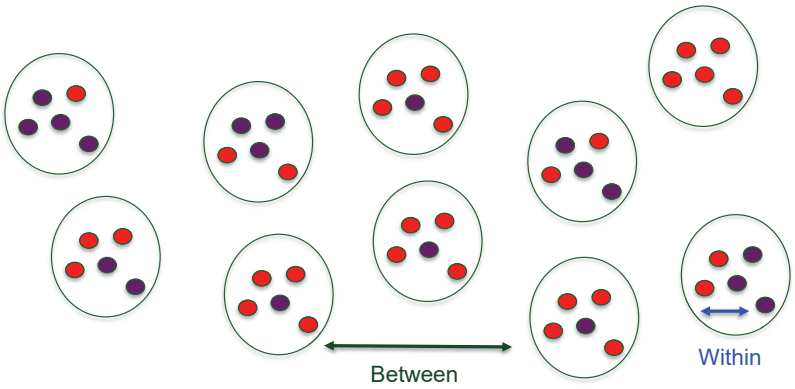
Intracluster correlation coefficient (ICC) = $\frac{\sigma_B^2}{\sigma_{\text{Total}}^2} = \frac{\sigma_B^2}{\sigma_B^2+\sigma_W^2} = \frac{\sigma_B^2}{\sigma_B^2} = 1$, because $\sigma_B^2 > 0$ & $\sigma_W^2 = 0$

σ_B^2 = between-cluster outcome variance; σ_W^2 = within-cluster outcome variance

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Understanding outcome clustering: some clustering (0 < ICC < 1)

- Screened
- Not screened



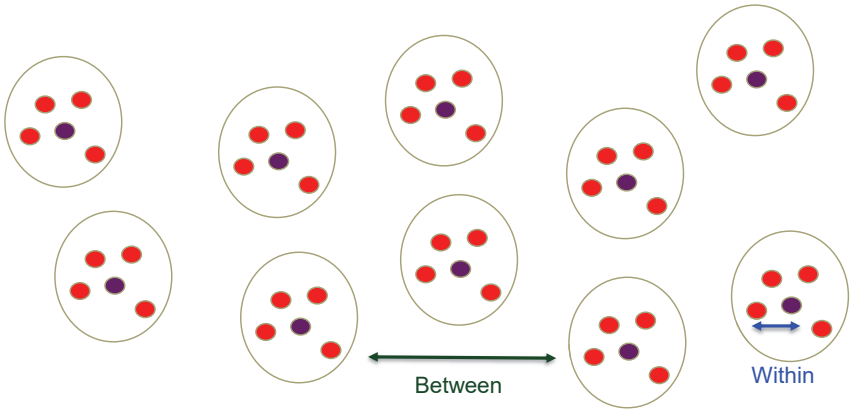
$$ICC = \frac{\sigma_B^2}{\sigma_B^2+\sigma_W^2};\quad 0 < ICC < 1, \text{ because } \sigma_B^2 > 0 \;\&\; \sigma_W^2 > 0$$

σ_B^2 = between-cluster outcome variance; σ_W^2 = within-cluster outcome variance

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Understanding outcome clustering: no clustering (ICC=0)

- Screened
- Not screened



$$ICC = \frac{\sigma_B^2}{\sigma_B^2+\sigma_W^2};\quad ICC = 0 \text{ because } \sigma_B^2 = 0 \;\&\; \sigma_W^2 > 0$$

σ_B^2 = between-cluster outcome variance; σ_W^2 = within-cluster outcome variance

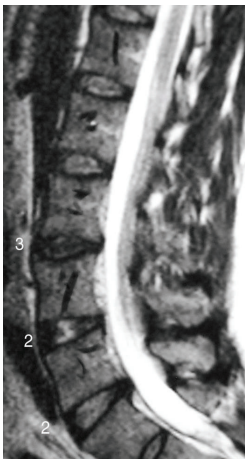
NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Summary of design issues for CRTs

- All the design features common to RCTs are available to CRTs with the added complication of an extra level of nesting:
  - Cohort and cross-sectional designs
  - Post only, pre-post, and extended designs
  - Single-comparison designs and factorial designs
  - Restricted randomization (stratification, constrained randomization, etc.)
- Most CRTs are “small”, ie, total # clusters (C) <50
  - Small number of independent units may result in low power
  - Randomization may not evenly distribute potential confounders
- The primary threats to internal and statistical validity are well known, and defenses are available.
  - Plan the study to reflect the nested design, with sufficient power for a valid analysis, and avoid threats to internal validity.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

NIH Collaboratory ePCT: LIRE

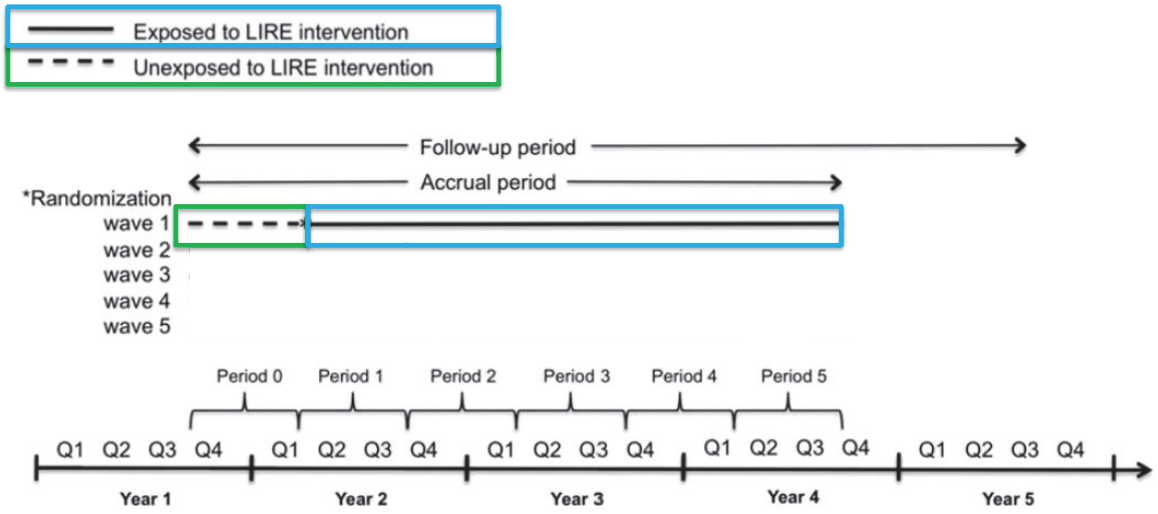


- Lumbar Imaging With Reporting of Epidemiology (LIRE)
- Goal: Reduce unnecessary spine interventions by providing info on prevalence of normal findings.
- Patients of 1700 PCPs across 100 clinics
- Clinic-level intervention → cluster randomization
- Unit of randomization: clinic
- Pragmatic trial
  - All clinics will eventually receive intervention
  - Stepped-wedge CRT (SW-CRT)

Jarvik JG et al. *Contemp Clin Trials*. 2015;45(Pt B):157-163.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

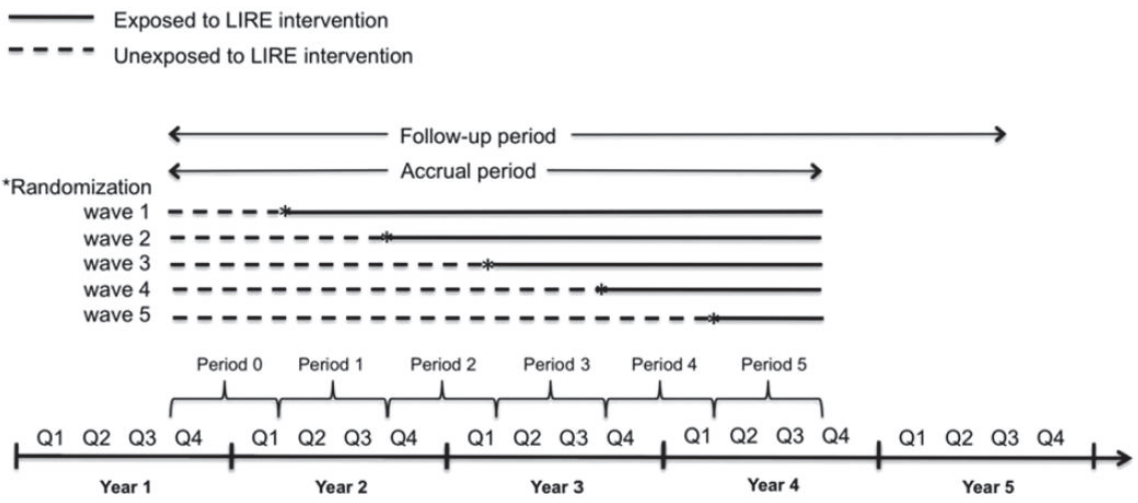
NIH Collaboratory ePCT: LIRE



Source: Jarvik JG et al. *Contemp Clin Trials*. 2015;45(Pt B):157-163.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

NIH Collaboratory ePCT: LIRE



Source: Jarvik JG et al. *Contemp Clin Trials*. 2015;45(Pt B):157-163.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

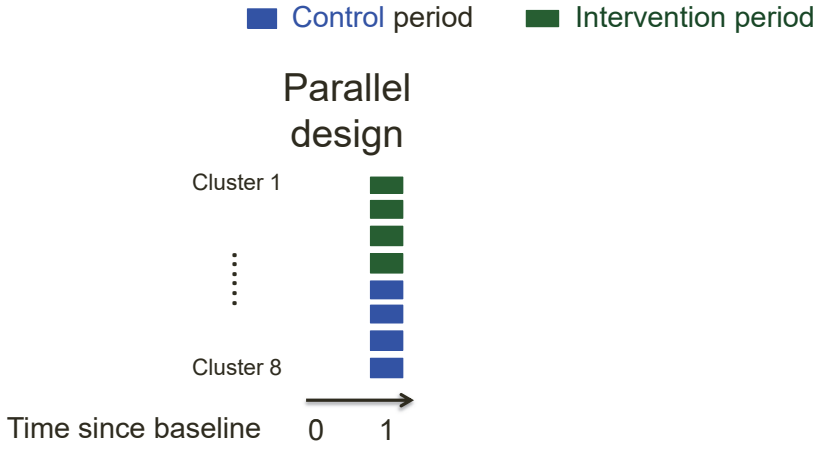
Summary of design issues for SW-CRTs

- Many design features common to RCTs are available to SW-CRTs:
  - Cohort and cross-sectional designs.
  - Single-comparison designs and factorial designs.
  - Restricted randomization to create comparable sequences.
- Clusters crossed with study condition, which minimizes confounding
  - Intervention effects confounded with time by design – always adjust for time!
  - SW-CRTs inherently more complicated than parallel CRTs.
- A SW-CRT may be an acceptable alternative to a parallel CRT if…
  - Intervention is being rolled out to all groups as part of system-wide implementation.
  - Cannot implement intervention in many groups at same time.
    - Consider a staggered start parallel CRT.
  - External events are unlikely to affect the outcomes (disruption!)
- Accounting for the pattern of the intervention effect over time:
  - The common assumption of an immediate, sustained intervention effect may yield biased estimates.
  - In the absence of evidence to the contrary, it is reasonable to assume intervention effect changes with exposure time.
  - Important to define intervention effect in this case – e.g., average at one point in time, average over more than one time.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Types of CRT designs

Examples with 8 clusters: 1-year intervention

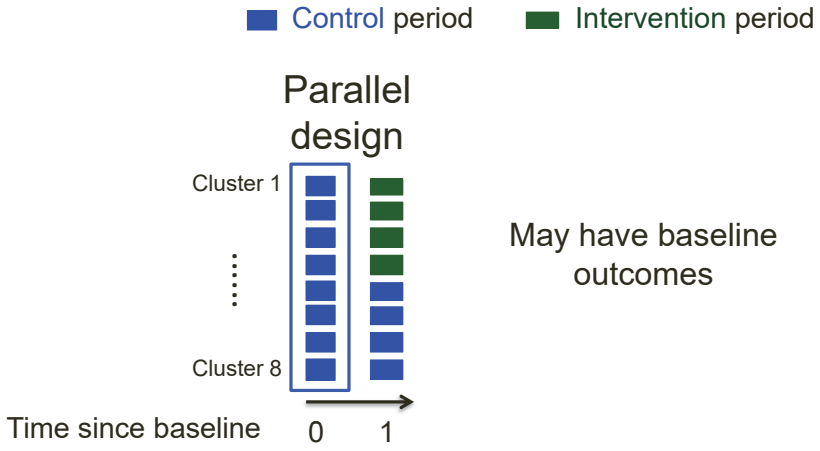


Based on: Hemming K et al. 2015. *Stat Med*. 34:181-196.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Types of CRT designs

Examples with 8 clusters: 1-year intervention

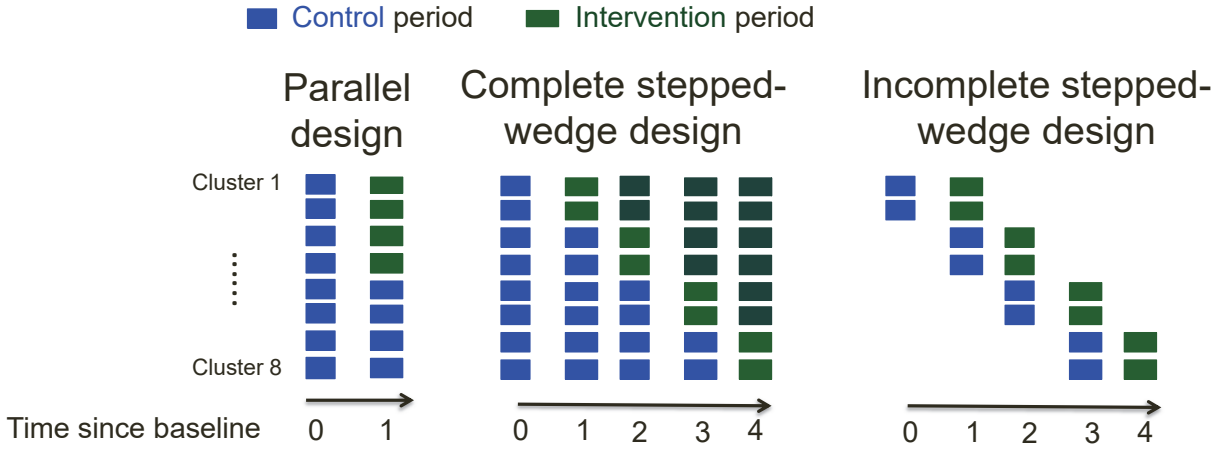


Based on: Hemming K et al. 2015. *Stat Med*. 34:181-196.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Types of CRT designs

Examples with 8 clusters: 1-year intervention



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

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

NIH Collaboratory ePCT: OPTIMUM

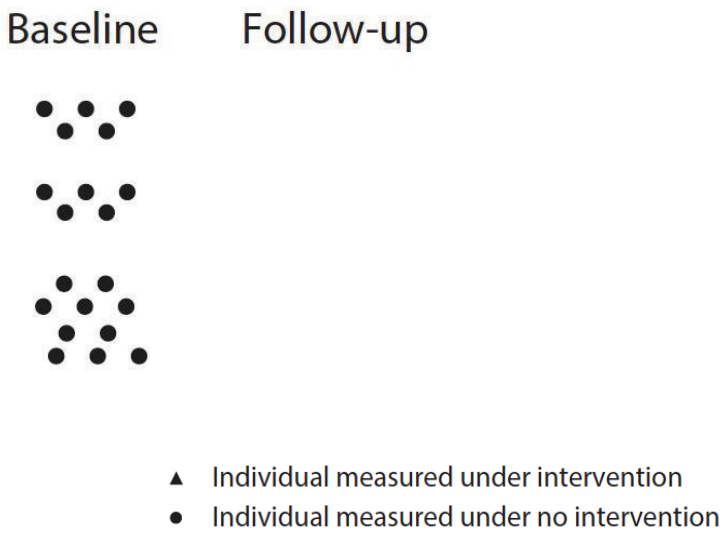
Optimum
Optimizing Pain Treatment in Medical Settings Using Mindfulness

- Optimizing Pain Treatment In Medical settings Using Mindfulness (OPTIMUM)
- Goal: to reduce pain and pharmacologic medications via a group-based mindfulness-based stress reduction (MBSR) program
- Study population: individuals with chronic lower back pain
- Unit of randomization: individual
  - Participants randomized to control and intervention conditions
  - No correlated outcomes before randomization
- Control condition: No post randomization correlation between outcomes for control participants
- Group-based online intervention → groups must be formed by study team
  - Post randomization interactions between participants!
- Individually-randomized group treatment (IRGT) trial
  - Post randomization groupings induce correlated outcomes

Greco CM et al. *Contemp Clin Trials*. 2021;109:106545.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

NIH Collaboratory ePCT: OPTIMUM



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

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Summary of design issues for IRGT trials

- Many design features common to RCTs are available to IRGT trials:
  - Cohort, but not easy to conceive of a cross-sectional design
  - Single-comparison designs and factorial designs
  - Restricted randomization procedures
- Clustering emerges post randomization
  - Could be due to a shared agent, participation in a group-based intervention, etc.
    - Fully Nested: Agents in both arms and nested within arm
    - Partially Nested: Agents in one arm only – participants in the other arm (usually control) are not clustered
    - Crossed: Agents interact with participants in both arms
  - Individual randomization, but ICC has a similar impact as it does for CRTs
  - Impact of ICC due to shared agent or group-based intervention often overlooked
- The primary threats to internal and statistical validity are well known, and defenses are available.
  - Plan the study to reflect the design, with sufficient power for a valid analysis, and avoid threats to internal validity

More information: Moyer JC et al. 2024. *Stat Med*. 43(25):4796-4818.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering: Impact on power

- Power and sample size
  - Account for clustering in CRTs (inc. SW-CRTs) & IRGT trials
  - Inflate RCT sample size
  - Work with statistician to do this correctly
- Use ICC for outcome
  - ICC often 0.01-0.05 in CRTs, larger in IRGT Trials
  - STOP CRC: ICC = 0.03 for primary outcome
  - OPTIMUM: ICC = 0.053 for primary outcome
  - Depends on outcome & study characteristics
  - Different outcome = different ICC, even in same CRT or IRGT trial
  - **More than 1 ICC in longitudinal study like SW-CRT!**

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering: Impact on power in STOP CRC

- "Assumed equal numbers of subjects per clinic and equal numbers of **clinics (n = 13) per [arm]**. In practice, the clinic sizes will not be equal, but since almost all clinics have at least **450** active age-eligible patients, we conservatively use this figure for all sites.

Source: Coronado GD et al. *Contemp Clin Trials*. 2014;38:344-9.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering: Impact on power in STOP CRC

- We based our calculations on the simple paradigm of comparing two binomial proportions with a **type I error rate of 5%**, and adjusted both for **intraclass correlation (ICC)** and the reduced **degrees-of-freedom (n = 24)** for the critical values. […] we expect the **ICC to be about .03**.

Source: Coronado GD et al. *Contemp Clin Trials*. 2014;38:344-9.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

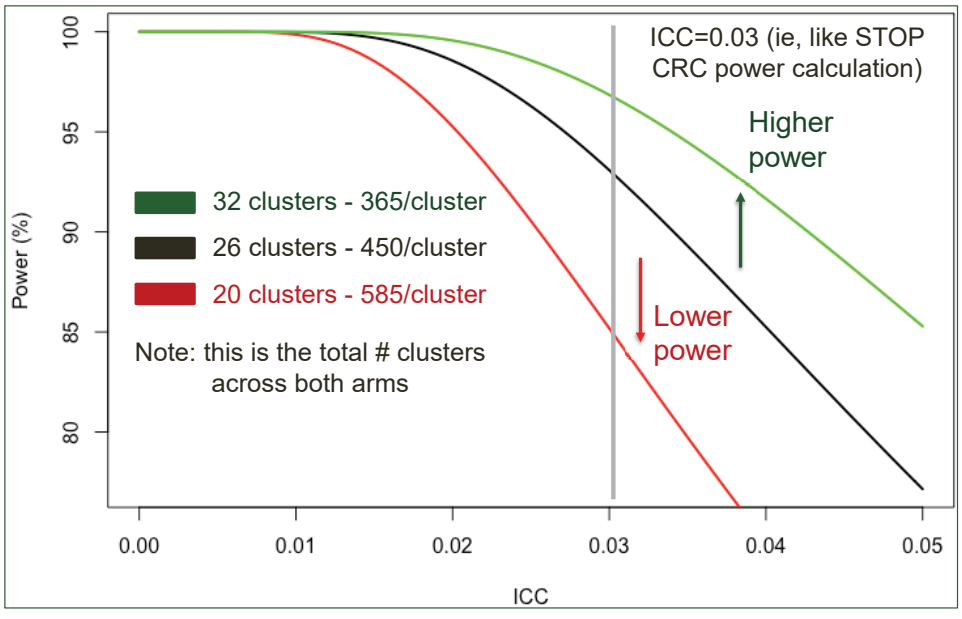
Clustering: Impact on power in STOP CRC

- "Using this figure, we will have **very good power (>91%) to detect absolute differences as small as 10 percentage points** even if the FIT [fecal immunochemical testing] completion rate in the **UC arm is as high as 15%** (fecal testing rates for 2013 for usual care clinics was 10%)."

Source: Coronado GD et al. *Contemp Clin Trials*. 2014;38:344-9.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

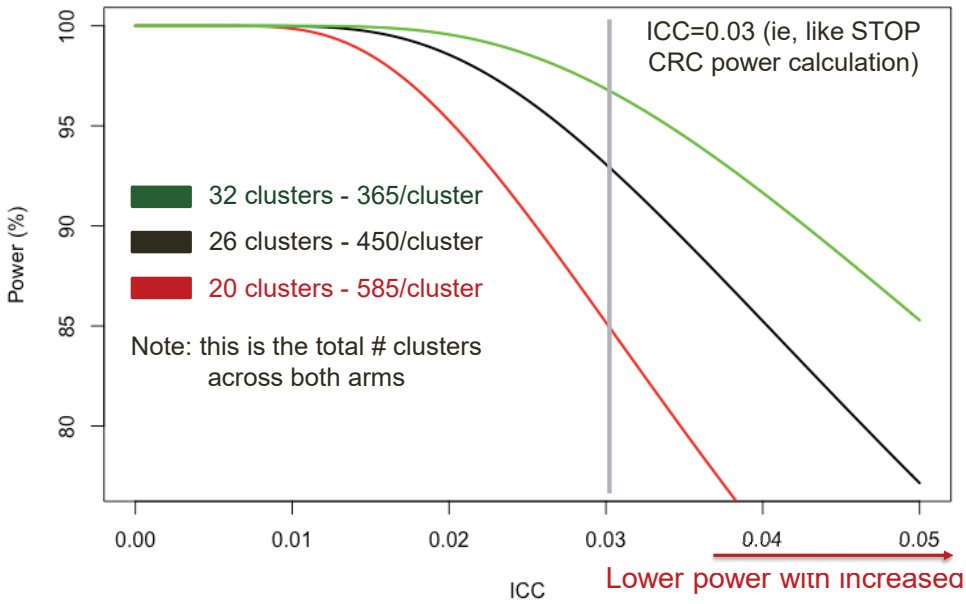
Clustering: Impact on power in STOP CRC



Power for parallel-arm CRT to compare two proportions of 15% vs 25% at two-tailed 5% significance (alpha) for an **overall sample of 11,700** (ie, like STOP CRC CRT)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering: Impact on power in STOP CRC



Power for parallel-arm CRT to compare two proportions of 15% vs 25% at two-tailed 5% significance (alpha) for an **overall sample of 11,700** (ie, like STOP CRC CRT)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Knowledge Checkpoint

Researchers are interested in the effect of participation in support groups vs. usual care on weight loss. Participants in the intervention attend group meetings, while usual care involves no group meetings. Out of 20 enrolled participants, 10 are randomly assigned to the intervention arm and attend support groups. Two therapists each lead a support group that meets on different weekday nights. Participant BMI will be measured at baseline (before randomization) and at 3 months.

- What design is this trial?
- Researchers powered this study assuming an RCT with 20 participants. How is the power likely to change if the IRGT nature of the trial is properly accounted for?
- What would be better approach to address correlated observations: increase the caseloads of the two therapists, or increase the number of therapists leading support groups?

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Question & Answer

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Analysis Considerations

Embedded Pragmatic Clinical Trials

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Learning goals

- Learn about cluster randomized and stepped-wedge study designs
- Recognize the analytical challenges and trade-offs of pragmatic study designs, focusing on what PIs need to know

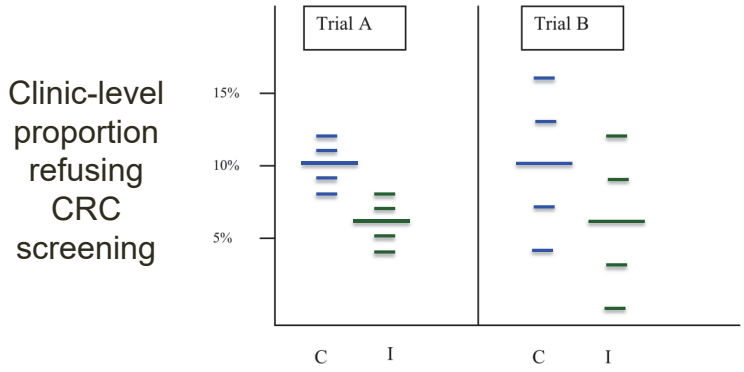
NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

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”

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering in CRTs: Implications for analysis



C=Control
I=Intervention

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

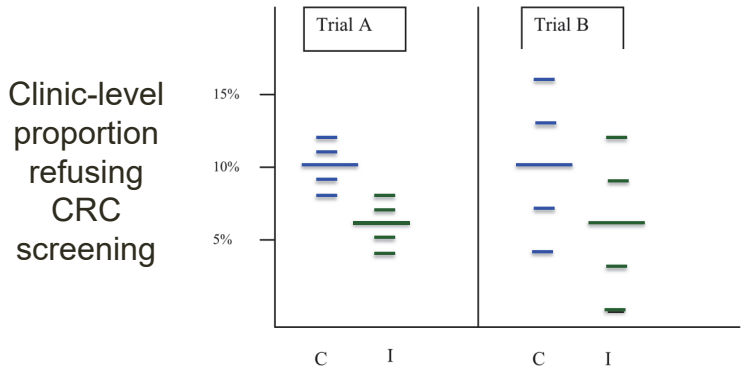
Overall screening refusal proportion in both trials: 10% vs 6%

Question: is intervention effective?

Adapted from Hayes & Moulton (2009)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering in CRTs: Implications for analysis



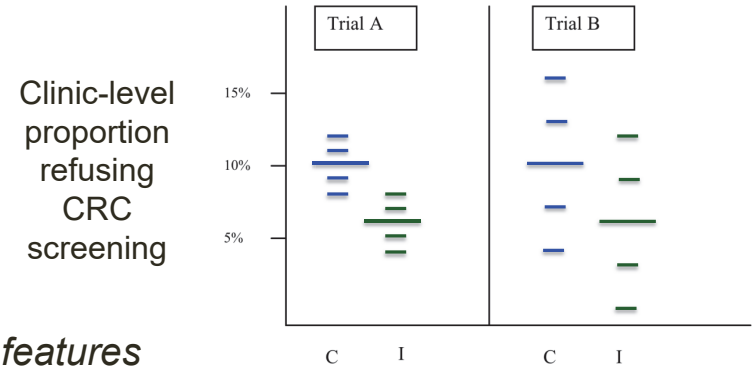
C=Control
I=Intervention

Which trial shows more evidence of benefit?

Adapted from Hayes & Moulton (2009)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering in CRTs: Implications for analysis



C=Control
I=Intervention

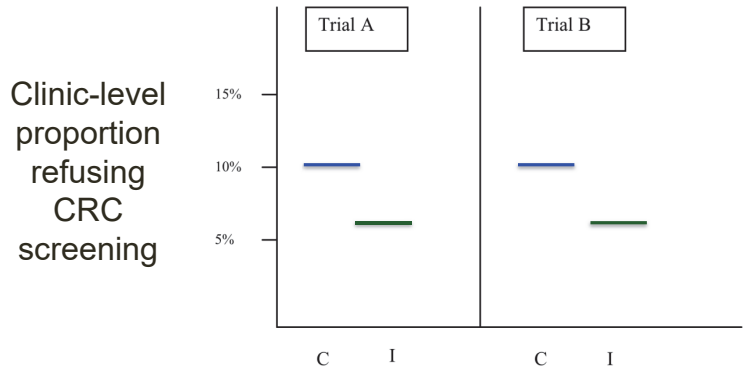
Study features

- 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)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering in CRTs: Implications for analysis



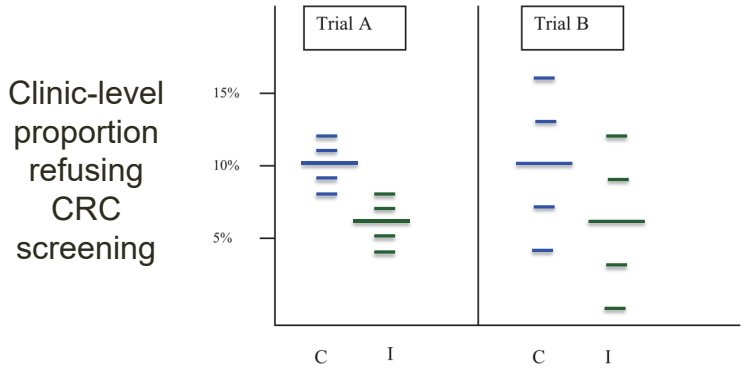
C=Control
I=Intervention

- 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)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering in CRTs: Implications for analysis



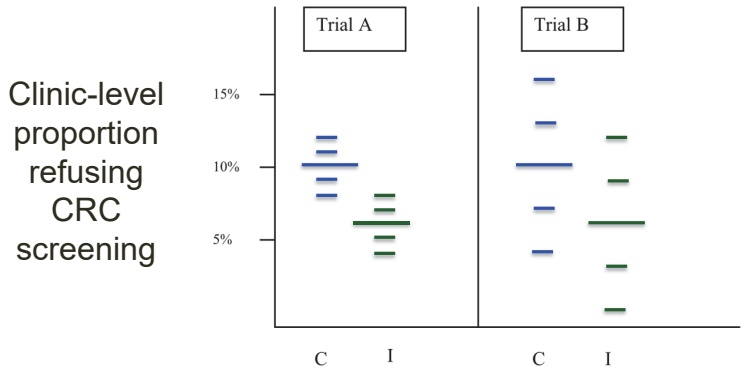
C=Control
I=Intervention

- Trial B p-value accounting for clustered design = **?**
- 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



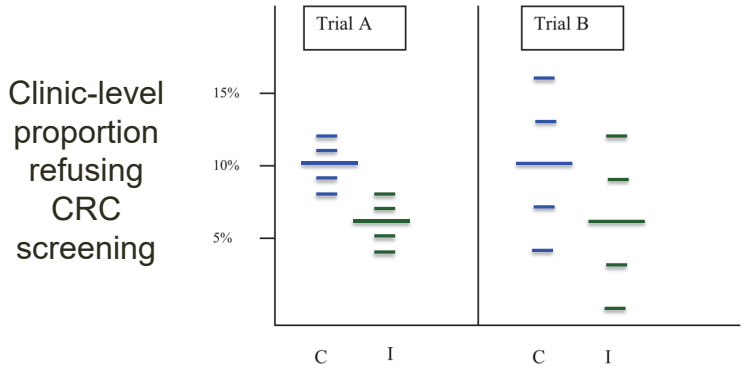
C=Control
I=Intervention

- 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



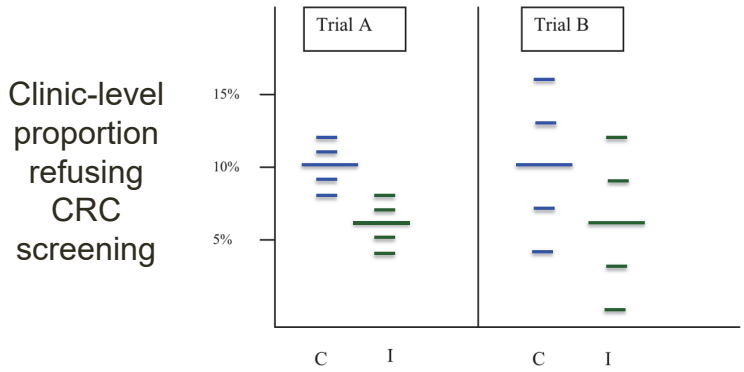
C=Control
I=Intervention

- 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)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering in CRTs: Implications for analysis



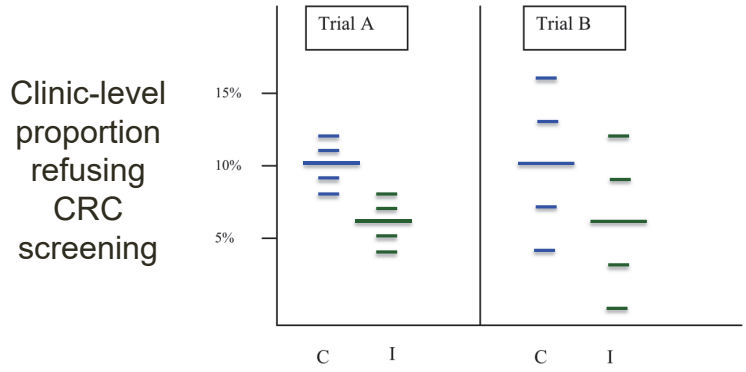
C=Control
I=Intervention

- 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)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Clustering in CRTs: Implications for analysis



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)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

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 hypothesis testing results (and confidence intervals) from each trial
  - Between-cluster variability (& clustering) in Trial A < Trial B
  - Important: if incorrectly ignore clustered design, could claim ‘significant’ when not (eg, Trial B)

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

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

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

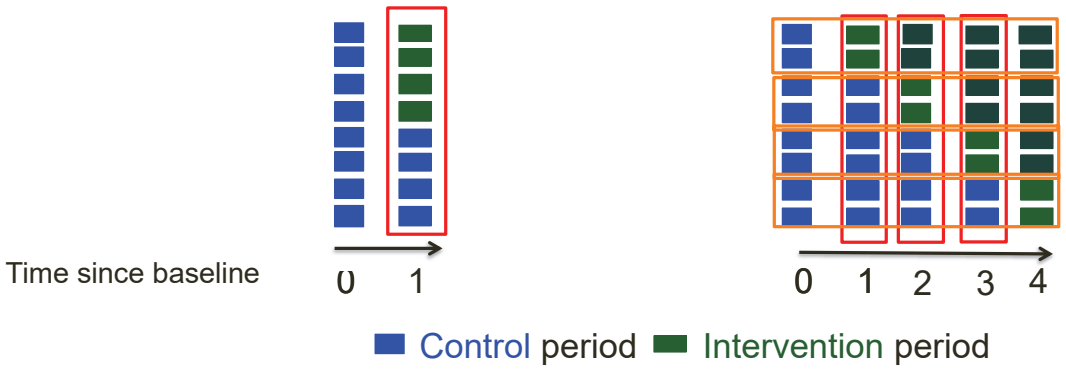
Analysis of CRTs, including SW-CRTs

Parallel design

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

Complete SW design

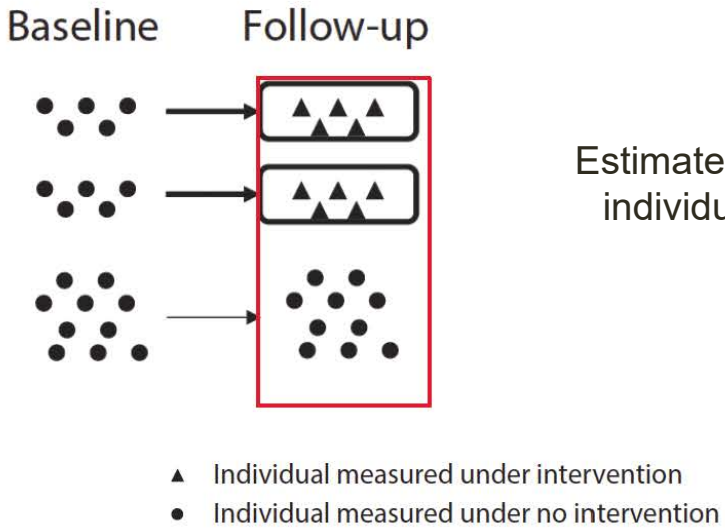
Estimated using both **vertical** & **horizontal** (ie, within-cluster) information



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

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Analysis of IRGT trials



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
  - Clustering due to shared agents or group-based intervention delivery often overlooked

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Analysis of CRTs, SW-CRTs, and IRGTs

- Clustering must be accounted for in analysis
- Challenges in “small” trials (# clusters < 50)
  - Intervention effect SE may be under-estimated
    - Mixed Models: degree of freedom
    - GEE: small sample adjustments corrections
  - Ignoring can lead 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 IRGT, even with small ICC

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

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

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

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

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

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 due to shared agents or group-based interventions in design and analysis.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Knowledge Checkpoint

Researchers are interested in the effect of participation in support groups vs. usual care on weight loss. Participants in the intervention attend group meetings, while usual care involves no group meetings. Out of 20 enrolled participants, 10 are randomly assigned to the intervention arm and attend support groups. Two therapists each lead a support group that meets on different weekday nights. Participant BMI will be measured at baseline (before randomization) and at 3 months.

- What are some sources of variation and correlation that should be accounted for in the analysis?

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Effectiveness-Implementation Hybrid Trial Designs

Embedded Pragmatic Clinical Trials

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

It all starts with a clear research question…

- Population
- Intervention
- Comparison
- Outcome(s)

From: European Medicines Agency
ICH E9 (R1)

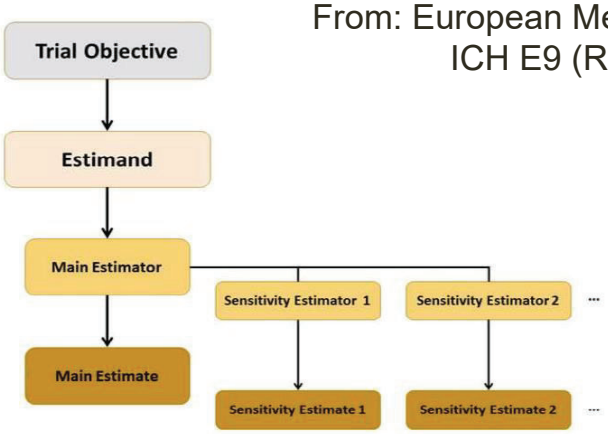


Figure 1: Aligning target of estimation, method of estimation, and sensitivity analysis, for a given trial objective

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Effectiveness and Implementation

- Trials often study both effectiveness and implementation outcomes.
- Effectiveness outcomes focus on how successful the trial was in addressing a health issue
  - Measured health outcomes, functional ability, quality of life, etc.
- Implementation outcomes focus on how the trial was implemented and delivered
  - Acceptability, adoption, appropriateness, cost, feasibility, fidelity, reach, etc.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Hybrid Designs

- Curran et al. (2012) introduced the hybrid effectiveness-implementation designs
  - Hybrid Type I tests a clinical intervention while gathering information on implementation
  - Hybrid Type II simultaneously tests a clinical intervention and an implementation intervention or strategy
  - Hybrid Type III tests an implementation intervention or strategy while gathering information on effectiveness
- "Hybrid Design" is in hindsight a somewhat unfortunate choice of words
  - Suggests that implementation research had different methods than other research and might not be held to the same standard as other research
  - The same rigorous methods for implementation research that we use for other research, changing only the focus

Hybrid Studies

- Curran et al (2022) updated their original description of hybrid designs, labeling them as **hybrid studies** without offering designs for each type.
- The usual trial evaluates a single intervention strategy delivered with a single implementation strategy as a package and it is not possible to distinguish the effects of the two strategies.
- In contrast, implementation trials compare intervention strategies and/or implementation strategies.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Hybrid Study Design Prototypes

- Stevens et al (2023) outline three design prototypes
- Type I (Effectiveness) requires at minimum a two-arm trial:
  1. No Intervention
  2. Intervention
  - ➢ Compare: No Intervention vs. Intervention
- Type II (both) requires at minimum a three-arm trial:
  1. No Intervention
  2. Intervention
  3. Intervention with Enhanced Implementation Strategy
  - ➢ Compare: No Intervention vs. Intervention vs. Intervention with Enhanced Implementation Strategy
- Type III (Implementation) requires at minimum a two-arm trial:
  1. Intervention
  2. Intervention with Enhanced Implementation Strategy
  - ➢ Compare: Intervention vs. Intervention with Enhanced Implementation Strategy

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Other Issues in Hybrid Studies

- Addressing clustered outcomes
  - Usual issues with ICC, small effective sample size, etc
  - Implementation outcomes are often cluster-level outcomes
- Masking of study arms
  - Routine in most clinical trials, helps guard against bias
  - However, many implementation outcomes serve as process variables (e.g. reach, adoption, fidelity)
  - Need to put into place practices that protect against bias but allow dedicated implementation staff to encourage adherence to study protocol and allow for feedback to stakeholders
- Adaptation of the intervention
  - Uncommon in most clinical trials
  - Adaptive interventions allow adaptations of the intervention using a prespecified process that describe what and when changes can be made
  - Limited guidance for implementation studies (Murray et al 2023):
    - Anticipated changes to the protocol should be pre-specified, as with any other adaptive intervention
    - Protocol changes should be approved of in advance

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®

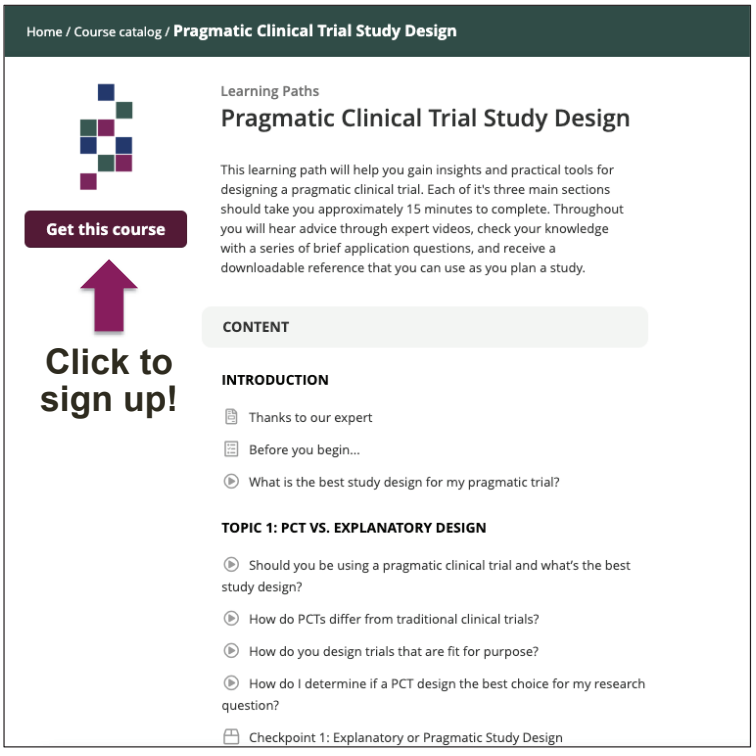
New Self-Paced Learning Path on Study Design

Free | Earn Certificate

1-Hour Course Includes

- Expert-led content, reference materials, and knowledge checkpoints
- Insights on how to:
  - Select the most appropriate study design for a pragmatic trial
  - Make decisions about randomization
  - Choose between parallel and stepped-wedge design

– Visit rethinkingclinicaltrials.org/training-resource/



Summary: Important things to know

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

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

NIH resources

- Pragmatic and Group-Randomized Trials in Public Health and Medicine
  - https://prevention.nih.gov/GRTcourse
  - 7-part online course on GRTs and IRGTs
- Mind the Gap Webinars
  - https://prevention.nih.gov/MindTheGap
    - Deconstruction of the Type 2 Hybrid Effectiveness-Implementation Study Design that Uses Two Randomized Controlled Trials (June Stevens, March 20, 2024)
    - Toward Causal Inference in Cluster Randomized Trials: Estimands and Reflection on Current Practice (Fan Li, November 3, 2022)
    - Robust Inference for Stepped Wedge Designs (Jim Hughes, May 17, 2022)
    - When is the Stepped Wedge Study a Good Study Design Choice? (Karla Hemming, January 21, 2022)
- Research Methods Resources Website
  - https://researchmethodsresources.nih.gov/
  - Material on GRTs, IRGTs, SWGRTs and a sample size calculator for each
  - Information on hybrid effectiveness-implementation studies

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Recommended reading

- Brown CH et al. Accounting for context in randomized trials after assignment. Prev Sci. 2022. PMID: 36083435.
- Curran GM et al. Reflections on 10 years of effectiveness-implementation hybrid studies. Front Health Serv. 2022. PMID: 36925811.
- Hemming K, Taljaard M. Reflection on modern methods: When is a stepped-wedge cluster randomized trial a good study design choice? Int J Epidemiol. 2020. PMID: 32386407.
- Hemming K, Taljaard M. Key considerations for designing, conducting and analysing a cluster randomized trial. Int J Epidemiol. 2023. PMID: 37203433.
- Hughes JP et al. Sample size calculations for stepped wedge designs with treatment effects that may change with the duration of time under intervention. Prev Sci. 2023. PMID: 37728810.
- 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.
- Murray DM et al. Essential ingredients and innovations in the design and analysis of group-randomized trials. Ann Rev Public Health. 2020. PMID: 31869281
- Murray DM et al. Implementation Research at NHLBI: Methodological and Design Challenges and Lessons Learned from the DECIPHeR Initiative. Ethn Dis. 2023. PMID: 38846726.
- Stevens J et al. Design of a dual randomized trial in a type 2 hybrid effectiveness-implementation study. Implement Sci. 2023. PMID: 37996884.

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®

Question & Answer

NIH PRAGMATIC TRIALS COLLABORATORY
Rethinking Clinical Trials®