NIH Collaboratory *Rethinking Clinical Trials®*

Health Care Systems Research Collaboratory

Living Textbook Grand Rounds Series

Demystifying Biostatistical Concepts for Embedded Pragmatic Clinical Trials

June 19, 2020

Elizabeth L. Turner, PhD, Duke University
Patrick J. Heagerty, PhD, University of Washington
David M. Murray, PhD, National Institutes of Health

For the NIH Collaboratory Coordinating Center
Biostatistics and Study Design Core Working Group

Overview

- Focus of this talk: demystifying design-related issues for embedded pragmatic clinical trials (ePCTs)
- Context: NIH Collaboratory–funded studies
- Three kinds of randomized trials
  - Randomized controlled trial (RCT)
  - Cluster randomized trial (CRT)
    - Parallel vs stepped-wedge
  - Individually randomized group treatment (IRGT) trial
- How to select amongst these designs?
- Other brief topics: clustering, power, and analytical issues

In the Living Textbook

EXPERIMENTAL DESIGNS AND RANDOMIZATION SCHEMES

1. Introduction
2. Statistical Design Considerations
3. Cluster Randomized Trials
4. Randomization Methods
5. Choosing Between Cluster and Individual Randomization
6. Alternative Cluster Randomized Designs
7. Concealment and Blinding
8. Designing to Avoid Identification Bias
9. Additional Resources

ANALYSIS PLAN

1. Introduction
2. Intraclass Correlation
3. Unequal Cluster Sizes
4. Accounting for Residual Confounding in the Analysis
5. Missing Data and Intention-to-Treat Analyses
6. EHR Data Extraction
7. Unanticipated Changes
8. Case Study: STOP CRC Trial

NIH Collaboratory ePCT: SPOT

SUICIDE PREVENTION OUTREACH TRIAL

- Suicide Prevention Outreach Trial (SPOT)
- Approximately 16,000 patients across 4 clinical sites
- Three-arm RCT to evaluate 2 individual-level interventions vs usual care
- Interventions
  - Skills training program
  - Care management program
- Intervention contact mostly though EHR
  - Low risk of “contamination”
  - Individual-level randomization appropriate
- Unit of randomization: patient

Simon GE et al. *Trials*. 2016;17(1):452.

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 rates
  - Applied to clinical site → cluster randomization
- Unit of randomization: clinical site
- Two-arm cluster randomized trial (CRT)
  - Also referred to as a group-randomized or community randomized trial

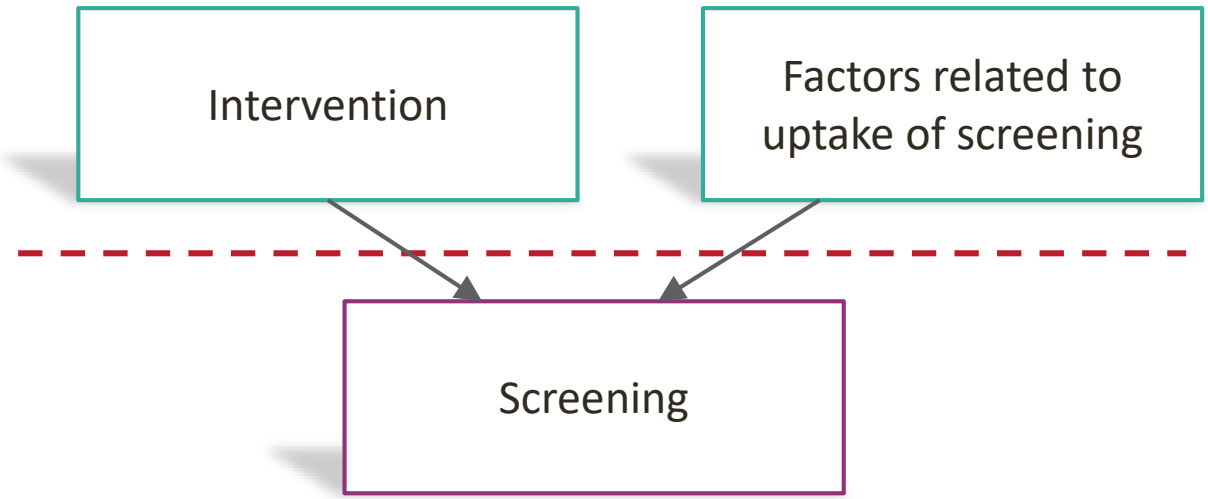
Coronado GD et al. *Contemp Clin Trials*. 2014;38(2):344-349.

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 at risk of contamination
  - Intervention adopted by 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

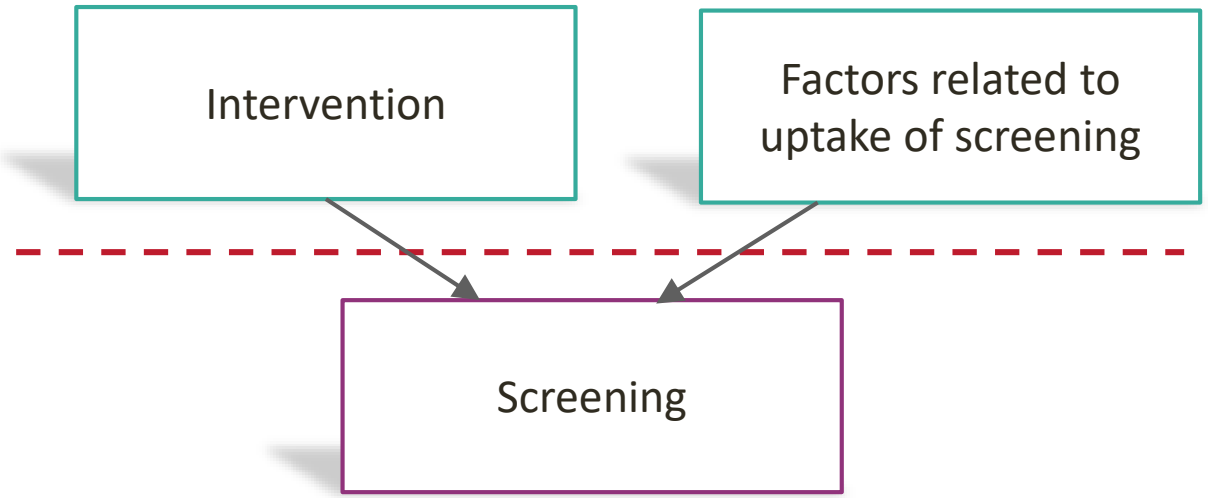
STOP CRC Cluster Randomization

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



Level 1: Individual-level outcomes nested within clinics

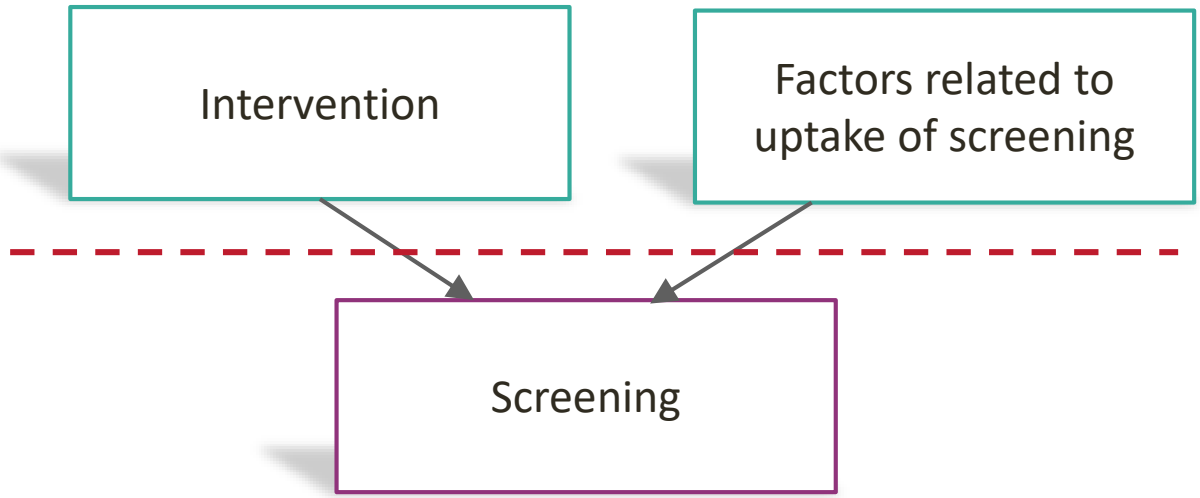
STOP CRC Cluster Randomization



Level 1: Individual-level outcomes nested within clinics

- Individual-level outcomes within same clinic expected to be correlated (ie, to *cluster*)

STOP CRC Cluster Randomization



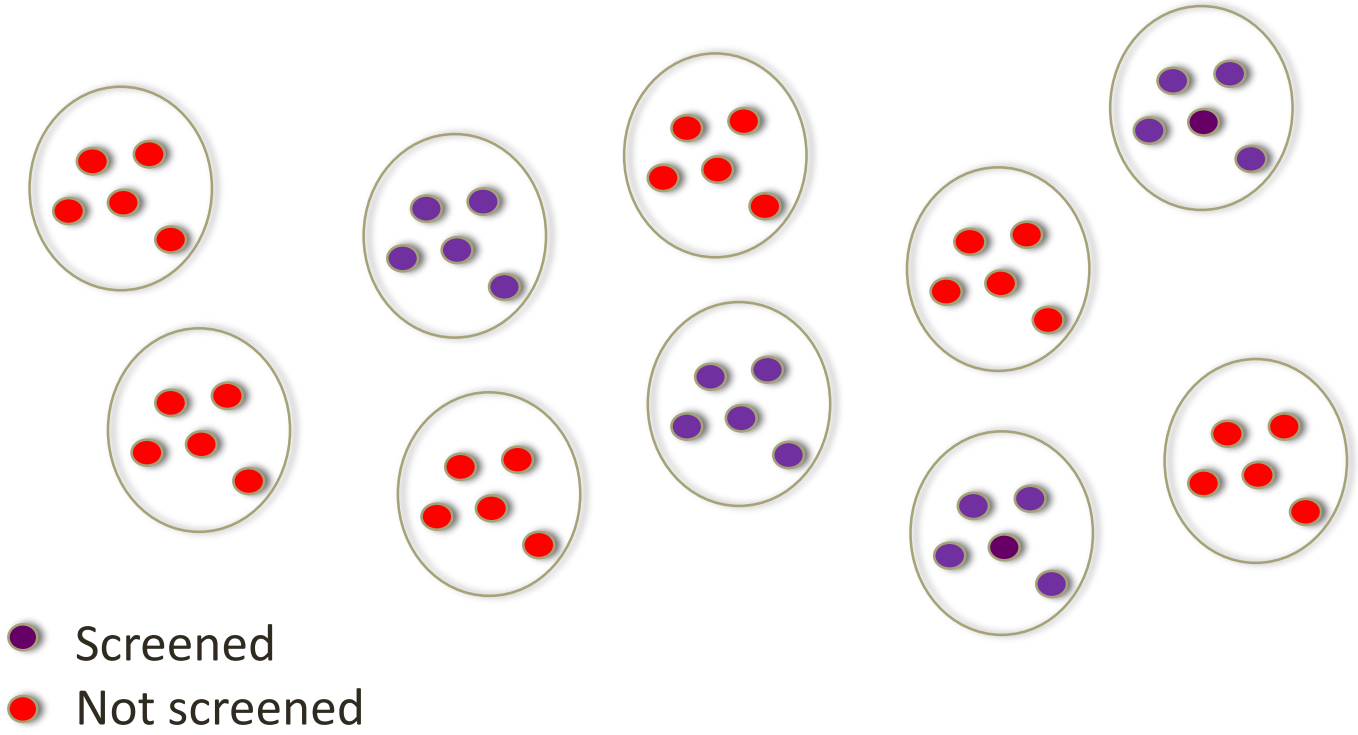
Level 1: Individual-level outcomes nested within clinics

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

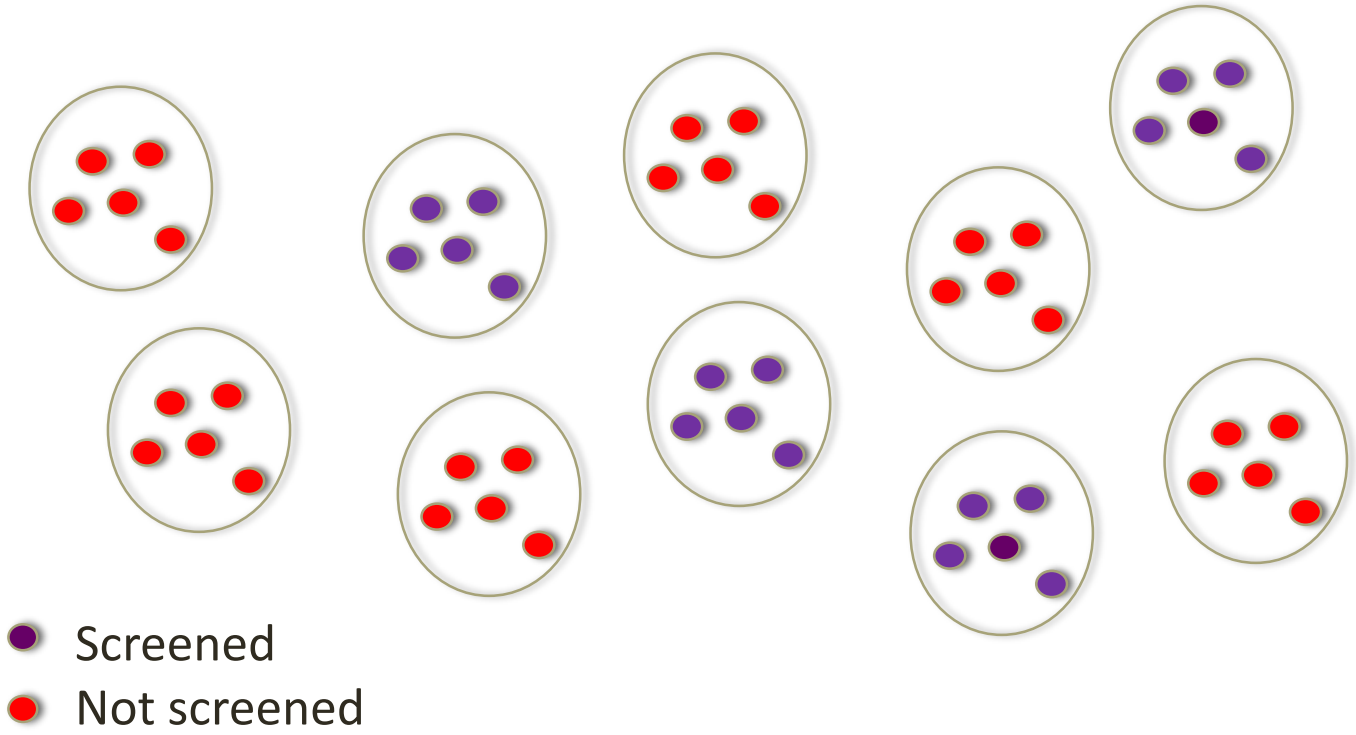
Understanding Outcome Clustering

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

Understanding Outcome Clustering: Complete Clustering

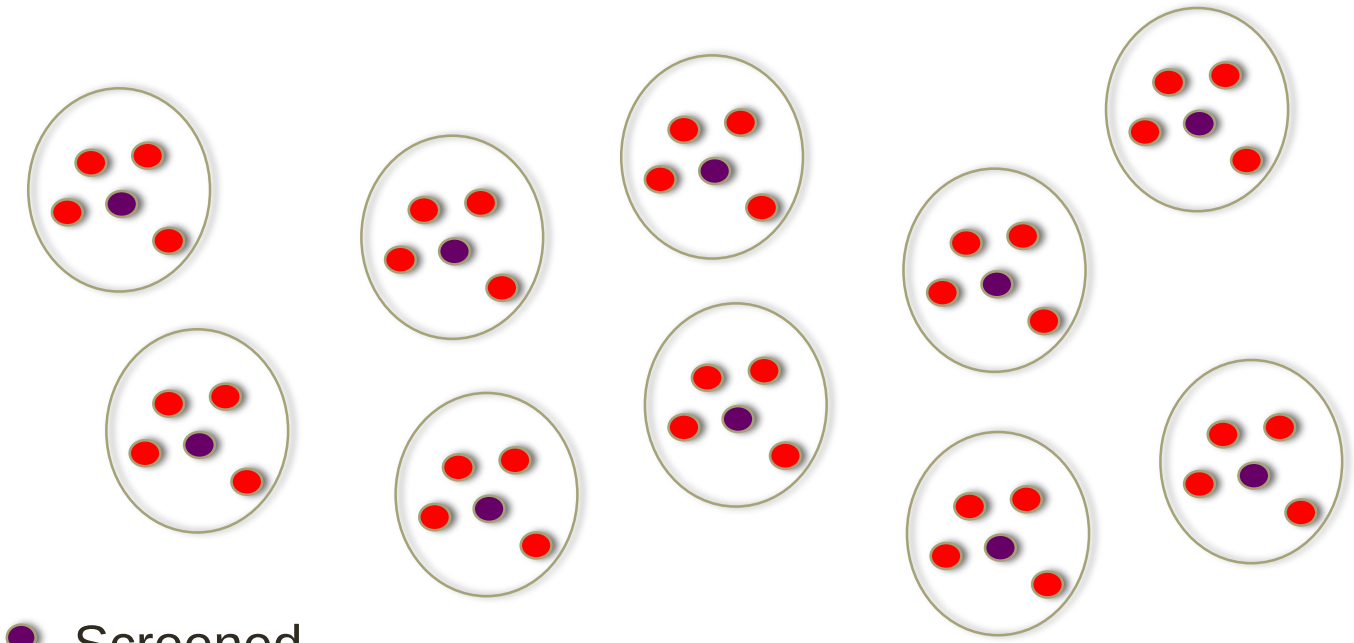


Understanding Outcome Clustering: Complete Clustering

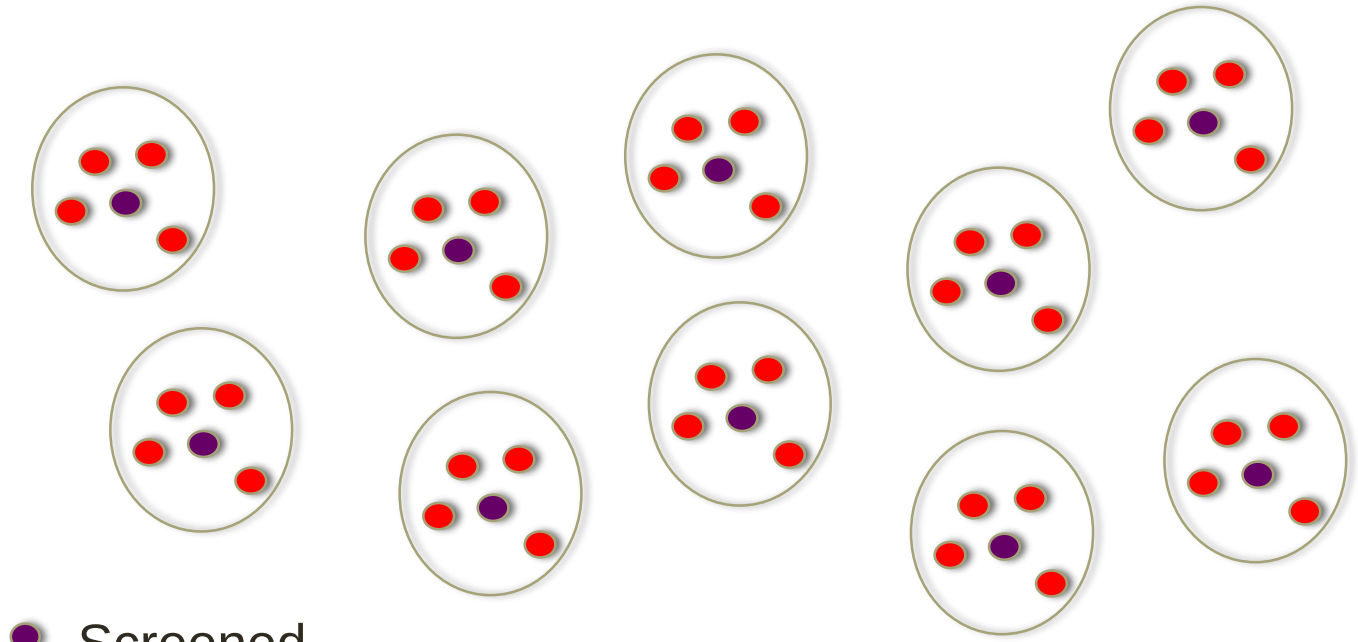


>1 participant/clinic gives no more information than a single participant/clinic since every participant in a given clinic has the same outcome

Understanding Outcome Clustering: No Clustering

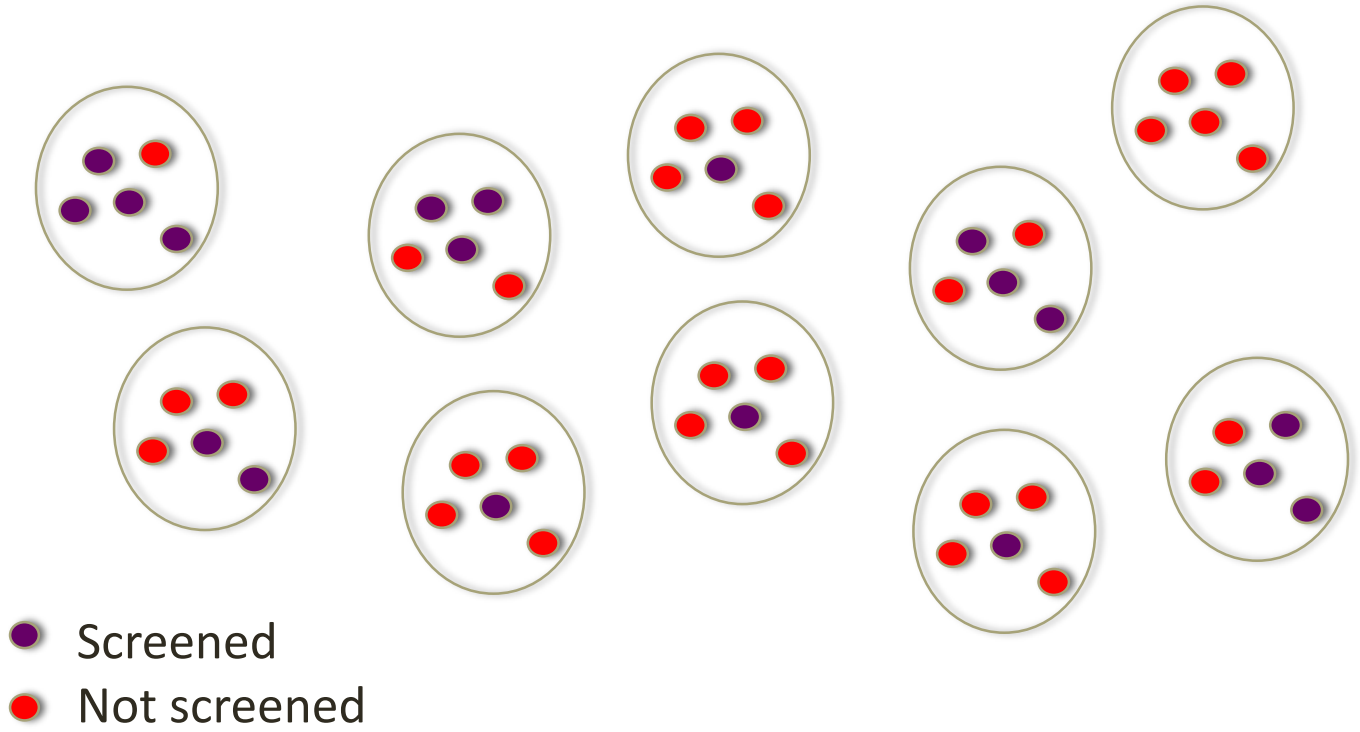


Understanding Outcome Clustering: No Clustering

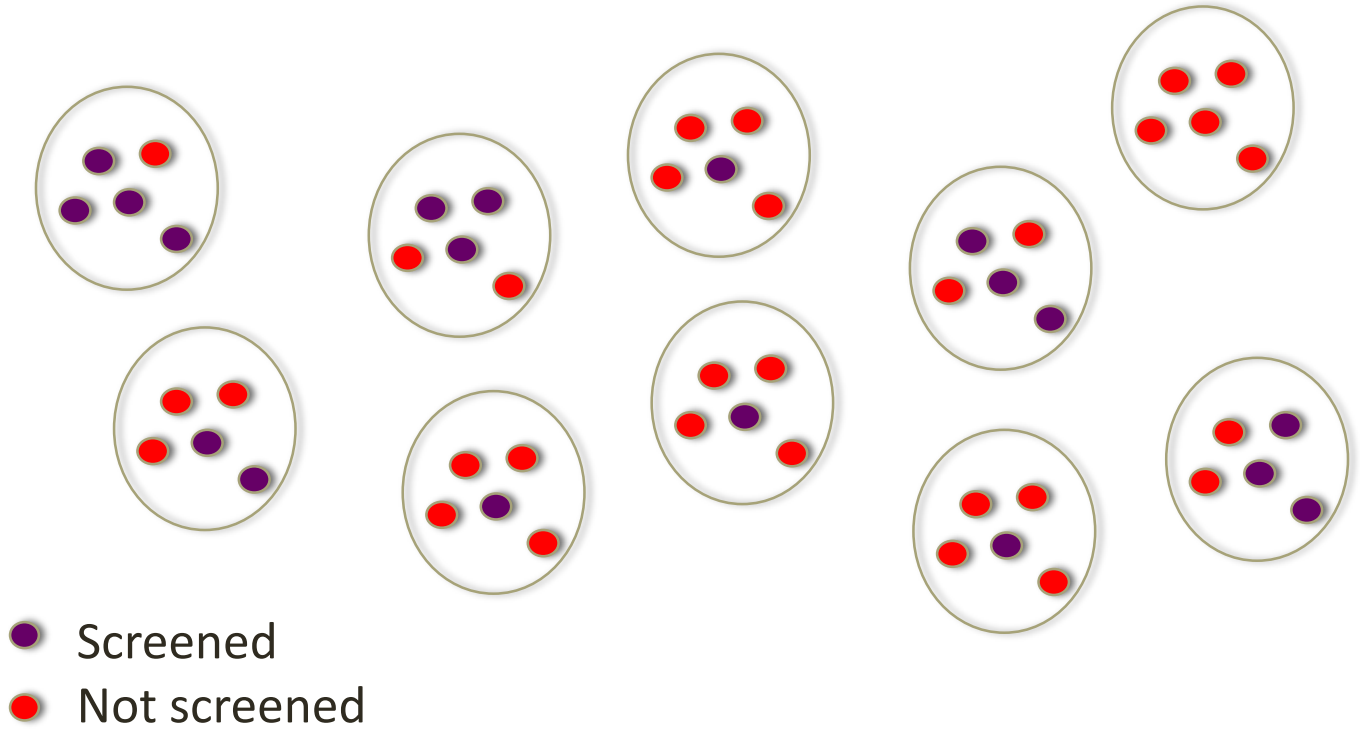


20% uptake of CRC screening in each clinic
No structure by clinic; more like a random sample of eligible participants

Understanding Outcome Clustering: Some Clustering



Understanding Outcome Clustering: Some Clustering



A more typical situation: proportion screened ranges from 0% - 80%

Measure of Outcome Clustering: Intraclass Correlation Coefficient (ICC)

- Needed for study planning and power
- Most commonly used measure of clustering
- Ranges: 0-1; 0 = no clustering; 1 = complete clustering
- Typically < 0.2; commonly around 0.01 to 0.05
- Between-cluster outcome variance vs total outcome variance

Measure of Outcome Clustering: Intraclass Correlation Coefficient (ICC)

- Needed for study planning and power
- Most commonly used measure of clustering
- Ranges: 0-1; 0 = no clustering; 1 = complete clustering
- Typically < 0.2; commonly around 0.01 to 0.05
- Between-cluster outcome variance vs total outcome variance

ICC for continuous outcomes:

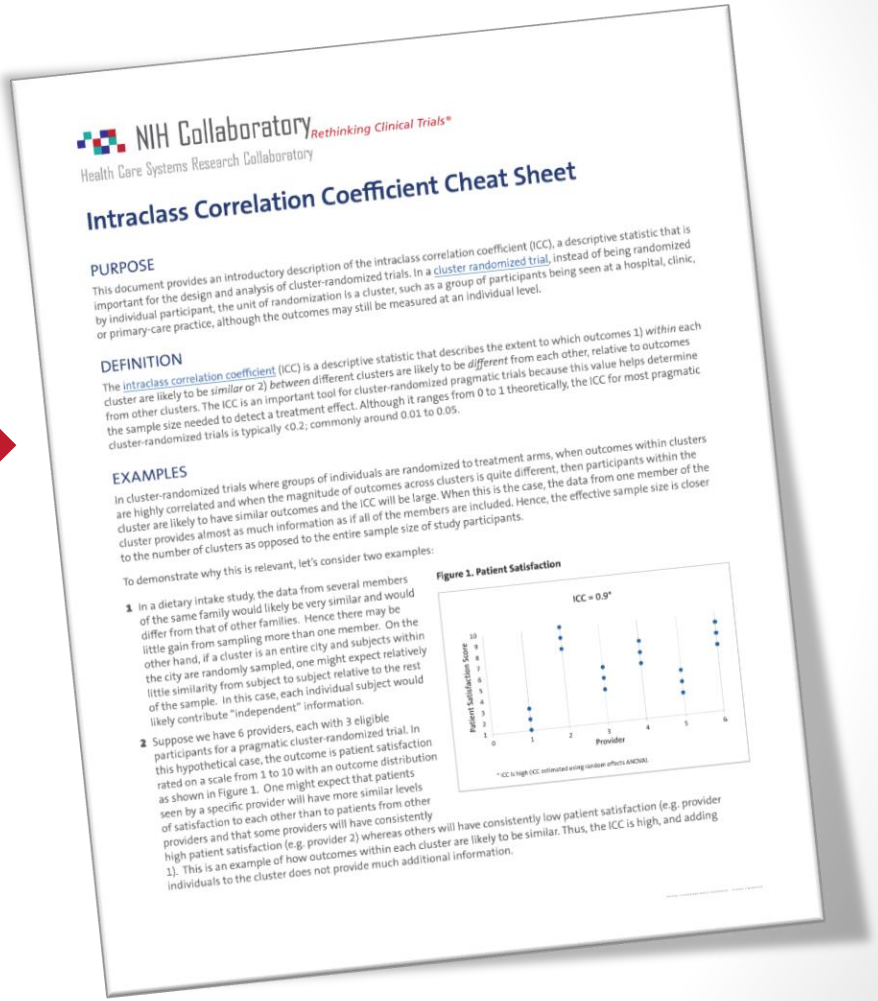
$$r = \frac{s_B^2}{s_B^2 + s_W^2} = \frac{s_B^2}{s_{Total}^2}$$

Involves both *between-cluster* and *within-cluster* variance

In the Living Textbook: ICC Cheat Sheet

DESIGN

ANALYSIS PLAN



NIH Collaboratory *Rethinking Clinical Trials*

Health Care Systems Research Collaboratory

Intraclass Correlation Coefficient Cheat Sheet

PURPOSE

This document provides an introductory description of the intraclass correlation coefficient (ICC), a descriptive statistic that is important for the design and analysis of cluster-randomized trials. In a cluster randomized trial, instead of being randomized by individual participant, the unit of randomization is a cluster, such as a group of participants being seen at a hospital, clinic, or primary-care practice, although the outcomes may still be measured at an individual level.

DEFINITION

The intraclass correlation coefficient (ICC) is a descriptive statistic that describes the extent to which outcomes 1) *within* each cluster are likely to be *similar* or 2) *between* different clusters are likely to be *different* from each other, relative to outcomes from other clusters. The ICC is an important tool for cluster-randomized pragmatic trials because this value helps determine the sample size needed to detect a treatment effect. Although it ranges from 0 to 1 theoretically, the ICC for most pragmatic cluster-randomized trials is typically <0.2; commonly around 0.01 to 0.05.

EXAMPLES

In cluster-randomized trials where groups of individuals are randomized to treatment arms, when outcomes within clusters are highly correlated and when the magnitude of outcomes across clusters is quite different, then participants within the cluster are likely to have similar outcomes and the ICC will be large. When this is the case, the data from one member of the cluster provides almost as much information as if all of the members are included. Hence, the effective sample size is closer to the number of clusters as opposed to the entire sample size of study participants.

To demonstrate why this is relevant, let's consider two examples:

1. In a dietary intake study, the data from several members of the same family would likely be very similar and would differ from that of other families. Hence there may be little gain from sampling more than one member. On the other hand, if a cluster is an entire city and subjects within the city are randomly sampled, one might expect relatively little similarity from subject to subject relative to the rest of the sample. In this case, each individual subject would likely contribute "independent" information.
2. Suppose we have 6 providers, each with 3 eligible participants for a pragmatic cluster-randomized trial. In this hypothetical case, the outcome is patient satisfaction rated on a scale from 1 to 10 with an outcome distribution as shown in Figure 1. One might expect that patients seen by a specific provider will have more similar levels of satisfaction to each other than to patients from other providers and that some providers will have consistently high patient satisfaction (e.g. provider 2) whereas others will have consistently low patient satisfaction (e.g. provider 1). This is an example of how outcomes within each cluster are likely to be similar. Thus, the ICC is high, and adding individuals to the cluster does not provide much additional information.

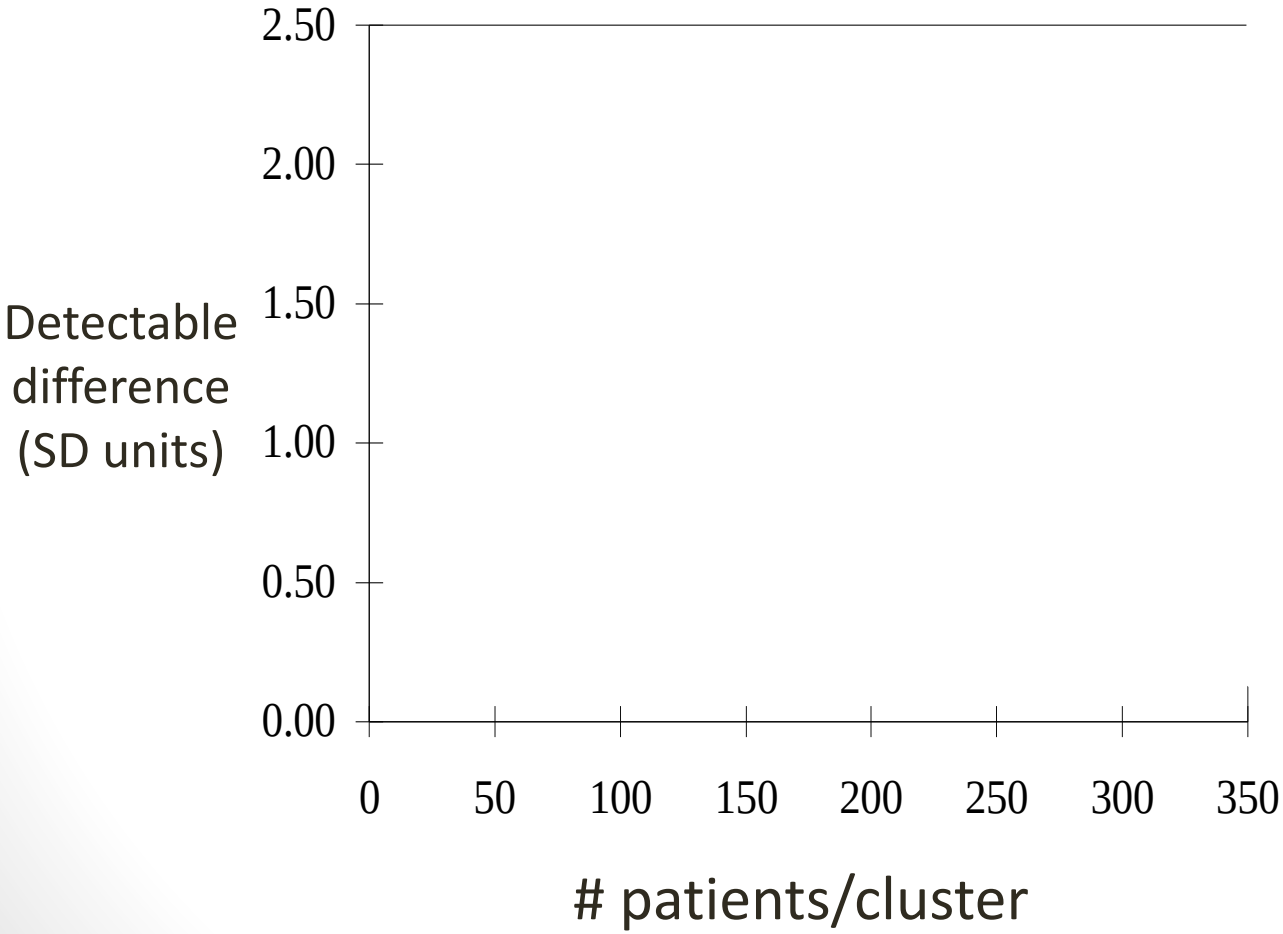
Figure 1. Patient Satisfaction

Accounting for Clustering Requires Larger Sample for Adequate Power

- Power and detectable difference is affected by…
  - Strength of the clustering effect (eg, size of ICC)
  - Number of clusters
  - Number of patients per cluster

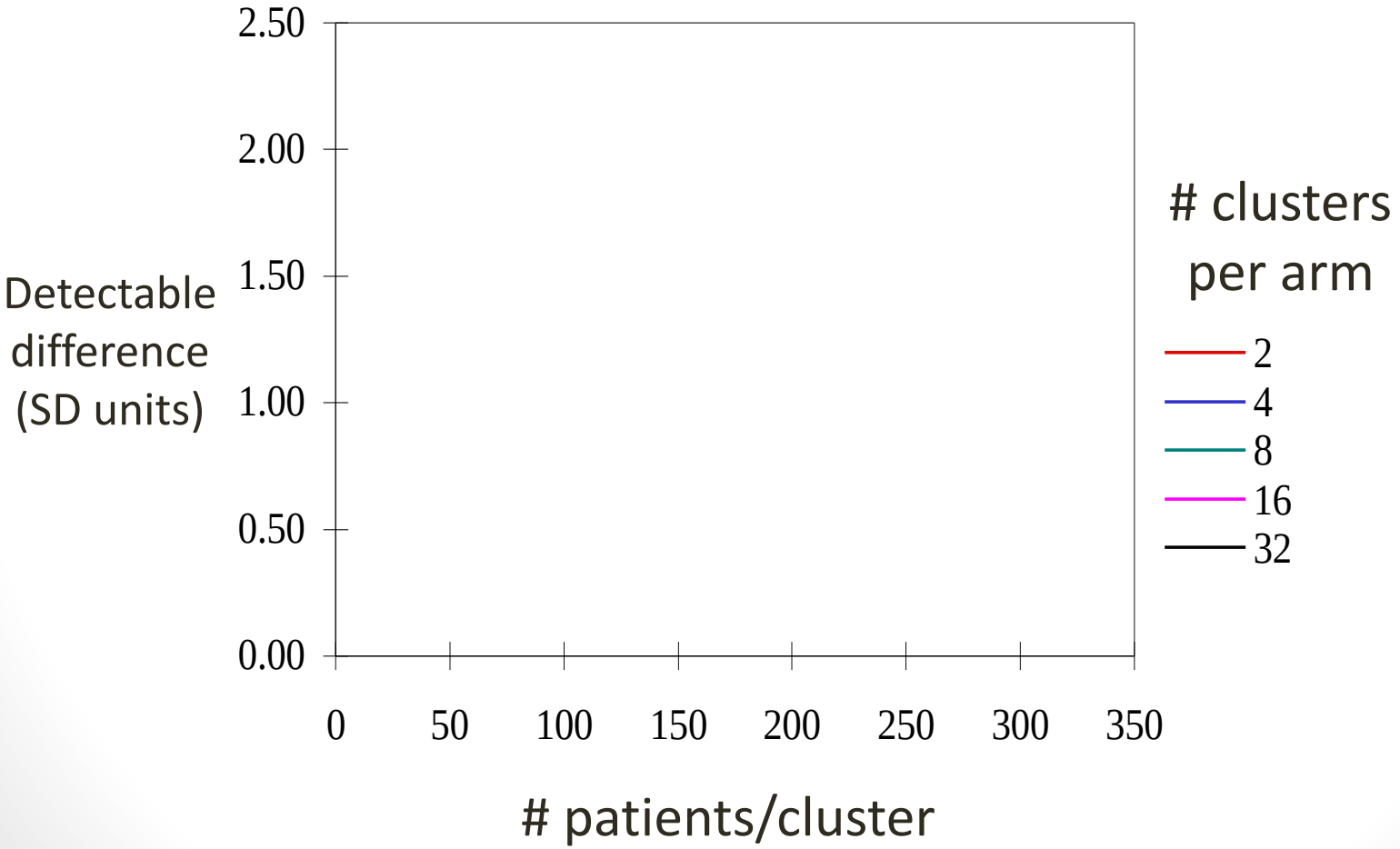
Impact of increasing # clusters

Example: CRT with **ICC=0.1** at fixed alpha & power



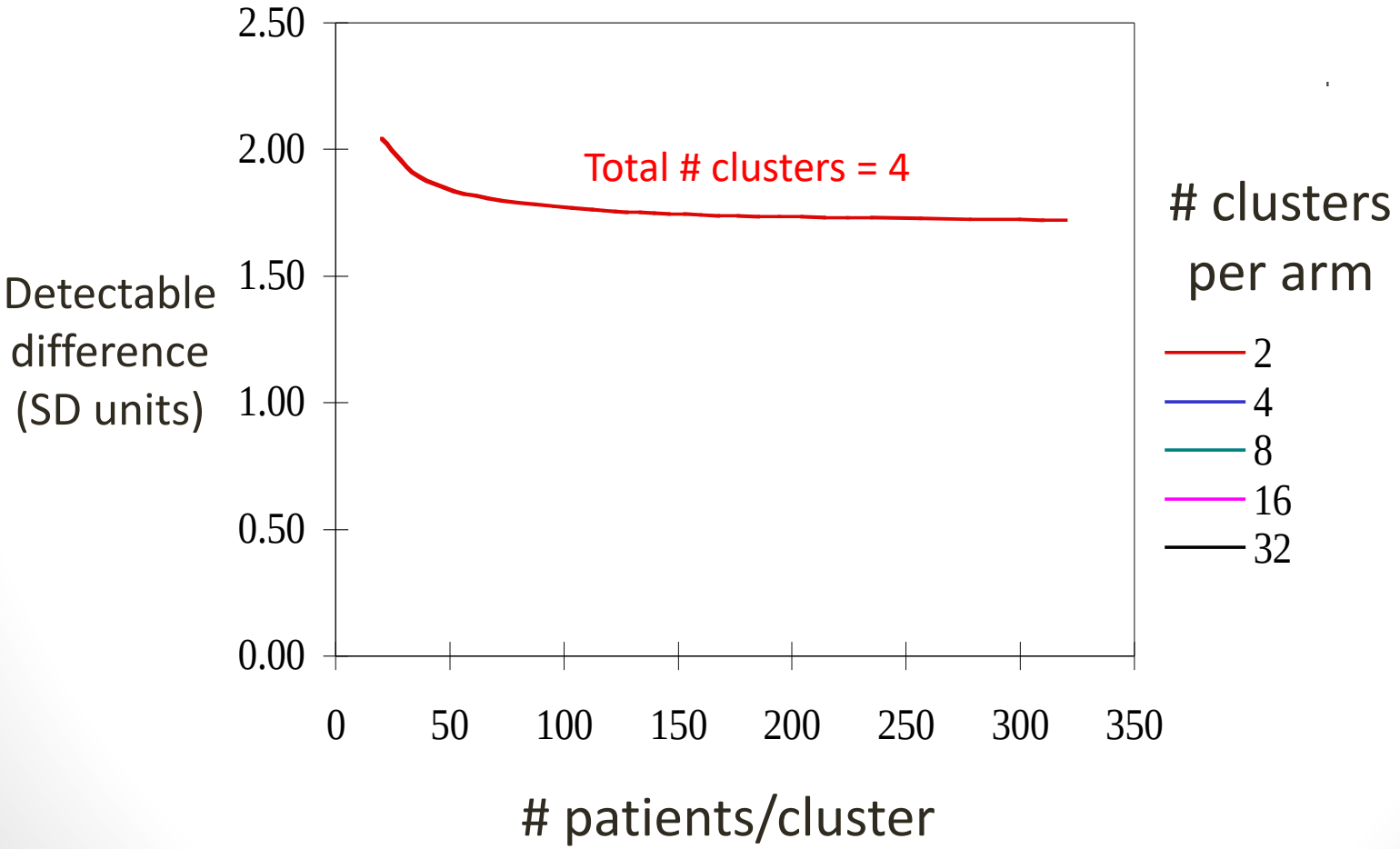
Impact of increasing # clusters

Example: CRT with **ICC=0.1** at fixed alpha & power



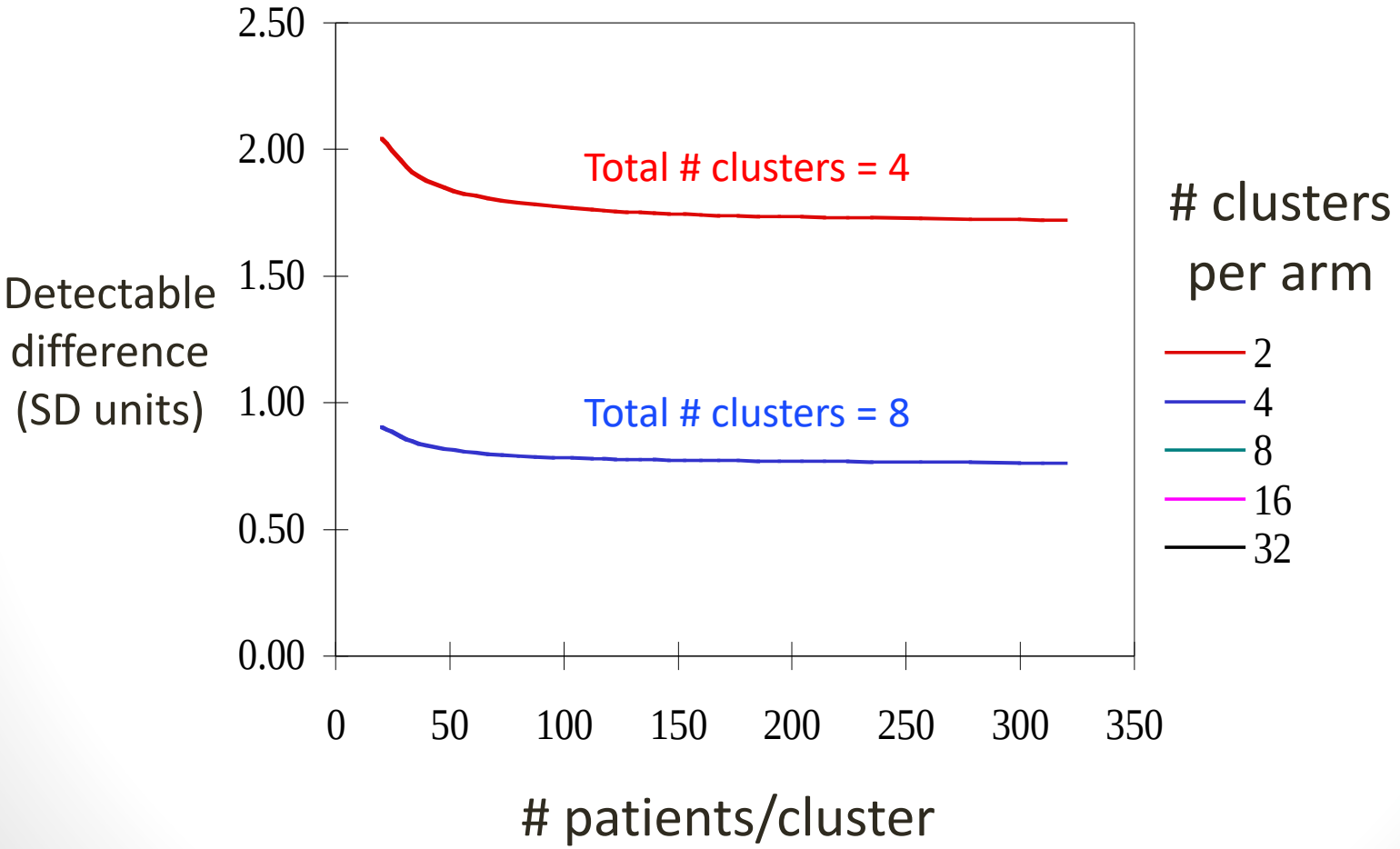
Impact of increasing # clusters

Example: CRT with **ICC=0.1** at fixed alpha & power



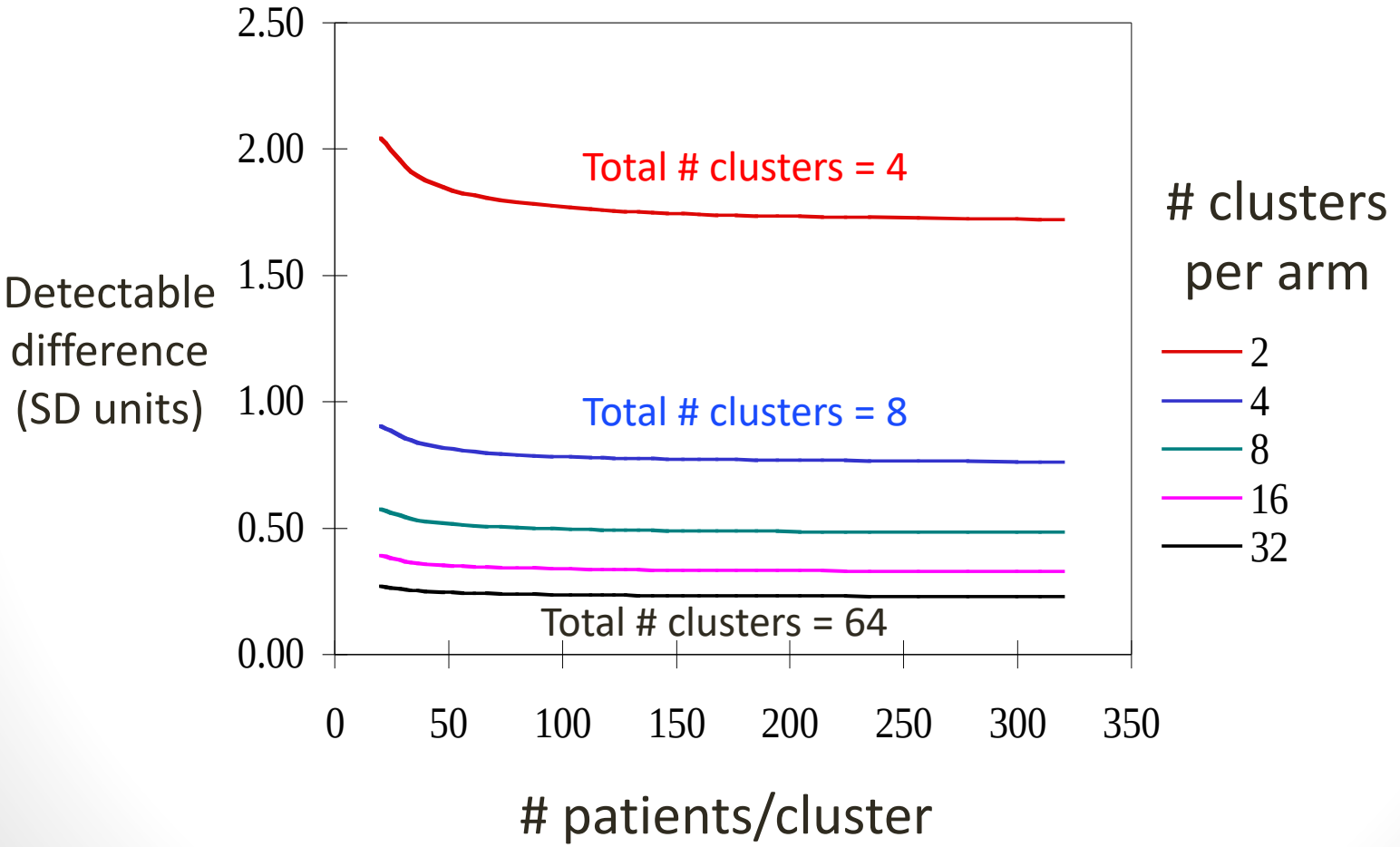
Impact of increasing # clusters

Example: CRT with **ICC=0.1** at fixed alpha & power



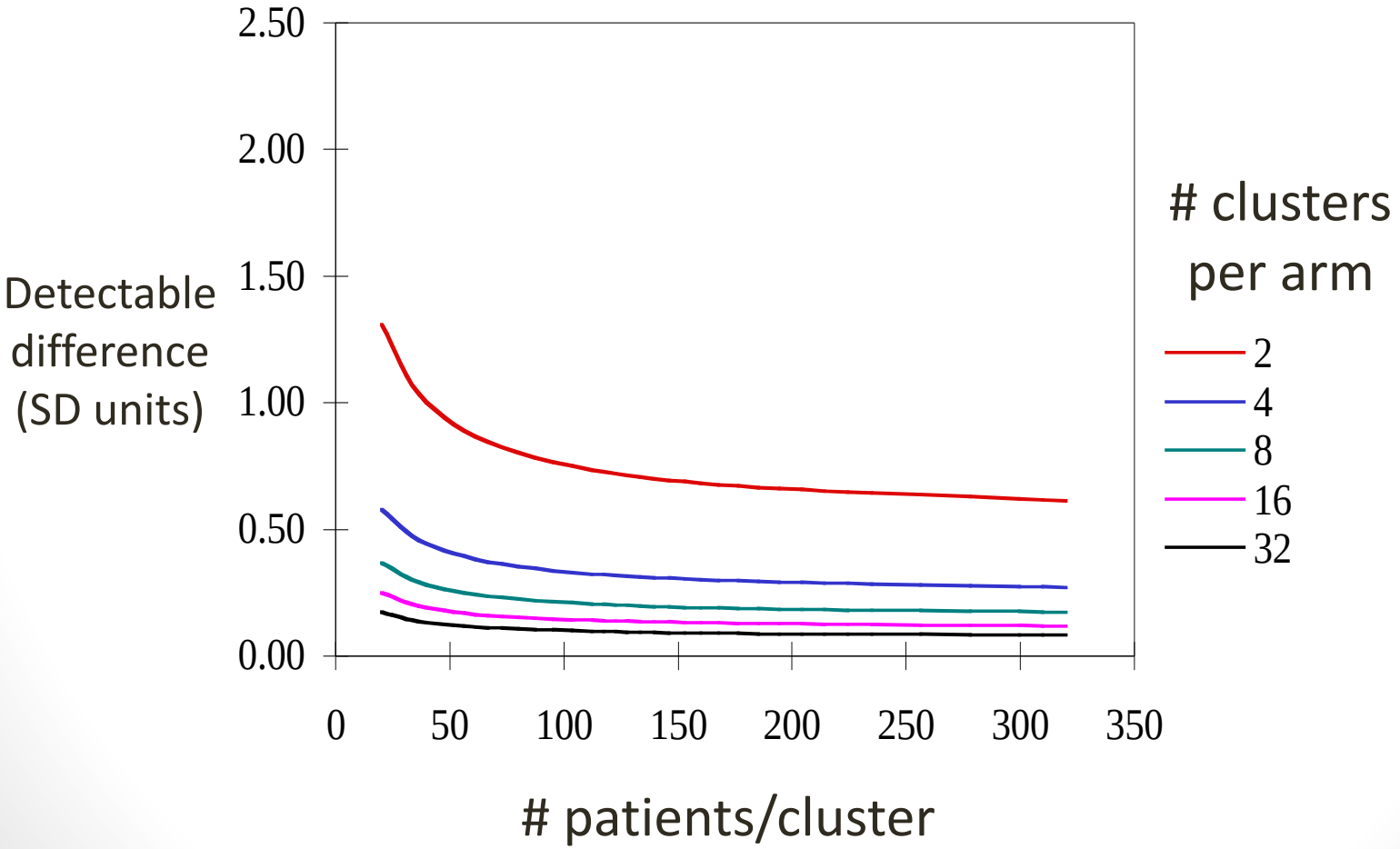
Impact of increasing # clusters

Example: CRT with **ICC=0.1** at fixed alpha & power



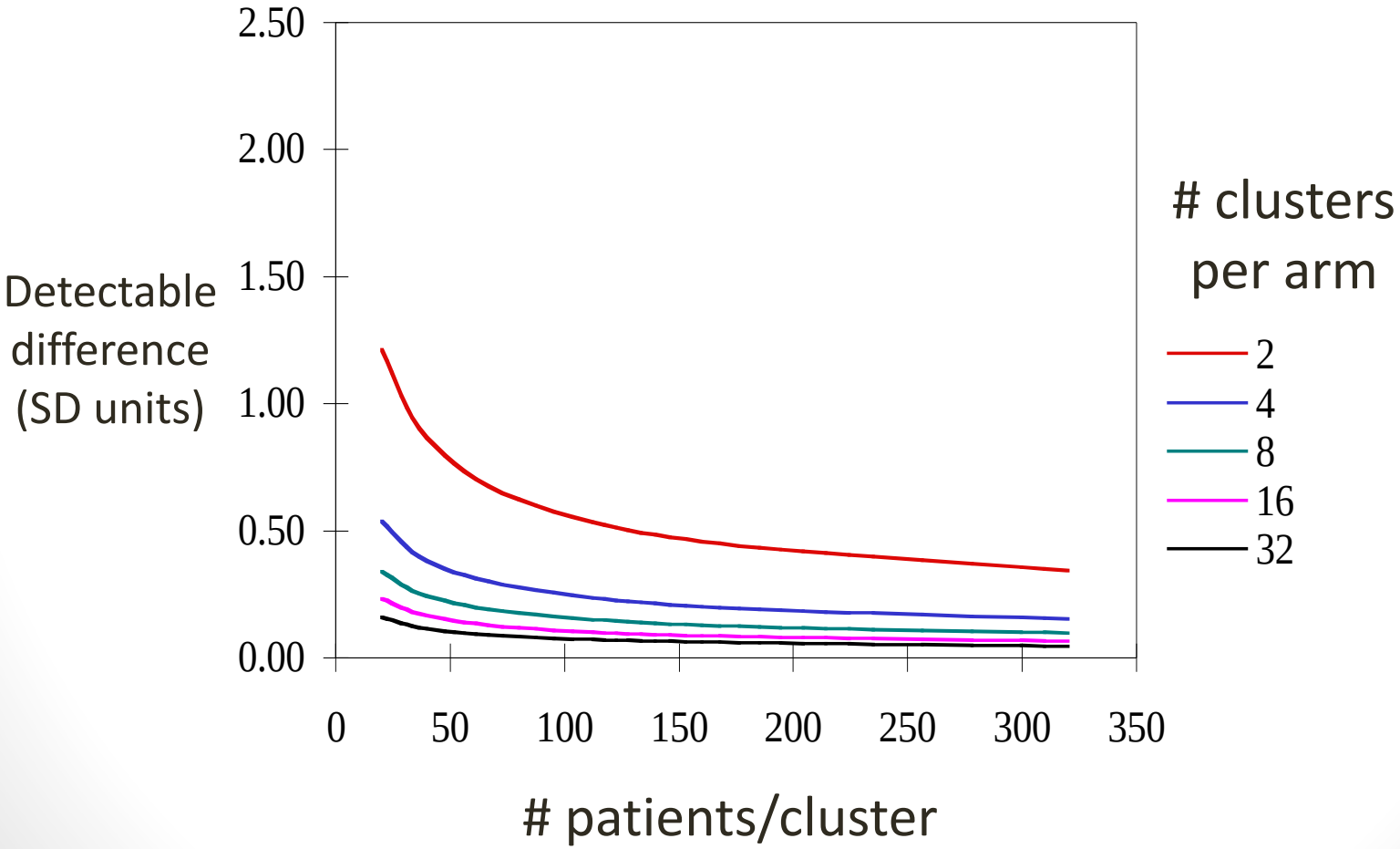
Impact of increasing # clusters

Example: CRT with **smaller ICC=0.01** at at fixed alpha & power



Impact of increasing # clusters/groups

Example: CRT with **even smaller ICC=0.001** at fixed alpha & power



Accounting for Clustering in Design

- Power and sample size for CRT
  - Account for anticipated clustering
  - Inflate RCT sample size
  - Work with statistician to do correctly
- Use ICC for outcome
  - ICC often 0.01-0.05
  - STOP CRC: ICC = 0.03 for primary outcome
  - Depends on outcome and study characteristics
  - Different outcome = different ICC, even in same CRT

Estimating ICC to Plan Study

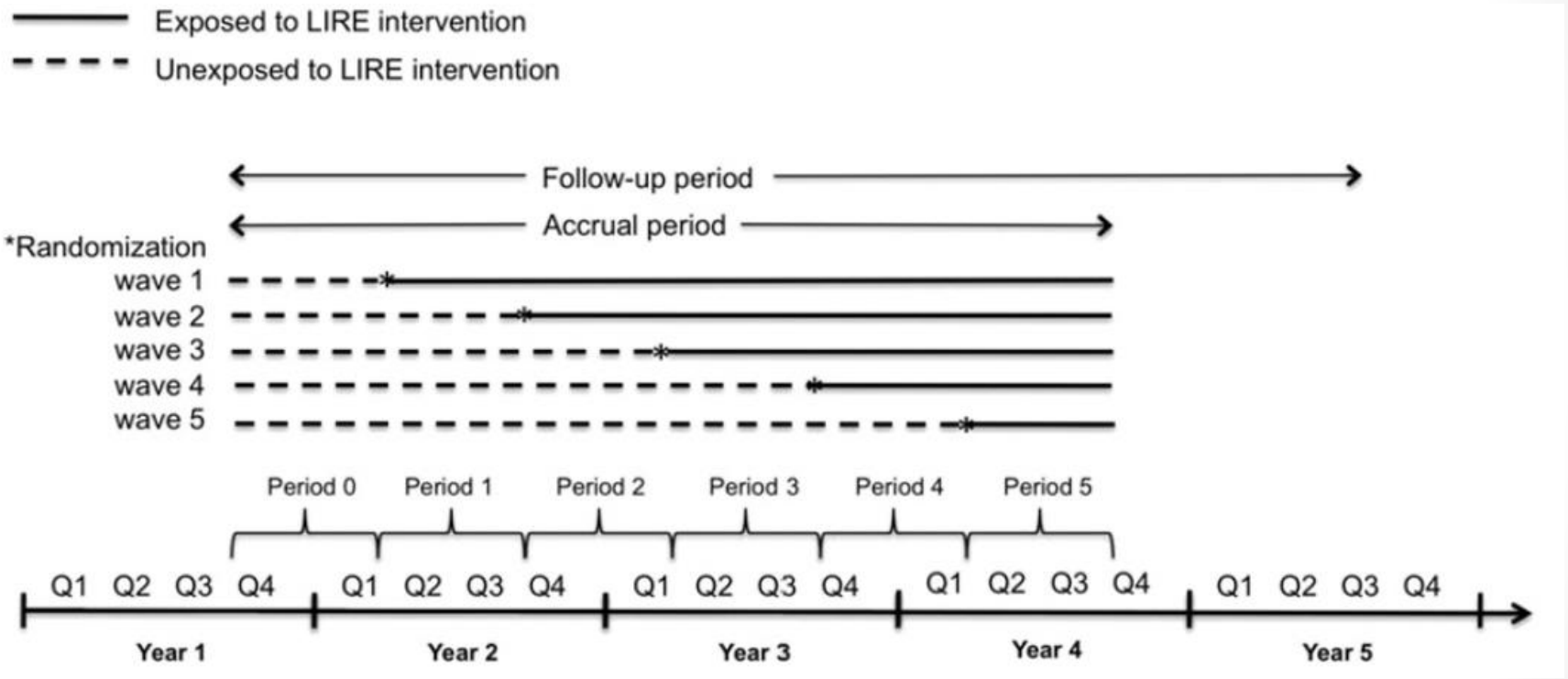
- How to get good estimate of ICC for a particular outcome?
  - Depends on outcome and study characteristics
  - CONSORT statement recommends ICC reported
  - Look at other articles with similar settings
  - Use available EHR data
- Be cautious when using pilot data from small study
  - ICC might have a wide confidence interval

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

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

NIH Collaboratory ePCT: LIRE



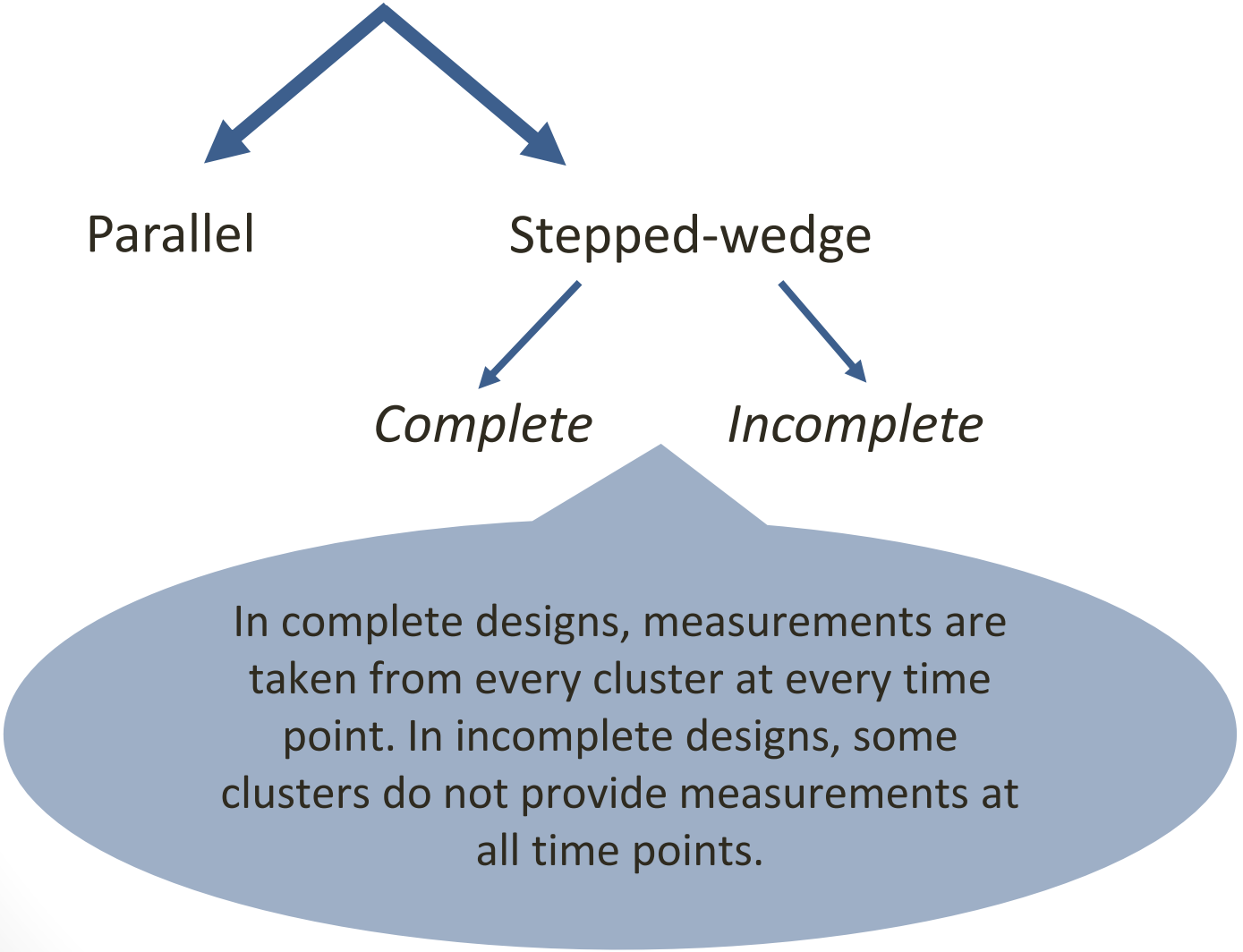
Source: Jarvik JG et al. *Contemp Clin Trials*. 2015;45(Pt B):157-163.

Types of CRT Designs

Parallel

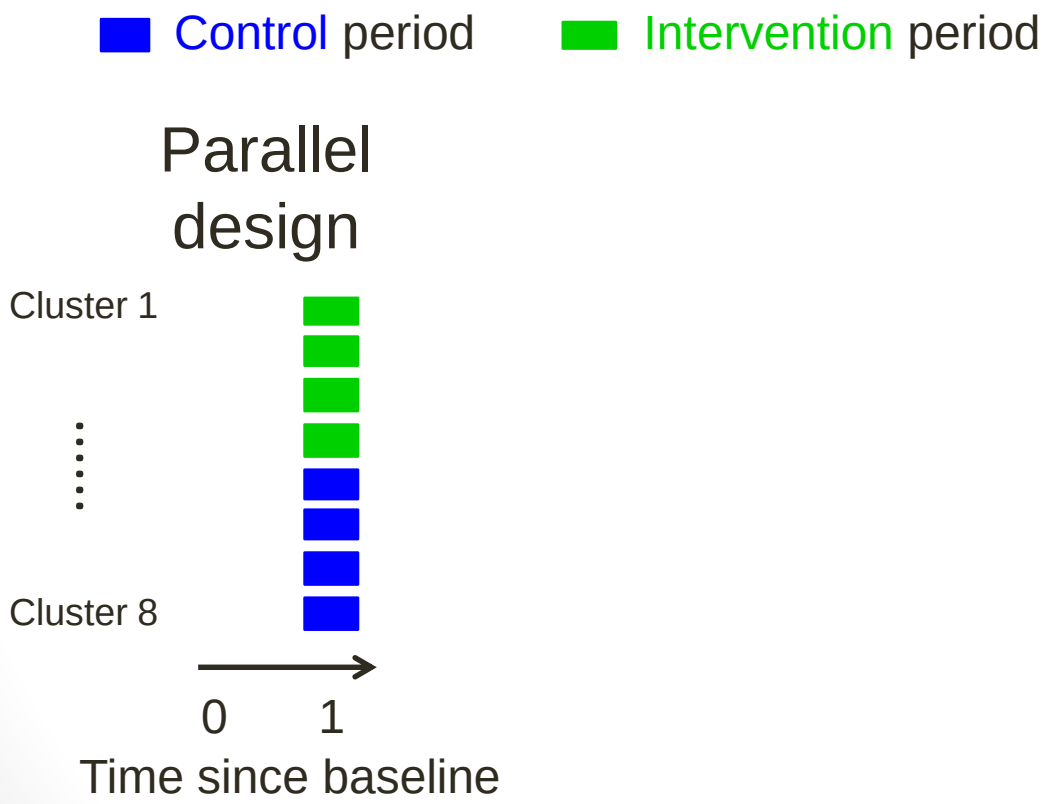
Stepped-wedge

Types of CRT Designs



Types of CRT Designs

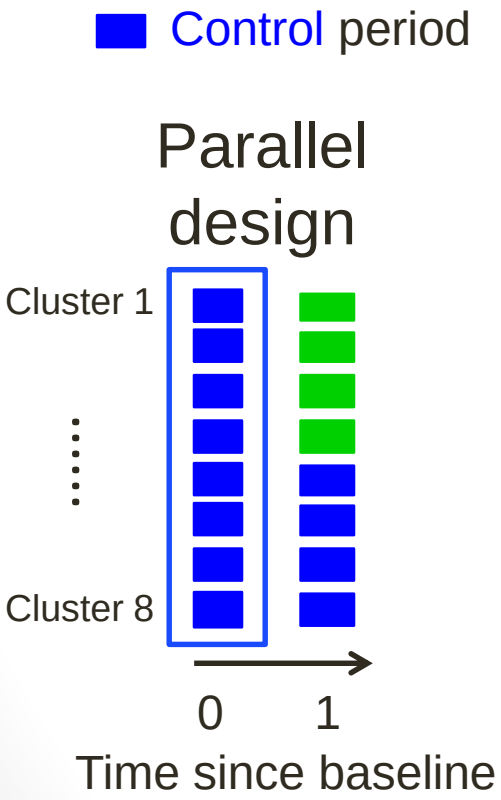
Examples with 8 clusters: 1-year intervention



Based on: Hemming K, Lilford R, Girling AJ. 2015. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. Stat Med. 34:181-196. doi:10.1002/sim.6325. PMID: 25346484

Types of CRT Designs

Examples with 8 clusters: 1-year intervention

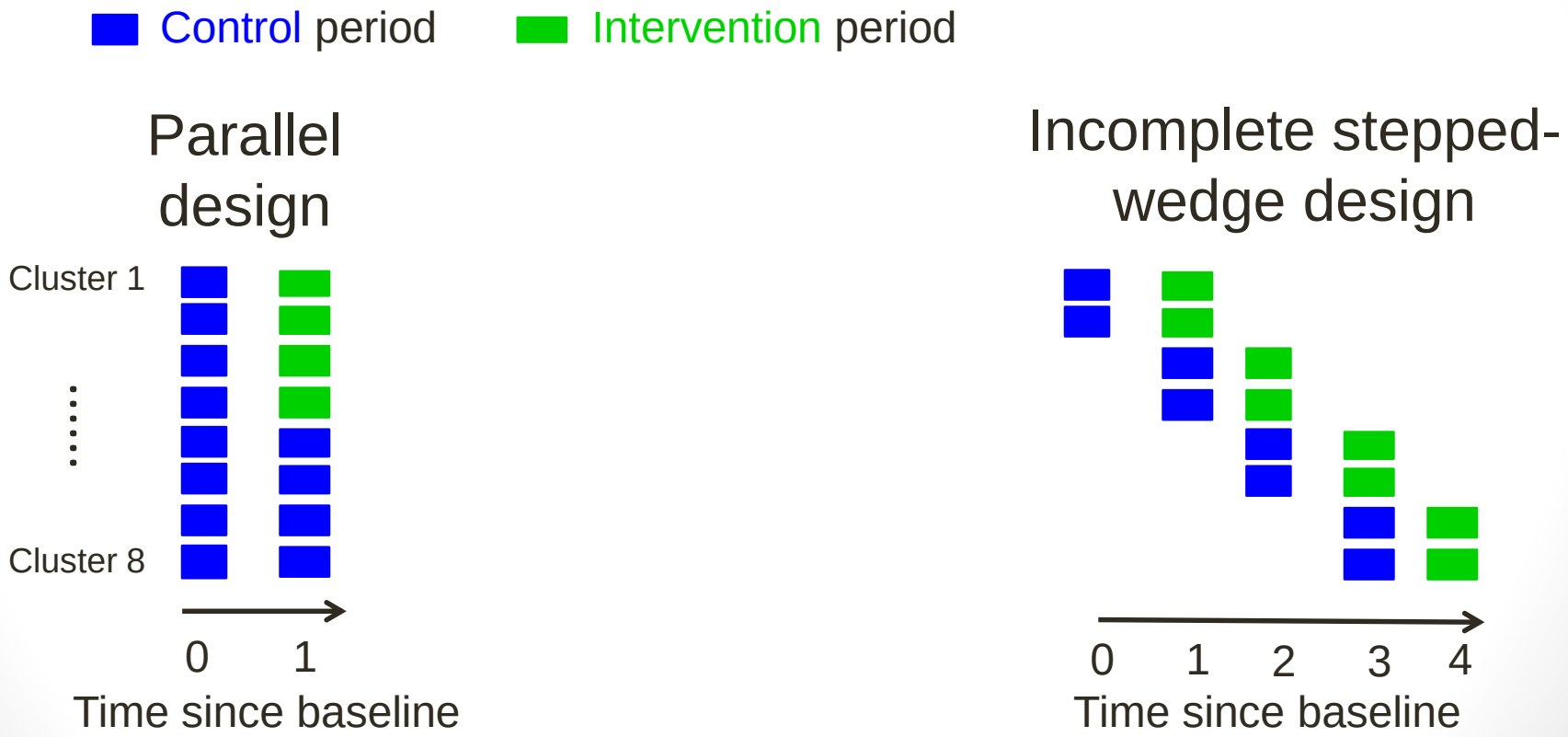


May have baseline outcomes

Based on: Hemming K, Lilford R, Girling AJ. 2015. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. Stat Med. 34:181-196. doi:10.1002/sim.6325. PMID: 25346484

Types of CRT Designs

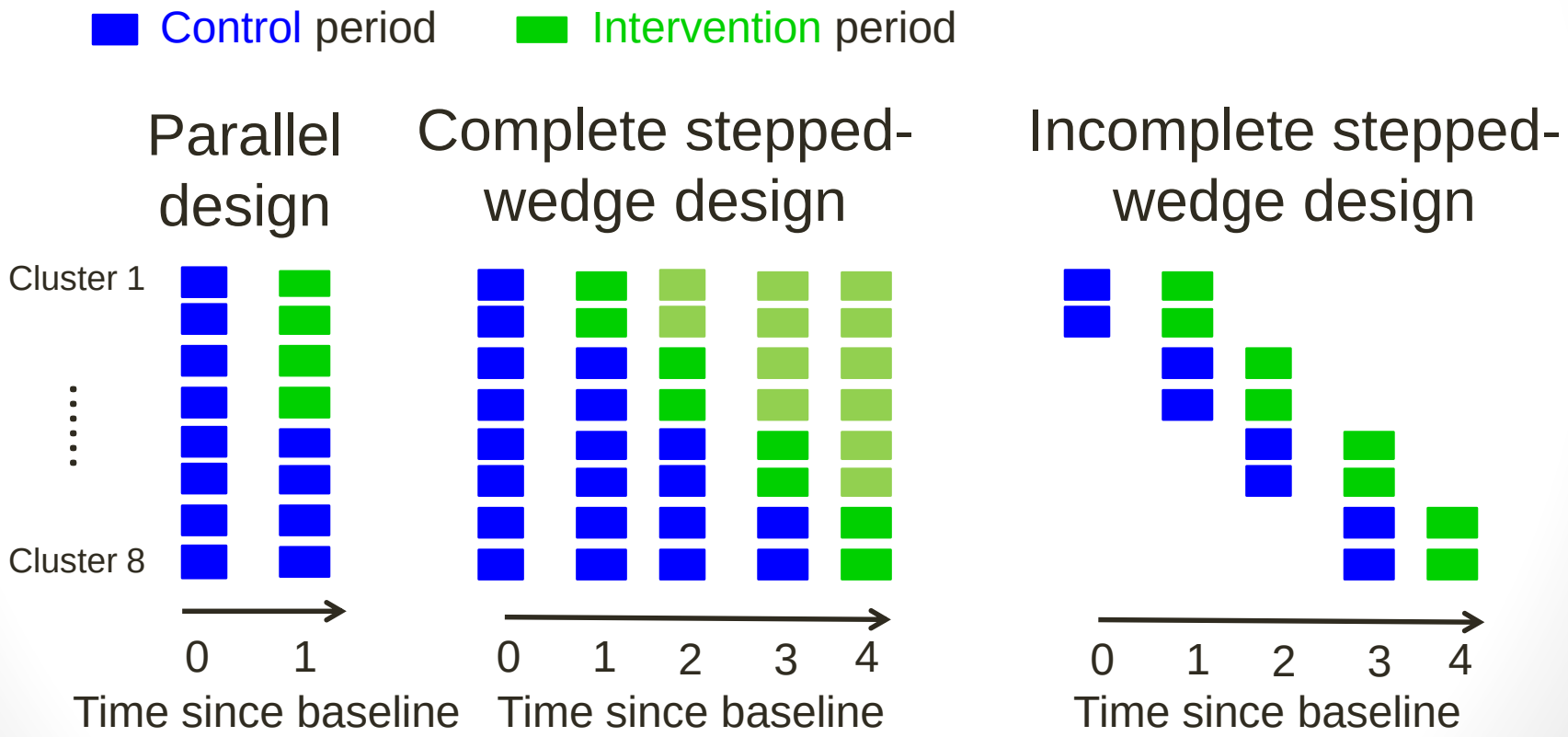
Examples with 8 clusters: 1-year intervention



Based on: Hemming K, Lilford R, Girling AJ. 2015. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. Stat Med. 34:181-196. doi:10.1002/sim.6325. PMID: 25346484

Types of CRT Designs

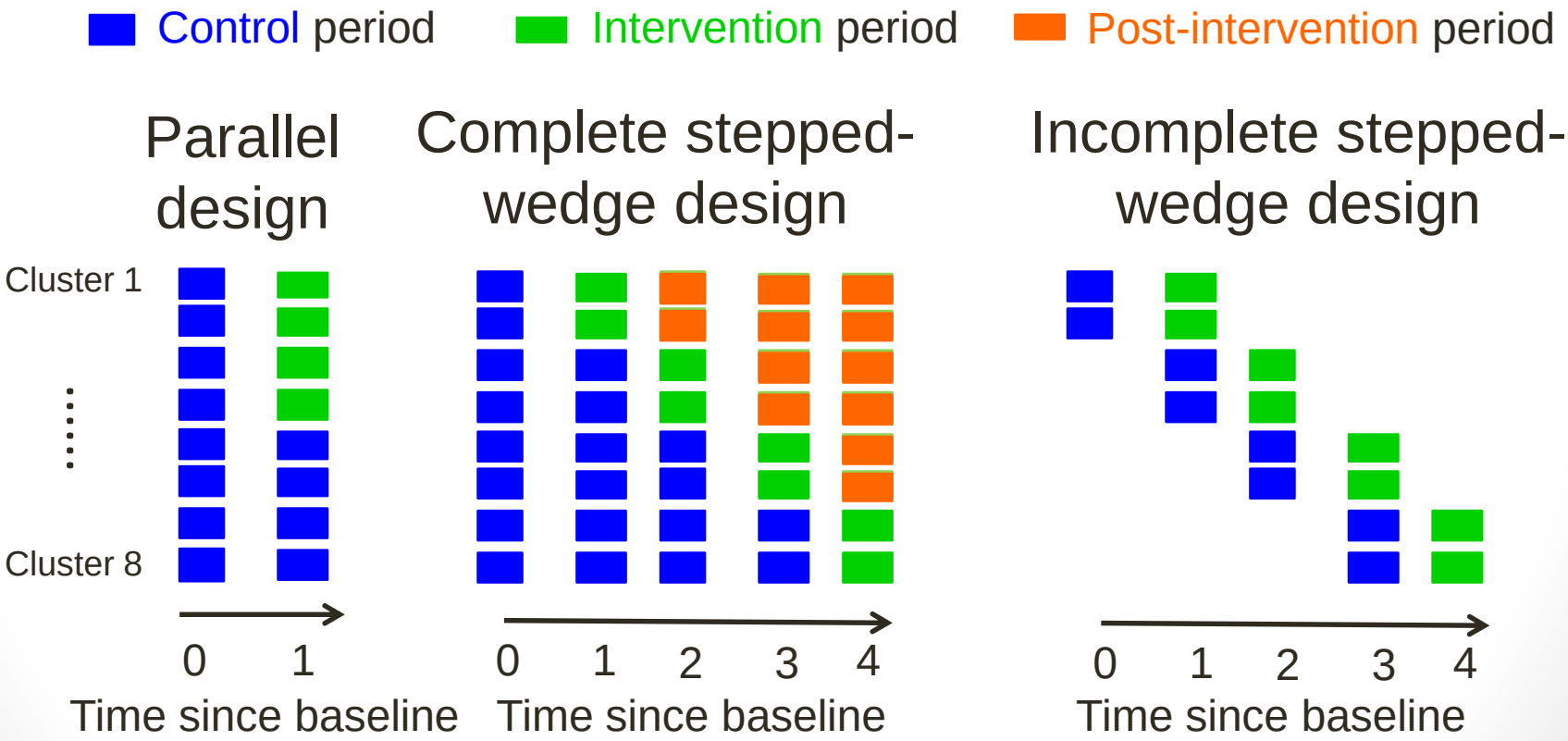
Examples with 8 clusters: 1-year intervention



Based on: Hemming K, Lilford R, Girling AJ. 2015. Stepped-wedge cluster randomised controlled trials: a generic framework including parallel and multiple-level designs. Stat Med. 34:181-196. doi:10.1002/sim.6325. PMID: 25346484

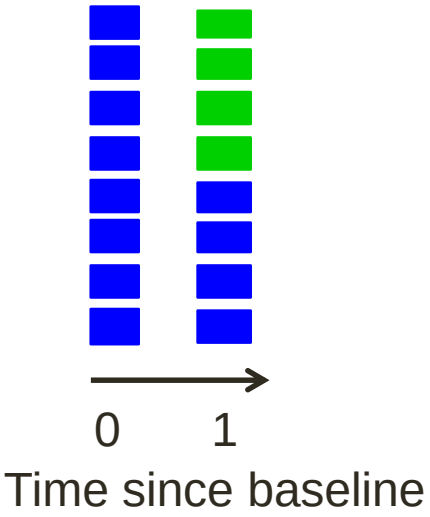
Types of CRT Designs

Examples with 8 clusters: 1-year intervention

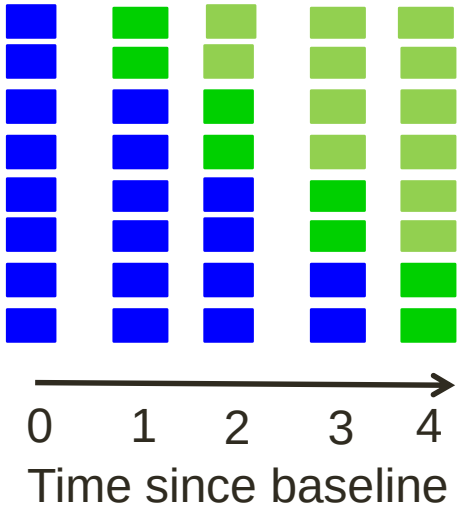


CRT Analysis: Treatment Effects

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



Parallel design

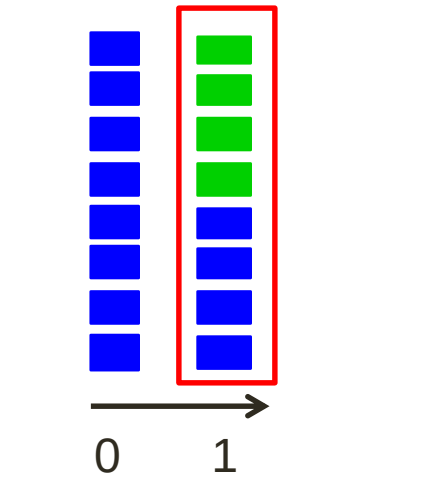


Complete SW design

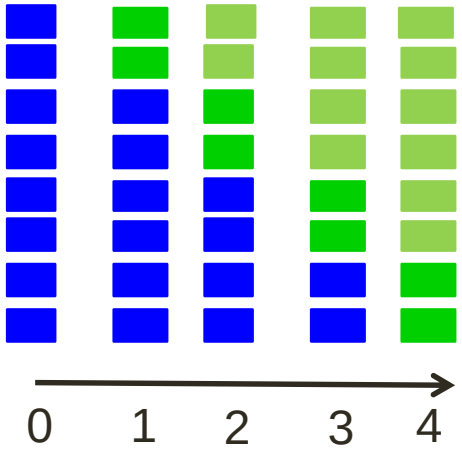
Control period Intervention period

CRT Analysis: Treatment Effects

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



Parallel design

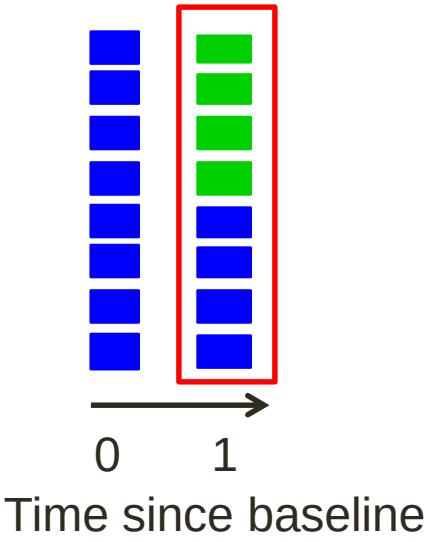


Complete SW design

Control period Intervention period

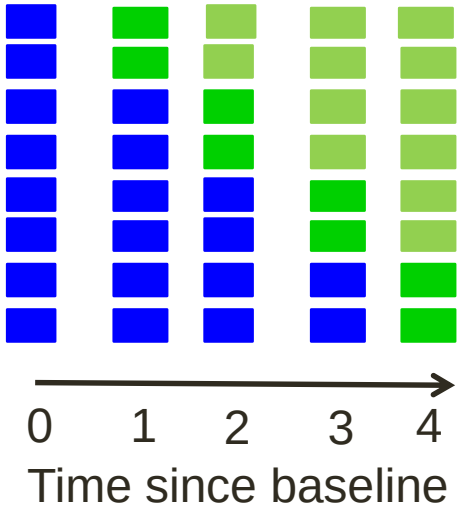
CRT Analysis: Treatment Effects

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



Parallel design

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

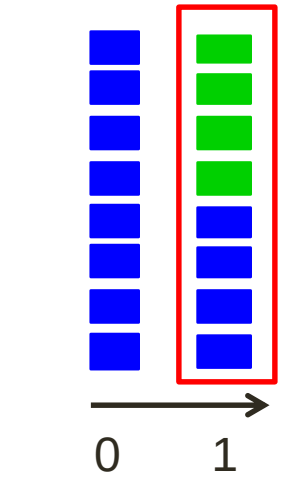


Complete SW design

Control period Intervention period

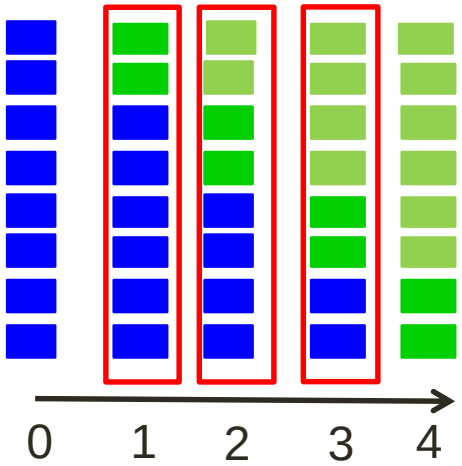
CRT Analysis: Treatment Effects

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



Parallel design

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

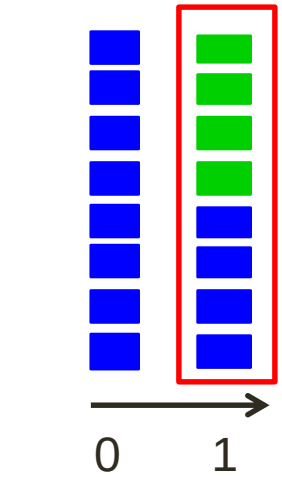


Complete SW design

Control period Intervention period

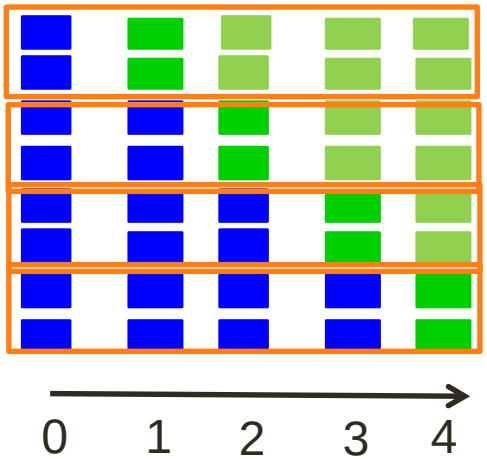
CRT Analysis: Treatment Effects

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



Parallel design

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

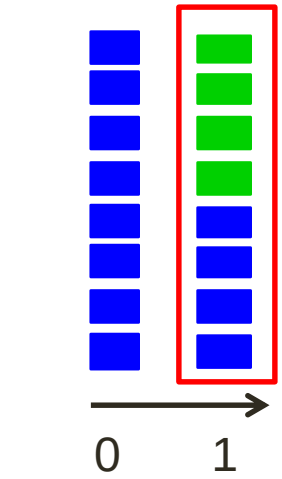


Complete SW design

Control period Intervention period

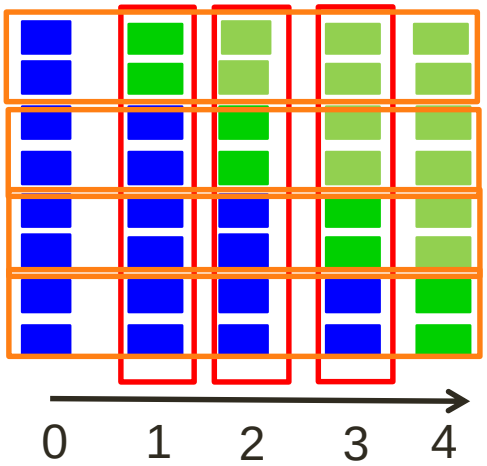
CRT Analysis: Treatment Effects

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



Parallel design

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



Complete SW design

Control period Intervention period

Choosing the Right Type of CRT

- Arguments **for** stepped-wedge CRT:
  - Cannot immediately implement intervention in 1/2 clusters
  - Pragmatic research: eventually implement in all clusters
  - Have few clusters and might gain power

Choosing the Right Type of CRT

- Arguments **for** stepped-wedge CRT:
  - Cannot immediately implement intervention in 1/2 clusters
  - Pragmatic research: eventually implement in all clusters
  - Have few clusters and might gain power
- Arguments **against** stepped-wedge CRT:
  - Risk confounding treatment effect with time effect
  - Risk of interruption or external events that could affect the outcome (eg, a pandemic!)

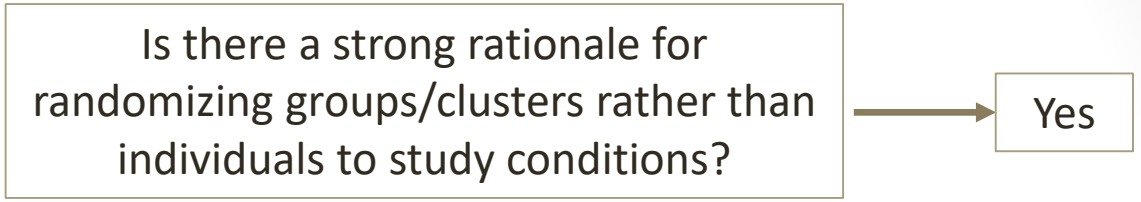
Recommendations for CRT Design

- Use a parallel CRT design if you can
- If stepped-wedge, plan for time effects in design & analysis
- Work with statistician to account for clustering in design and analysis of both designs

Choosing Study Design

Is there a strong rationale for randomizing groups/clusters rather than individuals to study conditions?

Choosing Study Design



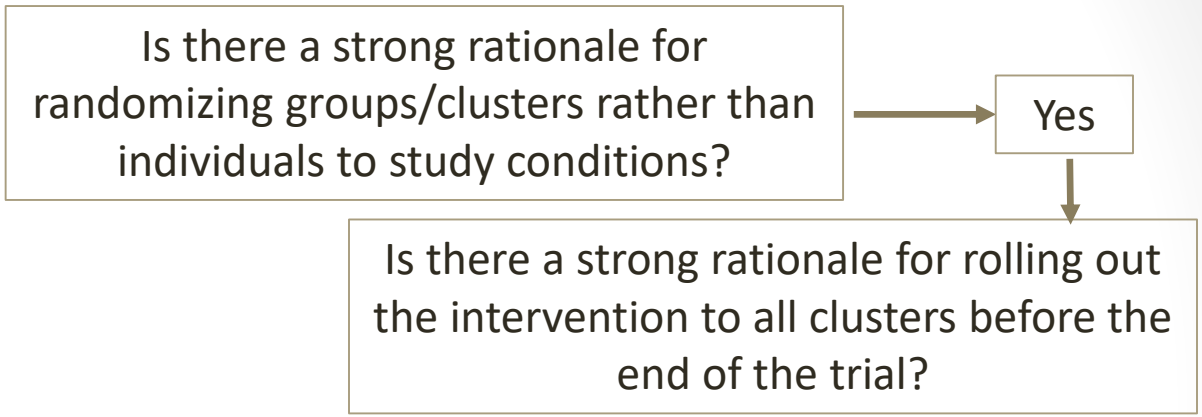
Choosing Study Design

Is there a strong rationale for randomizing groups/clusters rather than individuals to study conditions? → Yes

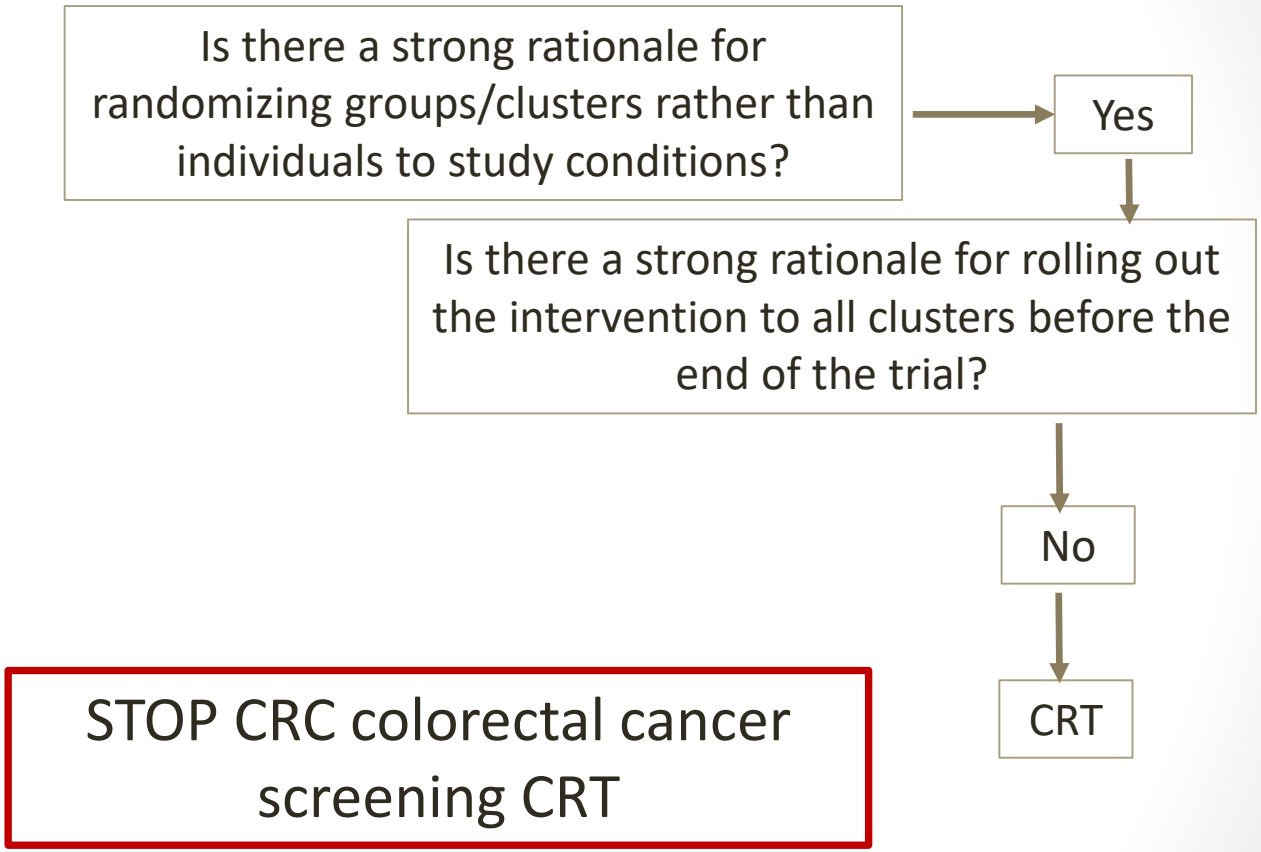
Examples with clinic/health-system-level interventions:

- STOP CRC colorectal cancer screening CRT
- LIRE lumbar imaging trial SW-CRT

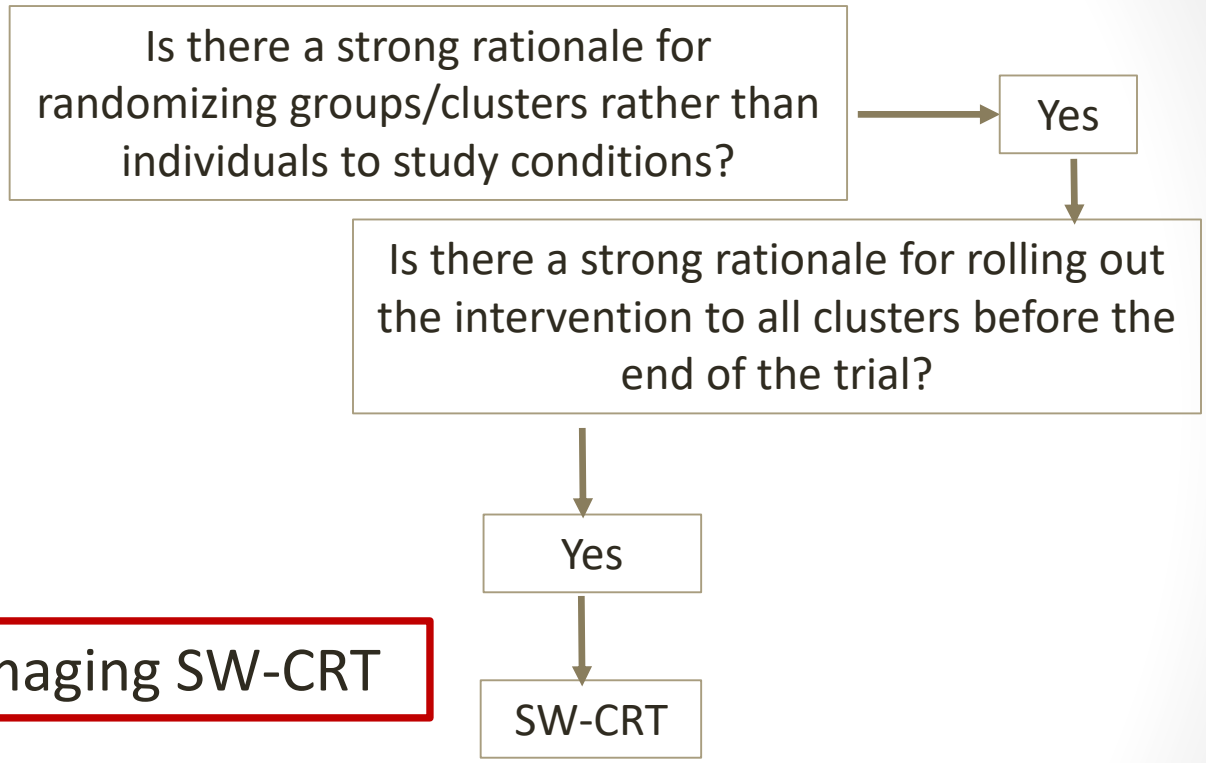
Choosing Study Design



Choosing Study Design

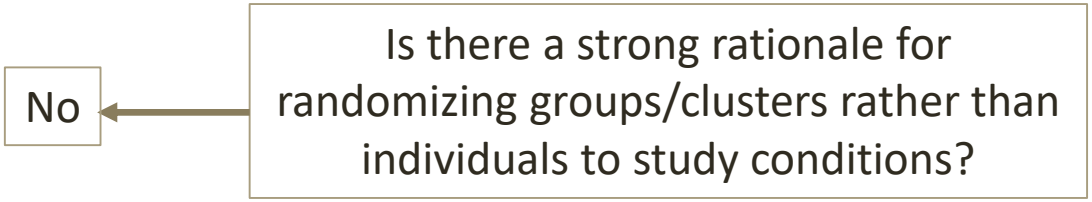


Choosing Study Design

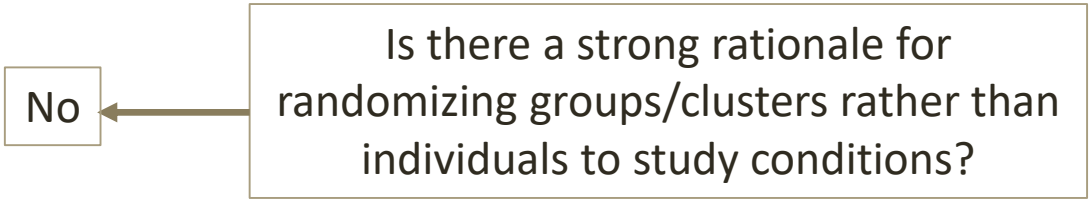


LIRE lumbar imaging SW-CRT

Choosing Study Design



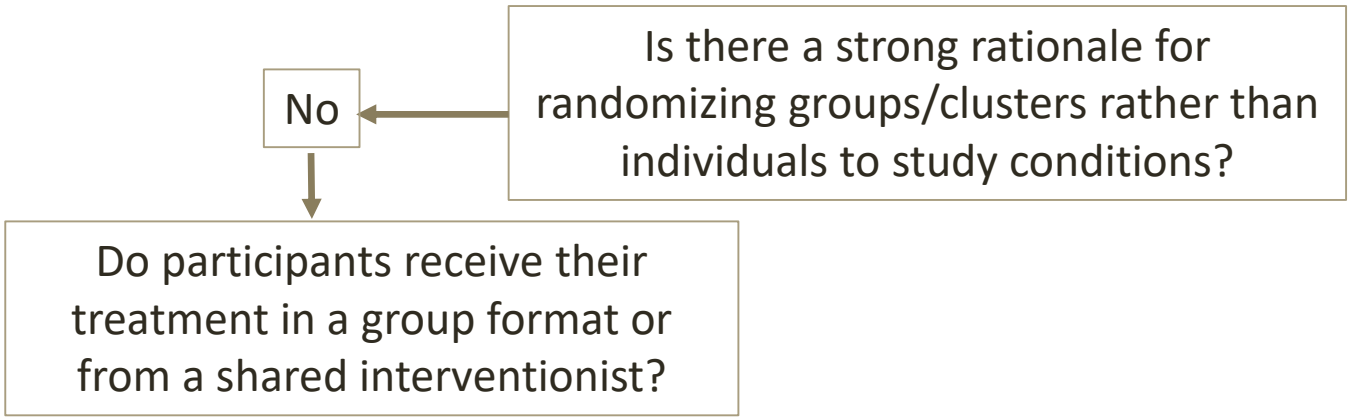
Choosing Study Design



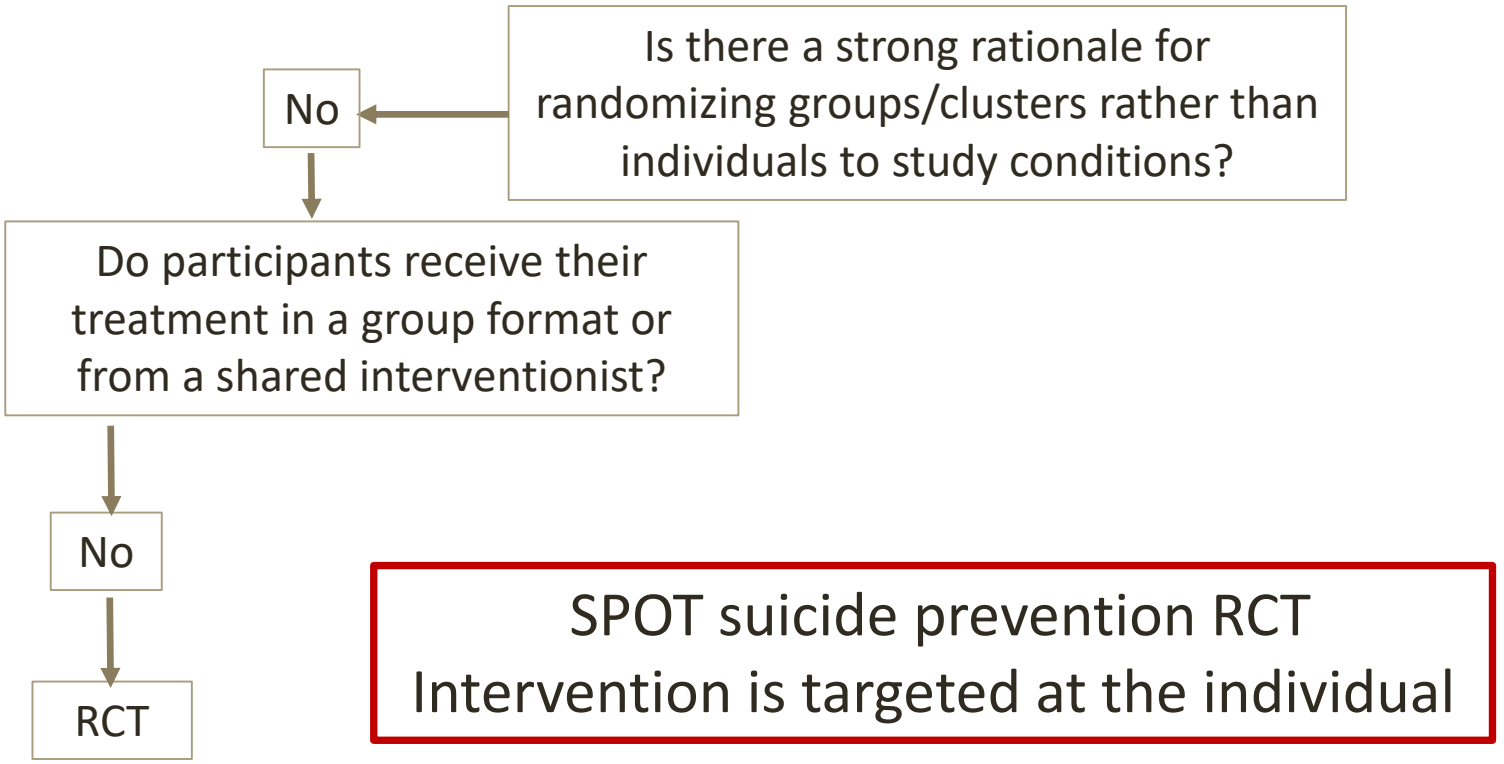
Examples with individual-level randomization:

- SPOT suicide prevention RCT
- OPTIMUM mindfulness for back-pain RCT

Choosing Study Design

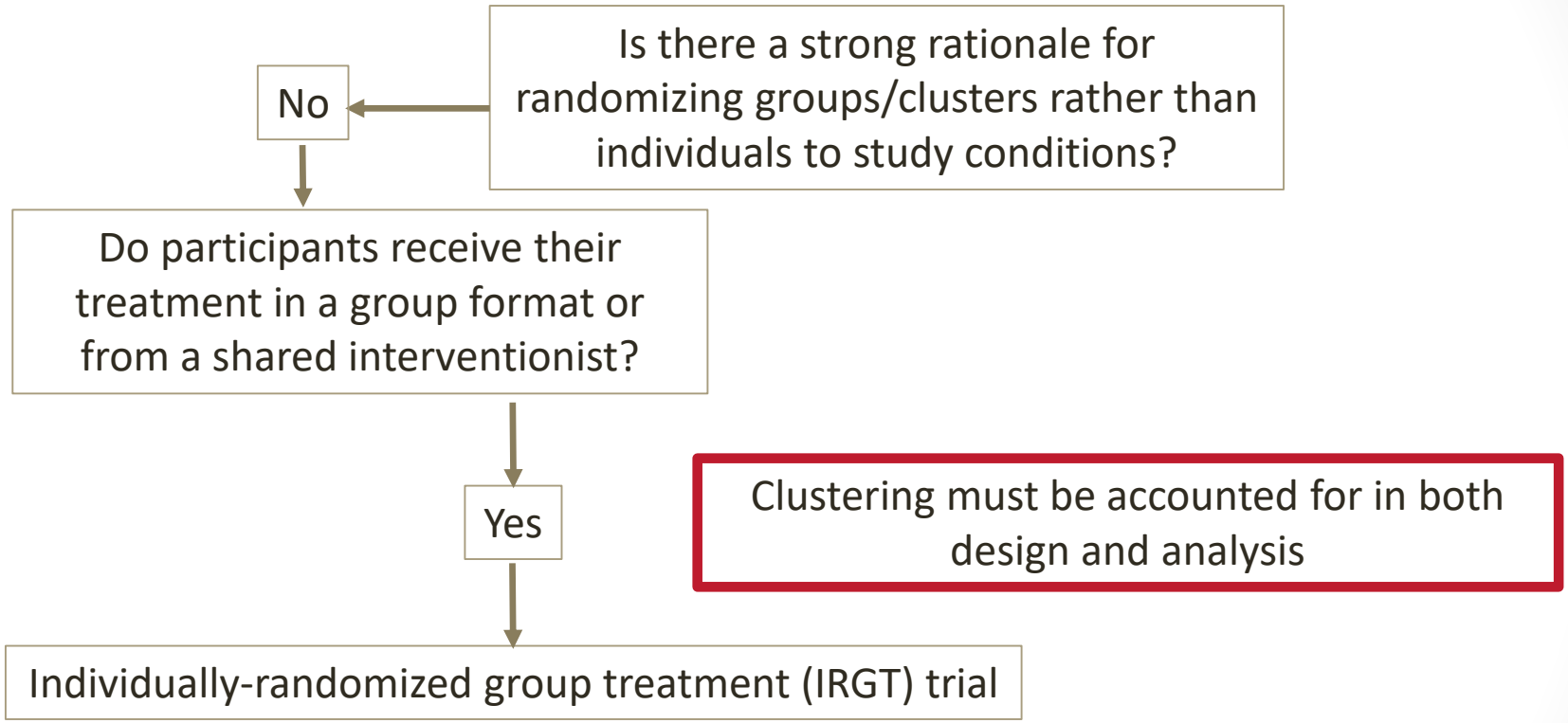


Choosing Study Design

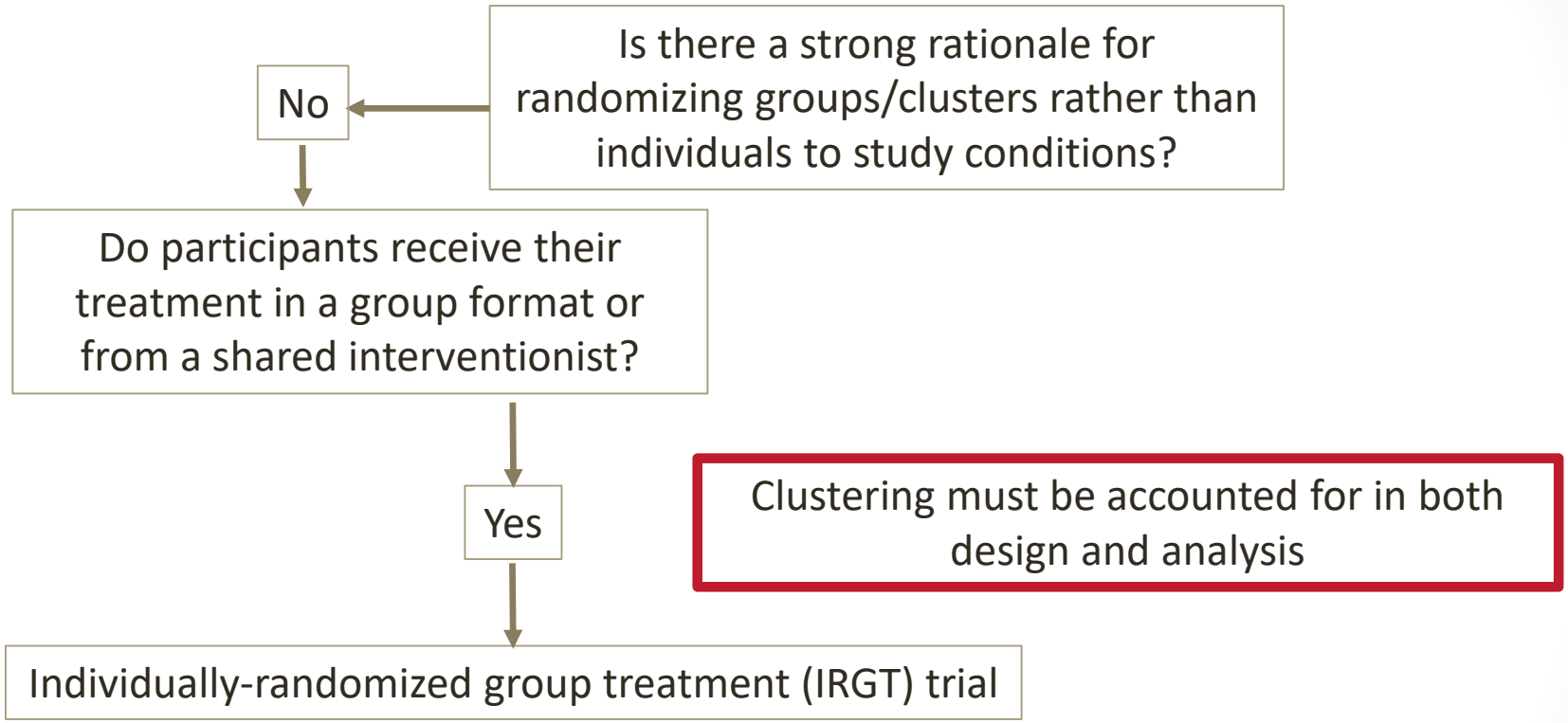


SPOT suicide prevention RCT
Intervention is targeted at the individual

Choosing Study Design



Choosing Study Design

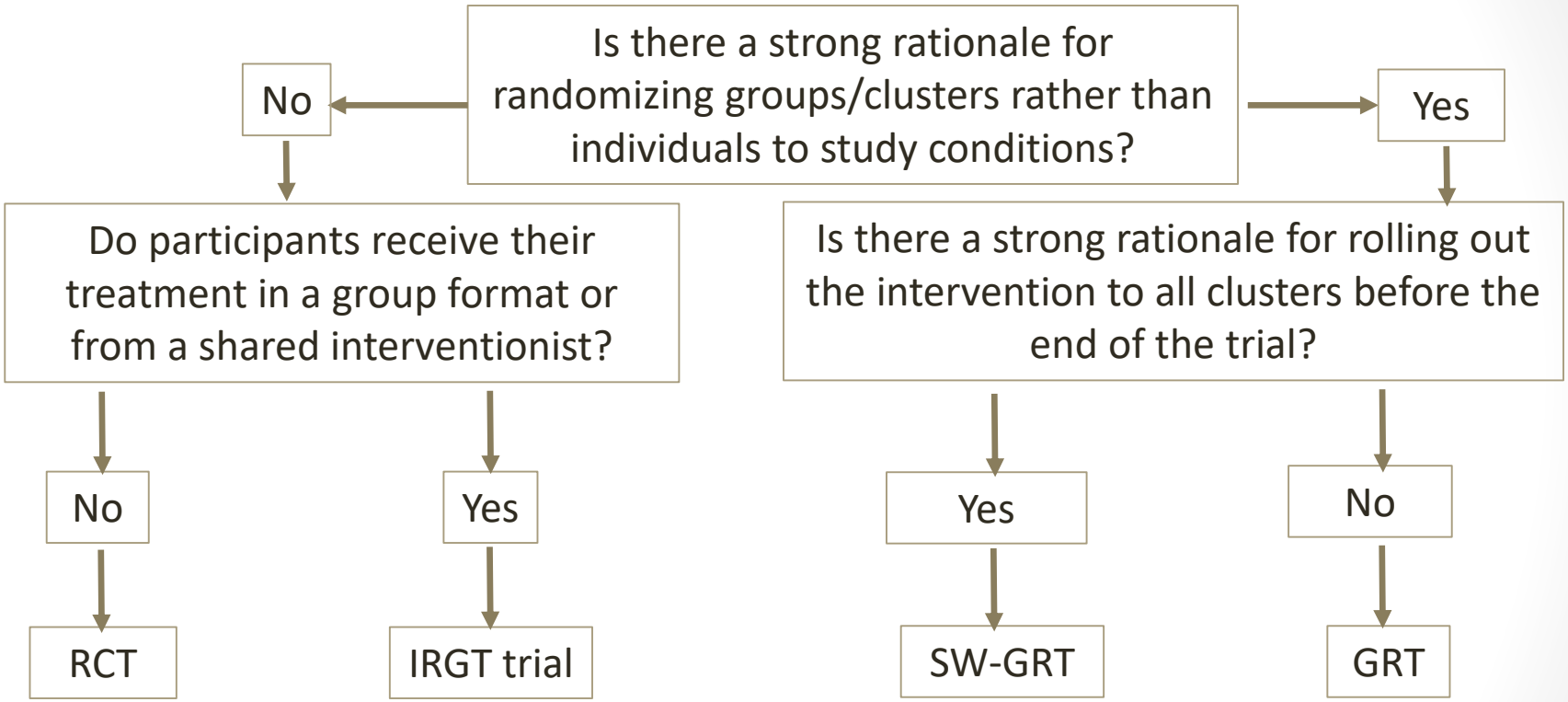


OPTIMUM mindfulness for back-pain RCT
Intervention is group-based

NIH Collaboratory ePCT: OPTIMUM

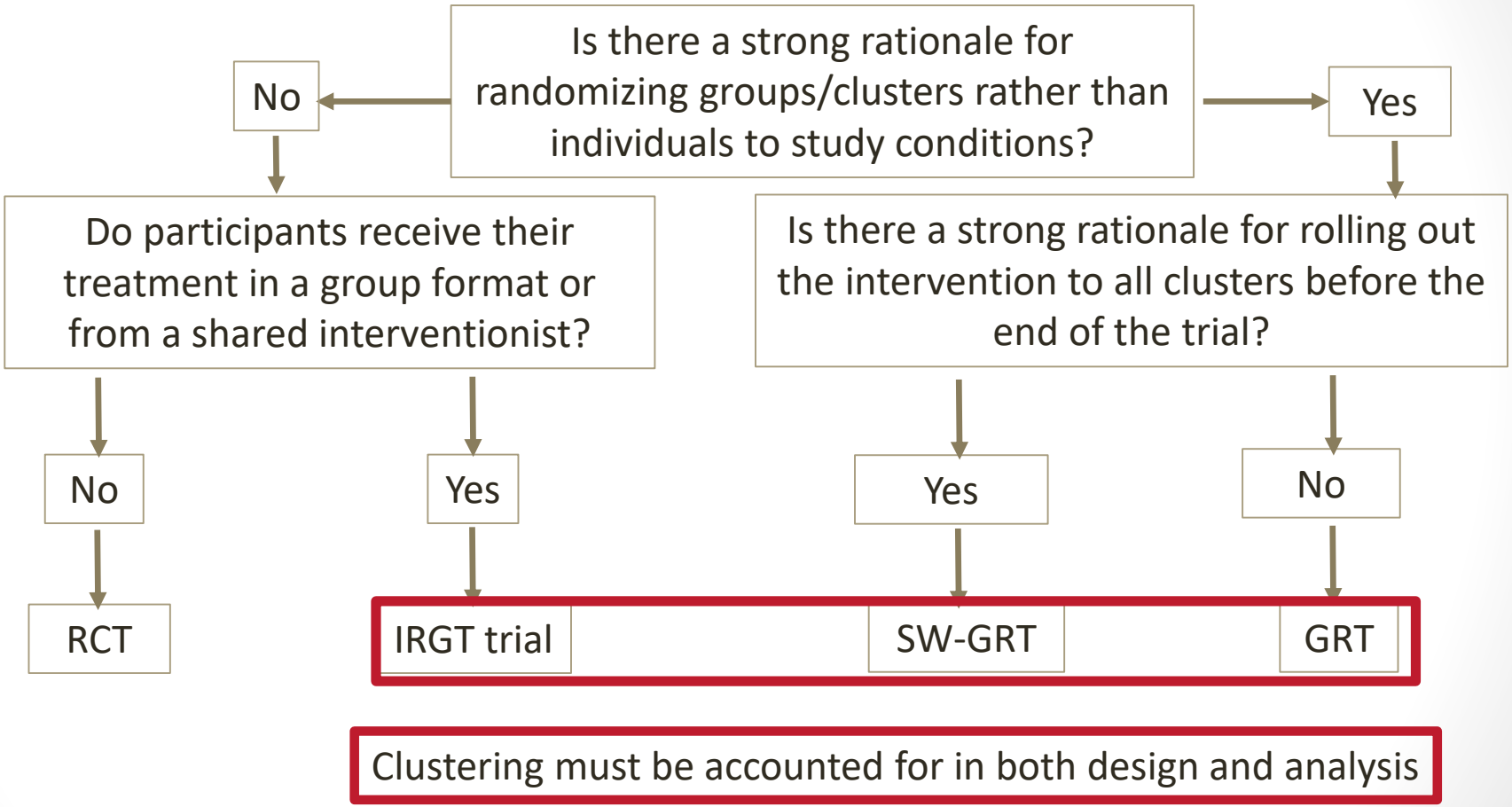
- OPTIMUM: optimizing pain treatment in medical settings using group-based mindfulness
- ~450 patients across 3 clinical sites
- Two-arm RCT
  - Intervention vs usual care
- Unit of randomization: individual
- Group-based intervention
  - Clustering of outcomes in intervention arm
  - Must be accounted for in both design and analysis
- "*Individually randomized group treatment (IRGT) trial*"

Choosing Study Design



See Figure: Murray DM, Taljaard M, Turner EL, George SM, Ann Rev Pub Health 2020. 41:1-19

Choosing Study Design



See Figure: Murray DM, Taljaard M, Turner EL, George SM, Ann Rev Pub Health 2020. 41:1-19

Important Things to Know

- Question drives design, design drives analysis
- Randomization
  - Individual-level preferred for statistical reasons
  - But cluster randomization often needed
- Account for clustering in design and analysis of:
  - CRT
  - IRGT trial
- Good design is difficult but critical
  - Need input from diverse team, including statistician
  - Analysis may not be able to overcome design flaws

Important Things to Do

- Focus on the research question
- Select design features with analysis in mind
- Collaborate early with a statistician
- Choose individual randomization, but only if possible
- Weigh statistical choices vs implementation challenges
- Write and publish a protocol paper

In the Living Textbook

EXPERIMENTAL DESIGNS AND RANDOMIZATION SCHEMES

1. Introduction
2. Statistical Design Considerations
3. Cluster Randomized Trials
4. Randomization Methods
5. Choosing Between Cluster and Individual Randomization
6. Alternative Cluster Randomized Designs
7. Concealment and Blinding
8. Designing to Avoid Identification Bias
9. Additional Resources

ANALYSIS PLAN

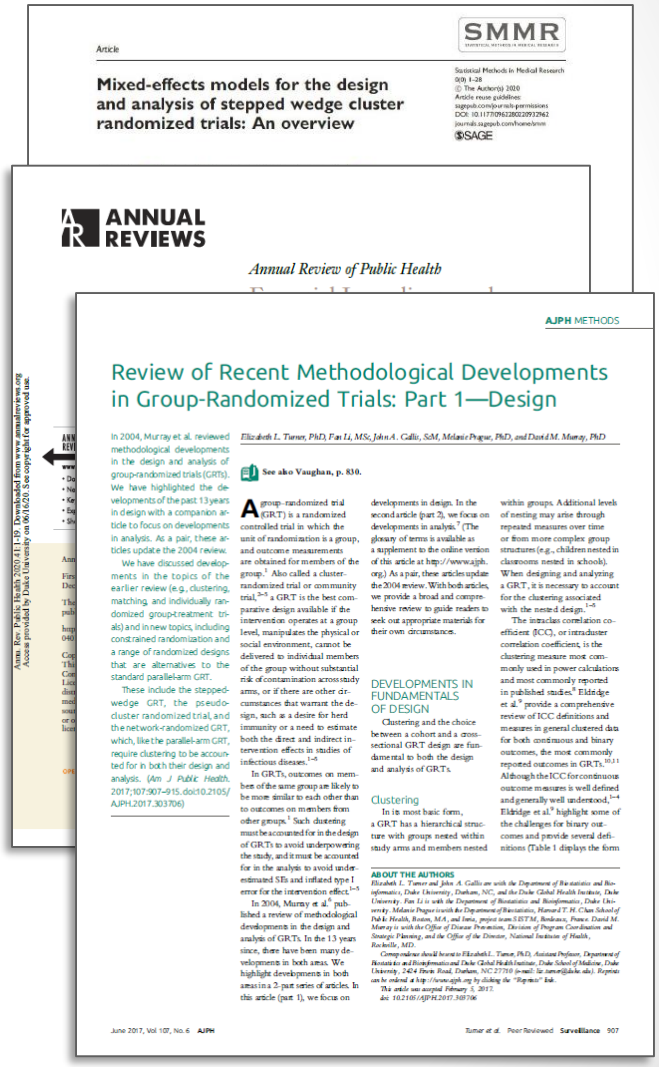
1. Introduction
2. Intraclass Correlation
3. Unequal Cluster Sizes
4. Accounting for Residual Confounding in the Analysis
5. Missing Data and Intention-to-Treat Analyses
6. EHR Data Extraction
7. Unanticipated Changes
8. Case Study: STOP CRC Trial

Summary

- Focus of this talk: demystifying design-related issues for embedded pragmatic clinical trials (ePCTs)
- Context: NIH Collaboratory–funded studies
- Three kinds of randomized trials
  - Randomized controlled trial (RCT)
  - Cluster randomized trial (CRT)
    - Parallel vs stepped-wedge
  - Individually randomized group treatment (IRGT) trial
- How to select amongst these designs?
- Other brief topics: clustering, power, and analytical issues

Design and Analysis Methods

- Turner EL et al. Review of recent methodological developments in group-randomized trials: part 1-design. *Am J Public Health*. 2017;107(6):907-915.
- Turner EL et al. Review of recent methodological developments in group-randomized trials: part 2-analysis. *Am J Public Health*. 2017;107(7):1078-1086.
- Murray DM et al. Essential ingredients and innovations in the design and analysis of group-randomized trials. *Annu Rev Public Health*. 2020;41:1-19.
- Li F et al. Mixed-effects models for the design and analysis of stepped wedge cluster randomized trials: an overview. *Stat Methods Med Res*. In press.
- Hemming et al. The Shiny CRT Calculator: Power and Sample size for Cluster Randomised Trials. https://clusterrcts.shinyapps.io/rshinyapp/



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
    - Analytic methods for SW-GRTs (Fan Li, July 14, 2020)
    - SW-GRTs for Disease Prevention Research (Monica Taljaard, July 11, 2018)
    - Design and Analysis of IRGTs in Public Health (Sherri Pals, April 24, 2017)
    - Research Methods Resources for Clinical Trials Involving Groups or Clusters (David Murray, December 13, 2017)
- Research Methods Resources Website
  - https://researchmethodsresources.nih.gov/
  - Material on GRTs and IRGTs and a sample size calculator for GRTs.

NIH Collaboratory *Rethinking Clinical Trials®*

Health Care Systems Research Collaboratory

Demystifying Biostatistical Concepts for Embedded Pragmatic Clinical Trials

June 19, 2020

Elizabeth L. Turner, PhD, Duke University

Patrick J. Heagerty, PhD, University of Washington

David M. Murray, PhD, National Institutes of Health

Thank you

Any questions or comments?