NIH Collaboratory

Health Care Systems Research Collaboratory

Rethinking Clinical Trials®

Topic 4: Design and Analytic Considerations

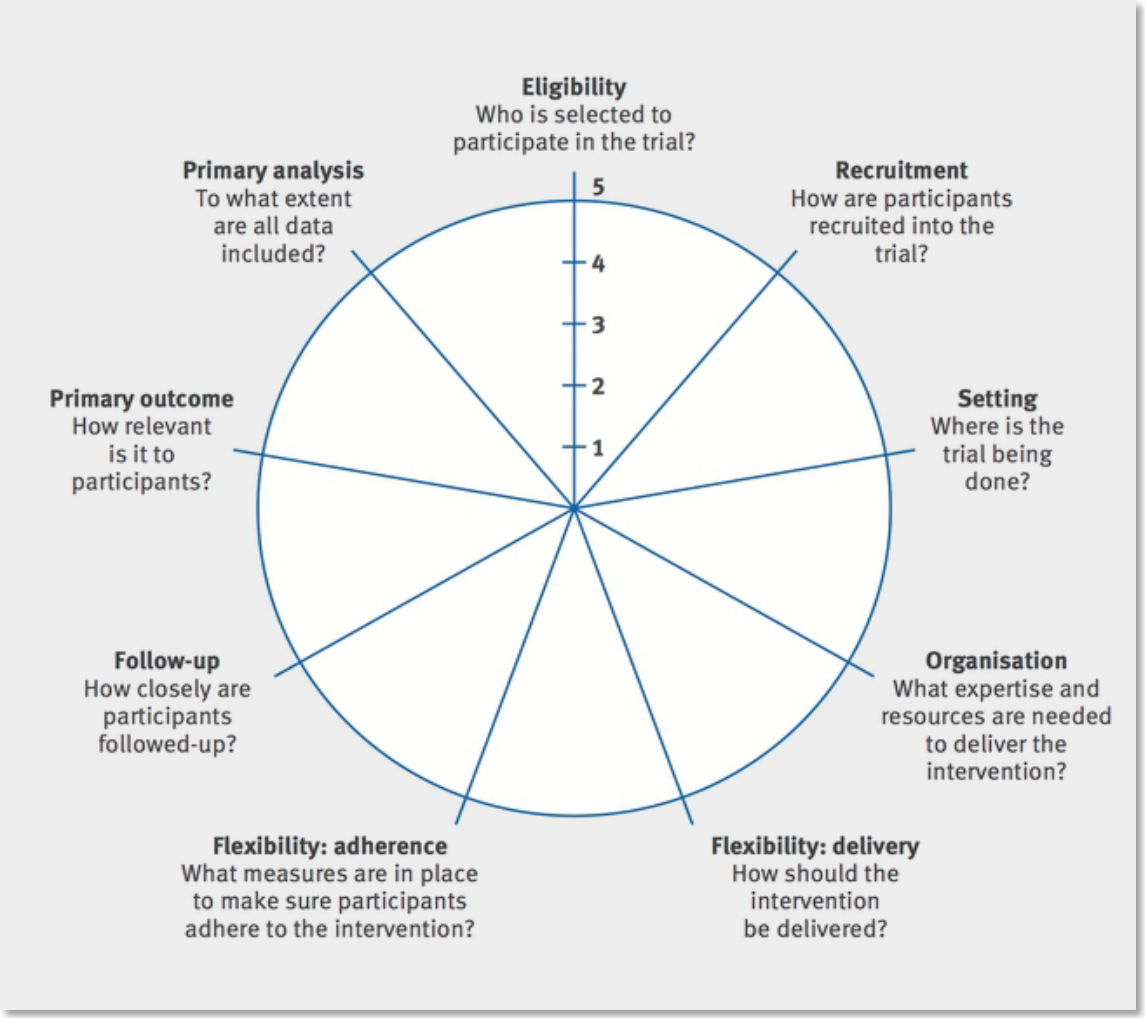
Liz Turner, PhD
Department of Biostatistics & Bioinformatics and Duke Global Health Institute
Duke University School of Medicine

Collaboratory ePCT Training Workshop

Overview

- Randomization schemes: cluster vs individual
- Cluster-randomized trials (CRTs)
  - 1: Special considerations for CRTs
    - Clustering of outcomes
    - Small # of clusters
  - 2: Varieties of cluster-randomized trials
    - Parallel
    - Stepped-wedge
- Other considerations
- How do I know I have the right statistician?

ePCTs: to inform decision-making



See: https://www.precis-2.org/

Considerations in ePCT design

- Why randomize?
  - Internal validity (ie, comparability of treatment and control arms)
- How to randomize?
  - Individual vs cluster
- Also want good external validity
  - Generalizability
  - Think carefully about eligibility

Overview

- Randomization schemes: cluster vs individual
- Cluster-randomized trials (CRTs)
  - 1: Special considerations for CRTs
    - Clustering of outcomes
    - Small # of clusters
  - 2: Varieties of cluster-randomized trials
    - Parallel
    - Stepped-wedge
- Other considerations
- How do I know I have the right statistician?

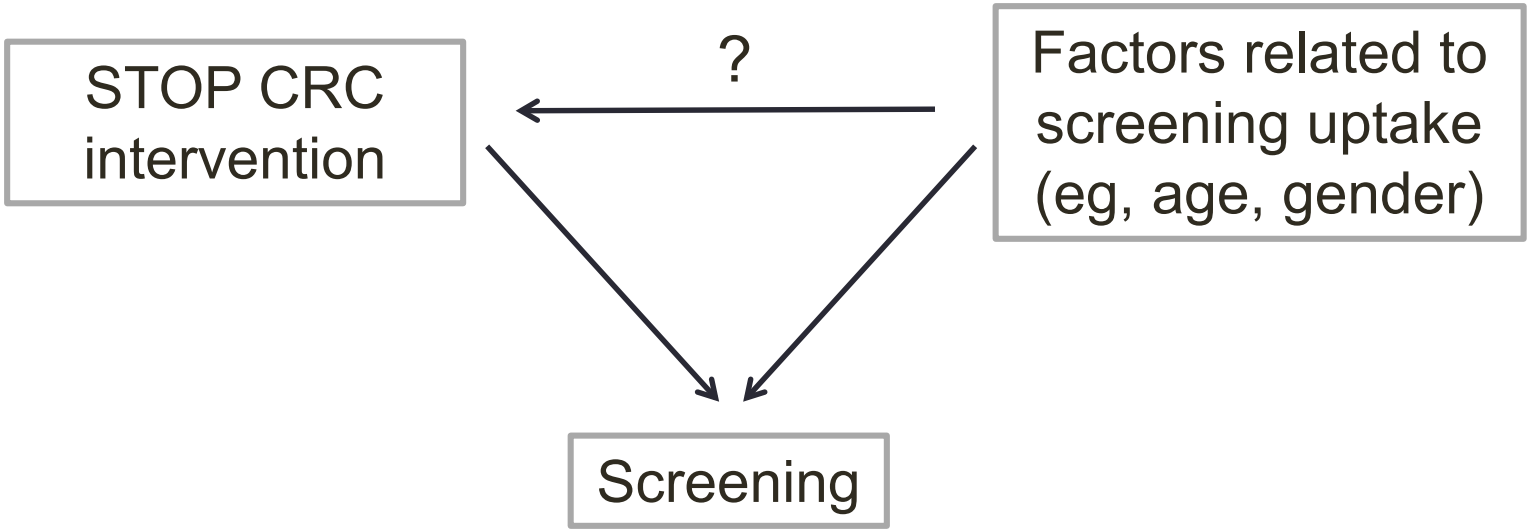
Randomization schemes

- Cluster vs individual
- Explanatory trials
  - Usually randomize individuals patient
- Pragmatic trials
  - Usually randomize clusters
  - Examples: practice, hospital, region

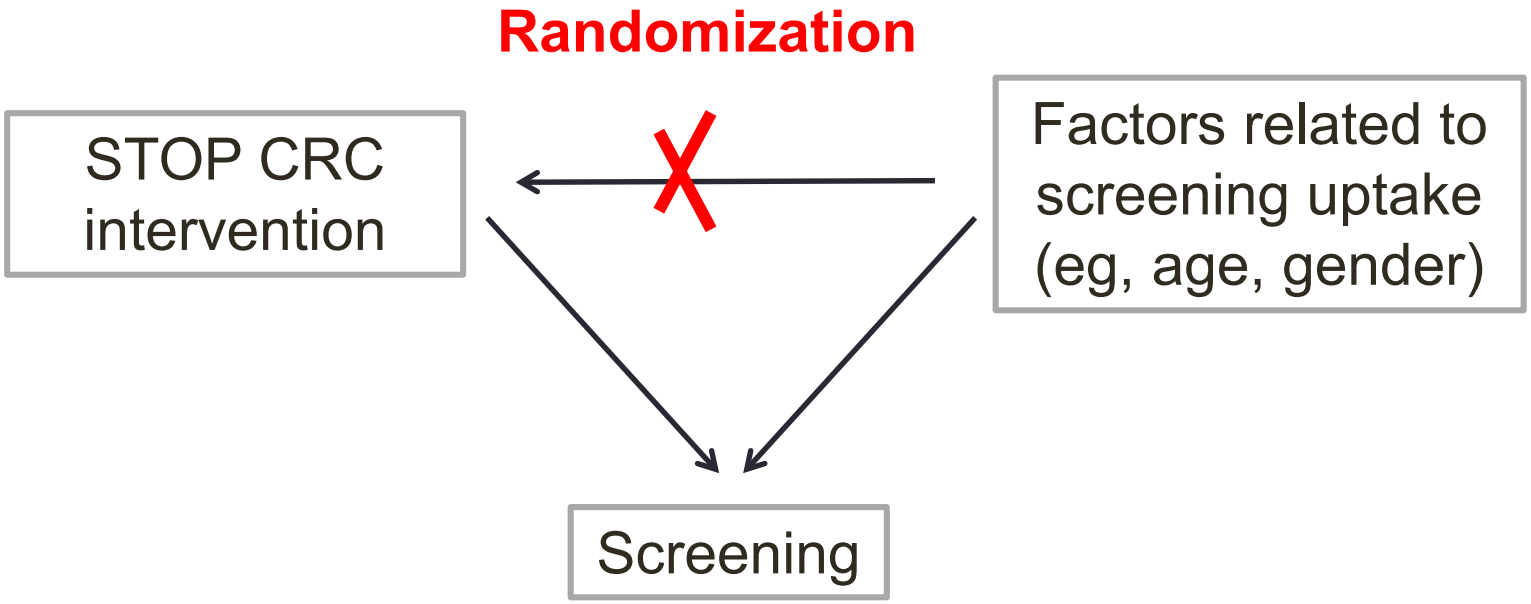
Cluster-randomized trial

- Cluster-randomized trial (CRT) definition
  - Unit of randomization is cluster of individuals
  - Unit of outcome measurement is individual
- 8 of 9 Demonstration Projects are CRTs
- Also known as:
  - Group-randomized trial
  - Community-randomized trial

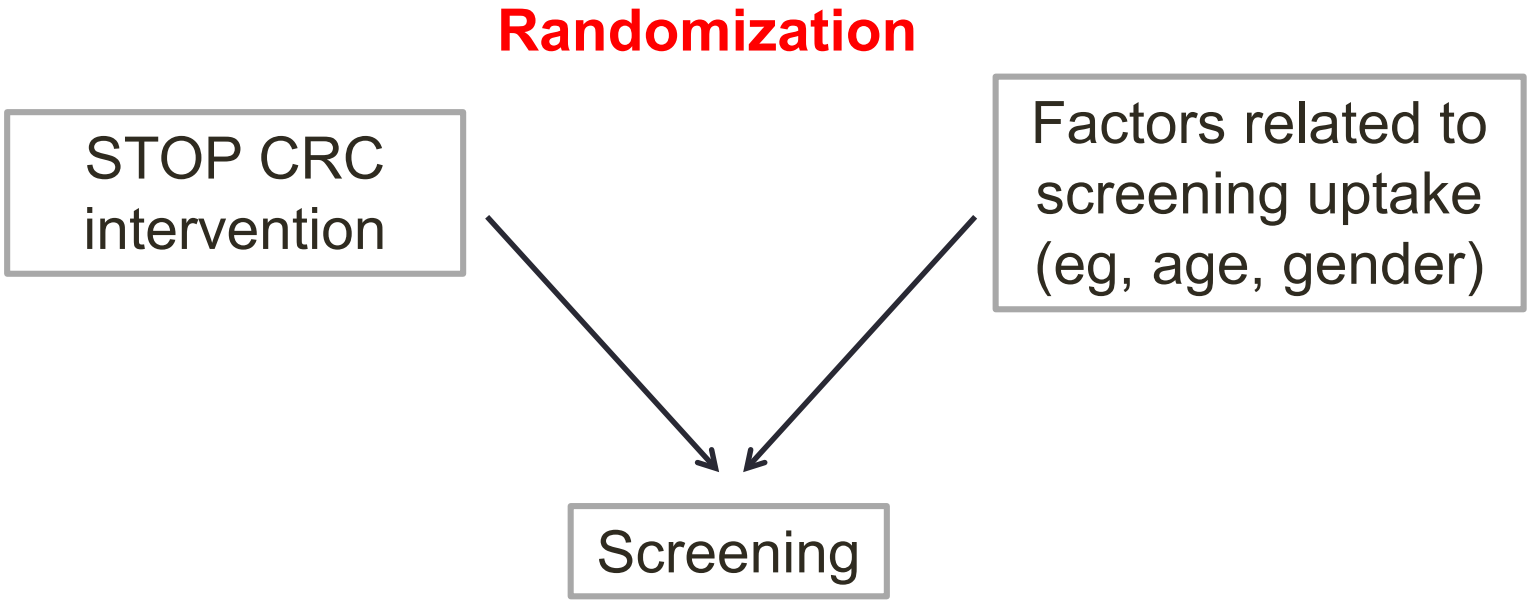
Example CRT: STOP CRC



Example CRT: STOP CRC

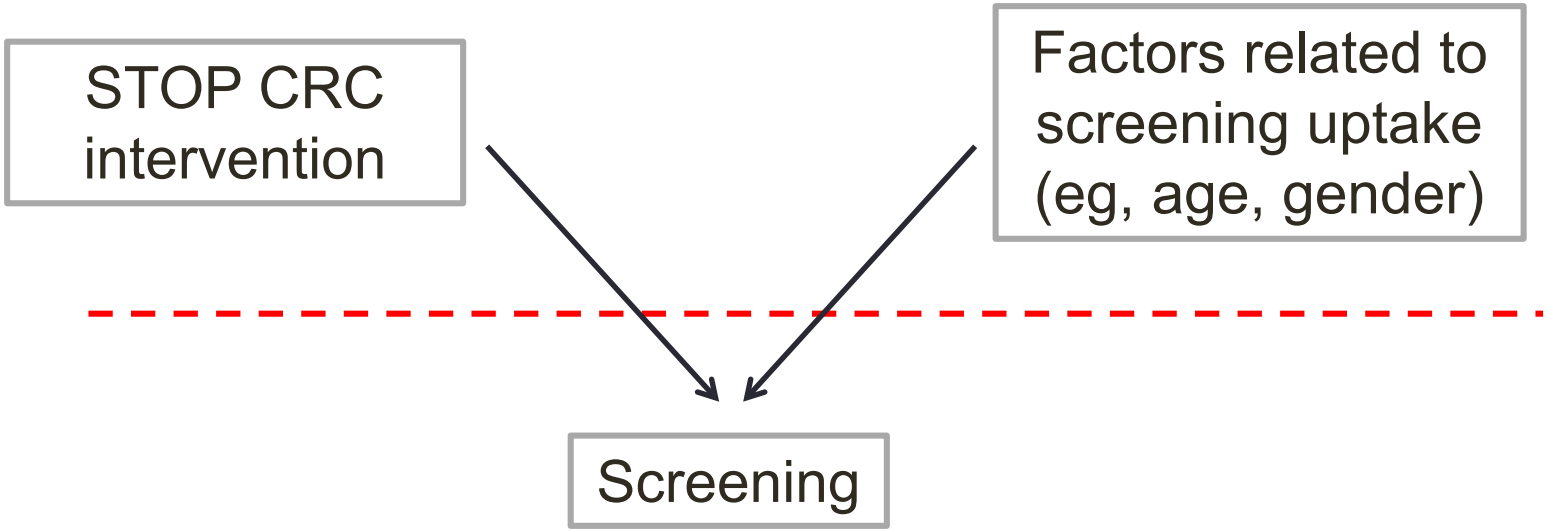


Example CRT: STOP CRC



Example CRT: STOP CRC

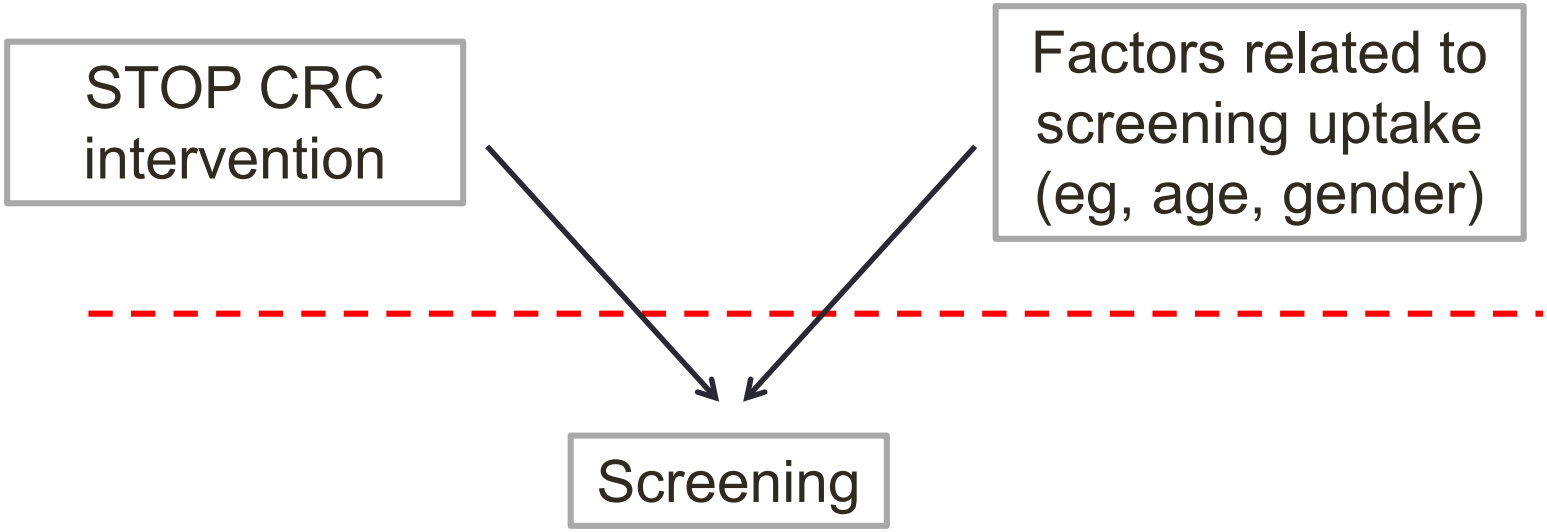
Level 2: Randomization at clinic (ie, cluster) level



Level 1: Individual-level outcomes nested in clinics

Example CRT: STOP CRC

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

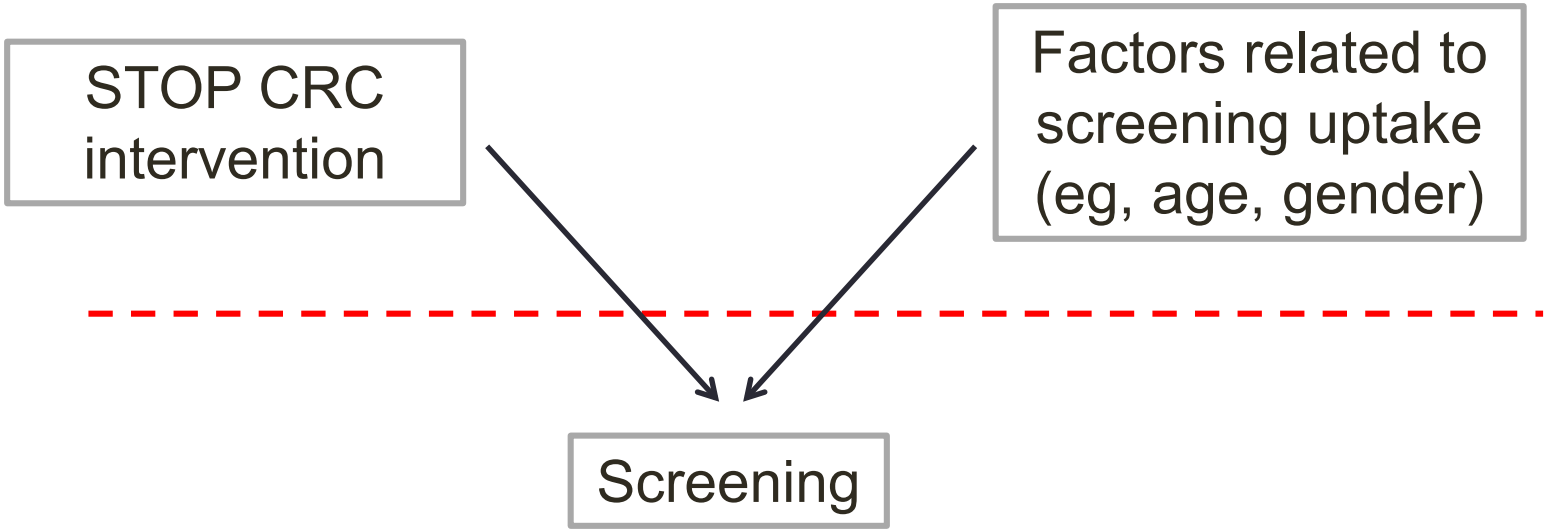


Level 1: Individual-level outcomes nested in clinics

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

Example CRT: STOP CRC

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



Level 1: Individual-level outcomes nested in clinics

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

Reduces power to detect treatment effect if same sample size used as under individual randomization

Implications of using CRT design

- CRT (statistical) price to pay
  - Lower power for same total sample size under individual randomization
  - Harder to detect an intervention effect
- So why use CRT design?
  - Intervention at cluster level (eg, STOP CRC)
  - To avoid treatment contamination under individual randomization
  - Logistically easier to implement trial

Rationale for CRT design

- STOP CRC
  - Clinic-level intervention
  - Any comments from Gloria?
- TSOS
  - Intervention at cluster level
  - Implementation science framework
  - Any comments from Doug?

Example RCT: SPOT RCT

- Only Demonstration Project with individual randomization
- Goal: suicide prevention
- Two active arms
  - Both interventions are individual-level
  - Intervention contact mostly through EHR, so expect low risk of contamination

Example RCT: SPOT study flow

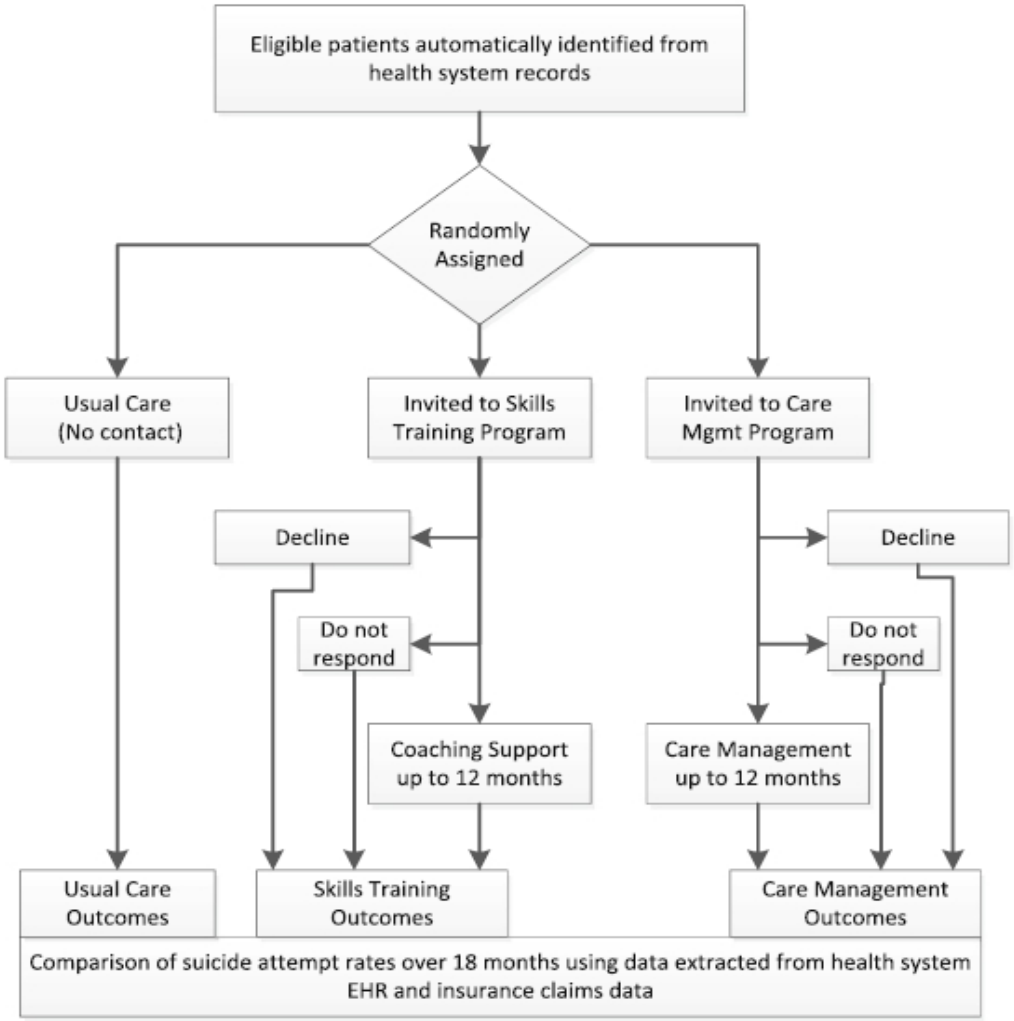


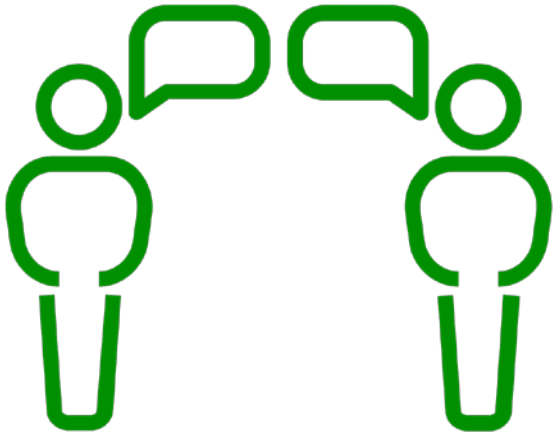
Fig. 1 Trial flow chart

Source: Simon G et al. Trials 2016;17:452

What unit of randomization makes the most sense for your study and why?

2 min

4 min



Overview: stats & design for ePCTs

- Randomization schemes: cluster vs individual
- Cluster-randomized trials (CRTs)
  - 1: Special considerations for CRTs
    - Clustering of outcomes
    - Small # of clusters
  - 2: Varieties of cluster-randomized trials
    - Parallel
    - Stepped-wedge
- Other considerations
- How do I know I have the right statistician?

Special considerations for CRTs

1. Clustering of outcomes
   - Clustering (of a particular outcome)
   - Accounting for clustering in analysis
   - Accounting for clustering in design
2. Small # of clusters
   - Potential for baseline covariate imbalance
   - How small is too small?

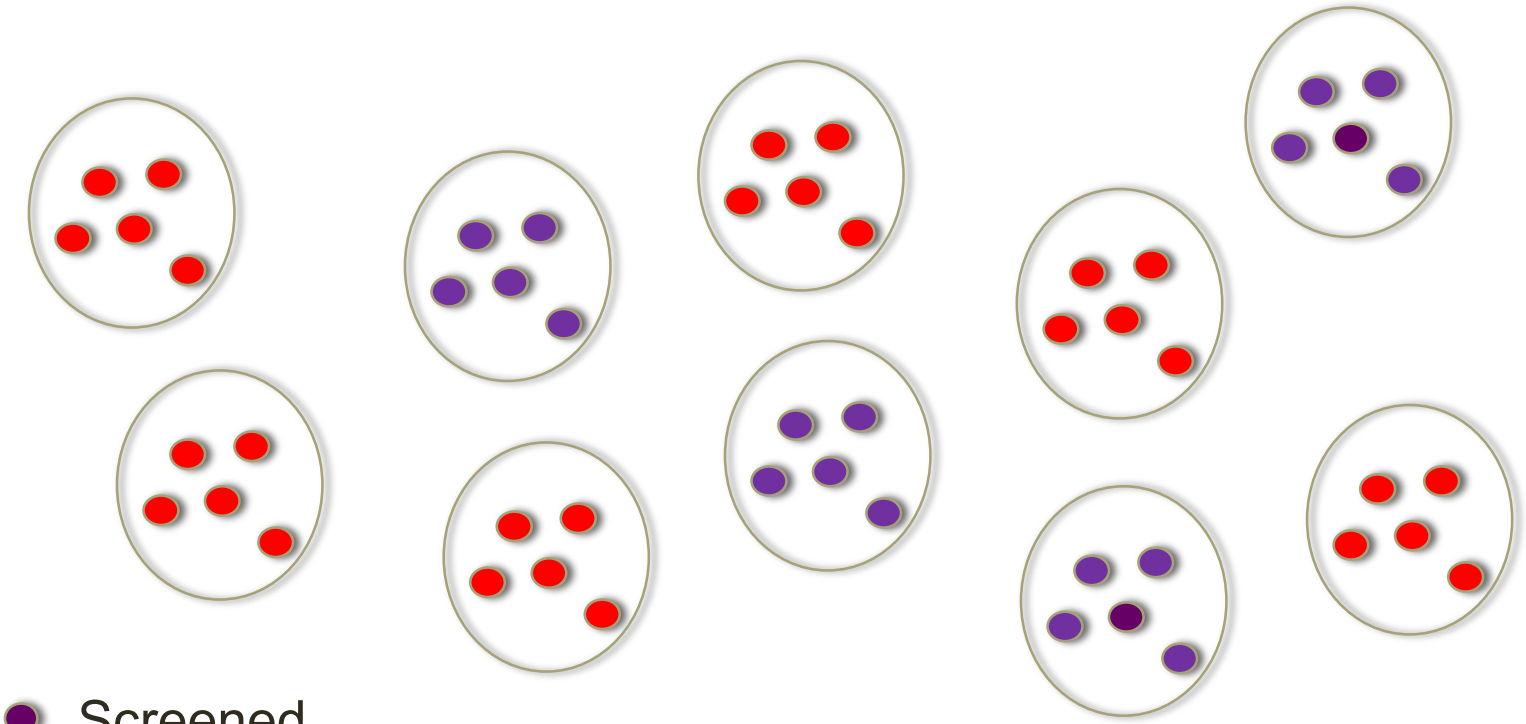
Special considerations for CRTs

1. Clustering of outcomes
   - Clustering (of a particular outcome)
   - Accounting for clustering in analysis
   - Accounting for clustering in design
2. Small # of clusters
   - Potential for baseline covariate imbalance
   - How small is too small?

Clustering example: motivated by STOP CRC

- Suppose 10 clinics
- Each with 5 age-eligible patients
  - ie, not up-to-date with CRC screening
- Outcome:
  - Binary outcome: refused screening
  - “No screening within year of enrollment”

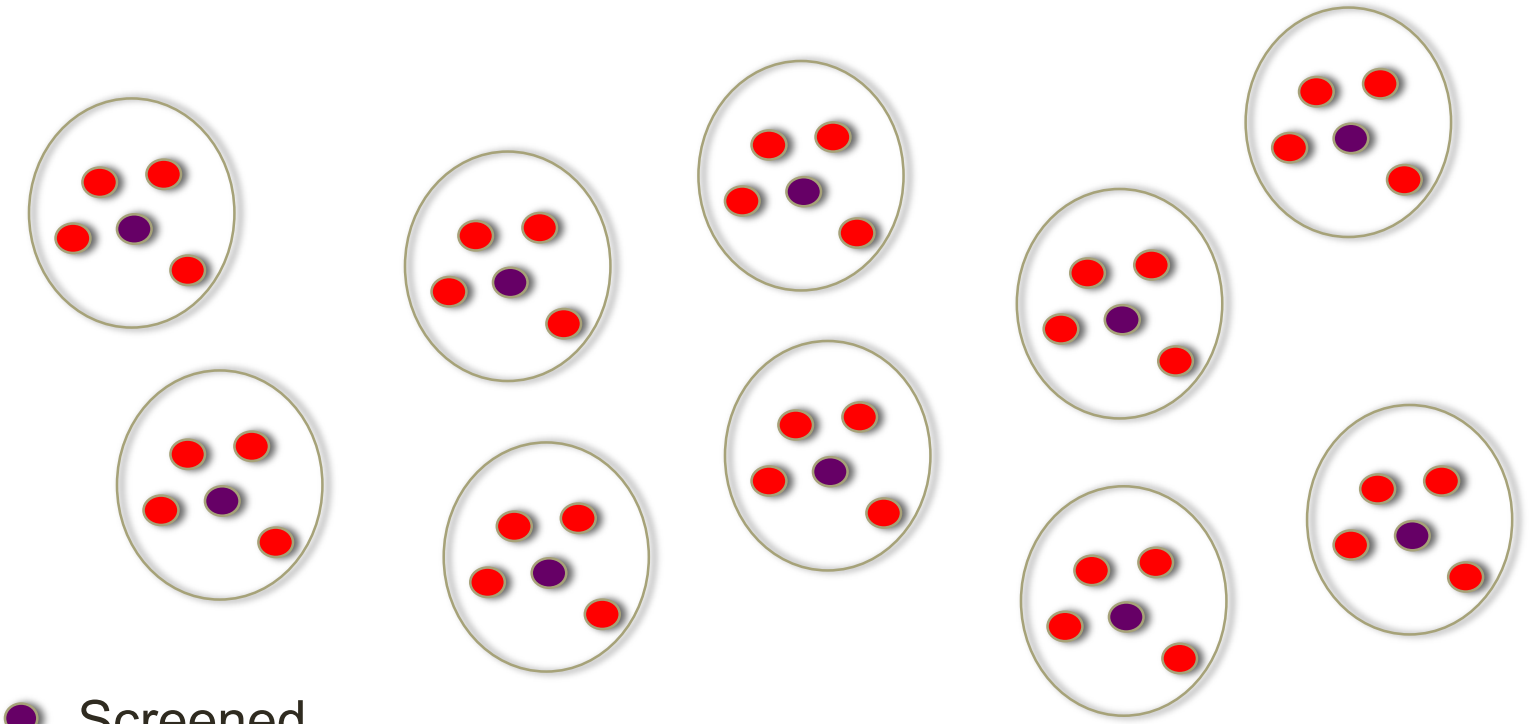
Complete clustering (ICC = 1)



- Screened
- Not screened

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

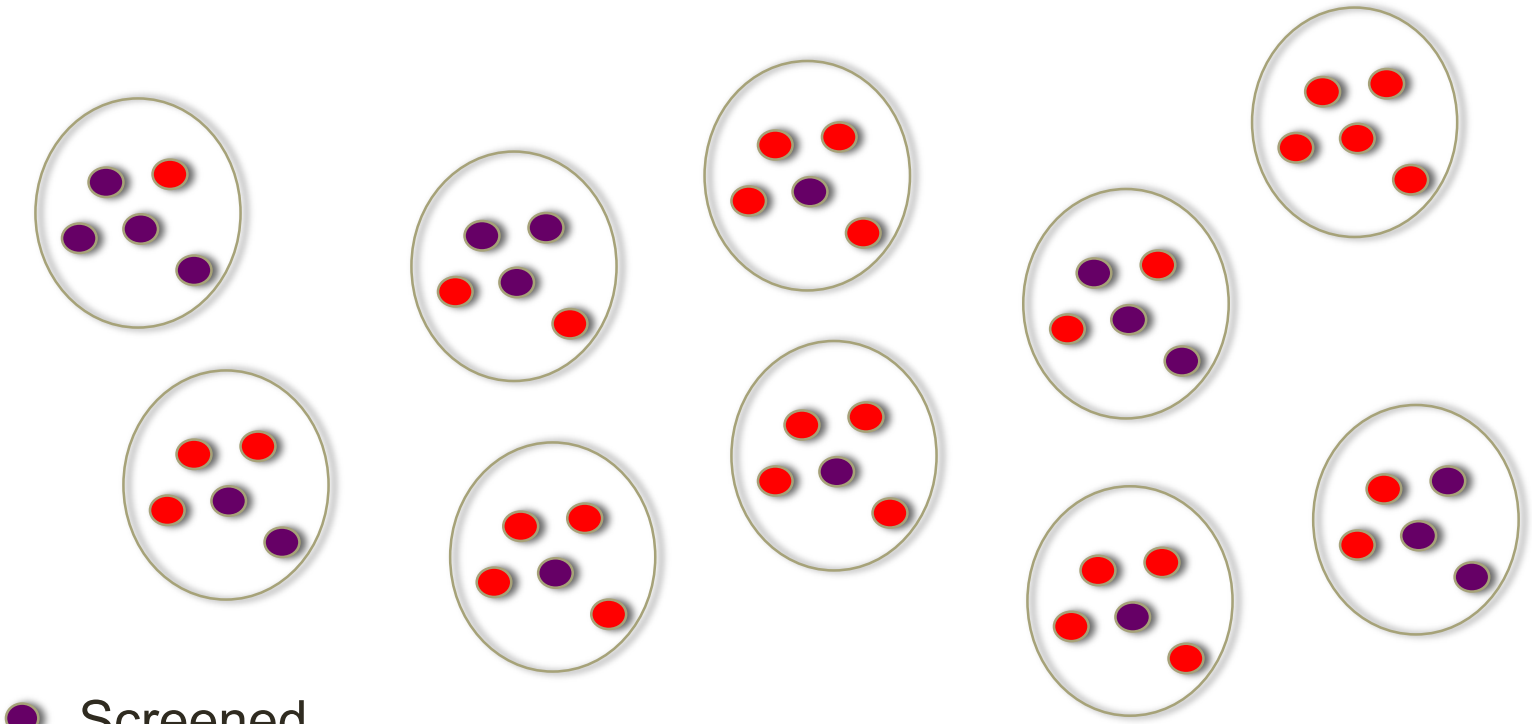
No clustering (ICC = 0)



- Screened
- Not screened

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

Some clustering (0 < ICC < 1)



- Screened
- Not screened

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

Clustering in CRTs

- Outcomes in same clusters more similar to each other than to outcomes in other clusters
- STOP CRC:
  - Planned: >450 participant/clinic in 26 clinics
  - Effective sample size: 26 – approx. 450
- Implications for statistical inference
- Major challenge in design & analysis

Measure of clustering: ICC

Intra-cluster correlation coefficient (ICC, ρ)

- Most commonly used measure of clustering
- Ranges: 0-1; 0= no clustering; 1= total clustering
- Typically < 0.2, commonly around 0.01 - 0.05
- "Between-cluster variance of outcome / total variance"

ICC for continuous outcomes:

$$\rho = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_W^2} = \frac{\sigma_B^2}{\sigma_{Total}^2}$$

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

Measure of clustering: ICC & CV

- Need measure of clustering for sample size
- Coefficient of variation (CV) alternative to ICC

$$k = \frac{\sigma_B}{\mu}$$

where μ is overall mean of outcome

- Multiple definitions of ICC for binary outcomes
  - Some authors prefer CV for binary

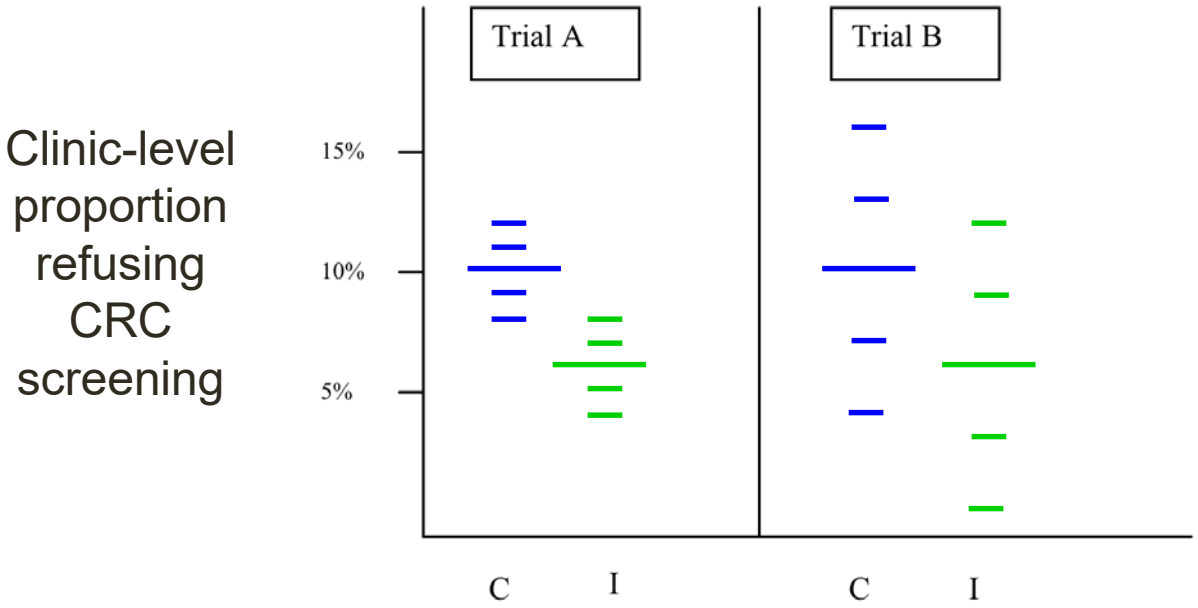
Special considerations for CRTs

1. Clustering of outcomes
   - Clustering (of a particular outcome)
   - Accounting for clustering in analysis
   - Accounting for clustering in design
2. Small # of clusters
   - Potential for baseline covariate imbalance
   - How small is too small?

Two example CRTs

- Inspired by STOP CRC
- 10 clinics/trial
  - 5 intervention (I) & 5 control (C)
  - 100 patients/clinic
- 1000 patients per trial
  - 500 intervention vs 500 control
- Binary outcome
  - Refused screening (yes/no)
  - “No screening within year enrollment”

Clustering in CRTs: implications for analysis



C=Control
I=Intervention

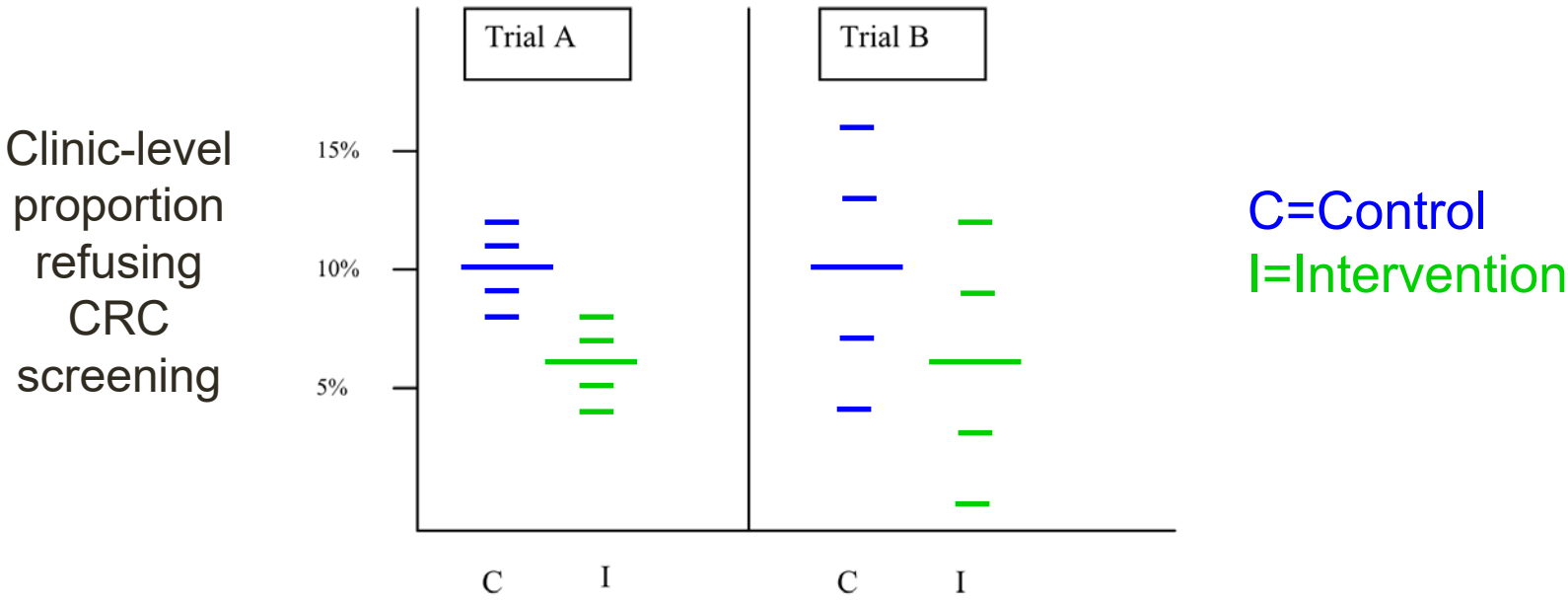
- 5 clinics each randomized to control and intervention
- 100 eligible participants per clinic measured

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

Question: is intervention effective?

Example from Hayes & Moulton (2009)

Clustering in CRTs: implications for analysis

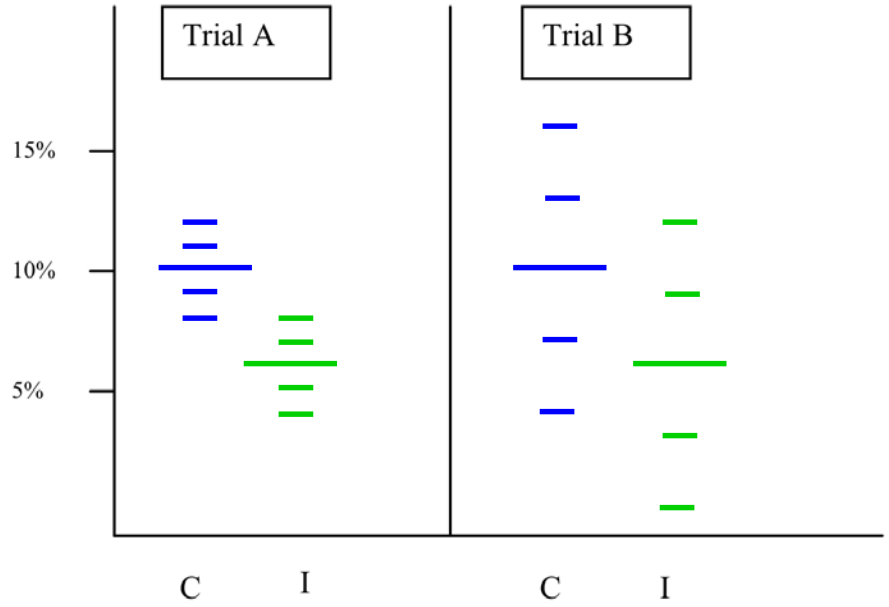


Which trial shows more evidence of benefit?

Example from Hayes & Moulton (2009)

Clustering in CRTs: implications for analysis

Clinic-level proportion refusing CRC screening



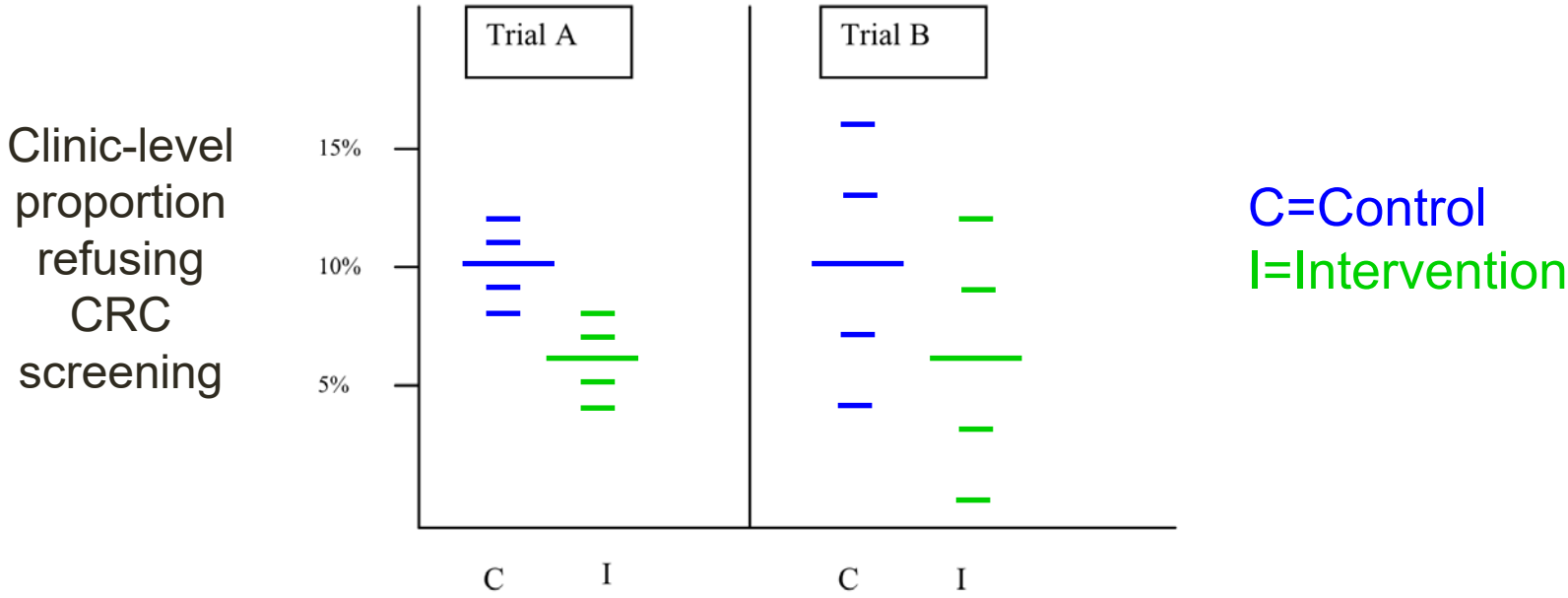
C=Control
I=Intervention

Study features

?

Example from Hayes & Moulton (2009)

Clustering in CRTs: implications for analysis

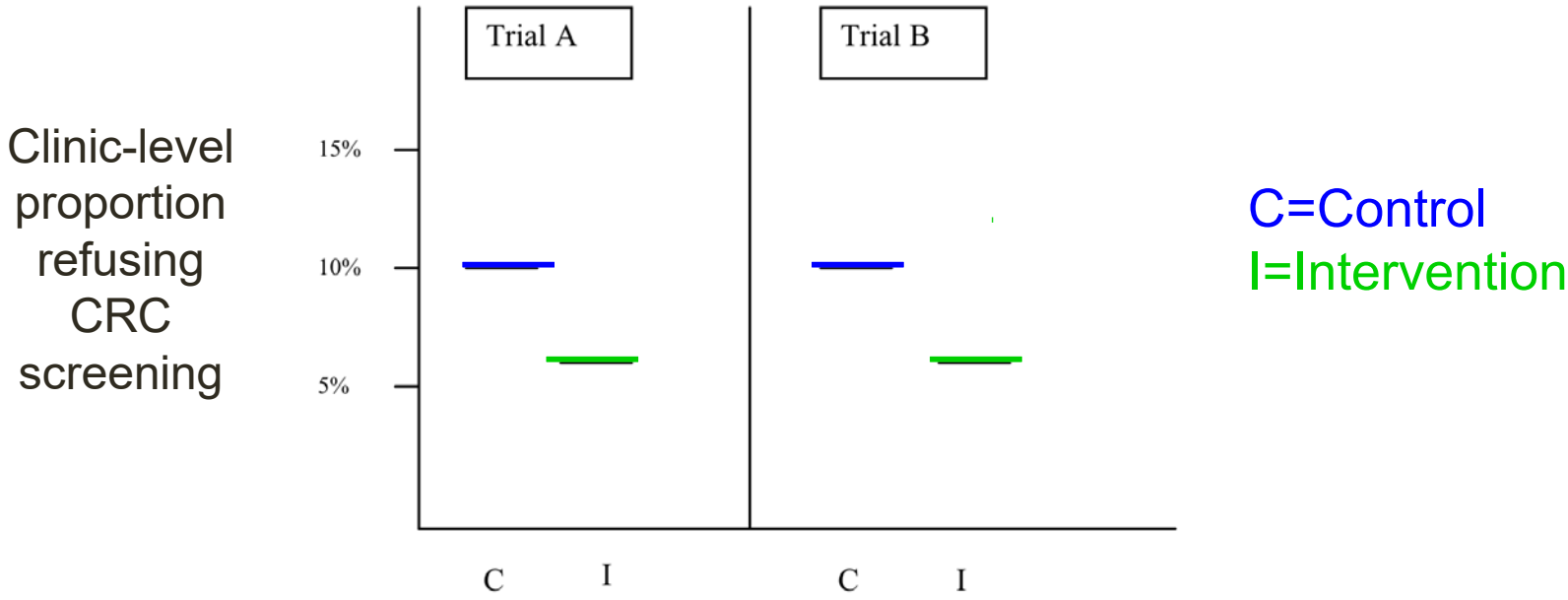


Study features

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

Example from Hayes & Moulton (2009)

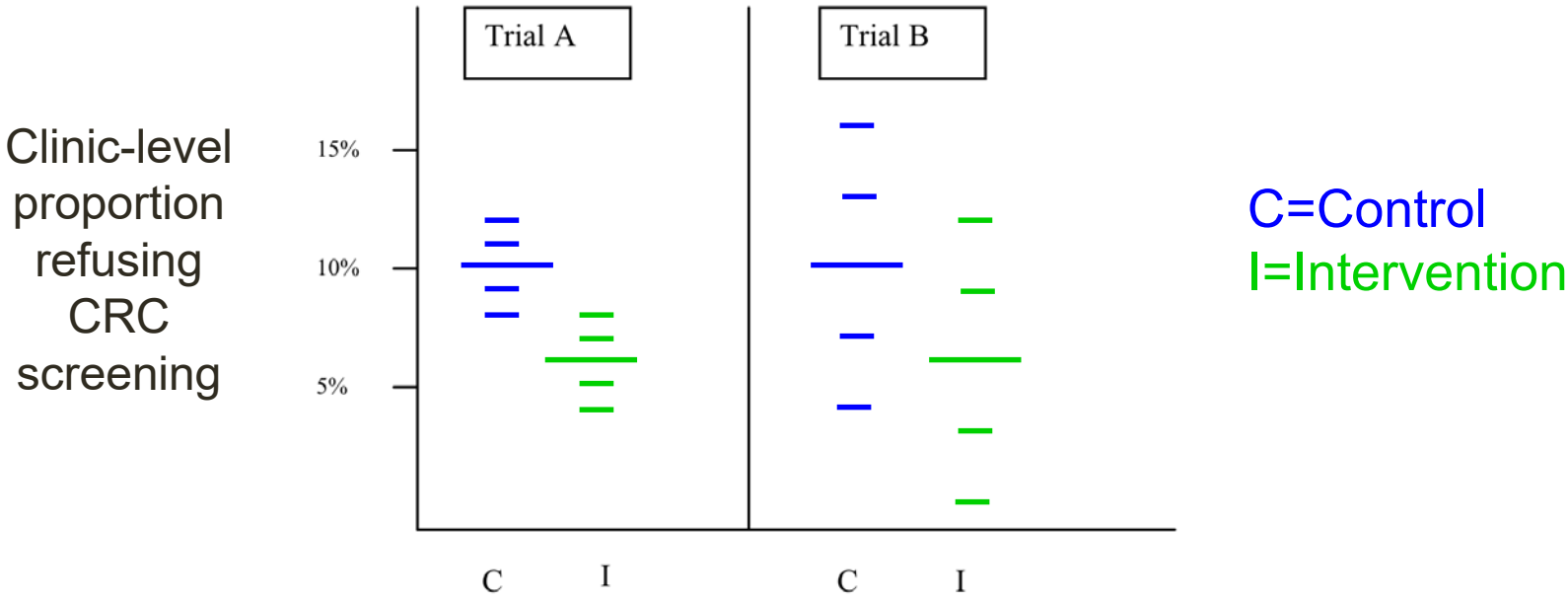
Clustering in CRTs: implications for analysis



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

Example from Hayes & Moulton (2009)

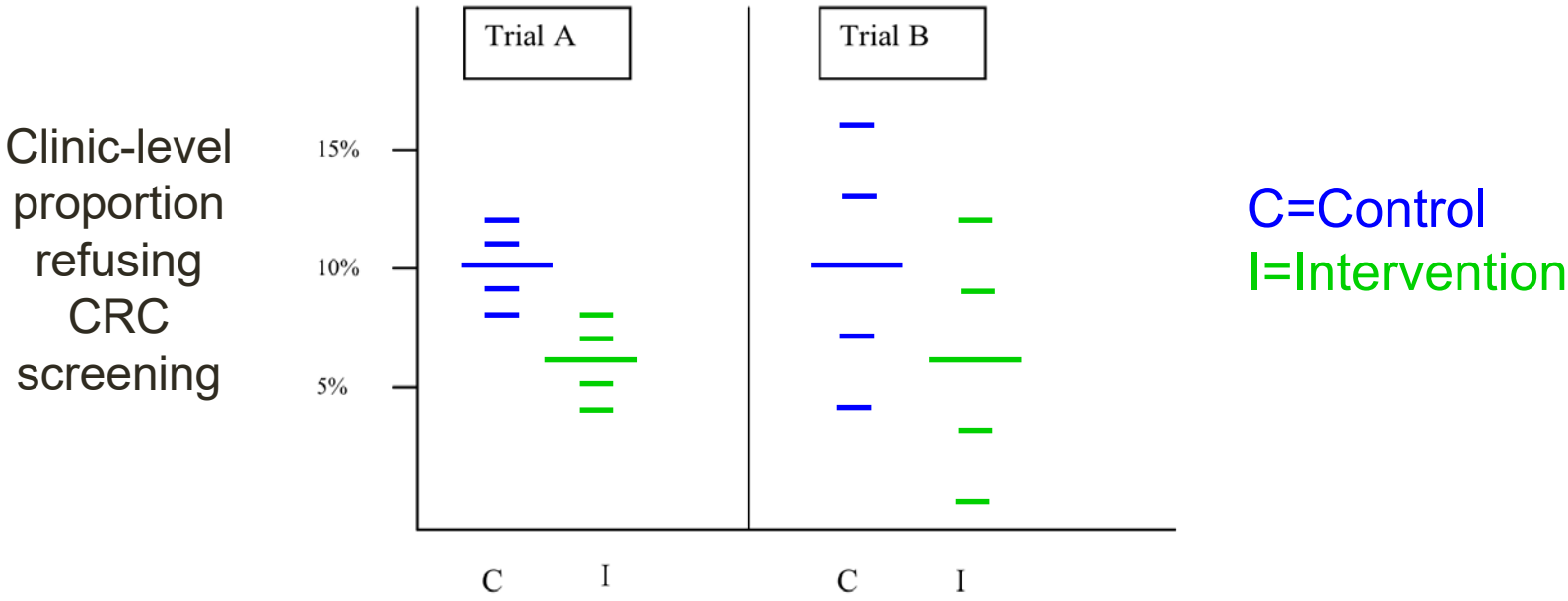
Clustering in CRTs: implications for analysis



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

Example from Hayes & Moulton (2009)

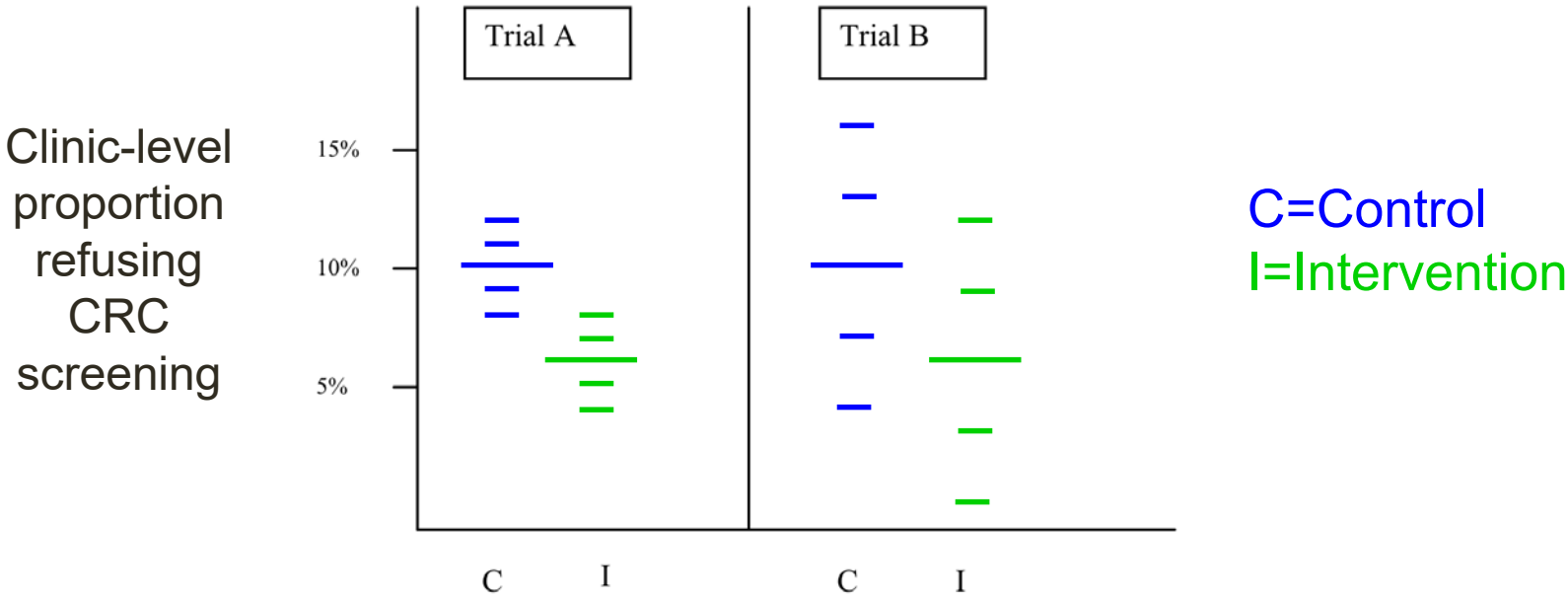
Clustering in CRTs: implications for analysis



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

Example from Hayes & Moulton (2009)

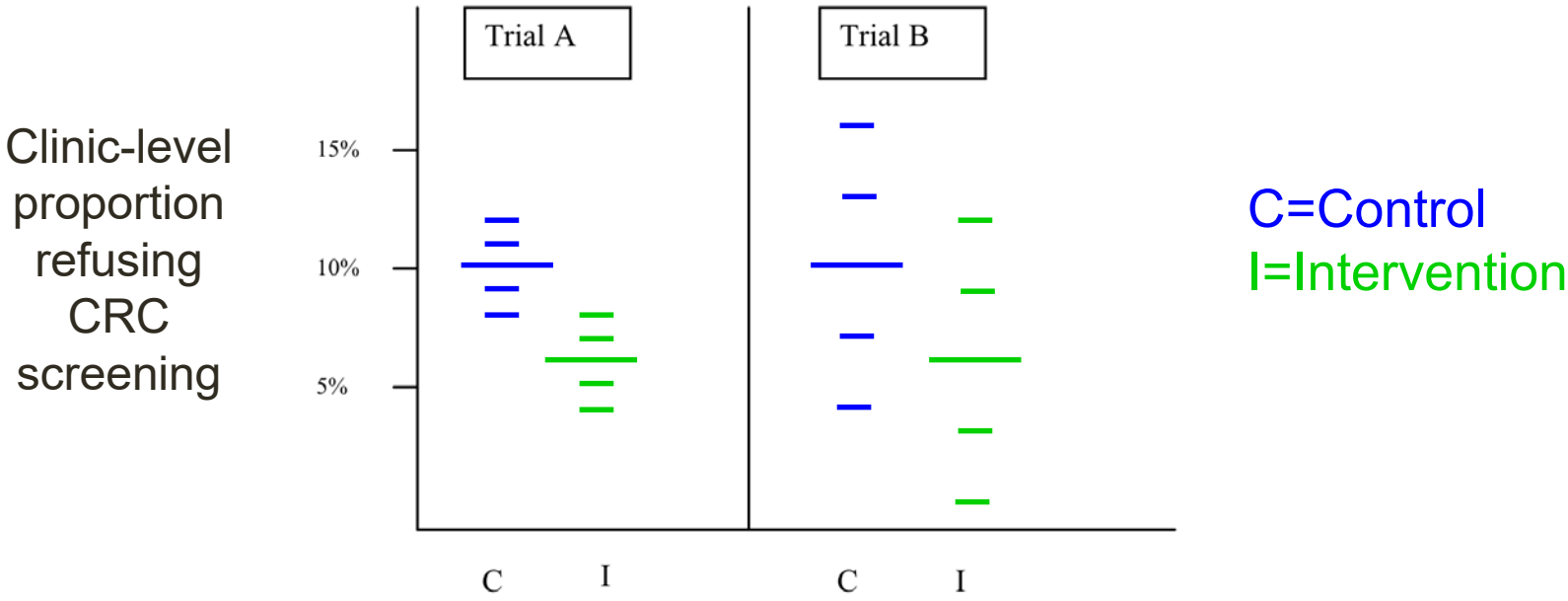
Clustering in CRTs: implications for analysis



- 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**

Example from Hayes & Moulton (2009)

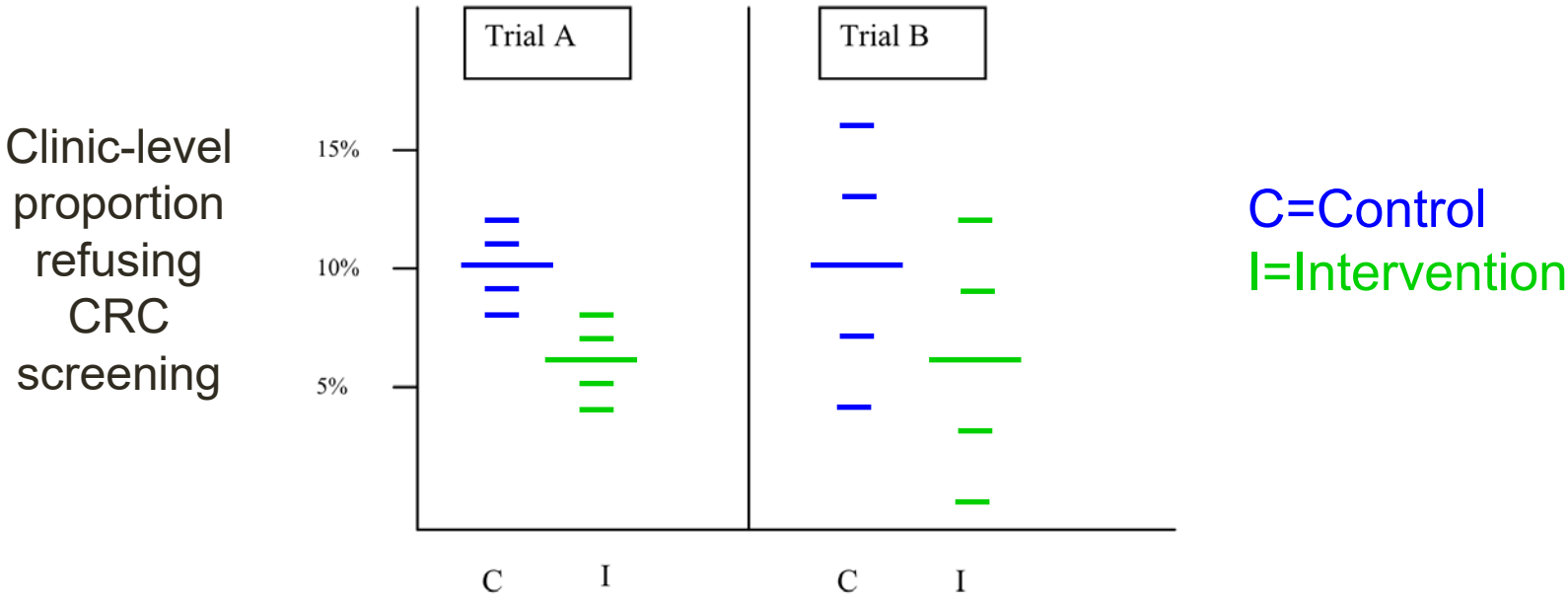
Clustering in CRTs: implications for analysis



- 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**

Example from Hayes & Moulton (2009)

Clustering in CRTs: implications for analysis

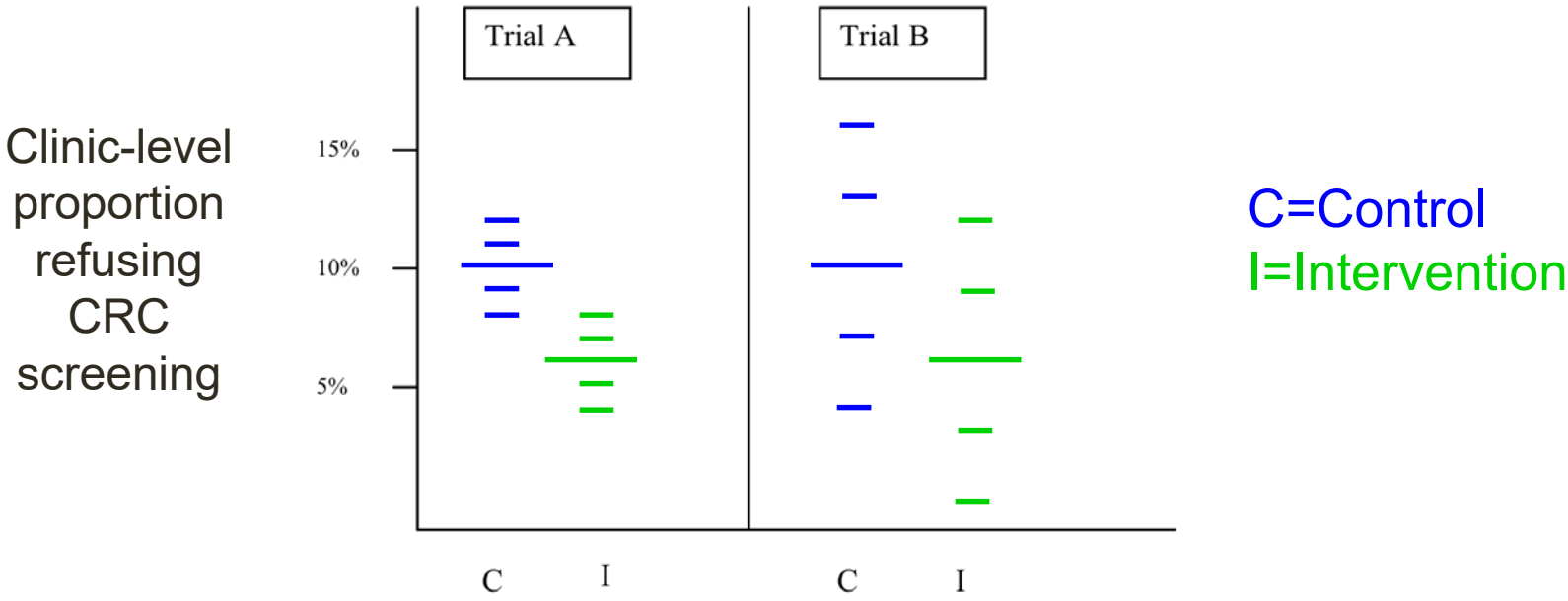


- Trial A p-value accounting for clustered design* = **0.01**
- Trial B p-value accounting for clustered design* = **0.17**

*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

Example from Hayes & Moulton (2009)

Clustering in CRTs: implications for analysis



- 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

Example from Hayes & Moulton (2009)

Summary: clustering & analysis

- 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 Trial A < Trial B
  - *P*-value Trial A < *P*-value Trial B
- Important
  - If ignore clustered design, could claim ‘significant’ when not (eg, Trial B)

Summary: clustering & analysis

- Cluster-level analysis rarely used
- Typically use regression methods
  - Analyze individual-level data, eg, data from 1000 participants/trial not only 10 clinics
  - Methods to account for clustering
    - Random effects / mixed effects models
    - Generalized estimating equations (GEE)
- Work with statistician to ensure properly account for clustering

Special considerations for CRTs

1. Clustering of outcomes
   - Clustering (of a particular outcome)
   - Accounting for clustering in analysis
   - Accounting for clustering in design
2. Small # of clusters
   - Potential for baseline covariate imbalance
   - How small is too small?

Clustering: design considerations

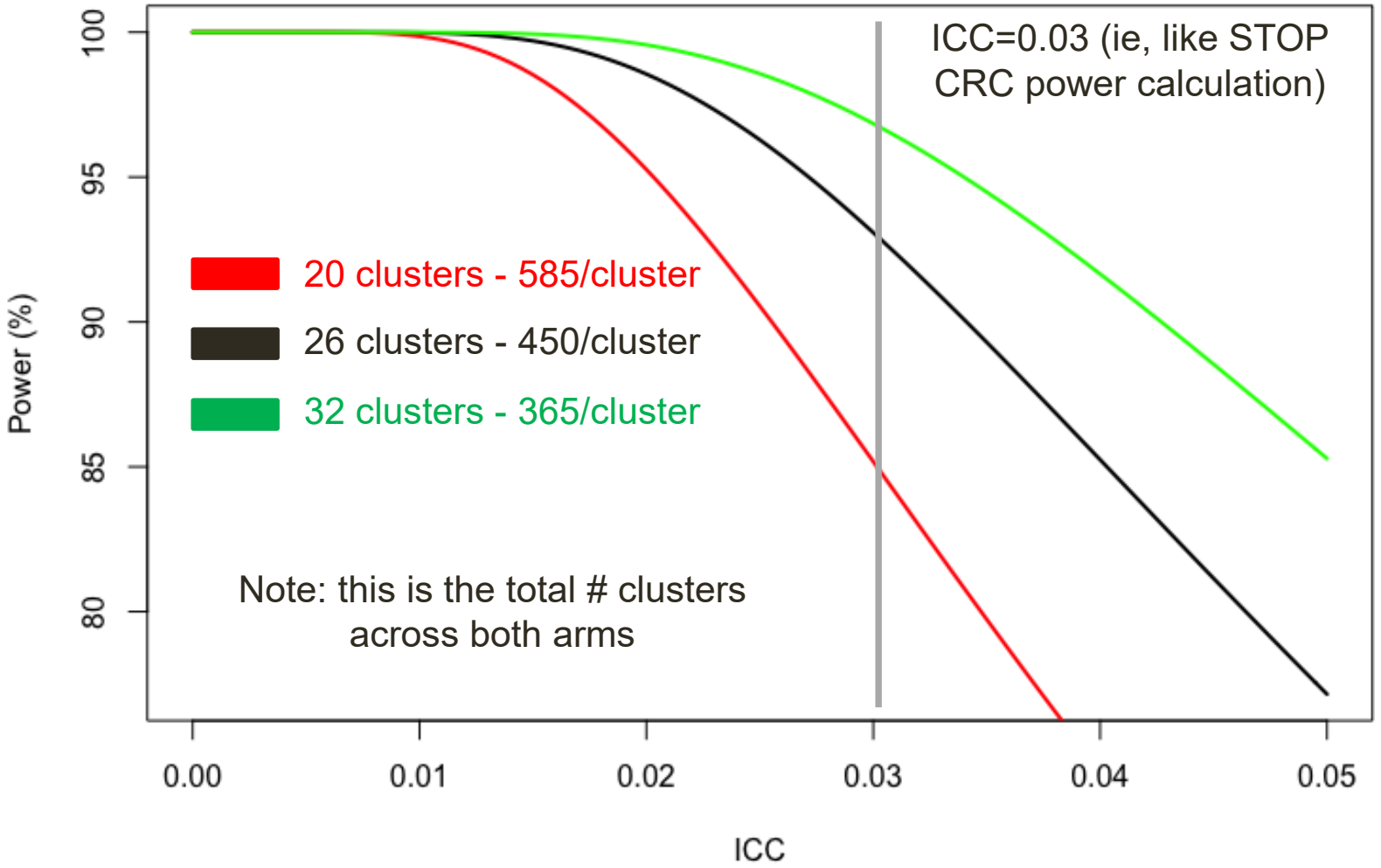
- Power & sample size
  - Account for anticipated clustering
  - Inflate RCT sample size
  - Work with statistician to do this correctly
- Use ICC (or CV) for outcome
  - ICC often 0.01-0.05
  - STOP CRC: ICC = 0.03 for primary outcome
  - Depends on outcome & study characteristics
  - Different outcome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

Clustering in STOP CRC: design considerations

"Assumed equal numbers of subjects per clinic and equal numbers of clinics (n = 13) per group. 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. 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. Based on analyses by Dr. Green using the data from her Systems Of Support study [12,28], we expect the ICC to be about .03. Using this figure, we will have very good power (>91%) to detect absolute differences as small as 10 percentage points even if the FIT completion rate in the UC arm is as high as 15% (fecal testing rates for 2013 for usual care clinics was 10%). For an ICC of .05 we would still have >91% power for detecting effect sizes of at least 13 percentage points."

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

Clustering: impact on power



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)

Clustering: design considerations

- Many references on CRT power and sample size
- Important to account for clustering
  - Some adjust RCT sample size by design effect: $1+(m-1)\rho$, where m = # participants/cluster
  - Better to be more explicit
    - eg, want to determine # clusters needed for fixed # participants/cluster or vice-versa?
- **Work with a statistician!**

Resources

- NIH: https://researchmethodsresources.nih.gov/
- 5 textbooks (see reference list)
- See reference list: Turner et al. (2017) and Rutterford et al. (2015)

Clustering: design considerations

- How to get good initial estimate of ICC for a particular outcome?[1]
  - It depends on outcome & study characteristics
  - CONSORT[2] statement on reporting of CRTs recommends ICC reported
  - Look at other articles with similar settings
- Be cautious when using pilot data from small study
  - ICC might have wide confidence interval

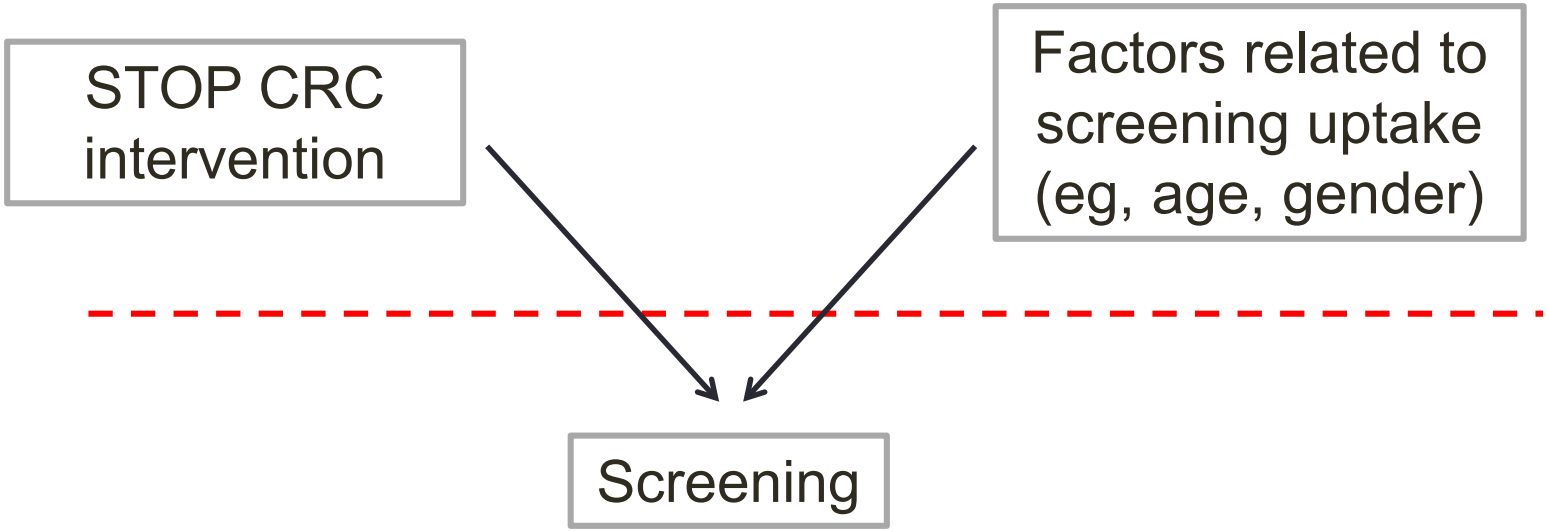
1. See FAQ 13 at: https://researchmethodsresources.nih.gov/
2. http://www.bmj.com/content/345/bmj.e5661

Special considerations for CRTs

1. Clustering of outcomes
   - Clustering (of a particular outcome)
   - Accounting for clustering in analysis
   - Accounting for clustering in design
2. Small # of clusters
   - Potential for baseline covariate imbalance
   - How small is too small?

Example CRT: STOP CRC

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



Level 1: Individual-level outcomes nested in clinics

- **Goal:** randomization → baseline balance of covariates
- **Challenge:** baseline imbalance may occur if not many clusters enrolled (eg, there are 26 clinics in STOP CRC)

Small # of clusters & baseline covariate imbalance

- Pragmatic CRTs often enroll small # (<40) clusters
- Randomization may not balance baseline covariates
- Baseline covariate imbalance threatens internal validity ie, comparability of treatment arms
  - Challenge: claim intervention effect is causal but there may be confounding due to non-comparability of treatment arms

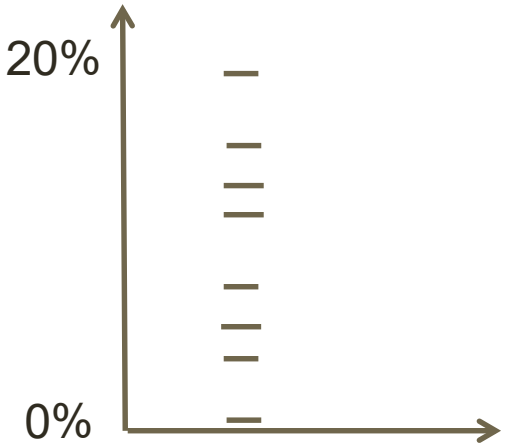
Baseline covariate imbalance

- Threat to internal validity of trial
- Could address with adjusted analysis
- Better to use design strategy
  - ‘Restricted randomization’
- Three types of restricted randomization
  - Pair-matching
  - Stratification
  - Covariate-constrained randomization

Baseline covariate imbalance

Example: 8 clinics (clusters)

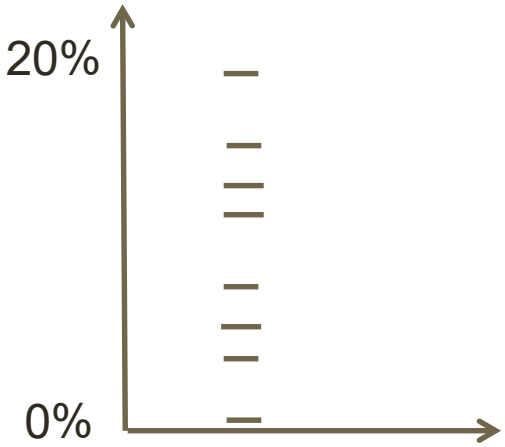
Baseline clinic-level proportion who refused screening in previous year



Baseline covariate imbalance

Example: 8 clinics (clusters)

Baseline clinic-level proportion who refused screening in previous year



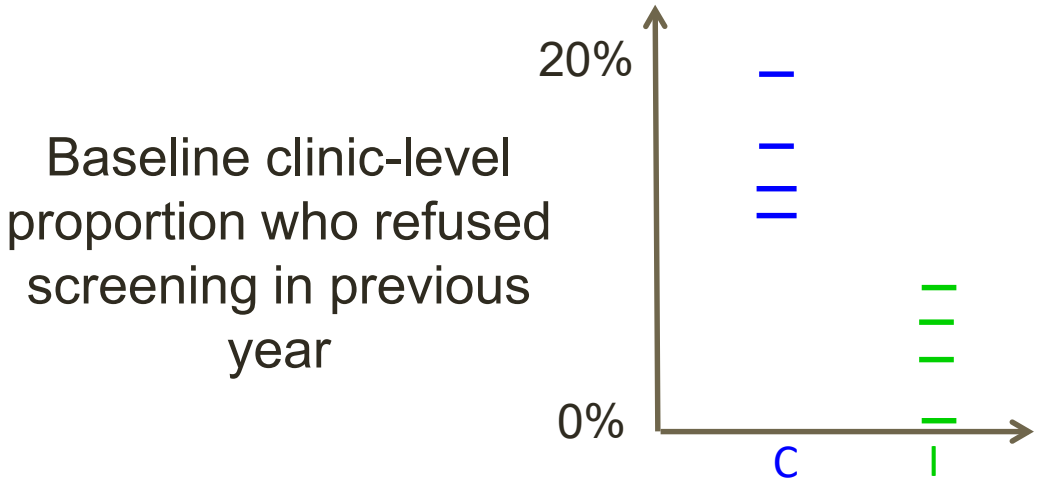
Question: Why do we care about getting balance between treatment arms on clinic-level proportion who refused screening in previous year?

It might be related to proportion in the next year!

Baseline covariate imbalance

Example: 8 clinics (clusters)

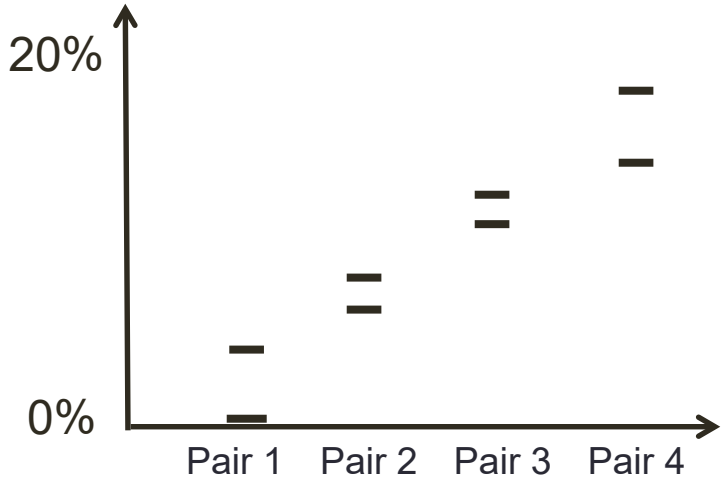
Example of extreme baseline imbalance using simple (ie, regular) randomization



Baseline covariate imbalance

Possible design solution 1: Pair-matching

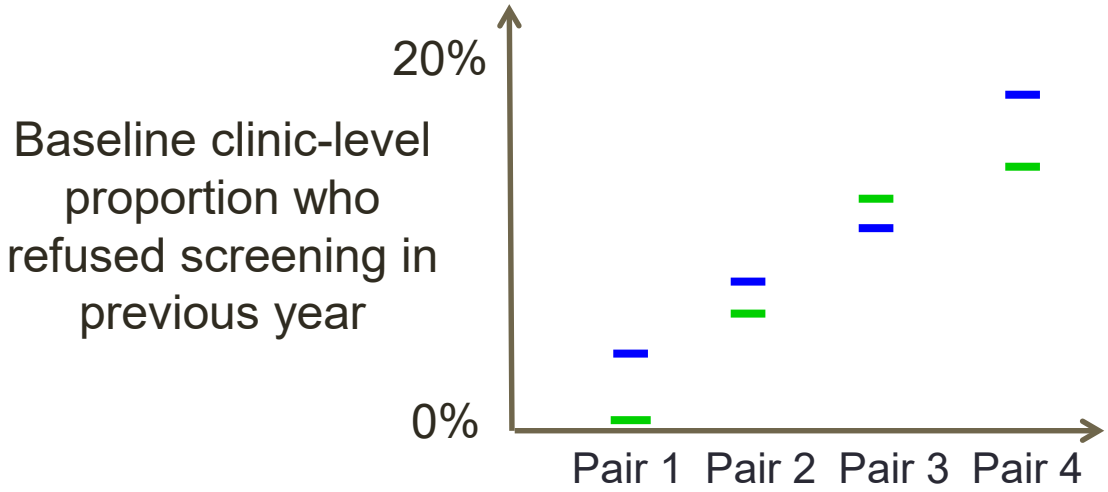
Baseline clinic-level proportion who refused screening in previous year



Baseline covariate imbalance

Possible design solution 1: Pair-matching

One example of pair-matched randomization to control & intervention arms

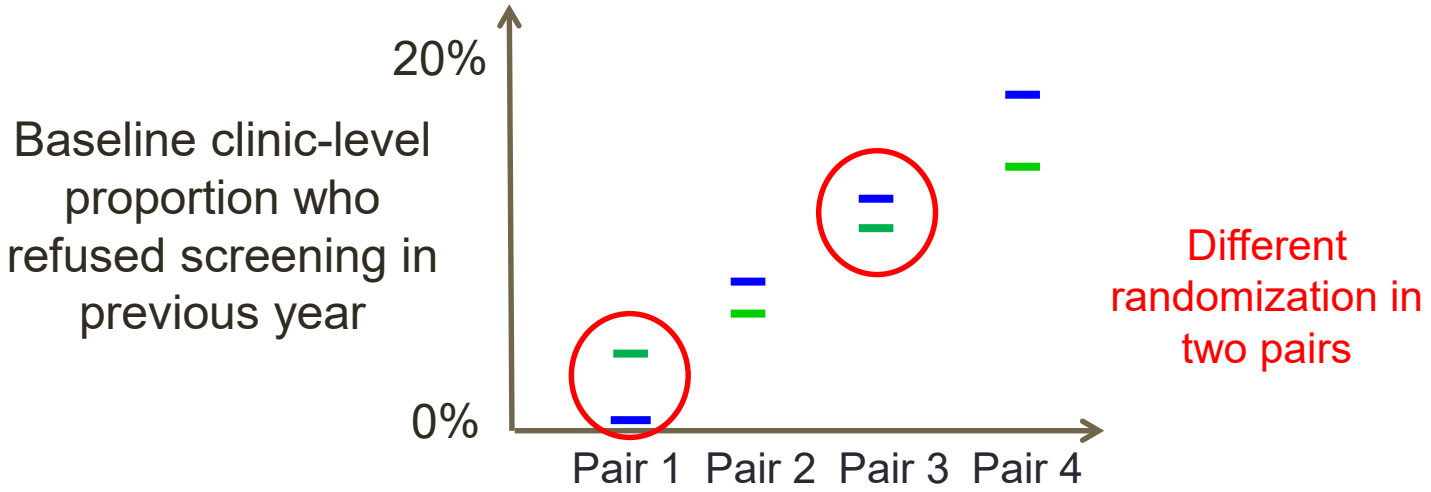


Intervention and control perfectly balanced on ”pair” ie, exactly 1 cluster from each pair in intervention and 1 in control

Baseline covariate imbalance

Possible design solution 1: Pair-matching

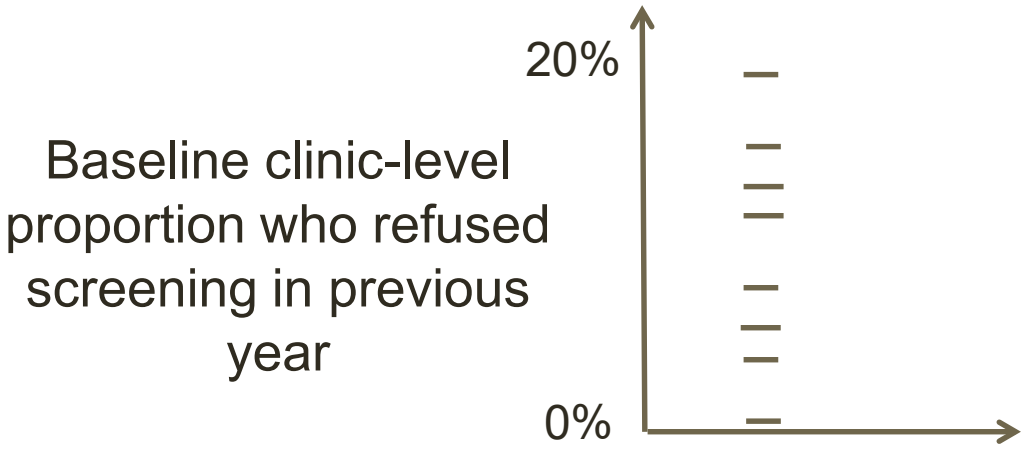
Another example of pair-matched randomization to control & intervention arms



Important: account for paired design in the analysis (eg, paired t-test or Wilcoxon signed rank test for cluster-level analysis or matched regression model)

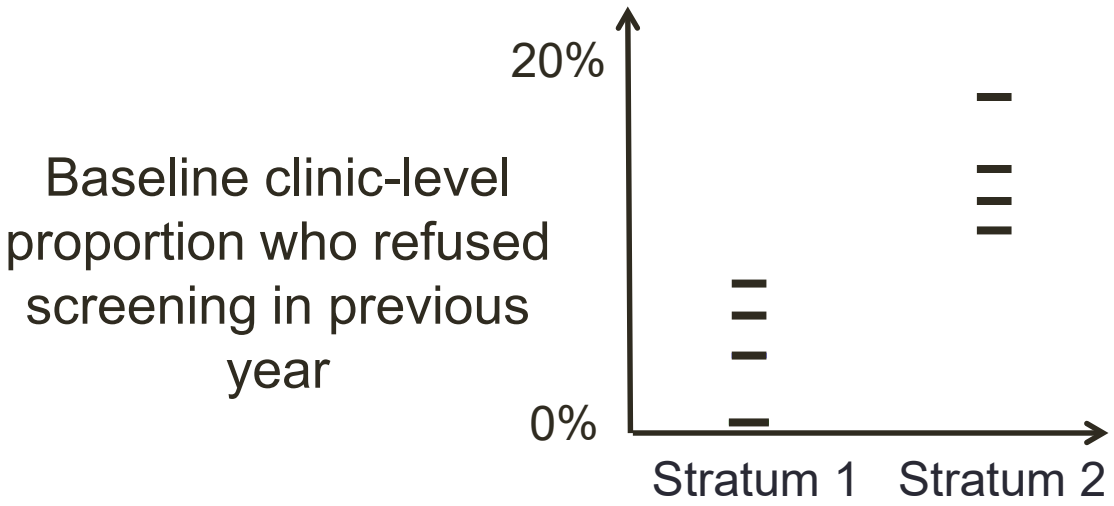
Baseline covariate imbalance

Example: 8 clinics (clusters)



Baseline covariate imbalance

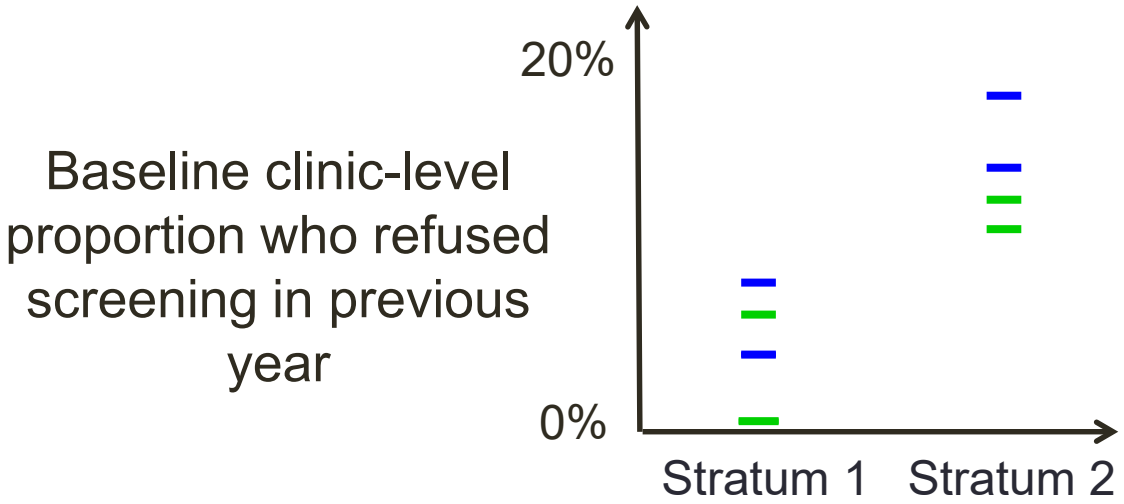
Possible design solution 2: Stratification



Baseline covariate imbalance

Possible design solution 2: Stratification

One example of stratified randomization to control & intervention arms

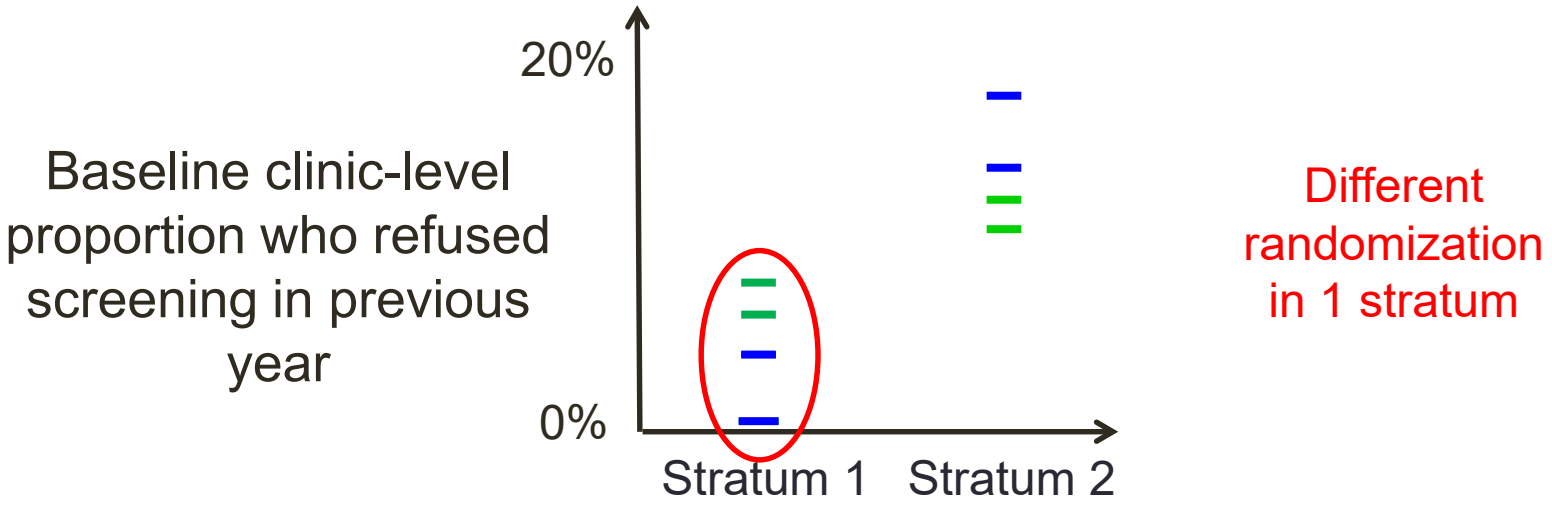


Intervention and control perfectly balanced on ”stratum” ie, exactly 2 clusters in intervention and 2 in control in each stratum

Baseline covariate imbalance

Possible design solution 2: Stratification

Another example of stratified randomization to control & intervention arms



Important: account for stratified design in the analysis (eg, stratified permutation test or fixed effect for strata in model-based analysis)

Baseline covariate imbalance

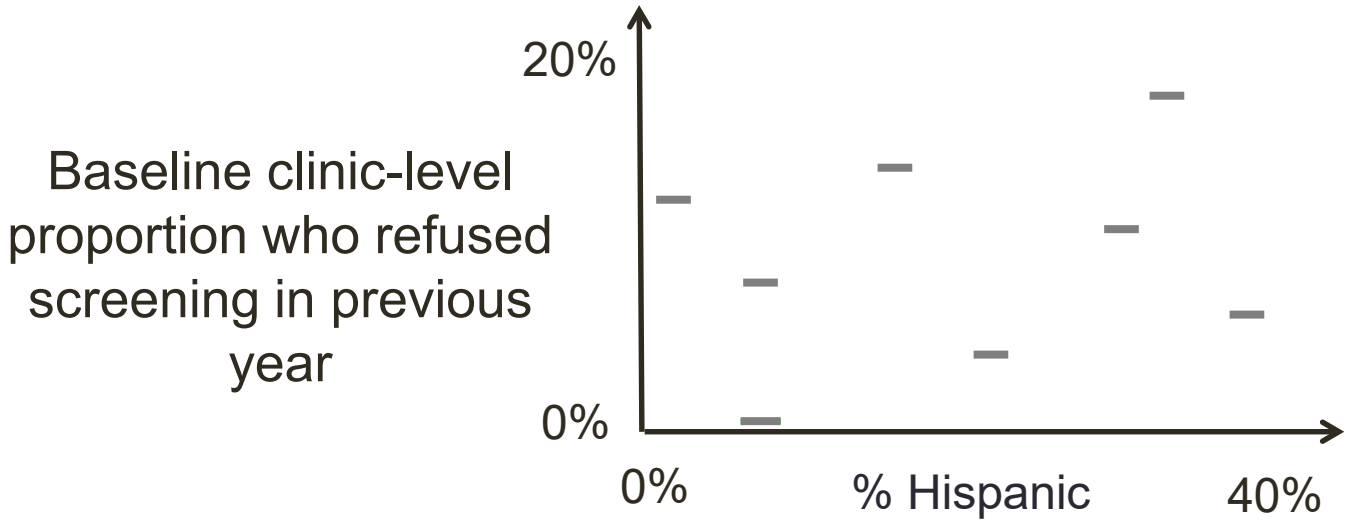
Possible design solution 3: Constrained randomization

- Previous examples
  - Baseline balance of 1 clinic-level covariate ie, % refused screening in previous year
- Often have multiple clinic-level covariates
  - Categorical & continuous
  - Pair-matching & stratification cannot easily handle this
- Need more general form of restricted randomization
  - Covariate-constrained randomization

Baseline covariate imbalance

Possible design solution 3: Constrained randomization

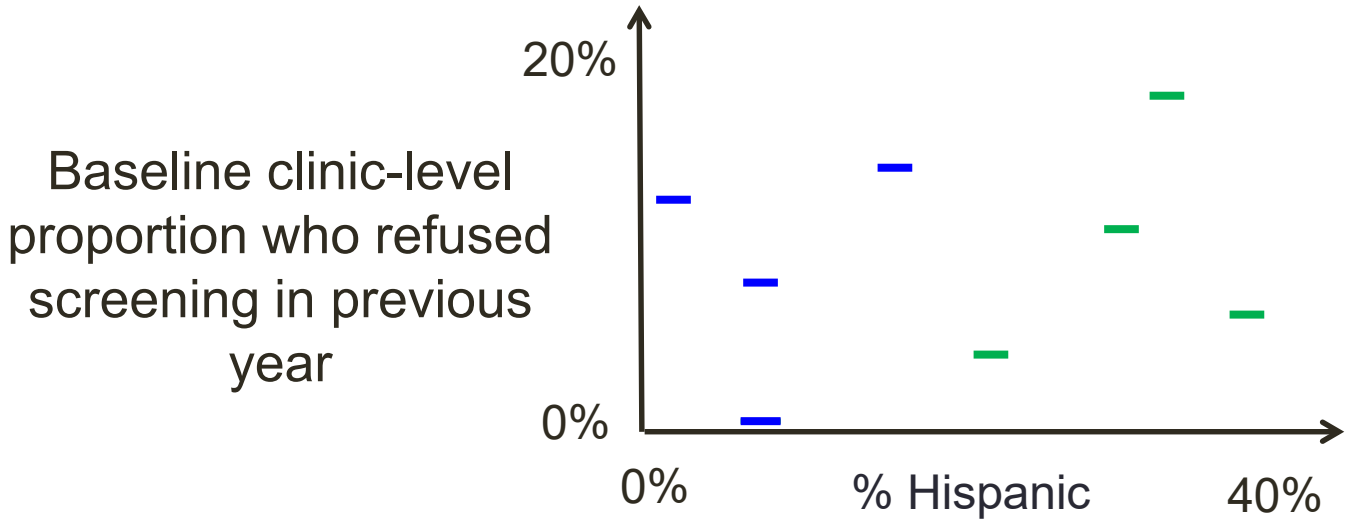
Example: balance two continuous cluster covariates



Baseline covariate imbalance

Possible design solution 3: Constrained randomization

One example of simple randomization to control & intervention arms

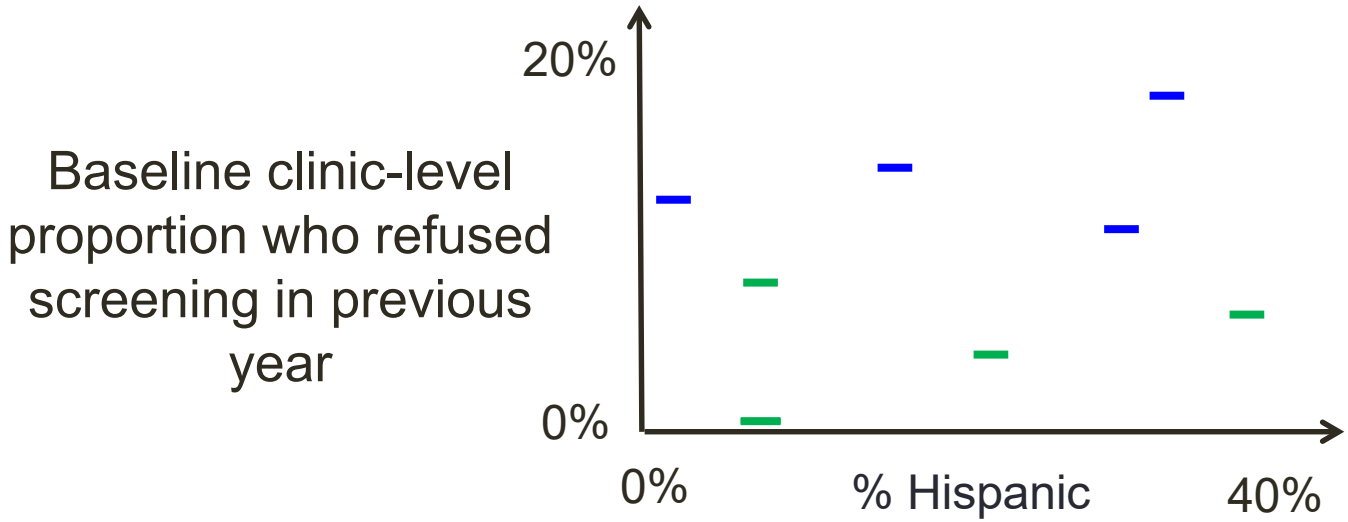


On average, % Hispanic in control < % Hispanic in intervention (ie, not well-balanced) but reasonable balance on proportion who refused screening

Baseline covariate imbalance

Possible design solution 3: Constrained randomization

Another example of simple randomization to control & intervention conditions

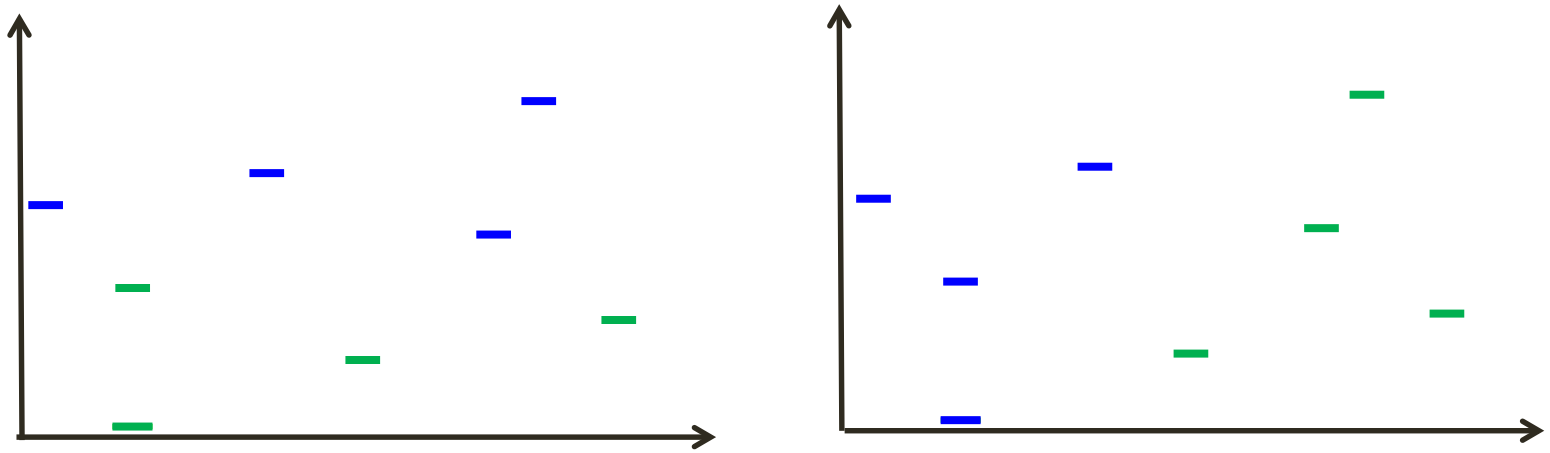


Not well-balanced on % refused screening but reasonable balance on % Hispanic

Baseline covariate imbalance

Possible design solution 3: Constrained randomization

Neither randomization has good balance of both covariates across trial arms.

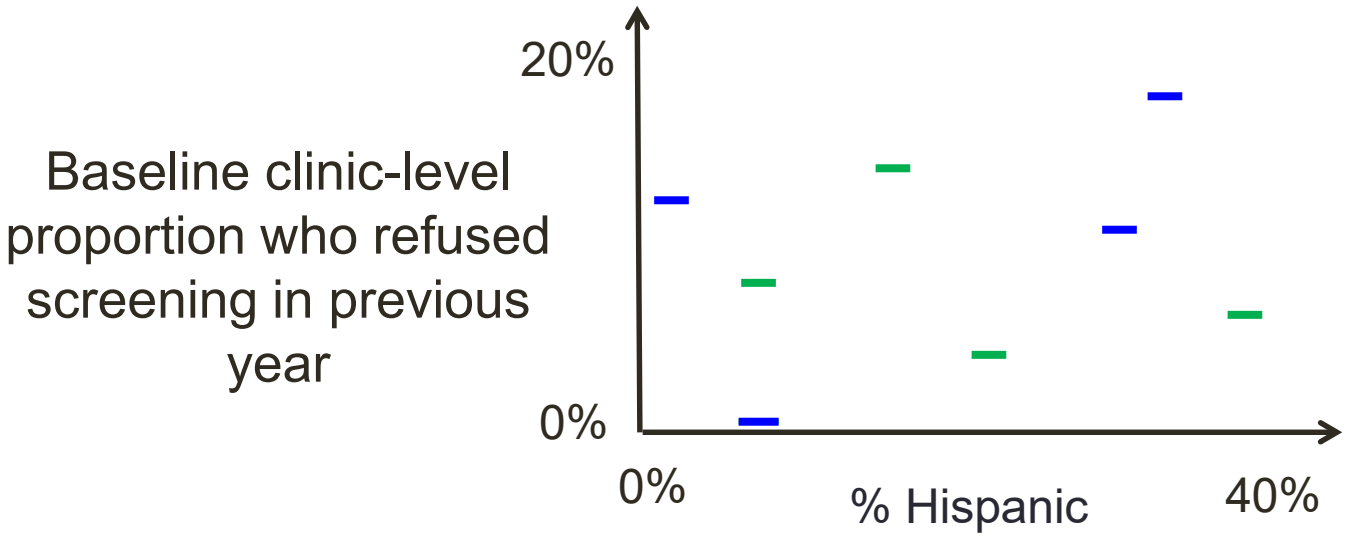


Solution: only allow randomizations that are “balanced enough” as measured by a “balance score” ie, use constrained randomization

Baseline covariate imbalance

Possible design solution 3: Constrained randomization

This randomization could be “balanced enough”



Work with a statistician!
Must account for the design in the analysis

Baseline covariate imbalance

Possible design solution 3: Constrained randomization

- More general than stratification
- Can include more cluster-level covariates
- Both continuous and categorical covariates
- Example:
  - % Hispanic
  - % refused screening in previous year
  - Rural/urban
- Measure “balanced enough” with a balance metric (no details here – use statistical rationale)

Baseline covariate imbalance

Possible design solution: Restricted randomization

- Three types of restricted randomization
  - Pair-matching
  - Stratification (sort-of a special case of CCR)
  - Covariate-constrained randomization (CCR)
- Recommendation
  - Use restricted randomization if total # clusters < 40 and know of predictive baseline covariates
  - Avoid pair-matching (for statistical reasons)
- In practice, analysis must account for whatever type of restricted randomization is used in design

Baseline covariate imbalance

Example: Restricted randomization

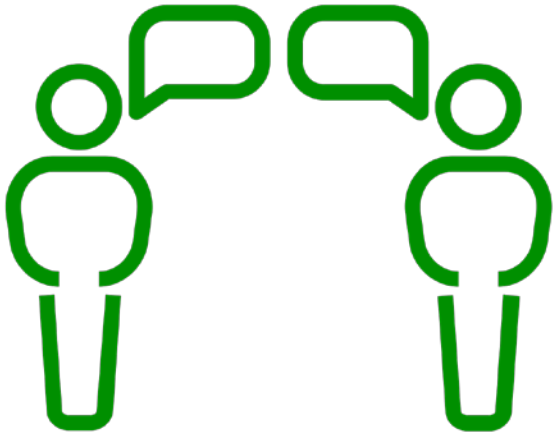
- For STOP CRC:
  - Used stratification by “clinic organization”
    - So “*each organization will have both intervention and control clinics*”
  - Considered using constrained randomization, but:
    - “*unpublished simulation models suggested that, for our relatively limited number of clusters, this approach might underperform relative to simple randomization*”

(If you are planning a cluster-randomized design)

What cluster-level covariates might be important to balance on?

2 min

4 min



Special considerations for CRTs

1. Clustering of outcomes
   - Clustering (of a particular outcome)
   - Accounting for clustering in analysis
   - Accounting for clustering in design
2. Small # of clusters
   - Potential for baseline covariate imbalance
   - How small is too small?

Few clusters: How low can you go?

- CONSORT extension for cluster RCTs
  - Recommends at least 4 clusters/arm
  - This is just a guide
- Statistical reasons may require much more than 8 clusters in total in a 2-arm trial!
- Remember: # clusters drives the power of trial more so than # participants
- CRTs require a lot of time and effort
  - Consider a pilot trial to get procedures in place*

* https://pilotfeasibilitystudies.biomedcentral.com/

Overview

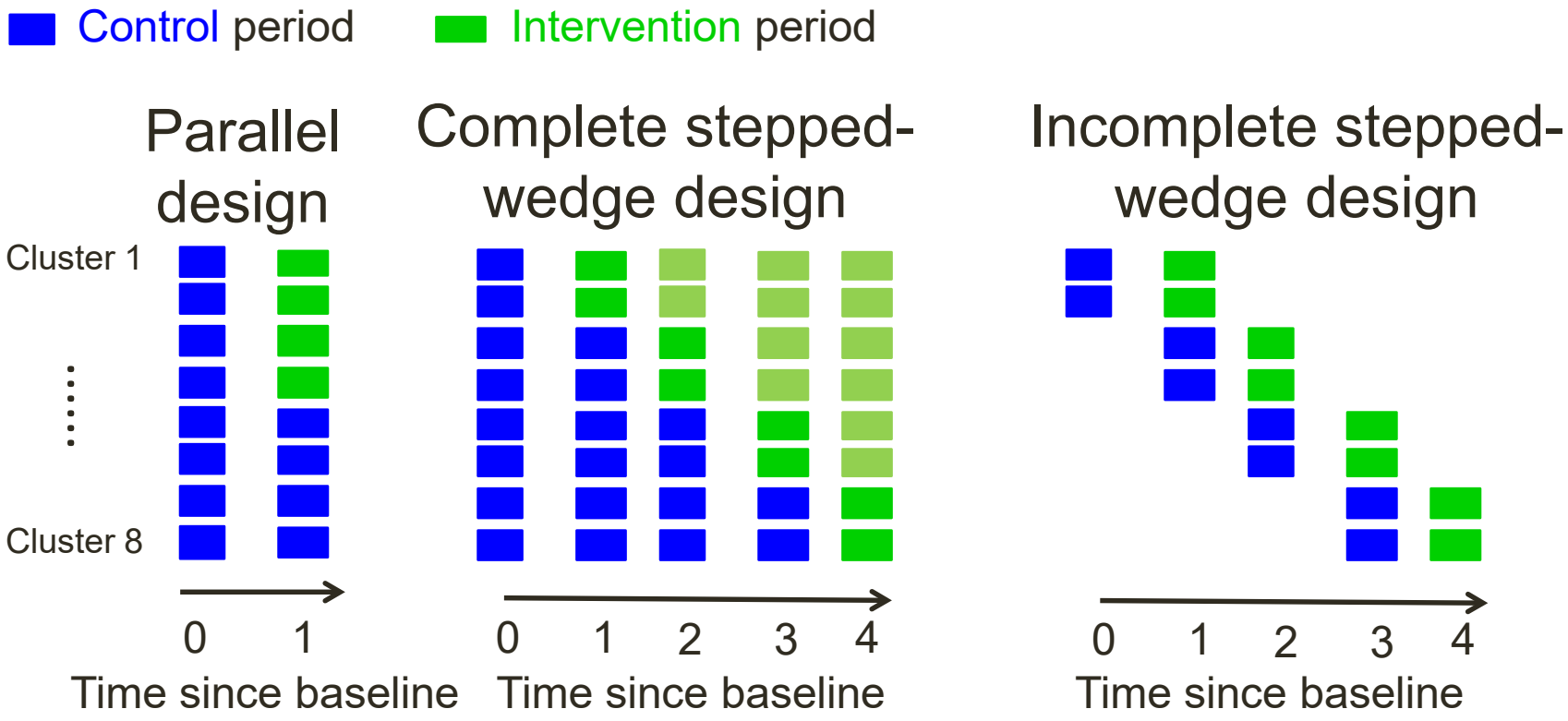
- Randomization schemes: cluster vs individual
- Cluster-randomized trials (CRT)
  - 1: Special considerations for CRTs
    - Clustering
    - Small # of clusters
  - 2: Varieties of cluster-randomized trials
    - Parallel
    - Stepped-wedge
- Other considerations
- How do I know I have the right statistician?

Varieties of CRT

1. Parallel
2. Stepped-wedge

Varieties of CRT

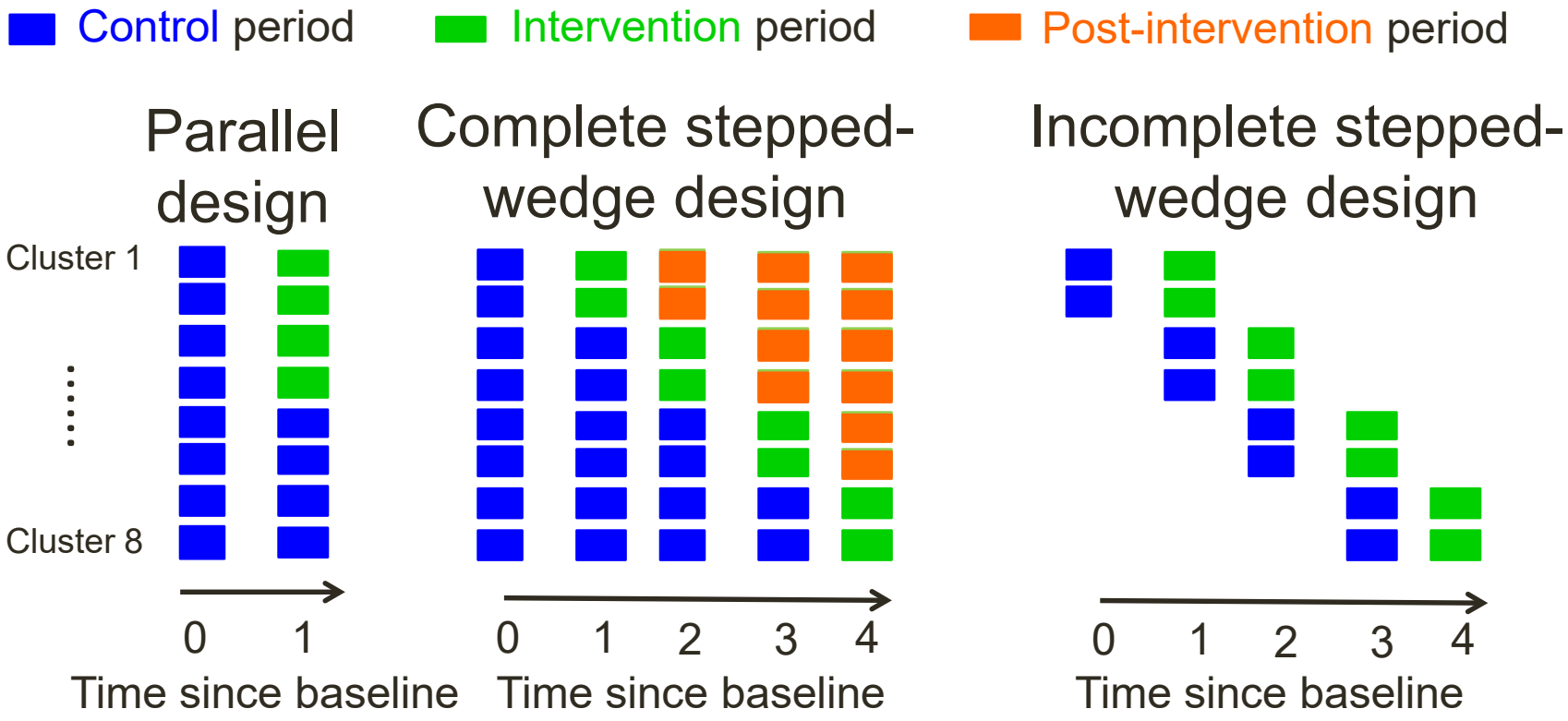
Examples with 8 clusters: 1-year intervention



Based on: Hemming (2015) Stat Med

Varieties of CRT

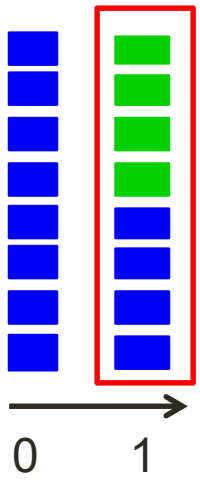
Examples with 8 clusters: 1-year intervention



Based on: Hemming (2015) Stat Med

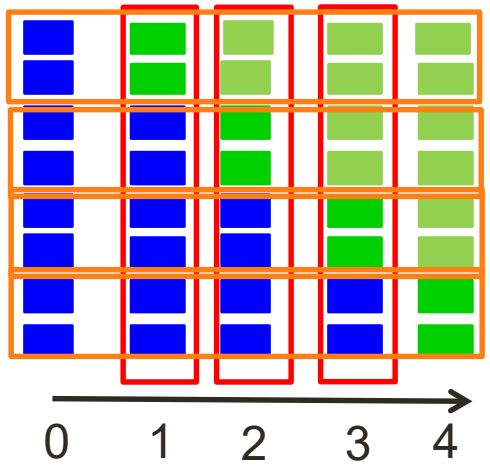
CRT analysis: treatment effects

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



Parallel design

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

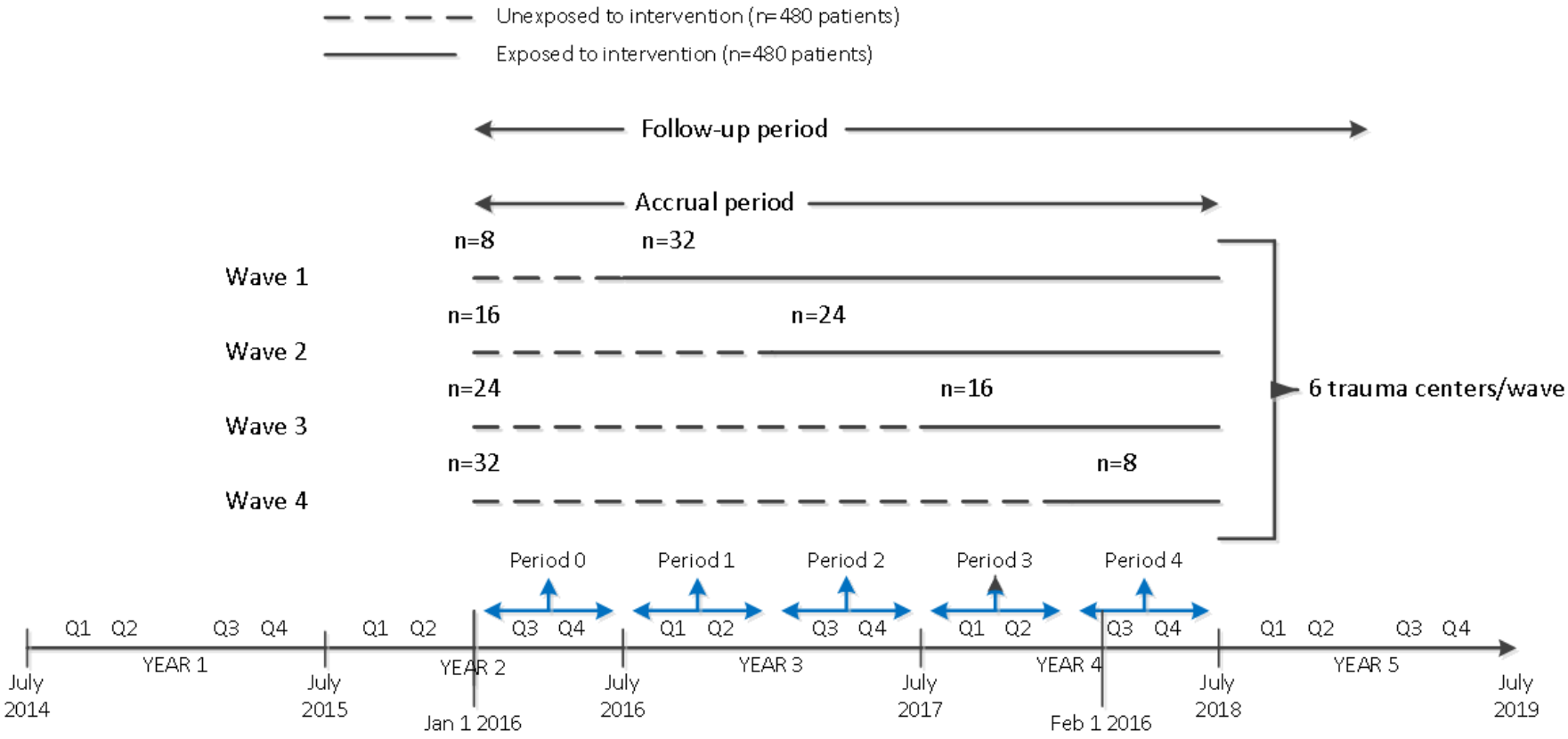


Complete SW design

Control period Intervention period

Based on: Hemming (2015) Stat Med

TSOS: SW-CRT



Zatzick D et al. Implementation Science 2016;11:58

Choosing CRT type: parallel vs SW

- Arguments **for** SW-CRT:
  - Can't immediately implement intervention in ½ clusters (eg, TSOS)
  - Pragmatic research: plan to implement in all clusters
  - Have few clusters + might gain power in SW-CRT

- Arguments **against** SW-CRT:
  - Risk confounding treatment effect with time effect
  - Could do staggered-start parallel-CRT if can't start implementation in ½ clusters immediately
    - Roll out to all clusters at end of evaluation, if effective

See: Hargreaves et al. Trials (2015). Five questions to consider before conducting a stepped wedge trial

Choosing CRT type: parallel vs SW

Statistical recommendations:

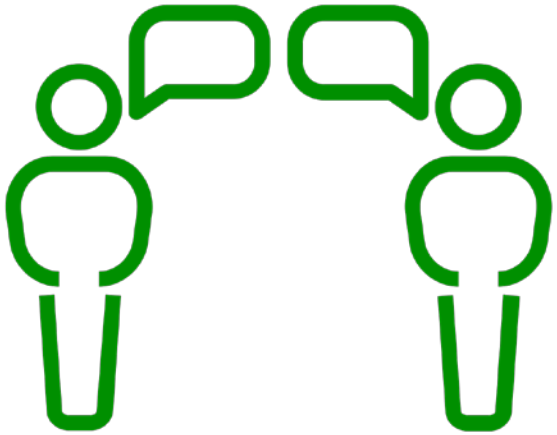
- Use a parallel CRT design if you can
- If not, plan for time effects in designing & analyzing SW-CRT
- Work with statistician to account for clustering in design & analysis of both designs

See: Hargreaves et al. Trials (2015). Five questions to consider before conducting a stepped wedge trial

(If you are planning a cluster-randomized design)
What are the pros and cons of using a parallel vs stepped-wedge design for your trial?

2 min

4 min



Overview

- Randomization schemes: cluster vs individual
- Cluster-randomized trials (CRT)
  - 1: Special considerations for CRTs
    - Clustering
    - Small # of clusters
  - 2: Varieties of cluster-randomized trials
    - Parallel
    - Stepped-wedge
- Other considerations
- How do I know I have the right statistician?

Other considerations for ePCTs

1. ITT vs PP analysis
2. Blinding and concealment
3. Monitoring and managing unexpected changes

Other considerations for ePCTs

1. ITT vs PP analysis
2. Blinding and concealment
3. Monitoring and managing unexpected changes

Intent-to-treat vs per protocol analysis

- Pragmatic nature → ITT commonly used
- PP often difficult to define
  - Screening yes/no is easy
  - Other interventions might have degrees of adherence to protocol
- Might be interested in other types of treatment effect
  - Average treatment effect on the treated

Other considerations for ePCTs

1. ITT vs PP analysis
2. Blinding and concealment
3. Monitoring and managing unexpected changes

ePCTs: blinding & concealment

- Concealment of randomization assignment to avoid selection bias
  - Less a problem in CRTs than RCTs if clusters all randomized together
- Blinding (masking)
  - May not be possible or practicable for CRTs
  - Objective assessment criteria should be consistently applied

Other considerations for ePCTs

1. ITT vs PP analysis
2. Blinding and concealment
3. Monitoring and managing unexpected changes

ePCTs: managing unexpected changes

- Study designs can be affected by:
  - Changes in study populations
  - Changes in coverage patterns
  - Changes in patient perceptions/decisions
  - Decisions by hospital/health system leadership
  - Changes in regulations or practice standards
  - Site turnover
- See examples of implications of ACA on STOP CRC (Vollmer et al, 2015)
- Careful planning and monitoring are needed

See: dx.doi.org/10.13063/2327-9214.1200

Overview

- Randomization schemes: cluster vs individual
- Cluster-randomized trials (CRT)
  - 1: Special considerations for CRTs
    - Clustering
    - Small # of clusters
  - 2: Varieties of cluster-randomized trials
    - Parallel
    - Stepped-wedge
- Other considerations
- How do I know I have the right statistician?

How do I know I have the right statistician?

- Someone who…
  - Wants to be involved from beginning of development of research proposal
  - Has experience with pragmatic trials & is familiar with the PRECIS tool
  - Has experience of EHR data?
  - Has experience of CRT design & analysis (if using a clustered design)

Important things to know

- Question drives design; design drives analysis
- Randomization
  - Individual preferred (for stat. reasons)
  - But cluster often needed (ie, a CRT)
- Considerations in both design and analysis
  - **Must** account for clustering (if CRT)
  - Best to account for baseline imbalance
- Good design is difficult, but critical
  - Need input from diverse team
  - Analysis may not be able to overcome design flaws

Important things to do

- Focus on the research question
- Collaborate with a faculty statistician – even when developing research question
- Choose individual randomization (but **only** if possible and defensible)
- Select design features with analysis in mind
- Weigh statistical choices vs implementation challenges
- Write a protocol paper and publish it!