ePCT Design and Analysis

Patrick Heagerty, PhD
University of Washington

Essentials of Embedded Pragmatic Clinical Trials
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 (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

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

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)

- Intervention
- Factors related to uptake of screening

**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

Screened
Not screened
>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

- Screened
- Not screened
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

- Screened
- Not screened
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

\[ r = \frac{s_{B}^2}{s_{B}^2 + s_{W}^2} = \frac{B}{B + W} = \frac{B}{Total} \]

Involves both *between-cluster* and *within-cluster* variance
In the Living Textbook: ICC cheat sheet

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
Accounting for clustering requires larger sample for adequate power

- Power and detectable difference is affected by...
  - Strength of the clustering effect (e.g., 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

- Detectable difference (SD units)
- # patients/cluster
Impact of increasing # clusters

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

Detectable difference (SD units)

# patients/cluster

# clusters per arm

- 2
- 4
- 8
- 16
- 32
Impact of increasing # clusters

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

Detectable difference (SD units)

Total # clusters = 4

# patients/cluster

# clusters per arm

- 2
- 4
- 8
- 16
- 32
Impact of increasing # clusters

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

![Graph showing impact of increasing clusters]

- Detectable difference (SD units)
- # patients/cluster
- # clusters per arm:
  - Red: 2
  - Blue: 4
  - Green: 8
  - Pink: 16
  - Orange: 32

Total # clusters = 4
Total # clusters = 8
Impact of increasing # clusters

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

Detectable difference (SD units)

<table>
<thead>
<tr>
<th># clusters per arm</th>
<th>2</th>
<th>4</th>
<th>8</th>
<th>16</th>
<th>32</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total # clusters</td>
<td>4</td>
<td>8</td>
<td>64</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

# patients/cluster
Impact of increasing # clusters

Example: CRT with smaller ICC=0.01 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

NIH Collaboratory ePCT: LIRE

Types of CRT designs

- Parallel
- Stepped-wedge
Types of CRT designs

- Parallel
- Stepped-wedge
  - Complete
  - Incomplete

In complete designs, measurements are taken from every cluster at every time point. In incomplete designs, some clusters do not provide measurements at all time points.
Types of CRT designs

Examples with 8 clusters: 1-year intervention

- **Control** period
- **Intervention** period

**Parallel design**

- Cluster 1
- ...
- Cluster 8

Time since baseline

Types of CRT designs

Examples with 8 clusters: 1-year intervention

- Control period
- Intervention period

Parallel design

Cluster 1

Cluster 8

Time since baseline

May have baseline outcomes

Types of CRT designs

Examples with 8 clusters: 1-year intervention

- **Control period**
- **Intervention period**

**Parallel design**

Cluster 1

Cluster 8

Time since baseline

**Incomplete stepped-wedge design**

Time since baseline

Types of CRT designs

Examples with 8 clusters: 1-year intervention

- Complete stepped-wedge design
- Incomplete stepped-wedge design
- Parallel design

Types of CRT designs

Examples with 8 clusters: 1-year intervention

- **Parallel design**
- **Complete stepped-wedge design**
- **Incomplete stepped-wedge design**

- **Cluster 1**
- **Cluster 8**

Time since baseline:
**CRT analysis: treatment effects**

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

- **Parallel design**
- **Complete SW design**

*Control period*  
*Intervention period*
CRT analysis: treatment effects

Estimated (primarily) using between-cluster information, i.e., vertical information

Parallel design

Complete SW design

Control period  Intervention period
CRT analysis: treatment effects

Estimated (primarily) using between-cluster information

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

Parallel design

Complete SW design

Control period  Intervention period
CRT analysis: treatment effects

Estimated (primarily) using between-cluster information

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

Parallel design

Complete SW design

- Control period
- Intervention period
**CRT analysis: treatment effects**

<table>
<thead>
<tr>
<th>Time since baseline</th>
<th>Control period</th>
<th>Intervention period</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Estimated (primarily) using between-cluster information

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

Parallel design

Complete SW design

- **Control period**
- **Intervention period**
CRT analysis: treatment effects

Estimated (primarily) using between-cluster information

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

Parallel design

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

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

Yes
Choosing study design

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

Examples with clinic/health-system-level interventions:
- STOP CRC colorectal cancer screening CRT
- LIRE lumbar imaging trial SW-CRT
Choosing study design

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

Is there a strong rationale for rolling out the intervention to all clusters before the end of the trial?
Choosing study design

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

Yes

Is there a strong rationale for rolling out the intervention to all clusters before the end of the trial?

No

STOP CRC colorectal cancer screening CRT

CRT
Choosing study design

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

Is there a strong rationale for rolling out the intervention to all clusters before the end of the trial? Yes

LIRE lumbar imaging SW-CRT
Choosing study design

Is there a strong rationale for randomizing groups/clusters rather than individuals to study conditions? No
Choosing study design

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

No

Examples with individual-level randomization:
• SPOT suicide prevention RCT
• OPTIMUM mindfulness for back-pain RCT
Choosing study design

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

No

Do participants receive their treatment in a group format or from a shared interventionist?
Choosing study design

- Is there a strong rationale for randomizing groups/clusters rather than individuals to study conditions? No
- Do participants receive their treatment in a group format or from a shared interventionist? No

SPOT suicide prevention RCT
Intervention is targeted at the individual
Choosing study design

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

No

Do participants receive their treatment in a group format or from a shared interventionist?

Yes

Cluster must be accounted for in both design and analysis

Individually-randomized group treatment (IRGT) trial
Choosing study design

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

- No
  
  Do participants receive their treatment in a group format or from a shared interventionist?

- Yes
  
  Clustering must be accounted for in both design and analysis

Individually-randomized group treatment (IRGT) trial

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

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

No

Do participants receive their treatment in a group format or from a shared interventionist?

No

RCT

Yes

IRGT trial

Is there a strong rationale for rolling out the intervention to all clusters before the end of the trial?

Yes

SW-GRT

No

GRT

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

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

No

Do participants receive their treatment in a group format or from a shared interventionist?

No

RCT

Yes

IRGT trial

Is there a strong rationale for rolling out the intervention to all clusters before the end of the trial?

Yes

SW-GRT

Yes

No

GRT

Clustering must be accounted for in both design and analysis

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


NIH resources

- Pragmatic and Group-Randomized Trials in Public Health and Medicine
  - [https://prevention.nih.gov/grt](https://prevention.nih.gov/grt)
  - 7-part online course on GRTs and IRGTs
- Mind the Gap Webinars
  - 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/](https://researchmethodsresources.nih.gov/)
  - Material on GRTs and IRGTs and a sample size calculator for GRTs.
Extra slides
Higher ICC = lower power

Example: CRT with 24 clinics & 40 patients/clinic & alpha = 2.5%

Standardized effect size i.e. detectable difference (SD units)
Higher ICC = lower power

Example: CRT with 24 clinics & 40 patients/clinic & alpha = 2.5%

Standardized effect size i.e. detectable difference (SD units)
Higher ICC = Lower Power

Example: CRT with 24 clinics & 40 patients/clinic & alpha = 2.5%

- Power
- Standardized effect size i.e. detectable difference (SD units)
Higher ICC = lower power

Example: CRT with 24 clinics & 40 patients/clinic & alpha = 2.5%

- RCT @ 80% power to detect effect size of approx. 0.2.
- CRT of same size with ICC of 0.03 approx. 40% power
Special considerations for CRTs

• Baseline covariate imbalance
• How many clusters is enough?
• Analytic approaches
Special consideration: baseline covariate imbalance

- Pragmatic CRTs often enroll small # of clusters (<40)
- Randomization may not balance baseline covariates
- Baseline covariate imbalance threatens internal validity; ie, comparability of treatment arms
  - There may be confounding due to noncomparability of treatment arms
Addressing imbalance in baseline covariates

- Prevent imbalance at design stage
- Only adjust for imbalance post-hoc at analysis stage
Addressing imbalance in baseline covariates

- Prevent imbalance at design stage
- Only adjust for imbalance post-hoc at analysis stage

Not recommended
Addressing imbalance in baseline covariates

Prevent imbalance at design stage

Recommended

Restricted randomization + adjustment for balancing variables in analysis

Only adjust for imbalance post-hoc at analysis stage

Not recommended
Using restricted randomization

• Use restricted randomization if
  • Total # clusters <40, and…
  • Know which baseline covariates are predictive of outcome

• Multiple approaches possible
  • Pair-matching
  • Stratification
  • Covariate-constrained randomization
  • Consult a statistician to choose!

• Analysis must account for whatever type of restricted randomization is used in design
In the Living Textbook: pair-matching and stratification in CRTs

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
Consider…

• If you are planning a cluster randomized design, what cluster-level covariates might be important to balance on?
Special consideration: 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 many more than 8 clusters in total in a 2-arm trial!
- # clusters drives the power of trial more than # participants
- CRTs require a lot of time and effort
  - Consider a pilot trial to get procedures in place
Special consideration: analytic approaches for clustered data

• Typically use regression models to analyze individual-level data
  • Random effects/mixed effects models
  • Generalized estimating equations
• Important
  • Work with a statistician to ensure correct accounting for clustering
Other considerations for ePCTs

• Intent-to-treat (ITT) versus per-protocol analysis
• Concealment and blinding
• Monitoring and managing unexpected changes
Intention-to-treat vs per-protocol analysis

- Pragmatic nature → ITT commonly used
- Per protocol 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
Concealment and blinding

- 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
Managing unanticipated 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
- Careful planning and monitoring are needed
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 and is familiar with the PRECIS-2 tool
• Has experience with using EHR data
• Has experience with CRT design and analysis (if using a clustered design)