Improving Pragmatic Clinical Trials: Lessons Learned from the NIH Collaboratory Biostatistics Core

Andrea J Cook, PhD

Associate Investigator
Biostatistics Unit, Group Health Research Institute
Affiliate Associate Professor
Dept. of Biostatistics, University of Washington

June 18, 2015
NIH Collaboratory
Acknowledgements

NIH Collaboratory Coordinating Center Biostatisticians
   Elizabeth Delong, PhD, Andrea Cook, PhD, and Lingling Li, PhD

NIH Collaboratory Project Biostatisticians
   Patrick Heagerty, PhD, Bryan Comstock, MS, Susan Shortreed, PhD, Ken Kleinman, PhD, and William Vollmer, PhD

NIH Methodologist
   David Murray, PhD

Funding
   This work was supported by the NIH Health Care Systems Research Collaboratory (U54 AT007748) from the NIH Common Fund.
NIH Collaboratory Pragmatic Trial Setting

UH2 Phase: What did we do?
- Common themes across studies
- How were the trials improved?

What are we doing now?
- UH3 Phase Issues
- New UH2 Trials
- Unanswered Questions?
A pragmatic–explanatory continuum indicator summary (PRECIS): a tool to help trial designers

Kevin E. Thorpe MMath, Merrick Zwarenstein MD MSc, Andrew D. Oxman MD, Shaun Treweek BSc PhD, Curt D. Furberg MD PhD, Douglas G. Altman DSc, Sean Tunis MD MSc, Eduardo Bergel PhD, Ian Harvey MB PhD, David J. Magid MD MPH, Kalipso Chalkidou MD PhD

Published at www.cmaj.ca on Apr. 16, 2009. An abridged version of this article appeared in the May 12 issue of CMAJ. This article was published simultaneously in the May 2009 issue of the *Journal of Clinical Epidemiology* (www.jclinepi.com).
Pragmatic vs. Explanatory Trials

- Pragmatic expertise (comparison)
- Practitioner expertise (experimental)
- Eligibility criteria
- Follow-up intensity
- Outcomes
- Participant compliance
- Practitioner adherence
- Primary analysis
- Flexibility of the comparison intervention
- Flexibility of the experimental intervention
- C

- D
Key features of most PCTs

Use of electronic health records (EHRs)
- EHRs allow efficient and cost-effective, recruitment, participant communication & monitoring, data collection, and follow up

Randomization at clinic or provider level
- Protocols can be tailored to local sites and can adapt to changes in a dynamic health care environment
Pragmatic Trials Concept

- **Size**: Large simple trials → precise estimates, evaluate heterogeneity

- **Endpoints**: patient oriented usually with minimal adjudication

- **Setting**: integrated into real world
  - Non-academic centers
  - Leverage electronic data
  - Patients as partners
NIH Collaboratory Pragmatic Trial Setting

UH2 Phase: What did we do?
- Common themes across studies
- How were the trials improved?

What are we doing now?
- UH3 Phase Issues
- New UH2 Trials
- Unanswered Questions?
## Round 1 Demonstration Projects

<table>
<thead>
<tr>
<th>Principal Investigator</th>
<th>Institution</th>
<th>Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gloria Coronado</td>
<td>Kaiser Foundation Research Institute</td>
<td>Strategies and Opportunities to Stop Colon Cancer in Priority Populations</td>
</tr>
<tr>
<td>Lynn DeBar</td>
<td>Kaiser Foundation Research Institute</td>
<td>Collaborative Care for Chronic Pain in Primary Care</td>
</tr>
<tr>
<td>Laura Dember</td>
<td>University of Pennsylvania</td>
<td>Pragmatic Trials in Maintenance Hemodialysis</td>
</tr>
<tr>
<td>Susan Huang</td>
<td>University of California--Irvine</td>
<td>Decreasing Bioburden to Reduce Healthcare-Associated Infections and Readmissions</td>
</tr>
<tr>
<td>Jeffrey Jarvik</td>
<td>University of Washington</td>
<td>A Pragmatic Trial of Lumbar Image Reporting with Epidemiology (LIRE)</td>
</tr>
<tr>
<td>Gary Rosenthal</td>
<td>University of Iowa</td>
<td>Nighttime Dosing of Anti-Hypertensive Medications: A Pragmatic Clinical Trial</td>
</tr>
<tr>
<td>Gregory Simon</td>
<td>Group Health Cooperative</td>
<td>Pragmatic trial of population-based programs to prevent suicide attempt</td>
</tr>
</tbody>
</table>
STUDY DESIGN
Mostly Cluster RCTs (except one)
- Randomization Unit:
  - Provider < Panel < Clinic < Region < Site

Average Size of Cluster
- Initial Proposals: Most large clinic level clusters
- Goal: Smallest Unit without contamination
  - More clusters are better if possible
- Smaller number of clusters increase sample size along with estimation issues (GEE)
- Potential Solutions: Panel-level or physician-level
Study Design: Variable Cluster Size

- **Variable Cluster Size**
  - Sample Size calculations need to take this into account
    - Design effects are different
    - Depends on the analysis choice
  - Analysis Implications: What are you making inference to?
    - Cluster vs Patient vs Something in-between
    - Marginal versus conditional estimates


Cluster Design: Which Cluster Design?

- **Cluster**
  - Randomize at cluster-level
  - Most common, but not necessarily the most powerful or feasible
  - Advantages:
    - Simple design
    - Easy to implement
  - Disadvantages:
    - Need a large number of clusters
    - Not all clusters get the interventions
  - Interpretation for binary and survival outcomes:
    - Mixed models within cluster interpretation problematic
    - GEE marginal estimates interpretation, but what if you are interested in within cluster changes?
Cluster with Cross-over

- Randomize at cluster but cross to other intervention assignment midway
- Feasible if intervention can be turned off and on without “learning” happening
- Alternative: baseline period without intervention and then have half of the clusters turn on
## Study Design: Which Cluster Design?

<table>
<thead>
<tr>
<th>Cluster Description</th>
<th>Period 1</th>
<th></th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple Cluster</td>
<td>1</td>
<td>INT</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>UC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>UC</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>INT</td>
<td></td>
</tr>
<tr>
<td>Cluster With Crossover</td>
<td>1</td>
<td>INT</td>
<td>UC</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>UC</td>
<td>INT</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>UC</td>
<td>INT</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>INT</td>
<td>UC</td>
</tr>
<tr>
<td>Cluster With Baseline</td>
<td>1</td>
<td>UC</td>
<td>INT</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>UC</td>
<td>UC</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>UC</td>
<td>UC</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>UC</td>
<td>INT</td>
</tr>
</tbody>
</table>
Cluster with Cross-over

Advantages:
- Can make within cluster interpretation
- Potential to gain power by using within cluster information

Disadvantages:
- Contamination can yield biased estimates especially for the standard cross-over design
- May not be feasible to switch assignments or turn off intervention
- Not all clusters have the intervention at the end of the study
Study Design: Which Cluster Design?

- **Stepped Wedge Design**
  - Randomize timing of when the cluster is turned **on** to intervention
  - Staggered cluster with crossover design
  - Temporarily spaces the intervention and therefore can control for system changes over time
# Study Design: Which Cluster Design?

## Stepped Wedge

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Baseline</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>UC</td>
<td>INT</td>
<td>INT</td>
<td>INT</td>
<td>INT</td>
</tr>
<tr>
<td>2</td>
<td>UC</td>
<td>UC</td>
<td>INT</td>
<td>INT</td>
<td>INT</td>
</tr>
<tr>
<td>1</td>
<td>UC</td>
<td>UC</td>
<td>UC</td>
<td>INT</td>
<td>INT</td>
</tr>
<tr>
<td>4</td>
<td>UC</td>
<td>UC</td>
<td>UC</td>
<td>UC</td>
<td>INT</td>
</tr>
</tbody>
</table>
Study Design: Which Cluster Design?

- **Stepped Wedge Design**
  - **Advantages:**
    - All clusters get the intervention
    - Controls for external temporal trends
    - Make within cluster interpretation if desired
  - **Disadvantages:**
    - Contamination can yield biased estimates
    - Heterogeneity of Intervention effects across clusters can be difficult to handle analytically
    - Special care of how you handle random effects in the model
    - Relatively new and available power calculation software is relatively limited
RANDOMIZATION
Randomization

- Crude randomization not preferable with smaller number of clusters or need balance for subgroup analyses

- How to balance between cluster differences?
  - Paired
    - How to choose the pairs best to control for important predictors?
    - Implications for analyses and interpretation
  - Stratification
    - Stratify analysis on a small set of predictors
    - Can ignore in analyses stage if desired

- Other Alternatives

Randomization: Constrained Randomization

- Balances a large number of characteristics
- Concept
  1. Simulate a large number of cluster randomization assignments (A or B but not actual treatment)
  2. Remove duplicates
  3. Across these simulated randomizations assignments assess characteristic balance
  4. Restrict to those assignments with balance
  5. Randomly choose from the restricted pool a randomization scheme.
  6. Randomly assign treatments to A or B
Is Constrained randomization better than unconstrained randomization?

How many valid randomization schemes do you need to be able to conduct valid inference?

Do you need to take into account randomization scheme in analysis?
- Ignore Randomization
- Adjust for variables in regression
- Permutation inference

=> Conduct a simulation study to assess these properties
Randomization: Constrained Randomization Simulation Design

- **Outcome Type:** Normal
- **Randomization Type:** Simple versus Constrained
- **Inference Type:** Exact (Permutation) versus Model-Based (F-Test)
- **Adjustment Type:** Unadjusted versus Adjusted
- **Clusters:** Balanced designs, but varied size and number
- **Correlation:** Varied ICC from 0.01 to 0.05
- **Potential Confounders:** Varied from 1 to 10

Randomization: Constrained Randomization Simulation Results

- Adjusted F-test and the permutation test perform similar and slightly better for constrained versus simple randomization.

- Under Constrained Randomization:
  - Unadjusted F-test is conservative
  - Unadjusted Permutation holds type I error (unless candidate set size is not too small)
  - Unadjusted Permutation more powerful than Unadjusted F-Test

- Recommendation: Constrained randomization with enough potential schemes (>100), but still adjust for potential confounders
What about Binary and Survival Outcomes??

Hypothesized Results (Mine not NIH Collaboratories):

- Constrained Randomization probably still wins
- Binary Outcomes: Likely less of a preference for adjusted versus unadjusted analyses (mean and variance relationship (minimal precision gains))
- Survival Outcomes: Depends on scenario and model choice (frailty versus robust errors)
OUTCOME

ASCERTAINMENT
Most trials use Electronic Healthcare Records (EHR) to obtain Outcomes
- Data **NOT** collected for research purposes

If someone stays enrolled in healthcare system - assume that if you don’t observe the outcome it didn’t happen
- In closed system this is likely ok
- Depends upon cost of treatment (likely to get a bill the more the treatment costs)
Do you need to validate the outcomes you do observe?
- Depends on the Outcome (PPV, sensitivity)
- Depends on the cost (two-stage design?)

How do you handle Missing Outcome Data?
- Leave healthcare system
  - Type of Missing Data: Administrative missingness (MCAR), MAR or non-ignorable?
  - Amount of Missing Data: how stable is your population being studied?
- Depends on the condition and population being studied.

NIH Collaboratory Pragmatic Trial Setting

UH2 Phase: What did we do?
- Common themes across studies
- How were the trials improved?

What are we doing now?
- Current UH3 Phase Issues
- New UH2 Trials
- Unanswered Questions?
Submitted new UH3 proposals last summer
- New design choices submitted
- Improved sample size calcs using pilot data collected in UH2 phase and modifications
- Improved and finalized analysis plans with feedback from all Collaboratory participants

Those funded moved to UH3 phase this Fall or Spring

Very early in the UH3 phase
- Most studies are already randomizing participants
- Some new issues have come up…
Are pragmatic clinical trials different?

- Depends on the study
- Main difference: how we collect, and timeliness of the collection, of adverse events and outcomes
- Formal Primary Outcome Monitoring
  - How do you handle the fact that you likely don’t have the validated outcome available in a timely manner?
  - IRB has restricted the population that the DSMB can monitor to those that receive the intervention in the intervention arm only (e.g. internet intervention if they passively refuse by not going to the website we can’t get their outcome data until the end of the study)
Data Safety Monitoring
Are pragmatic clinical trials different?

- Depends on the study
- Main difference: how we collect, and timeliness of the collection, of adverse events and outcomes
- Formal Primary Outcome Monitoring
  - How do you handle the fact that you likely don’t have the validated outcome available in a timely manner?
  - IRB has restricted the population that the DSMB can monitor to those that receive the intervention in the intervention arm only (e.g. internet intervention if they passively refuse by not going to the website we can’t get their outcome data until the end of the study)
### New UH2’s

<table>
<thead>
<tr>
<th>Principal Investigator</th>
<th>Institution</th>
<th>Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mor, Vincent; Volandes, Angelo; Mitchell, Susan</td>
<td>Brown University School of Medicine</td>
<td>Pragmatic Trial of Video Education in Nursing Homes</td>
</tr>
<tr>
<td>Vazquez, Miguel</td>
<td>UT Southwestern Medical Center</td>
<td>Improving Chronic Disease Management with Pieces (ICD-Pieces)</td>
</tr>
<tr>
<td>Zatzick, Douglas</td>
<td>University of Washington</td>
<td>A Policy-Relevant U.S. Trauma Care System Pragmatic Trial for PTSD and Comorbidity (Trauma Survivors Outcomes and Support [TSOS])</td>
</tr>
</tbody>
</table>
Conclusions

Pragmatic Trials are important to be able to move research quickly into practice

Pragmatic Trials add Complication

- First Question: Can this study be answered using a pragmatic trial approach??
- Study Design is essential and needs to be flexible
- Using EHR data is valuable, but understanding the performance of all measures is important
- Appropriate analysis taking into account design, randomization, and outcome ascertainment is key

Lot's of open statistical questions still to be addressed