Designing Cluster Randomized Trials for Health Policy Research

8th International Conference on Health Policy Statistics [ICHPS]
Washington, DC: January 20, 2010

Thomas E. Love, Ph.D. Thomas.Love@case.edu
Neal V. Dawson, M.D. nvd@case.edu
Randall D. Cebul, M.D. rdc@case.edu

Center for Health Care Research and Policy
Case Western Reserve University at MetroHealth Medical Center
Cleveland, Ohio www.chrp.org

Workshop Schedule

<table>
<thead>
<tr>
<th>Time</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3:30 – 3:45</td>
<td>Introduction to EMR-Centered Decision Support and Design of a Cluster Randomized Trial [Love, for Cebul]</td>
</tr>
<tr>
<td>3:45 – 4:00</td>
<td>Group Task: Designing a Cluster Randomized Trial</td>
</tr>
<tr>
<td>4:00 – 4:15</td>
<td>Discussion of Group Task / Strategic Concerns [Group]</td>
</tr>
<tr>
<td>4:15 – 4:40</td>
<td>Ethics and Implementation / “When Stuff Happens” [Dawson]</td>
</tr>
<tr>
<td>4:40 – 5:05</td>
<td>A Glimpse at Some Analytic Concerns [Love]</td>
</tr>
<tr>
<td>5:05 – 5:15</td>
<td>Additional Group Discussion and Workshop Evaluation</td>
</tr>
</tbody>
</table>
Group Task

Study: Improving the care and outcomes of patients with diabetes

Background: A health care system with a well-established electronic medical record (EMR) capable of providing various Clinical Decision Support (CDS) functions decides to undertake a controlled trial to determine whether CDS can improve the care and outcomes of its approximately 9,000 patients with diabetes. Selected characteristics of the system are described below, including the proportion of its patients by insurance class. The system asks for your help in planning and designing the study.

<table>
<thead>
<tr>
<th>Sites</th>
<th>Primary Care Providers</th>
<th>Diabetes Patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>115</td>
<td>8,804</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primary Insurance Class (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commercial / Medicare</td>
</tr>
<tr>
<td>Medicaid</td>
</tr>
<tr>
<td>Uninsured / Self-Pay</td>
</tr>
<tr>
<td>59</td>
</tr>
<tr>
<td>26</td>
</tr>
<tr>
<td>15</td>
</tr>
</tbody>
</table>

Task: As a group, you’ll have 15 minutes to select a reporter/scribe, discuss the key features of the trial that you would recommend they undertake, and then prepare a one-minute oral report for the rest of the workshop which describes your group’s views on the three concerns below, and any additional issues you come up with.

Specifically, it would be helpful to address the following concerns:

a. What would the intervention(s) be (i.e; what would you be comparing)?
b. What would the unit of assignment (to the interventions) be?
c. Identify other information that you would request in order to improve your recommendations.
A Few Terms/Acronyms We May Use

**CCHIT** Certification Commission for Healthcare Information Technology. [www.cchit.org](http://www.cchit.org/) Created to establish national standards in ambulatory and inpatient EMRs (functionality, interoperability, and security), CCHIT is a coalition of private sector organizations (the American Health Information Management Association, the Healthcare Information and Management Systems Society and The National Alliance for Health Information Technology; released list of first certified ambulatory EMR systems in July, 2006.

**CDS** Clinical Decision Support. In the context of this course, CDS is understood to mean "real-time" clinical decision support that is integrated with functions of electronic medical records systems.

**CRT** Cluster Randomized Trial. The subject of this course! In contrast to conventional RCT, in which the unit of assignment is typically the same (e.g., a patient) as the unit of analysis.

**CPOE** Computerized Physician Order Entry. A function of electronic medical records systems, may be integrated with clinical decision support functions to improve quality, safety, etc (e.g., warning of potential drug interactions).

**ICC** Intra-cluster Correlation Coefficient. Also described as rho (\(\rho\)), the ICC is the main metric of "clustering"; the amount of variance in a measure that is attributable to differences between clusters; ranges from 0-1. ICCs near zero imply virtually complete statistical independence of members across clusters in a CRT.

**MRC** United Kingdom's Medical Research Council. Established a classification system and made recommendations for types of informed consent in CRTs.
A Very Partial Bibliography on Methods for Cluster-Randomized Trials
Compiled by TE Love

Textbooks

Review and General Articles on CRTs
Designing CRTs: Epidemiologic Concerns (Power, etc.)


**Analyzing CRTs: Biostatistical Concerns**


**Ethical and Human Subjects Concerns in CRTs**


**Information on CRTs and Diabetes, CDS, and Related Fields**

Cluster Randomized Trials (CRTs) in Health Policy Research
(with a focus on CRTs of clinical decision support)

Thomas E. Love, PhD filling in for
Randall D. Cebul, MD
January 20, 2010
Center for Health Care Research & Policy
Case Western Reserve University at
MetroHealth Medical Center
rdc@case.edu

Cluster Randomized Trials?

As compared to individually randomized trials (RCTs),
• CRTs are more complex to design
• CRTs require more participants to obtain equivalent statistical power
• CRTs require more complex analysis

Overview of Workshop

• Introduction and Context
  – Electronic medical records (EMRs) to provide and examine clinical decision support (CDS)
• Group Task: Design a Study
• Task Review
• CRTs: Why they are necessary and challenging to execute
• Analytic and Strategic Concerns
• Discussion
Clinical Decision Support (CDS)

Consensus CDS Definition (AMIA): “Providing clinicians, patients, or individuals with knowledge and person-specific or population information, intelligently filtered or presented at appropriate times, to foster better health processes, better individual patient care, and better population health.”

Clinical Decision Support, informally

Giving the right information to the right person at the right time and place, making the right decision the easy decision.

Illustrative Real-time CDS – designed to change (improve) behavior

- Provider-centered CDS
  - Inpatient
    - Computerized Physician Order Entry, “Smart tools”, guidelines, etc.
  - Outpatient
    - Alerts, linked orders/referrals, links to patient/MD education, etc.
- Patient-centered
  - access to web-based records, reminders, input/feedback, etc.
**Encounter-based Alert to Improve care for Diabetic Patients with Leaking Kidneys**

(Links to Automated Order Set)

**What do we know about this patient?**
- She has diabetes and is visiting her PCP
- Her kidneys are leaking protein.
- She is not on an ACE inhibitor or ARB and has no documented allergies to them.
- She has no other contraindications (K, Cr)
- There are several alternative drugs/doses

---

**Automated Order Set**
**Linked to ACE/ARB Alert**

![Automated Order Set](image)

- Re-cap of indications
- Choice of Rx/doses
- Follow-up testing

---

**Comparative Performance Feedback**

![Comparative Performance Feedback](image)

- "My panel" vs. Comparator
CDS for Patients: Viewing their information, inputting data, getting feedback at home

CDS with EMRs: Untapped Promise?

"we conclude that effective EMR implementation...could save more than $81 billion annually - and that HIT-enabled management of chronic disease could eventually double those savings while increasing health benefits."

- Hillestad R et.al. (2004)

Or... increased mortality (with CPOE)?

- Pittsburgh Children's (Cerner)
  - 13 months pre-CPOE
  - 5 months post-CPOE
- 1,942 children admitted for specialty care
  - 75 deaths
    - 39 / 1394 pre-CPOE (2.80%)
    - 36 / 548 post (6.57%)
  - Multivariate OR 3.28 (CI 1.94 – 5.55)
- But: retrospective, short post-implementation...

- Han (2005)
Cluster Randomized Trials in Health Policy Research
Part One: Setting Up The Task of Designing A Study to Evaluate Clinical Decision Support

Need for Evidence

- What do we need to know?
  - Will providers adopt CDS?
  - Safety of implementation, System Effects
  - Safety and Quality Implications of Alerts
  - FPs, Alert Fatigue
  - Cost implications
    - Will they result in more cost-effective care?
    - What is most effective? Over what time horizon?
    - Etc – the possibilities are endless...
- How do we study CDS?

Group Tasks

1. Describe key features of a CDS-catalyzed trial to improve the care and outcomes of patients with diabetes.
2. Identify other information needed to improve the trial.

One-Minute Oral Reports

A health care system with an established electronic medical record (EMR) capable of providing various clinical decision support (CDS) functions decides to undertake a controlled trial to determine whether CDS can improve care and outcomes of its 8804 diabetic patients.

- 10 sites, 115 primary care providers, 8804 DM patients
- Insurance: 59% commercial or Medicare, 26% Medicaid, 15% Uninsured
### Task Review – Part A: Intervention
What Should We Be Comparing?

### Task Review – Part B: Unit of Assignment

### Task Review – Part C: Other Useful Information

---

### Comparing RCTs and CRTs

- **RCT (Randomized Controlled Trial)**
  - unit that is randomized is the unit of analysis
  - for large n, important attributes are likely to be distributed similarly across groups

- **CRT (Cluster-Randomized Trial)**
  - interventions focus on systems, prevention, behavior
  - useful when "contamination" is likely to be a problem and/or blinding is impossible
  - unit randomized is not the sole unit of interest
  - "clustering" within units is the critical issue
Patients within Physicians within Practices within Study Groups: A 4-Level CRT

Figure 10.3: A simple four-level cluster randomized trial

CRTs: Impact of Clustering

If there are important cross-group differences in important factors at baseline, this affects:

- Study Power [Effective Sample Size]
- Study Design:
  - Pre-randomization cluster “balancing” will improve balance of important prognostic factors
- Study Analysis:
  - Need analytic techniques that account for clustering e.g. GEE, hierarchical models, etc.
Ethics and Implementation

When CRT = Creative Responses to Trying Circumstances

“Stuff Happens”
Neal V. Dawson, MD
January 2010
nvd@case.edu

Ethical Issues in CRTs

• Wide spectrum of CRT designs
  – Can resemble individual RCTs where each subject decides whether to participate
  – May involve the randomization of whole practices, communities, or countries

UK MRC Clinical Trials Series

• Type A trials – are structured such that they do not allow participation decisions by individuals
  – Often CRTs

• Type B trials – allow individual subjects to decide about participation
  – Usually individual RCTs
Usual Ethical Concerns Apply

- Potential to produce findings that can improve human health or welfare
- Favorable balance of risks and benefits
- Conflicts: subjects' welfare prevails over interests of science and society
- Voluntary informed consent when possible
  - Alternative safeguards – cluster representation
- Alternative consent mechanisms
- Review by independent ethics review committee

Ethical Concerns for Type A Trials

- Mechanisms for representing the interests of the cluster – Representative (person or group)
  - Sufficient knowledge of circumstances, beliefs, and values of cluster
  - Delegated authority from or for the cluster
  - No conflicts of interest
- Cluster representative: ‘individual rights’
  - Suitably informed
  - Able to withdraw without adverse impact

Specific Trial Examples

- Many quality improvement studies
  - Interventions on groups rather than individuals
  - Interventions targeted at health care professionals
- Administrative trials
  - Studies that do not intrude into physician-patient decision making
  - Studies of aspects of health care activities about which patients never make decisions (e.g., patient scheduling schemes to improve patient flow)
Written Informed Consent

- May be impractical and produce important amounts of bias
  - Tu et al. NEJM 2004;350:1414-21
  - Attempted written informed consent of consecutive patients for stroke registry in Canada
  - Participation: Phase 1=39% (4285 eligible), Phase 2=51% (2823 eligible); Many died or were discharged before they could be approached
  - Selection bias: Inhospital mortality, Enrolled=6.9%; Not enrolled=21.7%;
  - Registry patients not representative

Alternative Consent Mechanisms

- Goldberg. Medical Care 1990;28:822-33
- Administrative trials: Firm trials and many CRTs where doctor-patient decision making is not compelled or constrained
- ‘Prior Notification’ – about studies that may be done to improve care; similar to commonly used notifications about possible use of patient records for epidemiological or biomedical research (HIPPA issues are separate)
- “You and your physician will always be able to determine which tests and treatments you will receive.”

Opt-out Strategies

- Mechanism for passive consent
  - Littenberg and MacLean. JGIM 2006;21:207-11
  - Quality improvement intervention: Vermont Diabetes Information System
  - Multi-state randomized trial
  - Patients were notified by mail that they were eligible
  - Were able to opt-out by calling a toll-free number
  - Of 7558 patients invited to participate, 210 (2.8%) opted-out
- Recruitment of a ‘broad and representative’ sample
  - Maintains appropriate protections for study subjects
Implementing CRTs: “Observational Studies”

- Some implementation challenges can make CRTs look like observational studies
  - Political issues
  - Temporal trends
- Challenges regarding
  - Randomization
  - Susceptibility
  - Performance
  - Detection
  - Transfer

Randomization

- Not all sites will allow randomization
  - e.g. they ‘need’ to be in one group or the other
- Sites may need to be randomized together
  - May be functionally linked
  - e.g. several clinicians may work regularly at 2 sites

Susceptibility to Outcome at Baseline

- Differential temporal trends
  - May threaten comparability across intervention/non-intervention clusters
  - e.g. in a study of patient health behaviors, a contract is lost and many patients with private insurance move to another health system leaving a disproportionate number of uninsured patients at one site
Performance

• Intervention fidelity
  – e.g., after the study starts, a third of physicians at one intervention site decide they no longer wish to participate

• Co-interventions
  – e.g., after the study starts, at only one site a pharmaceutical company provides at no cost a new efficacious but expensive drug that influences the outcome of interest

Detection of the Outcome

• e.g., study of ACE inhibitor use to slow diabetes related renal function decline
  – after the study starts, one group installs software that automatically records data that leads to more appropriate performance of the gold standard test; at other sites the recording is dependent on clinicians remembering to order the gold standard test

Transfer

• Transfer
  – Drop outs and crossovers
    • Sites
    • Subjects within a site

• Statistical comparisons that do not appropriately consider hierarchies and clusters
Summary

• When studying interventions designed to affect groups, individual-level RCTs risk contamination and Type II errors.
  – CRTs are often preferable for interventions to affect behavior
  – Clustering – differences in subjects across groups – affects power, design, and analyses of CRTs
  – EMRs facilitate CRTs – pre-assignment “balancing” of important prognostic attributes.
• Implementing CRTs – “stuff happens” and agile designs and analytic plans are called for...
Cluster Randomized Trials?

As compared to individually randomized trials (RCTs),
• CRTs are more complex to design
• CRTs require more participants to obtain equivalent statistical power
• CRTs require more complex analysis

Key Concerns when Doing CRTs

• Unit of randomization (assignment) is different than the unit of analysis
• Clustering has design (sample size) and analytic implications
  – Need larger samples than individual RCT
  – Need better pre-trial data for balancing
  – Need sophisticated statistical methods
The ICC (Intra-Cluster Correlation Coefficient)

Variance in the outcome attributable to differences BETWEEN clusters
TOTAL variance in the outcome

- ICC = 0.01 means 1% of the variance in the outcome is attributable to differences between clusters
- ICC is interpreted as if it were “R²”

**Toy Example #1**

Estimated ICC < 0.00001

<table>
<thead>
<tr>
<th></th>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>BETWEEN Clusters</td>
<td>1.8</td>
<td>1</td>
<td></td>
<td>1.8</td>
<td>0.01</td>
<td>0.93</td>
</tr>
<tr>
<td>WITHIN Clusters</td>
<td>22063</td>
<td>97</td>
<td></td>
<td>227.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Toy Example #2**

Estimated ICC = 0.569

<table>
<thead>
<tr>
<th></th>
<th>Source</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>BETWEEN Clusters</td>
<td>8961</td>
<td>1</td>
<td></td>
<td>8961</td>
<td>66.3</td>
<td>&lt; .0001</td>
</tr>
<tr>
<td>WITHIN Clusters</td>
<td>13102</td>
<td>97</td>
<td></td>
<td>136.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“Basic” Design Effect Formula

- Average Cluster Size:
  Mean # of subjects per cluster = $\bar{n}$
- ICC estimate for relevant outcome $\hat{ICC}$

Design Effect = $1 + \left(\frac{\bar{n} - 1}{\hat{ICC}}\right)$

This is the D. E. for continuous data. For comparing two proportions, there is an additional correction due to Cornfeld (1978). See Campbell (2001).

Design Effect

Total subjects required under CLUSTER randomization

Total subjects required under INDIVIDUAL randomization

- Design Effect $\geq$ 1.0
  - If $DE > 1.0$, CRT requires more patients than would a RCT with the same power.
- For larger ICC, Design Effect increases
  - CRT requires increasingly larger $n$ relative to individually randomized trial (RCT).

A Behavioral Intervention in General Practice to Lower Serum Cholesterol

- Practices randomized into two clusters
  - Intensive dietary intervention vs usual care
  - Suppose we recruit 50 patients per practice
  - Between practice estimate $s_b^2 = 0.046$
  - Within practice estimate $s_w^2 = 1.28$
- ICC estimate for this outcome …

$\hat{ICC} = \rho = \frac{0.046}{0.046 + 1.28} = 0.0036$

Kerry SM Bland JM (1998)
**Intervention to Lower Serum Cholesterol**

- Practices randomized into two clusters
  - With \( n = 50 \) pts / practice, and ICC = 0.0036, design effect is
    \[
    DE = 1 + (\bar{r} - 1) ICC = 1 + (50 - 1)(0.0036) = 1.17
    \]
- Individual RCT requires \( n = 5,364 \) pts to detect 0.1 mmol/L difference (90% power and \( \alpha = 0.05 \)). So the CRT needs…
  \[
  1.17(5364) = 6276 \approx 6300 \text{ pts}
  \]
  = 126 clusters of 50 patients each

---

**Detecting a 0.10 mmol/L cholesterol Difference (90% power, \( \alpha = 0.05 \), same ICC)**

<table>
<thead>
<tr>
<th>ICC</th>
<th>Pts / Practice</th>
<th>Practices</th>
<th>Patients</th>
<th>Design Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0036</td>
<td>10</td>
<td>558</td>
<td>5,580</td>
<td>1.04</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>234</td>
<td>5,850</td>
<td>1.09</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>126</td>
<td>6,300</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>74</td>
<td>7,400</td>
<td>1.38</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>32</td>
<td>16,000</td>
<td>2.98</td>
</tr>
</tbody>
</table>

Kerry SM Bland JM (1998)

---

**Are there fancy, free tools to estimate the sample size in CRTs?**

- [http://www.abdn.ac.uk/hsru/epp/cluster.shtml](http://www.abdn.ac.uk/hsru/epp/cluster.shtml)
  - University of Aberdeen sample size calculator with instructions
- Database of ICCs for use in planning
- [http://sitemaker.umich.edu/group-based/optimal_design_software](http://sitemaker.umich.edu/group-based/optimal_design_software)
  – “Optimal Design” from U. of Michigan
Patients within Physicians within Practices within Study Groups: A 4-Level CRT

Do I Have Enough Practices?

- Many studies: 5-10 practices / arm
  - ICC estimates vary quite a bit from study to study - consider 95% CI bounds?
  - DIG-IT: reasonable power for small ICCs
    - System A: 2 study groups of 5 practices
    - System B: 3 groups: 4, 6, and 4 practices
  - Design Effect increases with larger practice sizes, if ICC stays constant

Designing a CRT: More Patients Per Practice or More Practices?

<table>
<thead>
<tr>
<th># of Practices per Treatment Arm</th>
<th>ICC (rho)</th>
<th>Required # of Patients (Total)</th>
<th>Needed # of Pts/Practice</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>0.0</td>
<td>200</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>396</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>∞</td>
<td>∞</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>∞</td>
<td>∞</td>
</tr>
<tr>
<td>10</td>
<td>0.0</td>
<td>200</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>248</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>320</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>486</td>
<td>49</td>
</tr>
<tr>
<td>20</td>
<td>0.0</td>
<td>200</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>220</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>0.02</td>
<td>246</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>0.03</td>
<td>278</td>
<td>14</td>
</tr>
</tbody>
</table>
Allocating Practices to Study Groups: The DIG-IT study

- Want minimal differences across study groups on important predictors of:
  - CDS adoption
  - Response to the CDS intervention
- Balancing of practices across study groups (pre-randomization) is critical.
  - EMR plays a large role here
  - Most important for small # of practices

EMR-Based Balancing Procedure

- For all feasible clusterings of 10 practices into 2 study groups...
  - Assemble practice-level clinical and demographic data from EMR
  - Identify clusterings which appear to balance an array of baseline characteristics / trends
- (Blinded) consensus as to clustering with best balance. Intervention allocated by coin flip to one study group from that clustering.

Love TE Cebul RD Dawson NV et al. (in press) J Gen Internal Med special issue on Health IT

Which Table 1 do you want?
336 “clusterings” which split 10 practices into 2 study groups of sizes 4 & 6 or 5 & 5

Pre-Trial Balancing of Practices Using EMRs

• Result: 5 intervention practices and 5 “usual care” practices with excellent balance across study groups.

• EMRs provide new opportunities for state-of-the-art study design.
  – Could use this approach to create study groups for lots of community-based therapeutic or health care delivery trials.

ANALYSIS: Problems with Individual Level Analysis of Clustered Data

• Lack of independence among members of a cluster (cluster effect)
  – Need for larger sample sizes: Standard sample size formulas will lead to underpowered studies
  – Need for sophisticated statistical methods: Standard approaches will tend to bias p-values downward risking spurious claims of statistical significance

Analyzing a Cluster Randomized Trial
The Simple Approach

- Construct a summary statistic for each cluster, then analyze these summary values
  - As in repeated measures or meta-analyses
  - Simple, but doesn’t allow for covariates
- Example: 34 practices referring patients to St George’s Hospital for X-ray exams
  - 17 practices got nothing, 17 got a one-page laminated copy of referral guidelines
- Outcome: % of x-rays requested which conformed to the guidelines

Kerry SM Bland JM (1998)

The Wrong Approach

- Act as if we had randomly assigned individual patients to intervention groups
- Calculate difference in proportion of requests in each group that conform.

<table>
<thead>
<tr>
<th>Group</th>
<th>Total Requests</th>
<th>% Conforming</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intervention</td>
<td>429</td>
<td>79.5</td>
</tr>
<tr>
<td>Control</td>
<td>702</td>
<td>72.5</td>
</tr>
</tbody>
</table>

Difference is 7.0 % points, with SE of 2.6
95% CI is (2, 12) percentage points,
P value = 0.0008 (χ² test)

This is just wrong, despite what you see in many meta-analyses

X-Ray Requests Conforming to Guidelines
Two-Sample T
Mean (%C) Int. = 81.6
Mean (%C) Ctl. = 73.6
SE (diff) = 4.3, df = 32
95% CI: (-1, 17)
percentage points
P value = 0.07
Weights each practice equally
Designing Cluster Randomized Trials in Health Policy Research

Part Three: A Brief Overview of Some Analytic and Strategic Concerns

### X-Ray Requests Conforming to Guidelines

#### Two-Sample T

<table>
<thead>
<tr>
<th>ID</th>
<th>Intervention</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Req's</td>
<td>% Conf.</td>
</tr>
<tr>
<td>1</td>
<td>20</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>26</td>
<td>94</td>
</tr>
<tr>
<td>4</td>
<td>31</td>
<td>90</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>24</td>
<td>88</td>
</tr>
<tr>
<td>7</td>
<td>16</td>
<td>94</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>83</td>
</tr>
<tr>
<td>9</td>
<td>30</td>
<td>83</td>
</tr>
<tr>
<td>10</td>
<td>66</td>
<td>80</td>
</tr>
<tr>
<td>11</td>
<td>8</td>
<td>80</td>
</tr>
<tr>
<td>12</td>
<td>62</td>
<td>77</td>
</tr>
<tr>
<td>13</td>
<td>43</td>
<td>74</td>
</tr>
<tr>
<td>14</td>
<td>23</td>
<td>70</td>
</tr>
<tr>
<td>15</td>
<td>64</td>
<td>69</td>
</tr>
<tr>
<td>16</td>
<td>6</td>
<td>67</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>56</td>
</tr>
</tbody>
</table>

**Estimated Mean diff = 7.0**

**95% CI: (0.2, 14) percentage points**

**P value = 0.04**

Weights practices by # requests

---

### X-Ray Request Results

<table>
<thead>
<tr>
<th>Method</th>
<th>95% CI</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>WRONG</strong> – Treat as if individual RCT</td>
<td>(2, 12)</td>
<td>0.008</td>
</tr>
<tr>
<td><strong>CRT</strong> – T test (equal practice weights)</td>
<td>(0.2, 14)</td>
<td>0.07</td>
</tr>
<tr>
<td><strong>CRT</strong> – T test (weight by # of requests)</td>
<td>(1, 17)</td>
<td>0.04</td>
</tr>
</tbody>
</table>

- Ignoring clustering results in CIs which are too narrow and P values which are too small.
- Reported ICC in this trial turned out to be 0.019

---

### Other Analytic Approaches

- Adjust standard errors using the design effect – an approximation
- Robust variance estimates
- GEE
- Multi-level modeling
- Bayesian hierarchical models
- And much, much more …
Eight Methods Of Analyzing A CRT
CRT of 2 interventions designed to increase breast screening attendance.
Outcome: log (OR) of attendance for two intervention effects and their interaction
Three cluster-level analyses
1. Unweighted regression of practice log odds
2. Log odds Regression weighted by inverse variance
3. Random-effects meta-regression of log odds with practice as a random effect

Eight Methods Of Analyzing A CRT
CRT of 2 interventions designed to increase breast screening attendance.
Five individual-level analyses
4. Standard logistic regression (ignore clustering)
5. Logistic regression: robust standard errors
7. Random-Effects logistic regression
8. Bayesian random-effects logistic regression

Results of the Eight Analyses
• [4] was highly anti-conservative (i.e. wrong).
• The other (more valid) methods showed …
  – Large differences in parameter estimates
  – Large differences in standard errors
  – Some weren’t computationally stable
  – Some were more sensitive than others to between-cluster variation
  – GEE doesn’t work well with small # of clusters
• More Guidance Is Needed!
Additional Reporting Requirements for CRTs (CONSORT Statement)

1. **Rationale** for adopting a cluster design
2. How clustering was incorporated into the **sample size** calculations (include ICCs)
3. How clustering effects were incorporated into the **analysis**
4. The flow of **both** clusters and individuals through the trial, from assignment to analysis.

Campbell MK et al. for the CONSORT Group (2004)

Repeating Myself: Key Concerns when Doing CRTs

- **Unit of randomization** (assignment) is different than the unit of **analysis**
- Clustering has **design** (sample size) and **analytic** implications
  - Need larger samples than individual RCT
  - Need better pre-trial data for balancing
  - Need sophisticated statistical methods