

Objective

- Healthy People 2010
- Objective 11-3
- Research and Evaluation of Communication Programs

Cluster-Randomized Trials to Evaluate Health Communication Interventions

Paula Diehr, Ph.D.

National Conference on Health Communication,
Marketing, and Media

Atlanta, August 2007

Synonyms

- Group-randomized trials
- Community-randomized trials
- Cluster-randomized trials (CRTs)

Communication Interventions

- Interventions need to be evaluated
 - RCTs / drugs
- Messages are often delivered to groups or clusters of people
 - Media market (PSA)
 - Community
 - Workplace
- Randomize “clusters” of people rather than individual persons.

Outline

- What's different about CRTs?
- Improving Power of CRTs
- Recommendations

What's a CRT?

- Evaluation of an intervention that is applied to a cluster of people
- Evaluation data may be collected at either
 - The cluster level
 - From cluster members

Example 1

- Workplaces randomized to an intervention to improve workplace healthiness. One outcome is better insurance coverage for preventive services. (~ HMRC)
- CRT, cluster is the workplace, outcome data collected at the cluster level

Example 2

- PSAs used to encourage smoking cessation in 11 communities, vs. 11 matched communities in different media markets.
- Data collected from sample of smokers in each community. (~COMMIT)
- CRT, cluster is the community, data collected on individual level

CRT vs. Ordinary RCT

- If evaluation data are collected at the cluster level (example 1), there is not much difference.
- Concentrate on CRTs where data are collected at the individual (person) level (example 2)

Features of CRTs

- Intervention or evaluation may be expensive
- Number of clusters often small, for practical reasons
 - Low power?
- Number of persons per cluster are often large

Design Issues

Sample Size

Matching



HEALTH
MARKETING
RESEARCH CENTER
A CDC CENTER OF EXCELLENCE

Sample Size

Sample Size

- For individual randomization, need to choose the number of persons per treatment group
- In CRT, need to decide on two sample sizes

Two Sample Sizes

- K: # of clusters per tx group
 - Usually small (11 in COMMIT)
- N: # of people per cluster
 - Maybe be quite large (~1000 in COMMIT)
- Collect data on $2*N*K$ persons, but analyze K clusters
 - In COMMIT, data from 22,000 persons, analyzed the 22 cluster means.
- Potential small-sample problems, even though a lot of data collected

Variance in RCT

- σ^2_p is the variance of the outcome variable (change in smoking) among people
- If all alike, variance near zero
- If much variation, variance high

Variance of sample mean in a patient-randomized design

$$\text{Var}(\bar{Y}) = \frac{\sigma_P^2}{N}$$

Can estimate mean as accurately as desired by increasing N

Variations in CRTs

- σ^2_C is the variance among the true cluster means on the outcome variable (average smoking change)
- σ^2_P is the variance among people within a cluster

Variance of sample mean in a cluster-randomized design

$$V a r (\bar{Y}) = \frac{\sigma_c^2 + \frac{\sigma_p^2}{N}}{K}$$

Can always reduce variability
by increasing K;
Sometimes by increasing N.

Variability Among Cluster Means

- Rarely known
- Cluster level variances not usually reported
- Based on small # of clusters

Minimum Detectable Difference

- MDD
- The smallest difference between treatment and control that can be detected.
- Smaller is better
- Example for Sickness Impact Profile (SIP)

MDD if $\sigma^2_C = 0$

	N = # of persons per cluster				
K	250	500	1000	2000	4000
4	3.2	2.3	1.6	1.1	0.8
6	2.4	1.7	1.2	0.9	0.6
8	2.0	1.4	1.0	0.7	0.5
10	1.7	1.2	0.9	0.6	0.4
14	1.3	1.0	0.7	0.5	0.3

MDD if $\sigma^2_C = 7$

	N = # of persons per cluster				
K	250	500	1000	2000	4000
4	8.5	8.2	8.1	8.0	7.9
6	6.3	6.1	6.0	5.9	5.9
8	5.3	5.1	5.0	4.9	4.9
10	4.6	4.5	4.4	4.3	4.3
14	3.6	3.4	3.4	3.3	3.3

Sample Size Considerations

- σ^2_C , σ^2_P are both important
- Data usually not available
- Sample size calculations are probably inaccurate
- K is more important than N



HEALTH
MARKETING
RESEARCH CENTER
A CDC CENTER OF EXCELLENCE

Matched Design

Matching

- CRTs often create matched pairs of clusters
 - E.g., on community size
- Randomly assign one pair member to treatment, one to control
- Matched/paired analysis

Matching to Improve Power

- Effective matching variables should be strongly correlated with the outcome variable.
- Correlations rarely known at the cluster level
 - Change in smoking prevalence
- COMMIT study demonstrated effective matching after the study was over

Matching for Face Validity

- It would look bad if “big” communities were all in tx group
- Match even if effective matching variables are unknown
- But, if $K < 10$, a paired analysis can decrease power if matching variables are ineffective
- Unpaired analysis may be more powerful, even for matched clusters

Analysis

Cluster

Person

Unit of Analysis: Cluster

- Calculate a summary outcome measure for each cluster (e.g., cluster mean)
 - COMMIT, quit rate
- Perform a test on the $2 \times K$ cluster means

Non-parametric Test

- Sign test
- Permutation test (COMMIT)
- No assumptions of normality of cluster means
- Can not achieve statistical significance unless $K > 6$ or 7

T-test

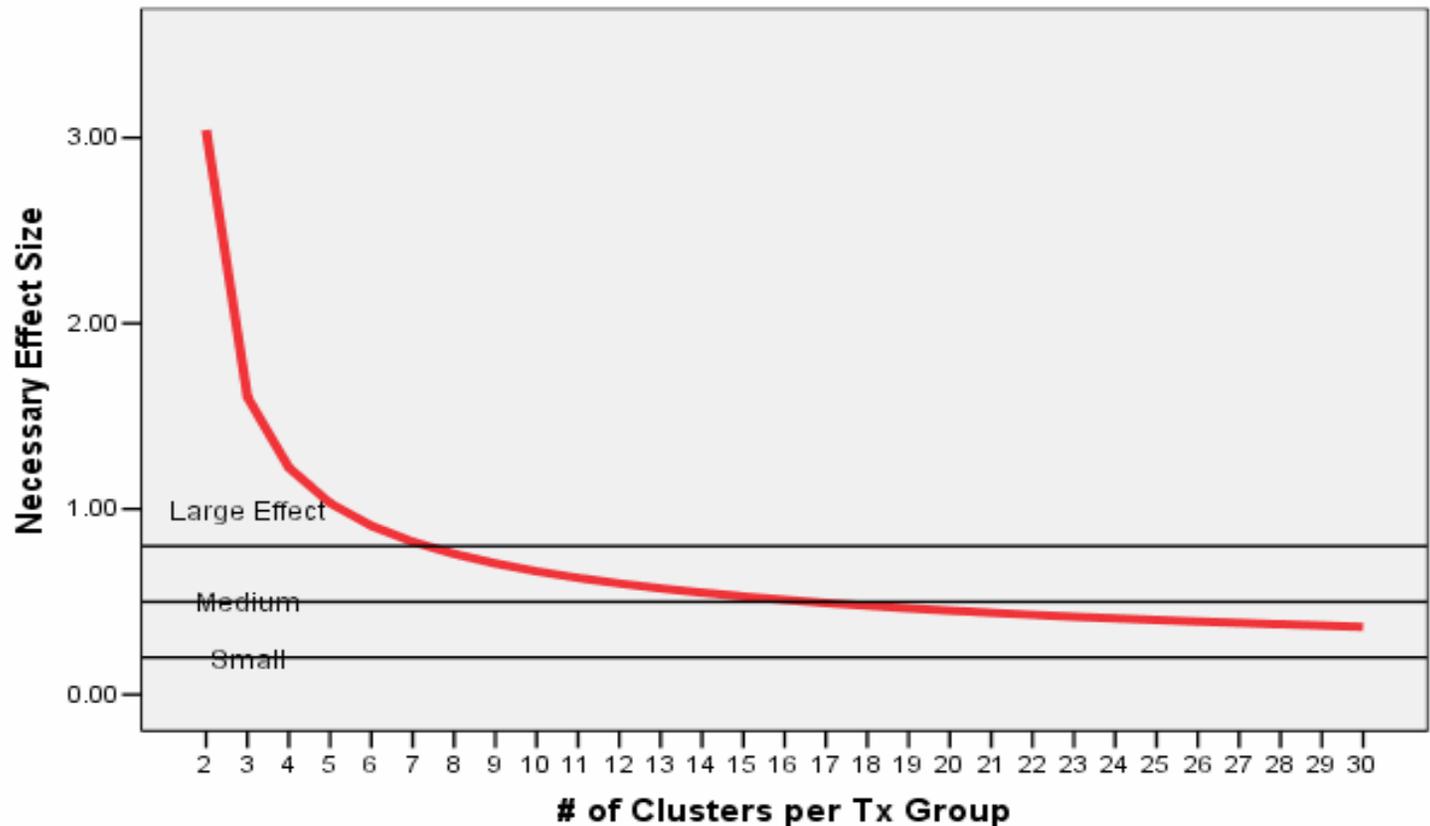
- Assumes that cluster outcome means are normally distributed
 - No way to test for normality, not enough data
 - And normality matters for small K
- “Works” for any number of pairs
- But effect size must be large for small K

Effect Size

- Expected difference between treatment and controls, in standard deviation units
 - s.d. among clusters
- Cohen's rule
 - Small effect .2
 - Medium effect .5
 - Large effect .8

Effect Size for 1-tail Test

Necessary Effect Size by # of Clusters (1 tail)



Effect Size

- CRT requires huge effect size if K is small
- And effect size is probably unknown
 - hence the CRT

Trials with small K

- Unlikely to achieve statistical significance
- May be useful for pilot study
- COMMIT
 - Mixed results
 - K too small?



HEALTH
MARKETING
RESEARCH CENTER
A CDC CENTER OF EXCELLENCE

Can Analysis Improve Power?

Analyze Clusters

Analyze Persons

Unit of Analysis: Cluster

- t-test
- Degrees of freedom
 - $2K-2$ (unpaired)
 - $K-1$ (paired)
- No fancier analysis will buy more degrees of freedom
- Person-level characteristics?
 - Adjusted cluster means (COMMIT)

Unit of Analysis: Person

- Not independent
- Repeated Measures
- Mixed Model ANOVA (various)
- Degrees of freedom should be the same as the cluster-level t-test
- If N same for each cluster, equivalent to the t-test
- Person-level covariates may improve power a little

New Regression Methods for Correlated Data

- Generalized Estimating Equations (GEE)
 - No d.f.
 - Requires “large” number of clusters
 - 50 clusters or more
- No help for small K

How to increase power in a CRT?

- Increase K , always helps
- Possible small improvements:
 - Increase N
 - Matching
 - Person-level covariates

References

- Murray D. Design and analysis of group randomized trials. 1998. Oxford Press.
- Donner A, Klar N. Design and analysis of cluster randomization trials in health research. Oxford Press. New York. 2000.

Conclusion

- CRTs are appropriate when an intervention is “communicated” to intact groups of persons
- CRTs with small K require an enormous effect size to achieve statistical significance.
- Best design should use as many clusters as possible



HEALTH
MARKETING
RESEARCH CENTER
A CDC CENTER OF EXCELLENCE

Thank-you

Intraclass Correlation (ICC)

- Measure of the amount of “clustering”
- Large when? Small when?
- If you know the ICC and σ_p^2 , can solve for estimate of σ_c^2 . 0.02?

$$ICC = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_p^2}$$

Paired analysis can cause a loss of power

- Suppose $K = 5$
- Paired t-test has 4 d.f.
 - Need $T > 3.37$ to achieve significance (1-tail)
- Unpaired t-test has 8 d.f.
 - Need only $T > 2.90$
- Hope that matching will make up for loss of degrees of freedom (hope r_{xy} is large)
 - But we will rarely know
- Martin et al., *Statistics in Medicine*

Match but ignore in the analysis?

- Unpaired t-test of matched data?
- For large number of clusters, may make power worse
- For small number of clusters, may actually improve power
- Diehr et al., *Statistics in Medicine*

Usual Sample Size Formula

- The # of subjects needed to detect a difference between treatment and control of size Δ , with 80% power, is:

$$N = \frac{(1.96 + .84)^2 (\sigma_{P(Tx)}^2 + \sigma_{P(Control)}^2)}{\Delta^2}$$

If variances are known, sample size calculation is straightforward

Paired t-test (K clusters per group)

$$t_{paired} = \frac{\bar{Y}_1 - \bar{Y}_2}{\frac{1}{\sqrt{K}} \sqrt{s_{Tx}^2 + s_{Control}^2 - 2r_{xy}^2 s_{Tx} s_{Control}}}$$

r_{xy} is the correlation between the outcome and the matching variable

COMMIT study

- Community Intervention Trial for Smoking Cessation