Design and Analysis of Stepped Wedge Cluster Randomized Trials

Michael A. Hussey & James P. Hughes

Biost 572

Navneet R. Hakhu

June 6, 2013

Randomized Trial (RT)

- Randomize (independent) subjects to intervention arm
  - Q: Why bother?
- Criteria for assessing intervention
  - Safety
  - Efficacy
  - Effectiveness
- Q: What is a different type of RT?

Cluster Randomized Trial (CRT)

- Randomize (independent) clusters to intervention arm
  - Subjects within clusters are correlated

- Q: Why are CRTs useful?

Examples of CRTs

- Goal: Administer intervention on cluster-specific basis

- Sommer et al. (1986)
  - Vitamin A supplementation and childhood mortality
  - 450 villages in Indonesia randomized
  - Reason: Individual randomization not feasible

- Zucker et al. (1995)
  - Child and adolescent trial for cardiovascular health (CATCH)
  - Goal: Prevention
  - Schools randomized to intervention
  - Reason: Implementation on school-wide basis

Partner Notification

- Public health authorities contact sex partner
  - Of potential exposure to sexually transmitted infection (STI)
  - To seek treatment
  - Drawback: Implementation expensive

- Alternative: Patient Delivered Partner Therapy
  - Infected patient brings treatment to sex partner
    - Drugs or drug vouchers

Expedited Partner Therapy (EPT)

- Individually randomized trial [Golden et al., 2005]
  - 1998 to 2003 in King County, WA
  - Notification strategies (Intervention arms)
    - Patient delivered partner therapy, referred to as EPT
    - Standard partner notification (control)
  - Goal: To compare effectiveness of notification strategies for treating chlamydia and/or gonorrhea
    - Primary outcome: "presence of persistent or recurrent infection in the original index patient 3 – 19 weeks after treatment"
  - Study results
    - Significantly increased proportion of partners treated
    - Decreased risk of infection in patients

- Q: Successful trial, but are we done?

Limitation of EPT

- Q: What about all the other counties in WA state?
  - King county is not representative of every county in WA
- Goal for WA: To implement EPT in every county
  - Q: How?
- Comments
  - Implementation of EPT on a county-wide basis motivates need for CRT
  - However, one can view EPT as each individual's choice

Motivation for CRT

- Individually randomized trial completed
  - But only for one county (King)

- New trial
  - Counties represent clusters
  - Q: What kind of CRT should we use?

Possible CRT Designs

Parallel		Time
		1
Cluster	1	1
	2	1
	3	0
	4	0

Crossover		Time	
		1	2
Cluster	1	1	0
	2	1	0
	3	0	1
	4	0	1

Stepped Wedge		Time				
		1	2	3	4	5
Cluster	1	0	1	1	1	1
	2	0	0	1	1	1
	3	0	0	0	1	1
	4	0	0	0	0	1

Comments on Designs

- Some argue that stepped wedge design is only preferable to no randomized trial [Kotz et al., 2012]
  - Takes longer
  - Stepped wedge only has higher power because more data than parallel
- Hussey and Hughes
  - Stepped wedge is not a design to always implement
  - But represents a viable option in some situations

Scientific Perspective

- Criteria for best design
  - Ethical
  - Logistical
  - Feasible

Statistical Perspective

- Criteria for best design
  - Power
    - Probability of rejecting null when alternative is true
    - For stepped wedge: Consider different effect sizes (i.e., number of clusters randomized at each time point)
  - Coefficient of Variation (CV)
    - Ratio of between-cluster standard deviation over mean prevalence

$$\mathrm{CV} = \frac{\tau}{\mu}$$

    - Intraclass correlation

$$\rho = \frac{\tau^2}{\tau^2 + \sigma^2} \neq 0$$

where $\sigma^2 = \mu\,(1 - \mu)$

Statistical Summary Measure

- Relative risk (RR)

$$\mathrm{RR} = \frac{\mu + \theta_A}{\mu}$$

  - μ : mean prevalence of outcome in control arm
  - $\mu + \theta_A$: effect in treatment arm
  - $\theta_A < 0$: we expect benefit in treatment arm (RR < 1)

- Note: With small prevalence μ, $\mathrm{OR} \approx \mathrm{RR}$

Generating Data: Individual-level

$$Y_{ijk} \sim \text{Binomial}(1, \mu_{ij})$$

$$g(\mu_{ij}) = \beta_0 + X_{ij}\beta_1 + \alpha_i$$

$$\alpha_i \sim \text{Normal}(0, \tau^2)$$

- $i = 1, \ldots, I$ (clusters)
- $j = 1, \ldots, T$ (time intervals)
- $k = 1, \ldots, N$ (individuals within cluster i at time j)
- $g(\,)$: link function (either identity or logit)
- X_{ij} : indicator of receiving treatment
- β_1 : treatment effect
- α_i : random effect for cluster

Choice of Scale

- β_0 and β_1 are different for identity versus logit link
- Generated random effects (α_i) and probabilities (μ_{ij}) are also different

Generating Data: Cluster-level

$$\bar{Y}_{ij} = \frac{1}{N} \sum_{k=1}^{N} Y_{ijk}$$

- $i = 1, \ldots, I$ (clusters)
- $j = 1, \ldots, T$ (time intervals)
- $k = 1, \ldots, N$ (individuals within cluster i at time j)

Predictor of Interest

Parallel		Time
		1
Cluster	1	1
	2	1
	3	0
	4	0

Crossover		Time	
		1	2
Cluster	1	1	0
	2	1	0
	3	0	1
	4	0	1

Stepped Wedge		Time				
		1	2	3	4	5
Cluster	1	0	1	1	1	1
	2	0	0	1	1	1
	3	0	0	0	1	1
	4	0	0	0	0	1

$$X_{ij} = \begin{bmatrix}1\\1\\0\\0\end{bmatrix} \qquad \begin{bmatrix}10\\10\\01\\01\end{bmatrix} \qquad \begin{bmatrix}01111\\00111\\00011\\00001\end{bmatrix}$$

$$(I,T) = (4,1) \qquad (4,2) \qquad (4,5)$$

(Approximate) Statistical Power

- Testing $H_0 : \theta = 0$ versus $H_1 : \theta = \theta_A$

$$Pwr(\theta_A) = \Phi\left(\frac{\theta_A}{\sqrt{Var(\hat{\theta})}} - z_{1-\alpha/2}\right)$$

  - Φ : Cumulative density function of $N(0,1)$
  - $z_{1-\alpha/2}$: $\left(1-\frac{\alpha}{2}\right)$-quantile of $N(0,1)$

Variance Formula

$$\mathrm{Var}(\hat{\theta}) = \frac{I\sigma^2(\sigma^2 + T\tau^2)}{(IU - W)\sigma^2 + (U^2 + ITU - TW - IV)\tau^2}$$

- $U = \sum_{ij} X_{ij}$
- $\mathrm{W} = \sum_j \left(\sum_i X_{ij}\right)^2$
- $\mathrm{V} = \sum_i \left(\sum_j X_{ij}\right)^2$
- $\sigma^2 = \mu\,(1 - \mu)$

Analysis of CRT

- Cluster-level
  - Linear Mixed Models (LMM)
- Individual-level
  - Generalized Estimating Equations (GEE)
  - Generalized Linear Mixed Models (GLMM)
- Goal: Compare power from LMM, GEE, GLMM

Simulation Study Design

Stepped Wedge		Time				
		1	2	3	4	5
Cluster	1	0	1	1	1	1
	⋮	0	1	1	1	1
	6	0	1	1	1	1
	7	0	0	1	1	1
	⋮	0	0	1	1	1
	12	0	0	1	1	1
	13	0	0	0	1	1
	⋮	0	0	0	1	1
	18	0	0	0	1	1
	19	0	0	0	0	1
	⋮	0	0	0	0	1
	24	0	0	0	0	1

N = 100 individuals in each cluster (i.e., 100 observations for each cell)

T = 5 time intervals

I = 24 clusters

Simulation Setup: Cluster-level

$$Y_{ijk} \sim \text{Binomial}(1, \mu_{ij})$$

$$\mu_{ij} = \beta_0 + X_{ij}\beta_1 + \alpha_i$$

$$\alpha_i \sim \text{Normal}(0, \tau^2)$$

$$\bar{Y}_{ij} = \frac{1}{N}\sum_{k=1}^{N} Y_{ijk}$$

- $\beta_0 = \mu = 0.05$
- $\tau = 0.015$
- $\text{RR} = \{1.0,\ 0.7,\ 0.6,\ 0.5\}$
  → $\beta_1 = \theta_A = \{0,\ -0.015,\ -0.020, -0.025\}$

Simulation Setup: Individual-level

$$Y_{ijk} \sim \text{Binomial}(1, \mu_{ij})$$

$$\text{logit}(\mu_{ij}) = \beta_0 + X_{ij}\beta_1 + \alpha_i$$

$$\alpha_i^* \sim \text{Normal}(0, \tau^{*2})$$

- $\mu = 0.05$
- $\tau = 0.015$
- $\text{RR} = \{1.0,\ 0.7,\ 0.6,\ 0.5\}$
  → $\theta_A = \{0,\ -0.015,\ -0.020, -0.025\}$
- $\beta_0 = \text{logit}(\mu)$
- $\beta_1 = \text{logit}(\mu + \theta_A) - \beta_0$
- $\tau^* = \text{logit}(\mu + \tau) - \beta_0$

Simulation Model: Cluster-level

- LMM – `lme` versus both `lme` and `lmer`
  - Fixed effects
    - Intervention effect
    - Time interval
  - Random intercepts only
    - Cluster
  - (Gaussian family with identity link)

Simulation Models: Individual-level

- GEE – `gee`
  - Fixed effects
    - Intervention effect
    - Time interval
  - Grouped by cluster
  - Exchangeable correlation structure
  - Binomial family with logit link

- GLMM – `glmmPQL` versus `glmer`
  - Fixed effects
    - Intervention effect
    - Time interval
  - Random intercepts only
    - Cluster
  - Binomial family with logit link

Simulation Study: Results

RR	Approximate Power	Cluster-level		Individual-level			
		LMM		GEE		GLMM	
		Paper	NRH*	Paper	NRH	Paper	NRH
1.0	0.050	0.056	0.056 (14)	0.084	0.074	0.076	0.058
0.7	0.412	0.697	0.690 (249)	0.719	0.736	0.716	0.711
0.6	0.659	0.907	0.891 (753)	0.907	0.939	0.917	0.940
0.5	0.951	0.988	0.985 (2154)	0.990	0.996	0.992	0.996

* includes number of re-sampled random intercepts to avoid negative probabilities

Critique

- Authors assume model with same fixed treatment effect for each cluster
  - Possible remedy: including random slopes
- Authors choose small tau to limit chances of negative probabilities for cluster-level approach
  - Q: What happens when CV ≠ 0.3 (with same μ)?
  - Resampling random effects might be a solution
  - Q: However, when resampling so often, do results have same interpretation?
    - (Not a normally distributed random effect)
- Authors do not compare power of stepped wedge to parallel design
  - Q: What would be a comparable way to compare designs?

Summary

- CRTs
  - Motivation: Implement on community-wide basis
  - Three designs: parallel, crossover, stepped wedge
- Stepped Wedge
  - Individually randomized trials ideal
  - Factoring in ethical, logistical, and feasibility issues
  - Phase IV effectiveness trials
  - Simulations of power based on Expedited Partner Therapy
- Next steps
  - Consider random intercepts and slopes to allow for different intervention effects for each cluster
  - Examine different sample sizes for each cluster
  - Extension: Compare Power for parallel versus stepped wedge

Thanks

- Jim Hughes
- Patrick Heagerty
- Jason Liang
- Everyone who provided feedback
- Everyone for listening