Cluster Randomized Trials and the Stepped Wedge Design

Jim Hughes

Dept of Biostatistics

Cluster Randomized Trials

- Randomization at group level; outcome measured on individuals
- Individual randomization not feasible, ethical, or potential contamination
- Usually, less efficient than individually randomized trial
- Intervention effect on a community may be greater than the sum of the parts (e.g. herd immunity)
- Clusters may be large (cities, schools) … or small (IDU networks, families)
- Key statistical challenge: individuals not independent

Key Considerations

- What is the unit of randomization?
- How is the intervention delivered?
- How is the outcome measured?
- Examples
  - PREVEN
  - HPTN037
  - HPTN041

Common Trial Designs

Parallel

Time
1
X
X
X
X
O
O
O
O

Crossover

Time	
1	2
X	O
X	O
X	O
X	O
O	X
O	X
O	X
O	X

The stepped wedge design

	Time			
1	**2**	**3**	**4**	**5**
O	X	X	X	X
O	O	X	X	X
O	O	O	X	X
O	O	O	O	X

- Time of crossover is randomized; crossover is unidirectional
- Need to be able to measure outcome on each unit at each time step
- Observations need to be “in sync” to control for time trends
- Individuals at each time can be same (cohort) or different (cross-sectional)

Reasons for choosing the Stepped Wedge Design

- Logistical or financial - cannot introduce the intervention in all units at once
- Efficiency: Units act as their own control, so fewer units needed
- Operations Research - evaluate the community effectiveness of an intervention previously shown to be efficacious in an individually randomized trial or in a different setting
- Disadvantages: lengthy, effect of intervention must be “immediate”, more complex analysis

Some Examples

- Effect of routine Isoniazid preventive therapy on tuberculosis incidence in HIV+ men in S. Africa (Grant et al, 2005)
- Individually randomized
- Due to constraints on clinic capacity employees of a mining company were invited to enroll in the study in a random sequence
- Analysis compared tuberculosis episode rate before and after clinic enrollment and adjusted for calendar time and baseline disease severity

Some Examples

- Introduction of HBV vaccination in infants in The Gambia (The Gambia Hepatitis Study Group, 1987)
- Cluster randomized (Health districts)
- 18 health districts, but program could not be implemented in all districts at the same time
- Immediate outcome: HBV antibody titre
- Longterm outcome: Hepatocellular cancer and other liver disease

Some Examples

- HPTN054: Comparison of combined versus targeted provision of Nevirapine to HIV+ pregnant women
- Cluster randomized (Health clinics); but only 8 clinics available
- Intervention: Combined vs targeted NVP provision during antenatal care
- Endpoint: Nevirapine in cord blood at delivery
- Time

1	2	
T	T	
T	C	× 2
T	C	
C	C	

- "Washout" period between times 1 and 2 to allow women to deliver

Some Examples

- Expedited partner treatment for Gc and Ct in WA state
- EPT shown to be effective in reducing reinfection in IRT (Golden et al., 2005) in a previous UW STDCRC project
- EPT to be implemented throughout Washington state; logistically difficult to implement the program in all counties simultaneously
- Solution: use a SW design; (24) counties are the randomization units; randomize 6 per time period
- Outcome (STI) measured in sentinel sites
- Six month intervals – 3 to implement, 3 to assess outcome

WA State EPT

	Time (mo)					
county	0	6	12	18	24	
1	O	X	X	X	X	
2	O	O	X	X	X	× 6
3	O	O	O	X	X	
4	O	O	O	O	X	

Statistical Issues - Model

Model:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + X_{ij}\theta + e_{ijk}$$

$$\alpha_i \sim N(0,\tau^2)$$
$$e_{ijk} \sim N(0,\sigma^2)$$

Key issue in a CRT: $\text{Corr}(Y_{ijk}, Y_{ij'k'}) = \tau^2/(\tau^2 + \sigma^2) \neq 0$

Note: Some authors express the correlation in terms of the *coefficient of variation* (CV) between clusters – $CV = \tau/\mu$

Statistical Issues - Power

$H_o: \theta = 0$

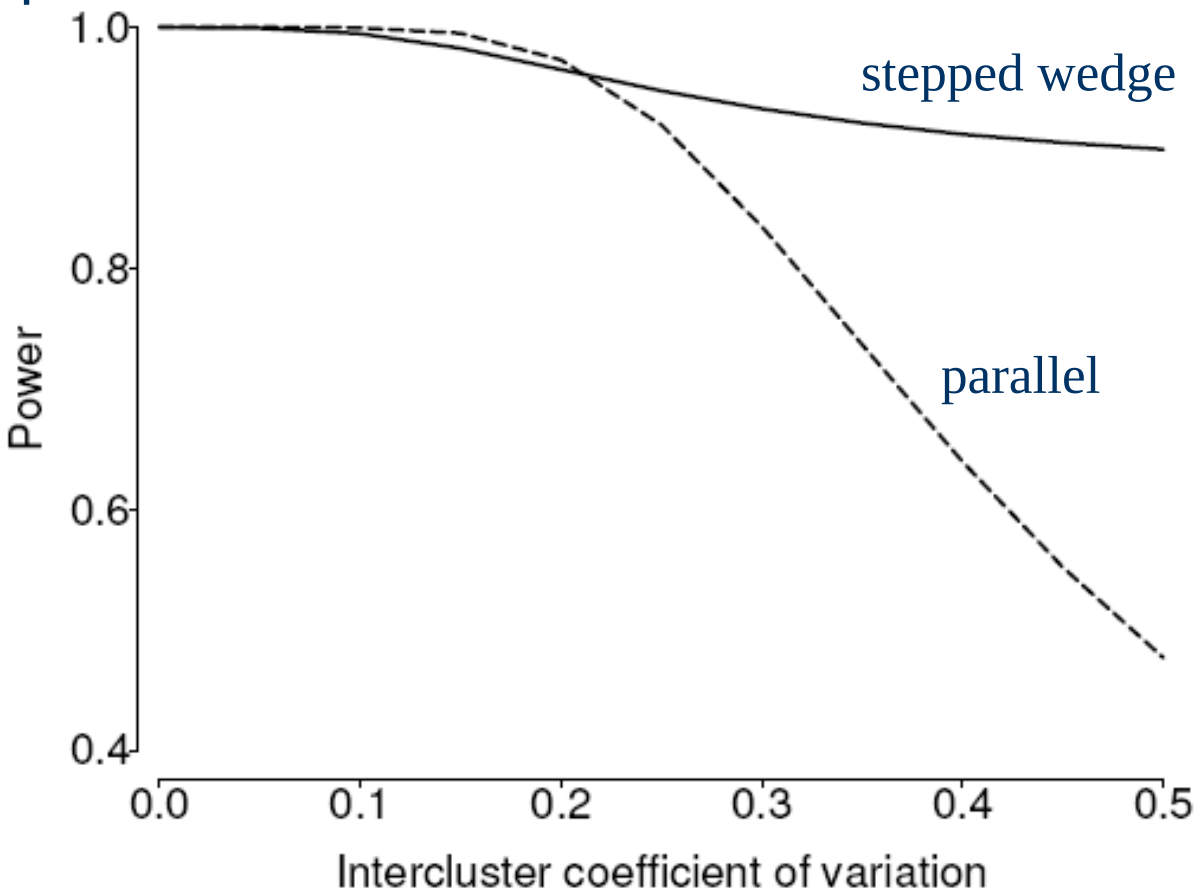
$H_A: \theta = \theta_A$

$$\text{Power} = \Phi\left(\sqrt{\frac{\theta_A^2}{Var(\hat{\theta})}} - Z_{1-\alpha/2}\right)$$

- A closed form expression for $\text{Var}(\hat{\theta})$ is available.
- Depends on number of units (clusters), number of steps, number observations per step, τ and σ.

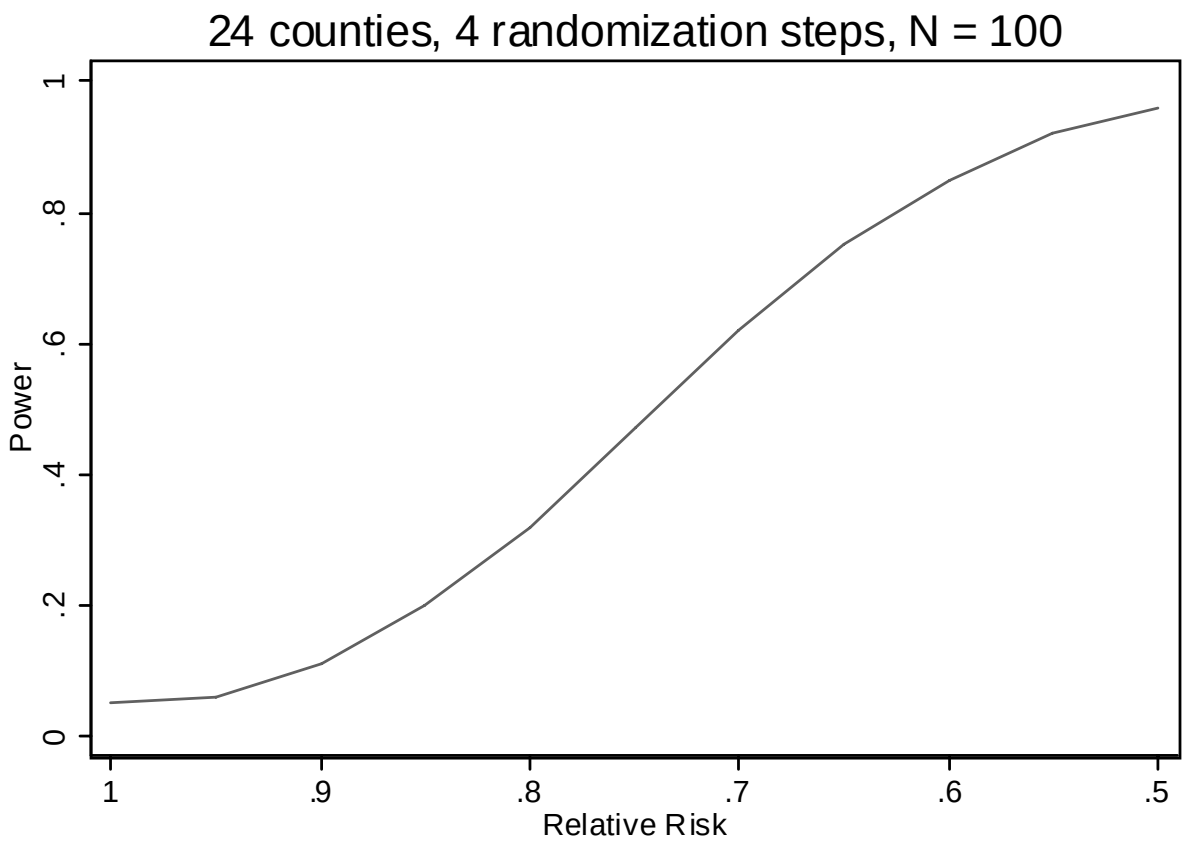
Power – SW vs parallel

HPTN054



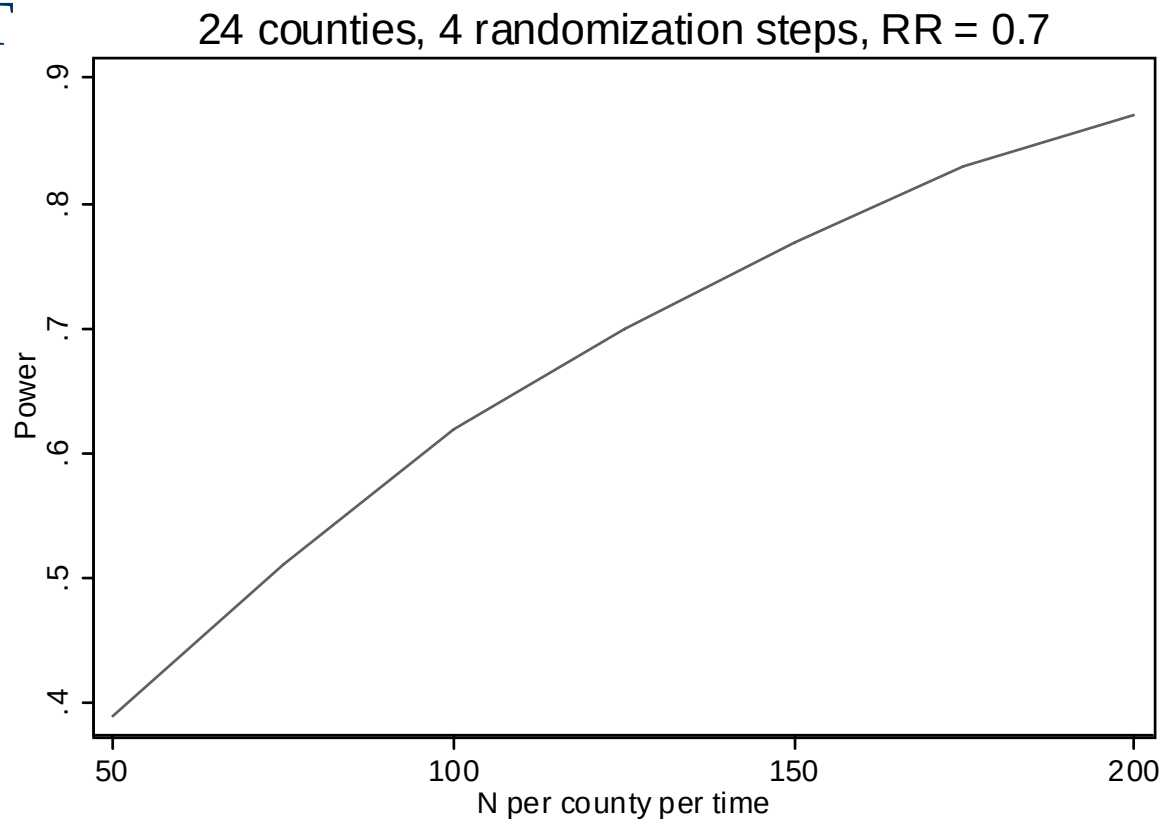
Power vs RR

WA State EPT



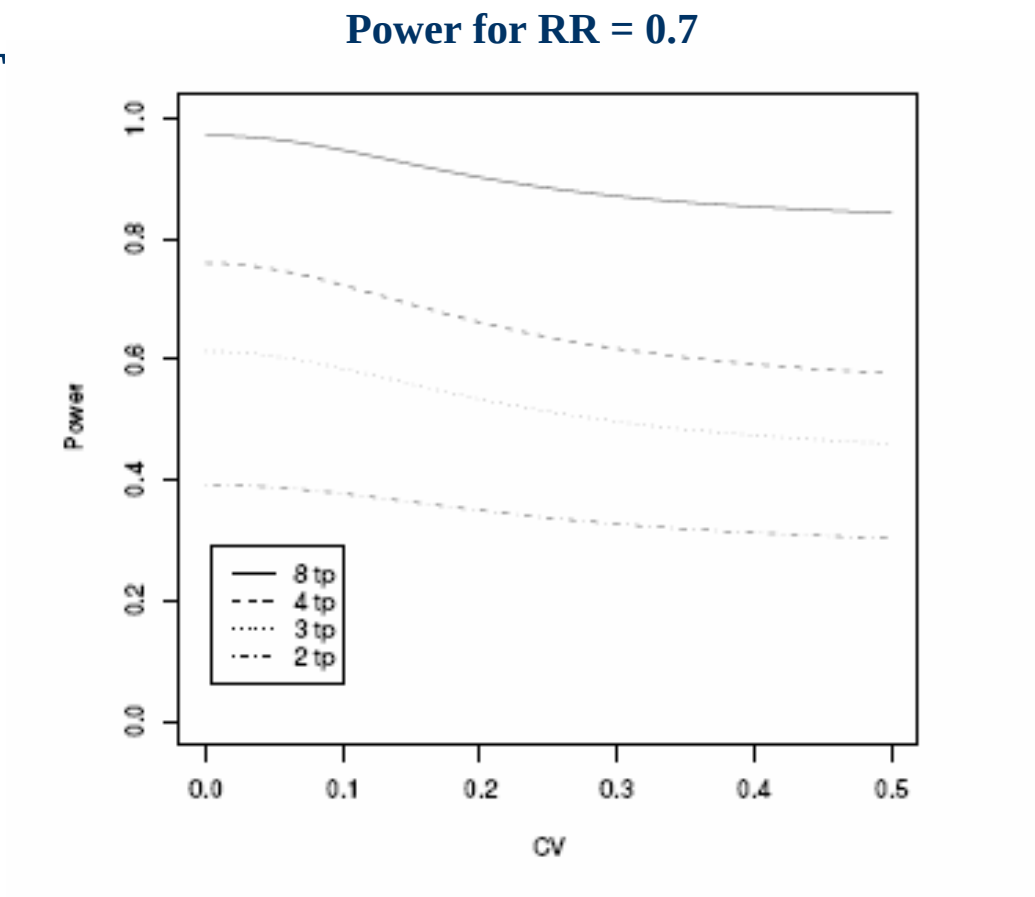
Power vs N per cluster

WA State EPT



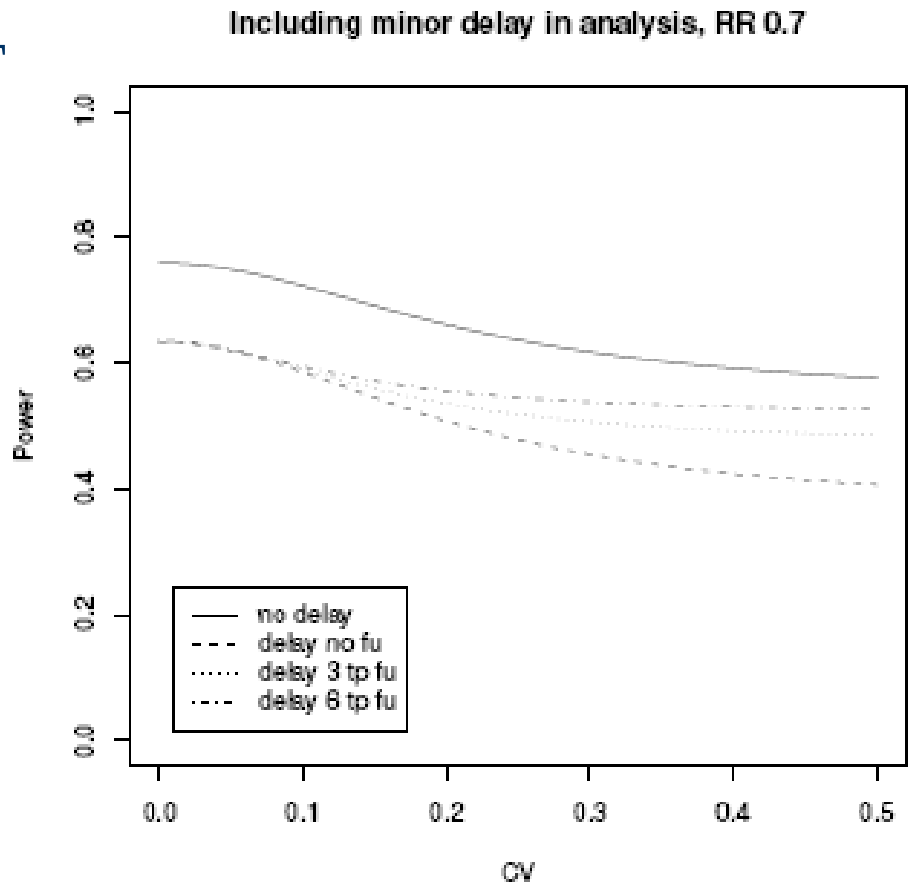
Power vs # of randomization steps

WA State EPT



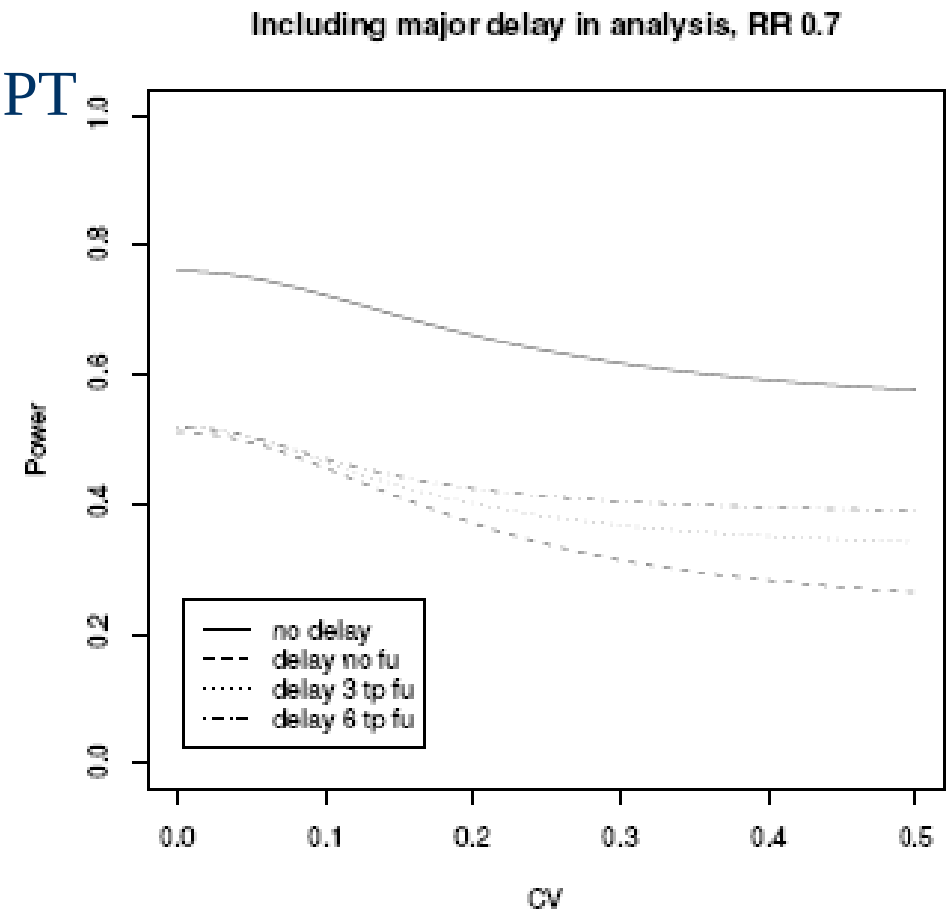
Power – Delayed treatment effect

WA State EPT



Power – Delayed treatment effect

WA State EPT



Statistical Issues - Analysis

- Paired t-test
  - Analyze cluster means, before vs after
  - Likely biased if there are time trends
- Repeated cross-sectional (in time) comparisons
  - Loses strength of within-unit comparisons; how to combine?
- LMM
  - Analyze cluster means using both within & between info
  - Must have equal cluster sizes (if binary outcome)
- GEE, GLMM
  - Analyze individual level data
  - Unequal cluster sizes ok

Summary

- Stepped wedge designs are useful for “phase IV” trials and as a way of dealing with logistic difficulties of implementing the intervention everywhere at once
- Power is relatively insensitive to CV
- Maximize the number of steps
- Intervals should be long enough to capture the full treatment effect
- Individual level analyses are necessary if cluster sizes vary
- Variations on this theme are possible

Thanks

Mike Hussey, MS

Matt Golden, MD

Jeff Stringer, MD

Analysis - Simulations

- Simulate WA State EPT trial – 24 clusters, 4 randomization times, CV = 0.3, baseline prevalence = .05, average 100/cluster-time; 1000 simulations
- Compare LMM, GEE, GLMM; equal and unequal cluster sizes
- Use jacknife estimate of variance (not necessary for equal cluster sizes)

	Equal cluster sizes				Unequal cluster sizes		
RR	LMM	GEE	GLMM		LMM	GEE	GLMM
1.0	.057	.052	.053		.038	.053	.049
0.7	.658	.644	.580		.307	.577	.559
0.6	.884	.866	.820		.503	.807	.805
0.5	.984	.981	.948		.653	.946	.942