Cluster Randomized Trials and The Stepped Wedge

Jim Hughes

UW Biostatistics

Cluster Randomized Trials

- Randomization at group level; outcome measured on individuals within the group
- Clusters may be large (cities, schools) … or small (IDU networks, families)
- Why? Individual randomization not feasible, potential contamination, or want to measure community effect
- Usually, less efficient than individually randomized trial (unless intervention effect on the community is greater than the individual effects)
- Key statistical challenge: individuals not independent

Cluster Randomized Trials

- A common error: two communities, flip a coin, one gets intervention; other gets control
- Underlying differences between communities confounded with treatment effect
- “Change from baseline” doesn’t solve the problem
- Key: Effective sample size is number of clusters, not number of individuals measured (though both are important)

Key Considerations

- What is the unit of randomization?
- How/to whom is the intervention delivered?
- How/on whom is the outcome measured?
- Examples
  - PREVEN
  - HPTN037
  - Mwanza HIV prevention trial

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
- Multiple observations per unit; observations need to be “in sync” to control for time trends (assumed similar across clusters)
- If CRT, then individuals at each time can be same (cohort) or different (cross-sectional)

Reasons for choosing the Stepped Wedge Design

- Efficiency: Units act as their own control, so fewer units needed (same as cross-over design)
- Logistical or financial - cannot introduce the intervention in all units at once
- Evaluate the community effectiveness of an intervention previously shown to be efficacious in an individually randomized trial or in a different setting; systematically evaluate new program
- To study the effect of time on intervention effectiveness (i.e. seasonality, time since introduction)

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

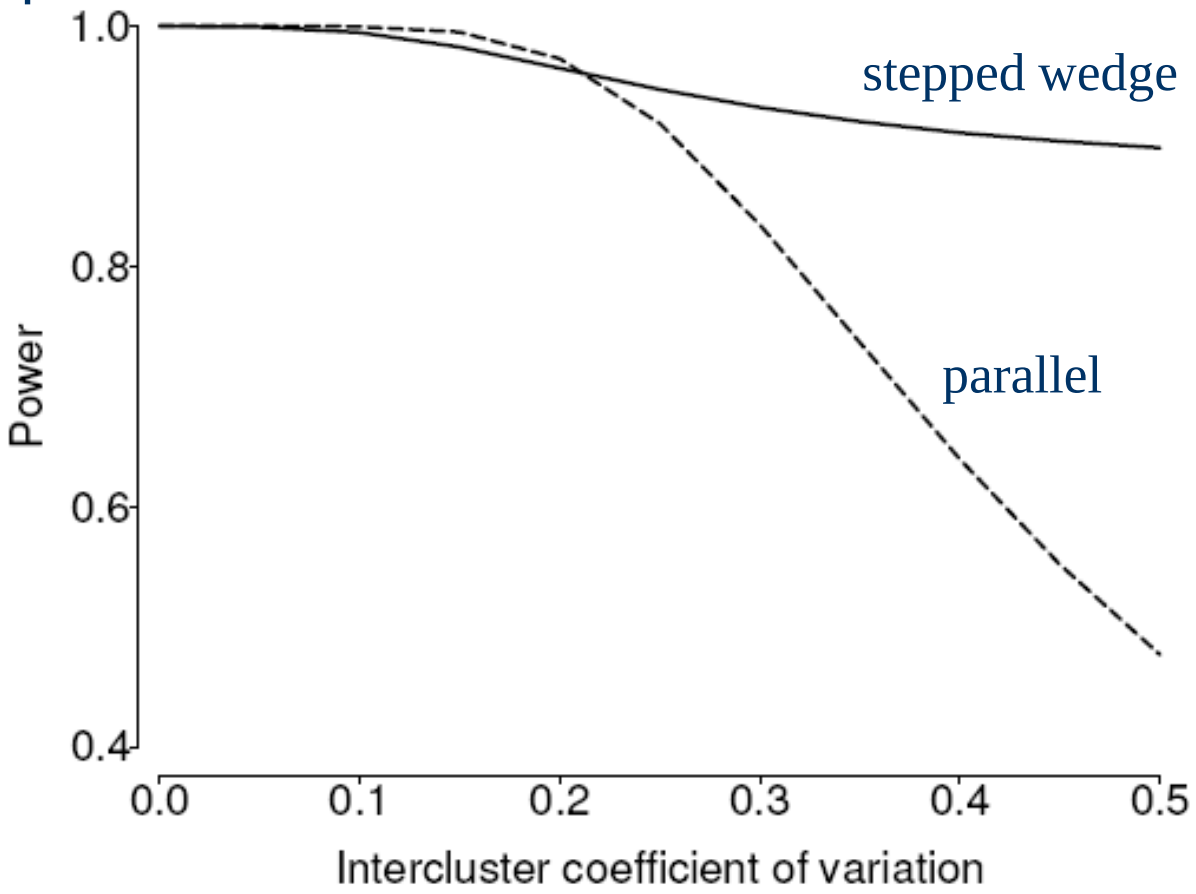
Note: Some authors express the correlation in terms of the *coefficient of variation* (CV) between clusters – CV = τ/μ

Statistical Issues - Power

- Power = Probability of detecting a treatment effect when the treatment really works
- Depends on …
  - strength of treatment effect
  - number of clusters
  - number of steps
  - number participants per cluster per step,
  - variance components: σ^2 (easy to know) ,τ^2 (hard to know).

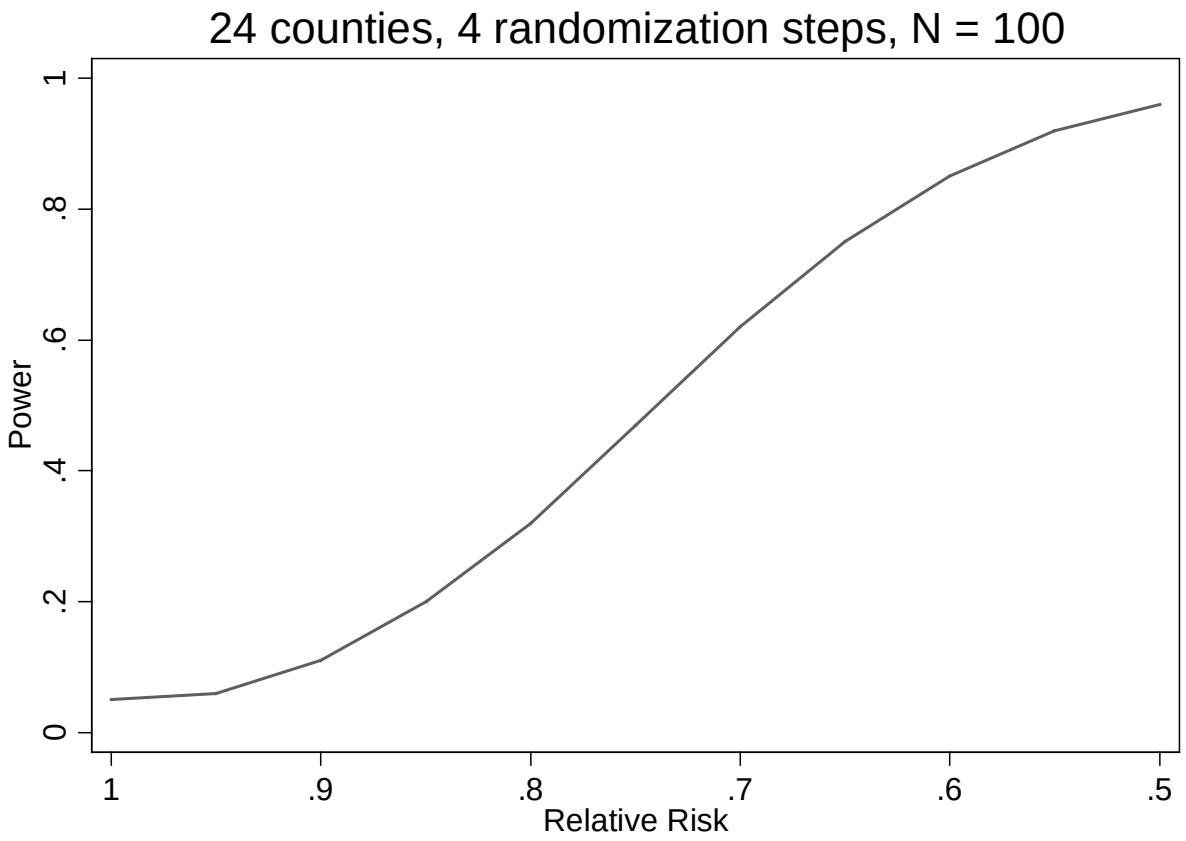
Power – SW vs parallel

HPTN054



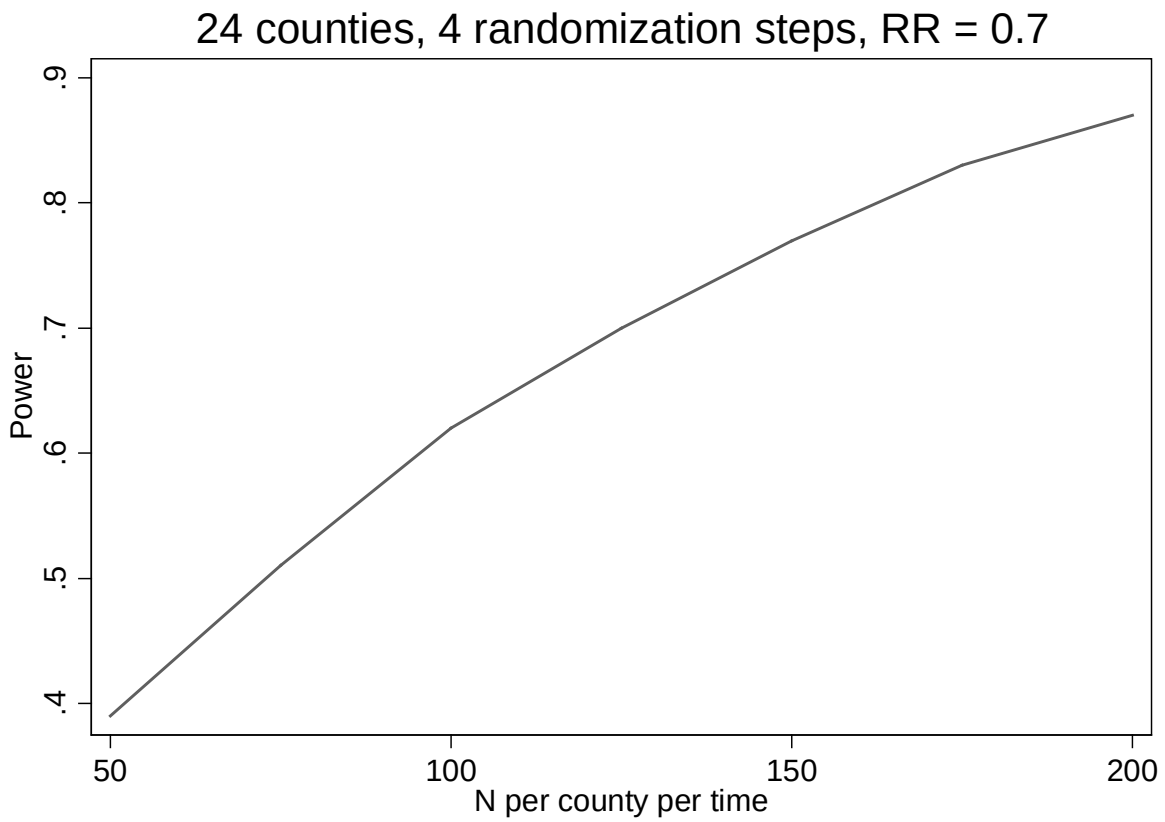
Power vs RR

WA State EPT



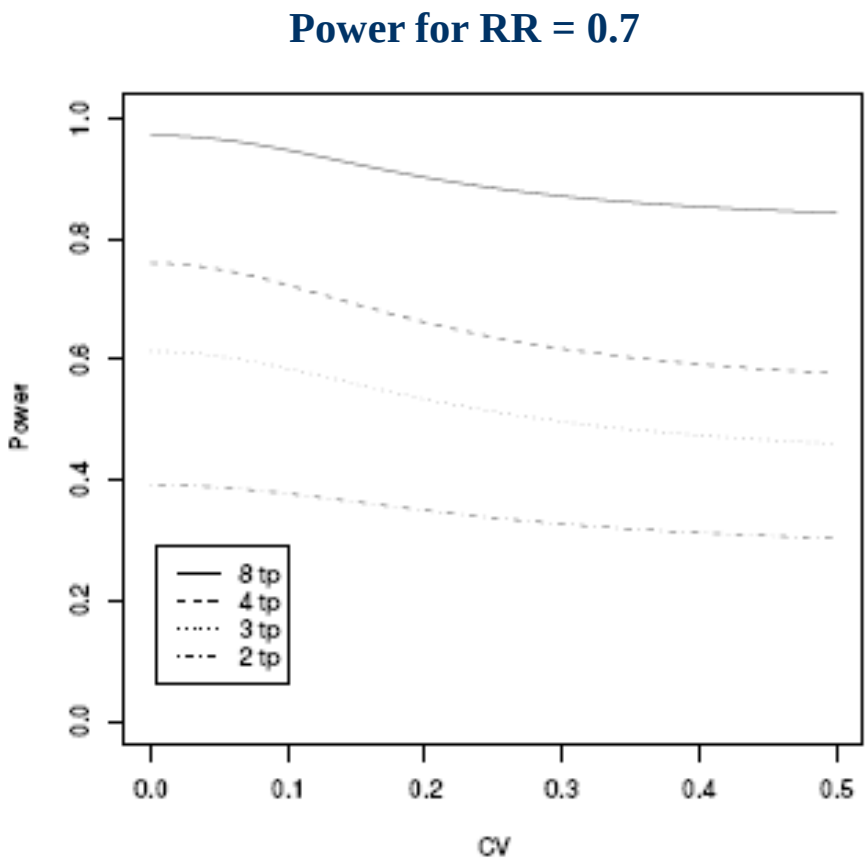
Power vs N per cluster

WA State EPT



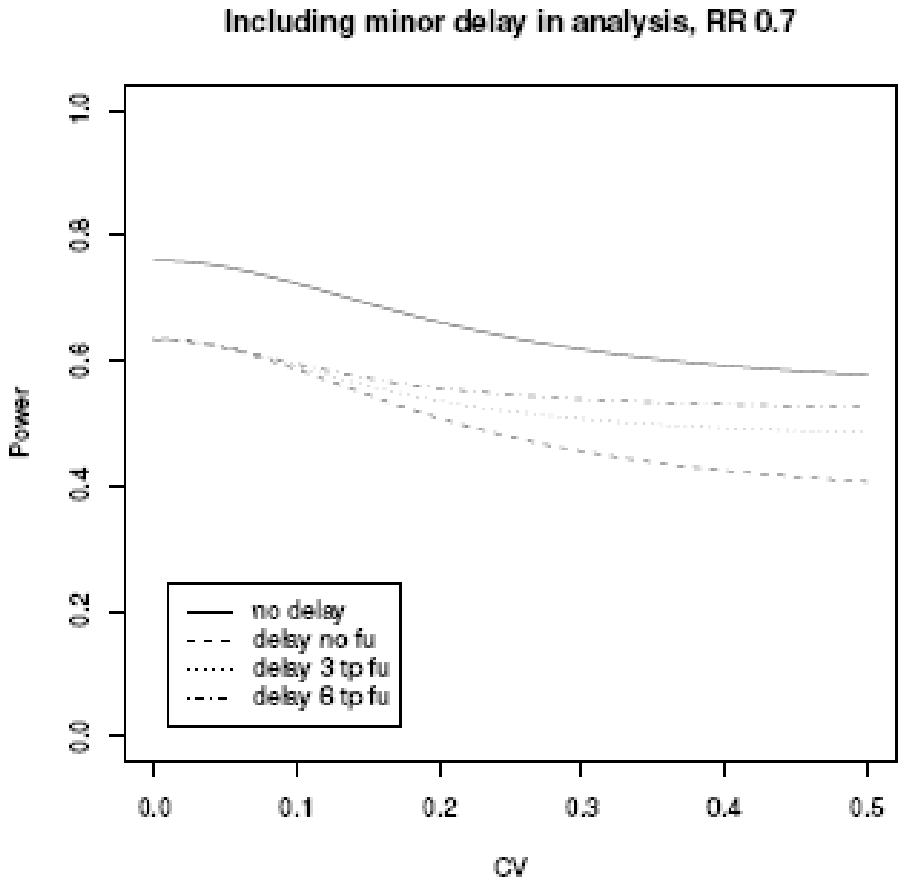
Power vs # of randomization steps

WA State EPT



Power – Delayed treatment effect

WA State EPT



Statistical Issues - Analysis

- Paired t-test (easy)
  - Analyze cluster means, before vs after
  - Likely biased if there are time trends
- Repeated cross-sectional (in time) comparisons (sorta' easy)
  - Loses strength of within-unit comparisons; how to combine?
- LMM (advanced, but standard)
  - Analyze cluster means using both within & between info
  - Must have equal cluster sizes
- GEE, GLMM (advanced)
  - Analyze individual level data
  - Unequal cluster sizes ok

Research Directions

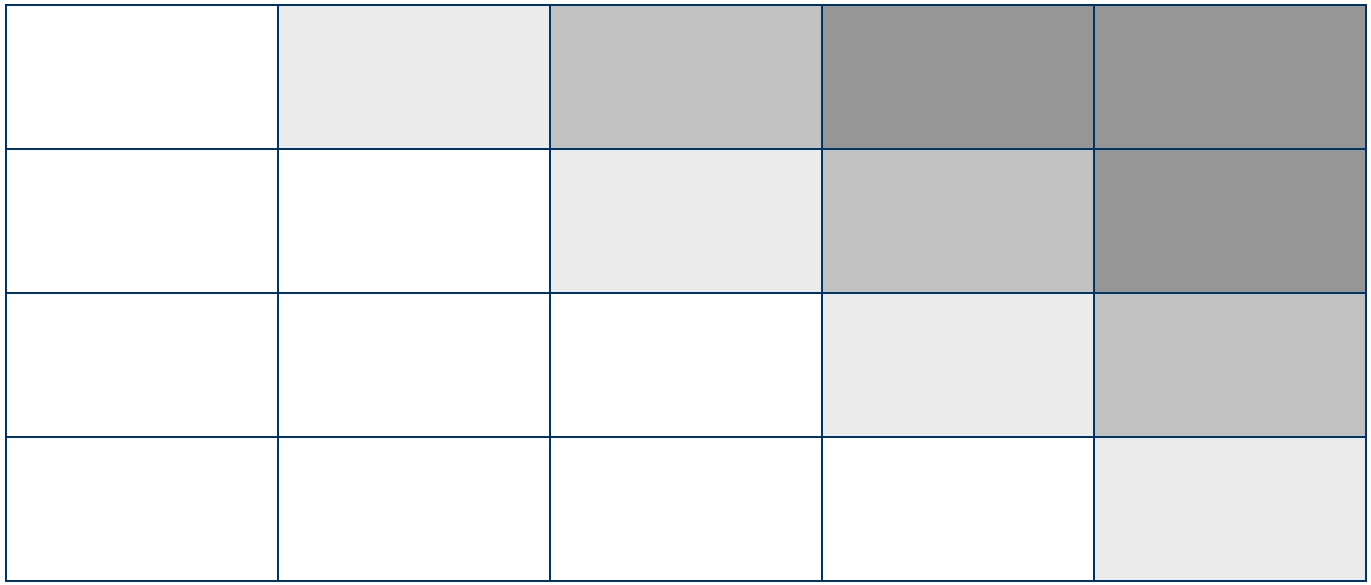
- Multicomponent interventions
- Various possibilities

Time				
1	2	3	4	5
O	1	1+2	…	
O	2	1+2	…	
O	O	1	1+2	…
O	O	2	1+2	…
O	O	O	1	
O	O	O	2	

Time				
1	2	3	4	5
O	1	1+2	…	
O	2	1+2	…	
O	1+2	1+2	…	
O	O	1	1+2	…
O	O	2	1+2	…
O	O	1+2	1+2	…

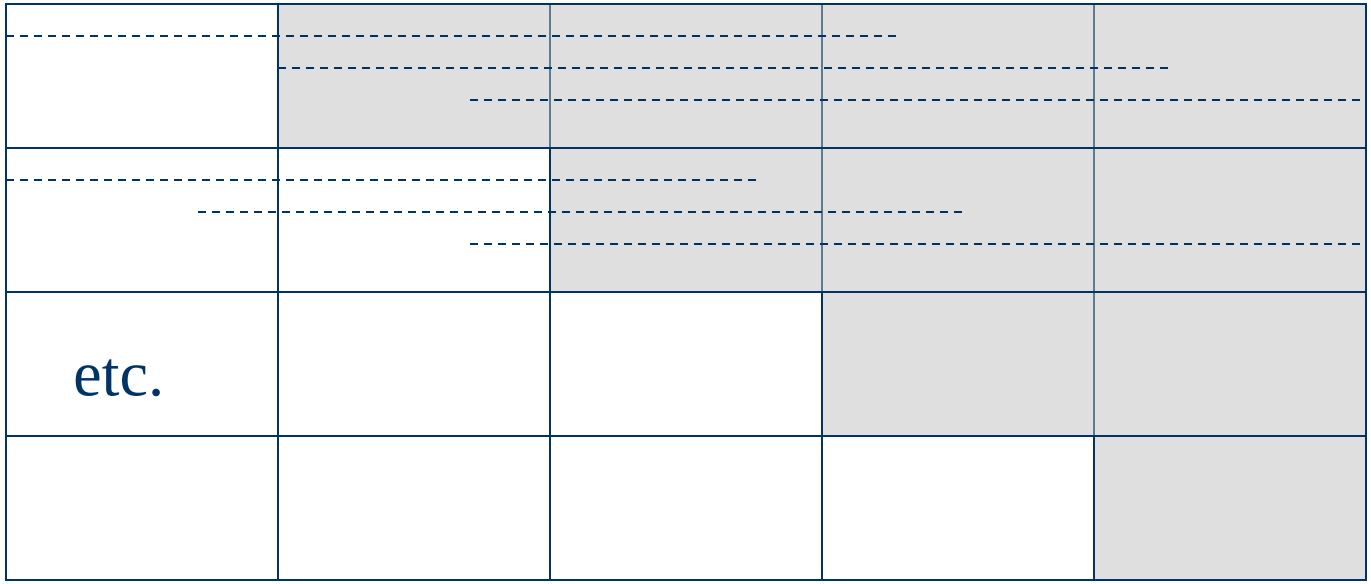
Research Directions

- Delayed intervention effects
  - How to estimate
  - Powering trial if delayed effect anticipated



Research Directions

- Rolling cohorts for evaluation



Summary

- Stepped wedge designs are useful for “phase IV” trials, to evaluate the effect of time on the intervention, 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 (Hussey and Hughes, CCT 28:182 – 191, 2007)

Matt Golden, MD

Jeff Stringer, MD

Alternative models

Also possible to write models for …

- Cluster by Time interaction
- Cluster by Treatment interaction (treatment effect varies by cluster)
- Treatment by Time interaction (treatment effect varies with time)
- Treatment effect varies with time since introduction of intervention