DESIGN AND ANALYSIS OF STEPPED WEDGE TRIALS

Jim Hughes
Professor of Biostatistics
University of Washington

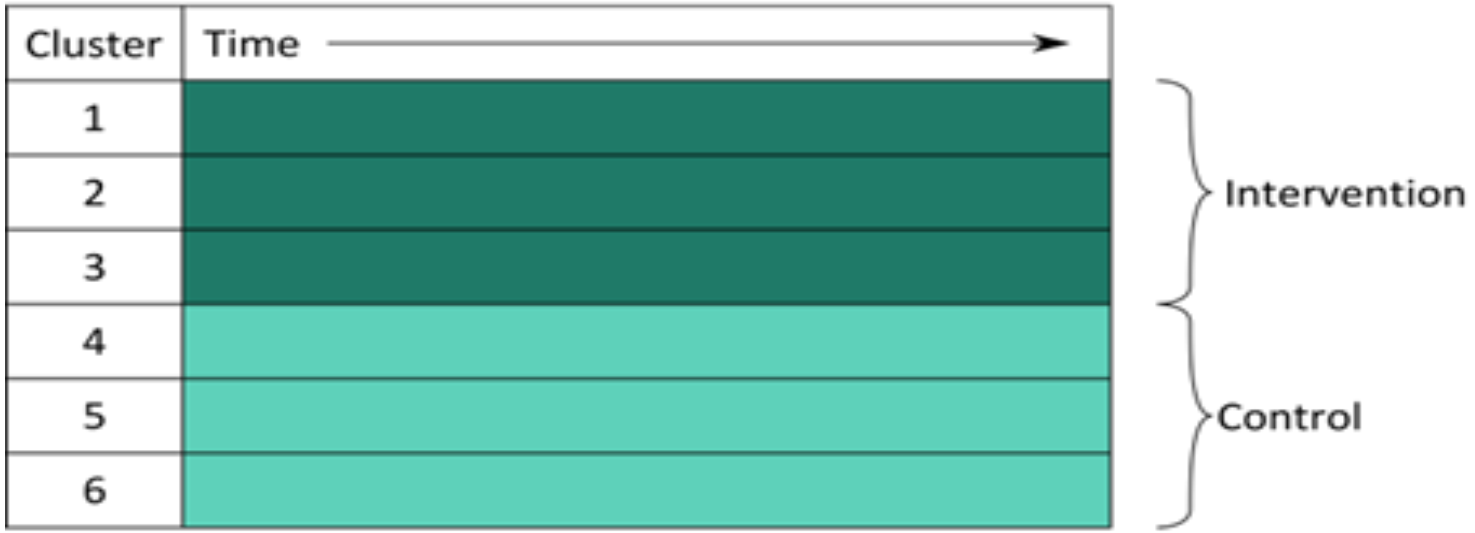
2nd Seattle Symposium on
Health Care Data Analytics
Oct 23-25, 2016

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, larger, more complex than individually randomized trial

Common Trial Designs

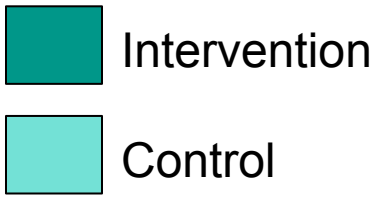
(a) Parallel Cluster Study



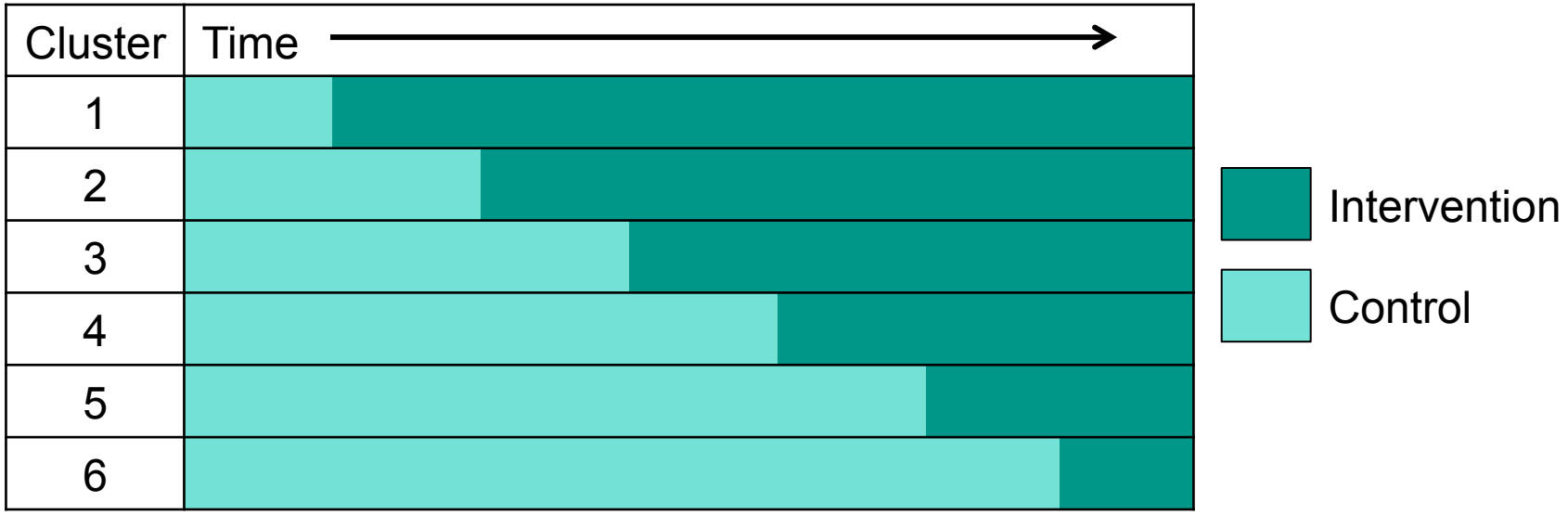
Common Trial Designs

(b) Matched Pair Parallel Cluster study

Pair	Cluster	Time →
1	1	
1	2	
2	3	
2	4	
3	5	
3	6	



Stepped wedge design



- Clusters are randomized as to <u>when</u> intervention is received
- All clusters receive intervention eventually

Stepped wedge design

- Time in NOT balanced between intervention and control periods
- Need to be able to measure outcome on each cluster, at each time step (to control for time trends)
- Cross-sectional or cohort sampling possible
    - Repeated measurements on members of a cohort may result in significant participant burden

Advantages

- Logistical or financial - cannot introduce the intervention in all units at once
- Units act as their own control, so (likely) fewer clusters needed
- Possible to study the effect of time on intervention effectiveness (i.e. seasonality, time since introduction)
- Acceptability (social, political, ethical)
  - All clusters receive the intervention
  - Intervention never removed

Disadvantages

- Long time to completion
  - Increased potential for contamination
  - Increased potential for external events to influence study
  - Potential for clusters scheduled for a later start to “jump the gun”
- Relatively complex analysis
  - Intentional confounding of time and treatment must be resolved using e.g. regression analysis
    - Dependent on assumptions

Disadvantages

- Inefficient compared to other row-column designs

	SWD	R-C
Design	0 1 1 1 1 1 0 0 1 1 1 1 0 0 0 1 1 1 0 0 0 0 1 1 0 0 0 0 0 1	1 0 1 0 1 0 0 1 0 1 0 1 1 0 1 0 1 0 0 1 0 1 0 1 1 0 1 0 1 0
Scaled coefficient matricies (fixed row and col effects) $\theta = \sum \uparrow \square c \downarrow ij \; Y \downarrow ij$	-10 14 8 2 -4 -10 -5 -11 13 7 1 -5 0 -6 -12 12 6 0 5 -1 -7 -13 11 5 10 4 -2 -8 -14 10	2 -2 2 -2 2 -2 -3 3 -3 3 -3 3 2 -2 2 -2 2 -2 -3 3 -3 3 -3 3 2 -2 2 -2 2 -2
$Var(\theta)$	$0.43\sigma^2$	$0.14\sigma^2$

Statistical Issues - Model

Model:

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

$a_i \sim N(0,\tau^2)$ – variation in mean between clusters
$c_i \sim N(0,\eta^2)$ – variation in tx effect between clusters
$e_{ijk} \sim N(0,\sigma^2)$ – random variation

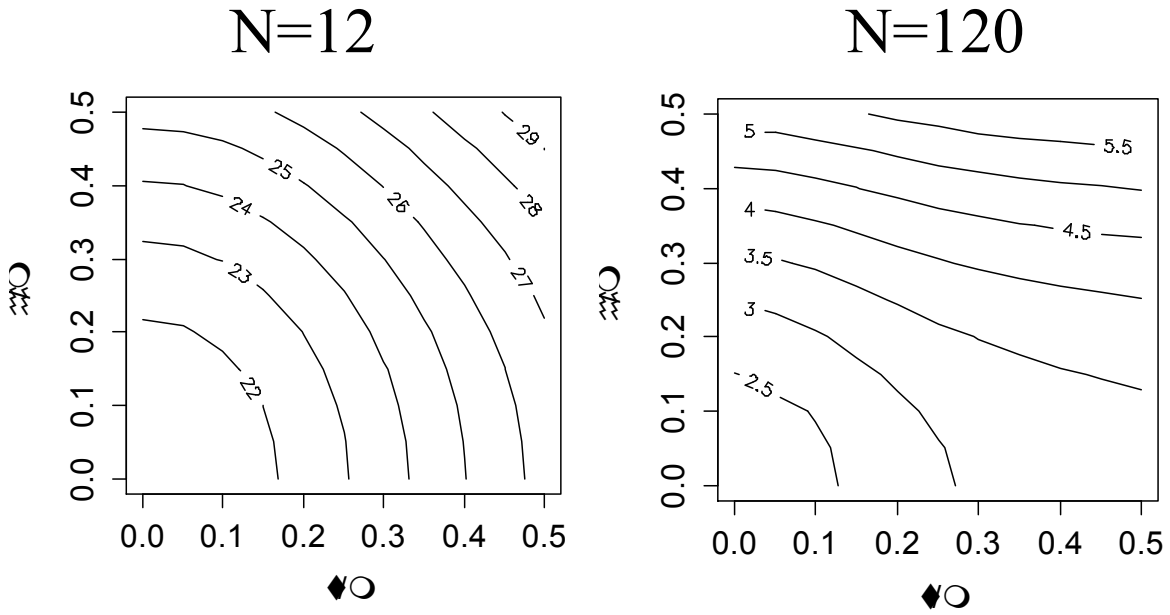
Notes:

1) "Standard" SW model does not include treatment heterogeneity
2) Model shown above assumes same time effect in all clusters
3) Assumes repeated cross-sectional sampling

Statistical Issues - Power

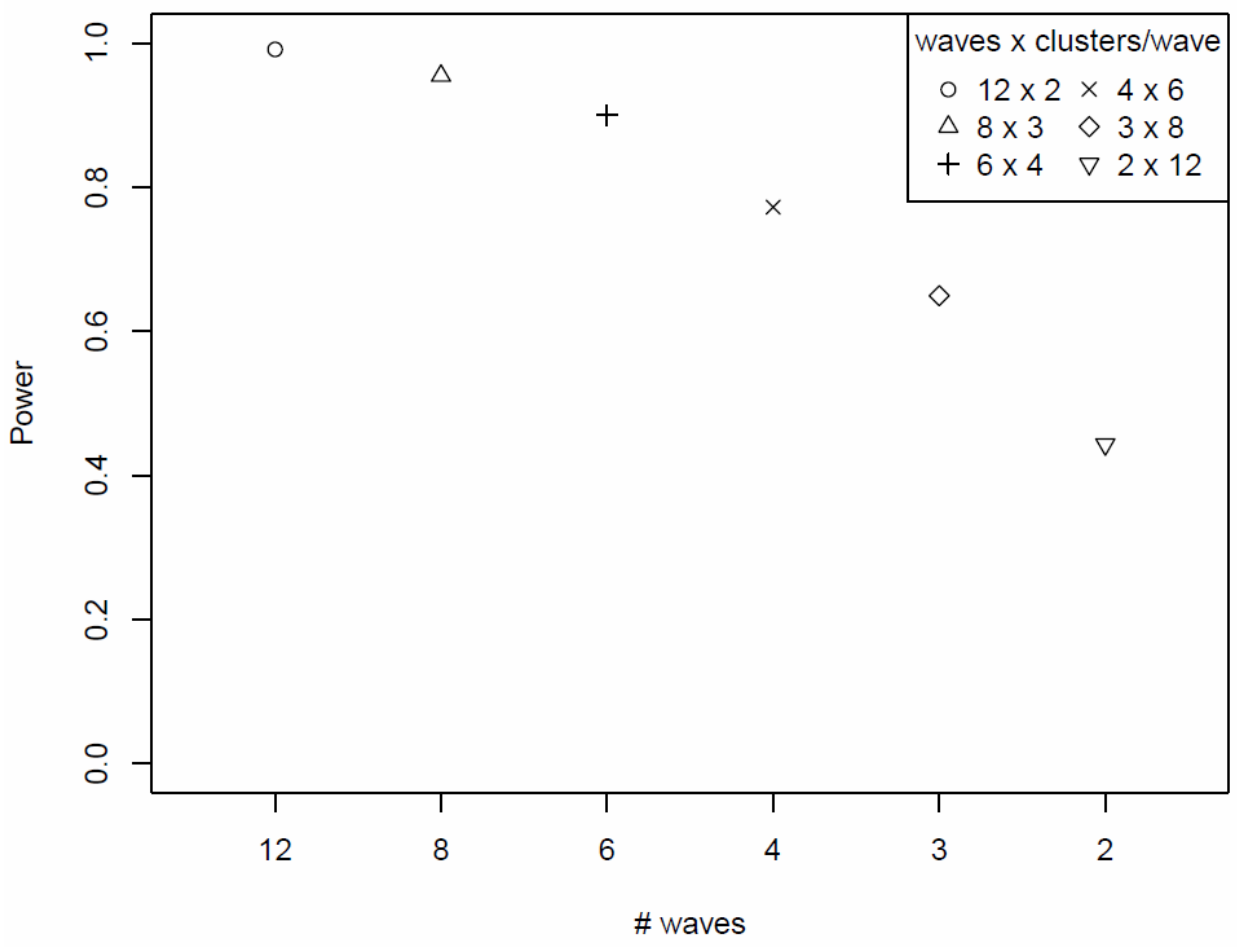
- Power = Probability of detecting a treatment effect when the treatment really works
- Depends on …
  - strength of treatment effect
  - number of clusters, steps, participants
  - variance components: σ^2 (easy to know) , η^2 ,τ^2 (hard to know).

Power – Variance Components

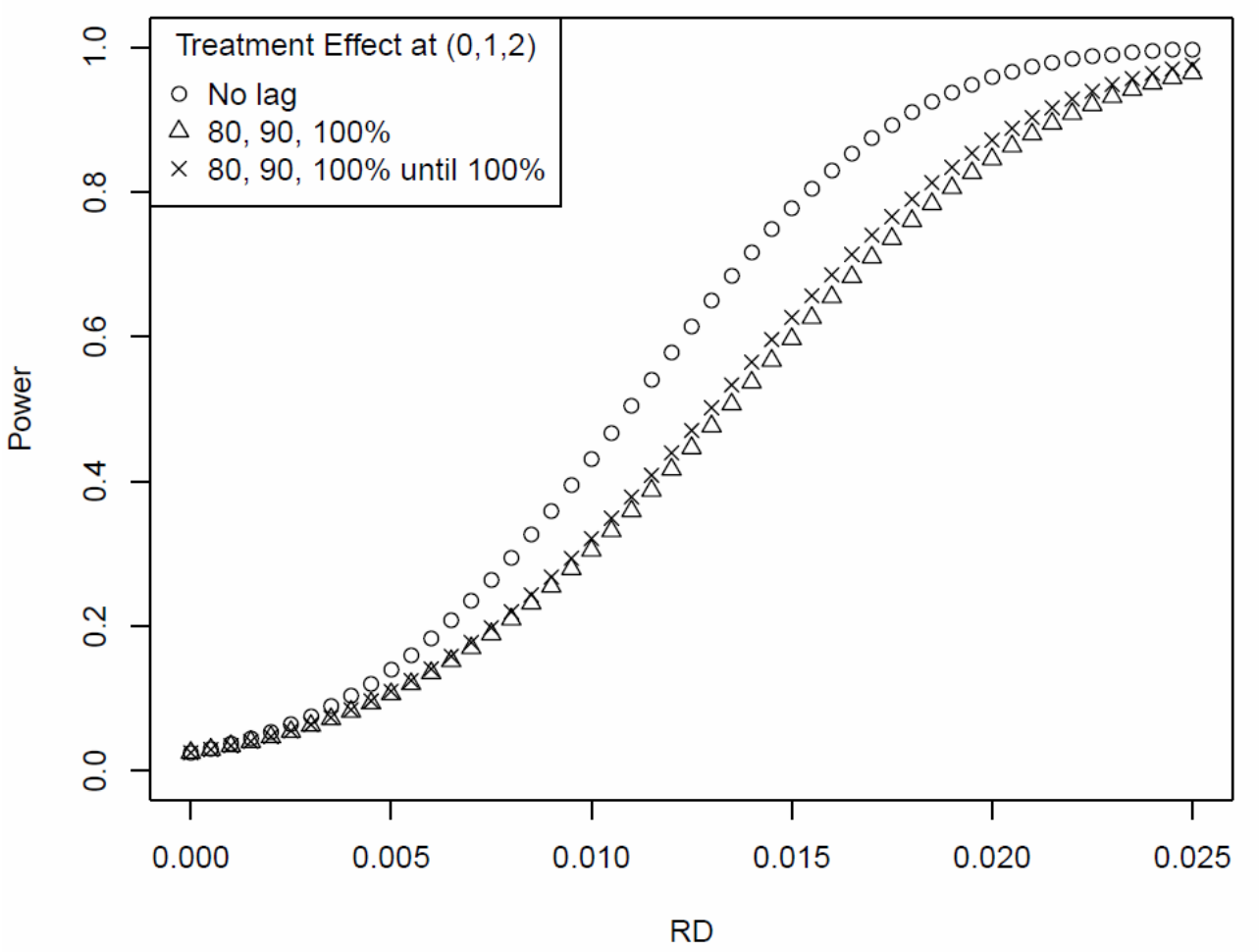


Contours of Var(θ) ($\times 10^5$) as a function of τ and η

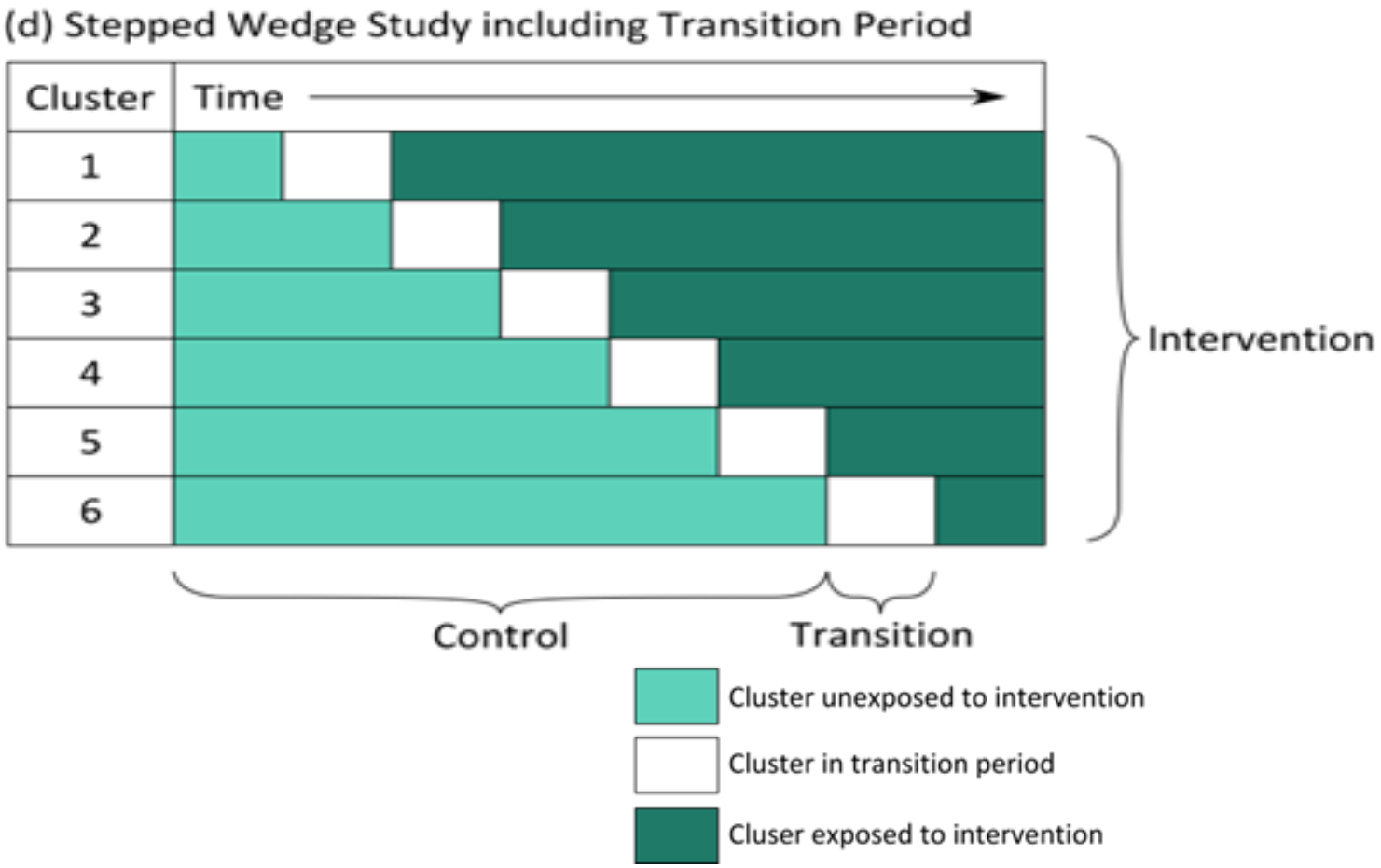
Power vs # waves



Power – Delayed treatment effect



Stepped wedge with transition period



Statistical Issues - Analysis

- Use regression based analysis (GEE, GLMM)
  - Controls for time trends and correlated data
  - Uses both within and between cluster info
  - Dependent on modelling assumptions (esp GLMM)
  - GEE w/ independent working correlation inefficient
- "Vertical" analyses
  - Compare intervention and SOC at each time point and combine
  - Valid, more robust, but potentially less efficient

Stepped Wedge extension

Concurrent

0	1	1
0	0	1
0	2	2
0	0	2

Supplementation

0	1	1	1+2	1+2	1+2
0	0	1	1	1+2	1+2
0	0	0	1	1	1+2

Factorial

0	1	1+2	1+2
0	0	1	1+2
0	0	2	1+2
0	2	1+2	1+2

Is the SW design the right design?

- Consider logistical and ethical issues, social and political acceptability
- SW useful for rollout/implementation studies
  - → For intervention A vs intervention B, parallel cluster RCT (perhaps matched) may be better
- SW confounds time trends with the intervention effect
  - → ALWAYS need to control for time trends (possibly within strata)
- SW power is sensitive to cluster variation in intervention effect
- Lag (time delay) in intervention effect reduces power
  - → Design step length > time lag
- Consider potential for changes in policy, other external factors not under investigator control

Resources

Recent Reference

- Hughes JP, Granston TS, Heagerty PJ. On the design and analysis of stepped wedge trials. Contemporary Clinical Trials. 45(Pt A):55-60, 2015.

Software: http://faculty.washington.edu/jphughes/pubs.html

- Excel spreadsheet for power calculations (does NOT include cluster to cluster variation in treatment effect)
- R package for power calculation (including cluster to cluster variation in treatment effect), data tabulation, plotting