SCHOOL OF PUBLIC HEALTH · UNIVERSITY of WASHINGTON

Department of Biostatistics

Robust Inference for Stepped Wedge Designs

Jim Hughes

2ND INTERNATIONAL CONFERENCE ON STEPPED WEDGE DESIGNS
MARCH 20, 2018

Stepped Wedge Design

		Time 1	Time 2	Time 3	Time 4
cluster	1	0	1	1	1
	2	0	0	1	1
	3	0	0	0	1

- N clusters, T time periods
- Interest is in intervention effect

Usual Analysis Approach

- Regression to disentangle time and intervention effect
- Mixed effect model

$$Y_{ijk} = \mu + \beta_j + x_{ij}\delta + \mathrm{a}_\mathrm{i} + e_{ijk}$$

  - Fixed effects for time (β_j) and intervention (δ)
  - Random cluster effect $\mathrm{a}_\mathrm{i} \sim N(0, \tau^2)$
  - Possible random time effect $\mathrm{b}_\mathrm{ij} \sim N(0, \gamma^2)$
  - Possible random intervention effect $\mathrm{d}_\mathrm{i} \sim N(0, \eta^2)$

Highly Model Dependent!

Permutation test

- Permute rows (intervention patterns)

0	1	1	1
0	0	1	1
0	0	0	1

0	1	1	1
0	0	0	1
0	0	1	1

0	0	1	1
0	1	1	1
0	0	0	1

0	0	1	1
0	0	0	1
0	1	1	1

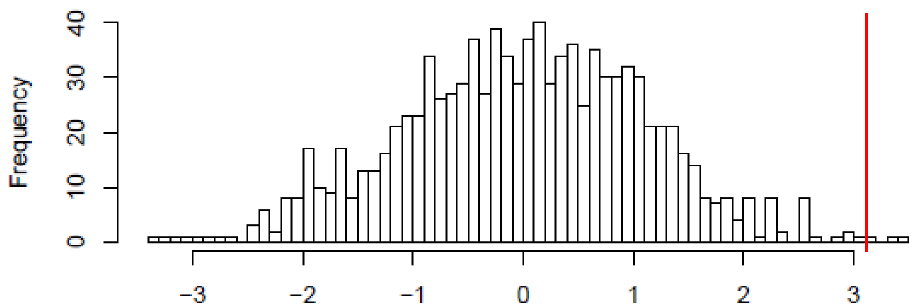
0	0	0	1
0	1	1	1
0	0	1	1

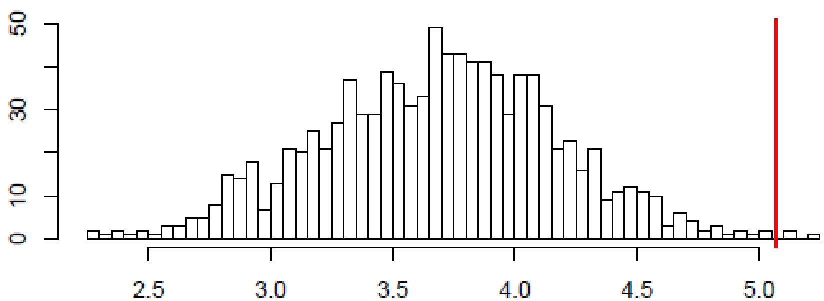
0	0	0	1
0	0	1	1
0	1	1	1

- Compute intervention effect for each permutation
- Compare observed intervention effect to permutation distribution

A surprising result …

- True model: $Y_{ijk} = \mu + \beta_j + x_{ij}\delta + \alpha_i + e_{ijk}$





Analysis Model: $Y_{ijk} = \mu + \beta_j + x_{ij}\delta + \alpha_i + e_{ijk}$ $\quad\quad$ $Y_{ijk} = \mu' + x_{ij}\delta' + e_{ijk}$

- Similar Z values ⇒ similar p-values!
- Why does it work?

An analytic result …

- Consider case where $n_{ij} = n$
- Fit model $Y_{ij} = \mu' + x_{ij}\delta' + e_{ij}$ to observed data
  - $\hat{\delta}' = \frac{4}{NT}\sum_{ij} Y_{ij}(x_{ij} - .5)$
- Repeat for all permutations
  - $E_P(\hat{\delta}') = \frac{4}{NT}\sum_{ij} Y_{ij}(\bar{x}_j - .5)$
- Define $\Delta = \hat{\delta}' - E_P\left(\hat{\delta}'\right) = \frac{4}{NT}\sum_{ij} Y_{ij}(x_{ij} - \bar{x}_j)$

An analytic result …

- Take expectation wrt Y …

$$\mathrm{E}_{\mathrm{Y}}(\Delta) = \delta \frac{4}{T} \sum_j \bar{x}_j (1 - \bar{x}_j)$$

- … suggesting the following estimate

$$\hat{\delta} = \frac{\Delta}{\frac{4}{T} \sum_j \bar{x}_j (1 - \bar{x}_j)} = \frac{\sum_{ij} Y_{ij} (x_{ij} - \bar{x}_j)}{N \sum_j \bar{x}_j (1 - \bar{x}_j)}$$

Variance of $\widehat{\delta}$

$$\hat{\delta} = \frac{\sum_{ij} Y_{ij}(x_{ij} - x_j)}{N \sum_j \bar{x}_j(1 - x_j)}$$

$$V_Y(\hat{\delta}) = \frac{\sum_i(\sum_j \sigma_{ij}^2 \left(x_{ij} - \bar{x}_j\right)^2 + 2\sum_{j<j'} \sigma_{ij,j'}^2 (x_{ij} - \bar{x}_j)(x_{ij'} - \bar{x}_{j'}))}{(N \sum_j \bar{x}_j(1 - \bar{x}_j))^2}$$

$$V_Y^*(\hat{\delta}) = \frac{\sum_i(\sum_j \sigma_j^2 \bar{x}_j(1 - \bar{x}_j) + 2\sum_{j<j'} \sigma_{j,j'}^2 \bar{x}_j(1 - \bar{x}_{j'}))}{(N \sum_j \bar{x}_j(1 - \bar{x}_j))^2}$$

* If Var(Y_i) does not depend on i

- Interesting but not useful - $V_Y(\hat{\delta})$ depends on true (but unknown) var-cov matrix

Permutation Variance of $\hat{\delta}$

- Define $Y_{ij}(\delta) = Y_{ij} - x_{ij}\delta$

$$V_\delta^1(\hat{\delta}) = \frac{\left\{ \begin{array}{c} \sum_i \left[\sum_j Y_{ij}(\delta)^2 \bar{x}_j \left(1 - \bar{x}_j\right) + 2 \sum_{j<j'} Y_{ij}(\delta) Y_{ij'}(\delta) \bar{x}_j \left(1 - \bar{x}_j'\right) \right] - \\ \frac{2}{N-1} \sum_{i<i'} \sum_{i,j'} Y_{ij}(\delta) Y_{i'j'}(\delta)\, \bar{x}_{\min(j,j')} \left(1 - \bar{x}_{\max(j,j')}\right) \end{array} \right\}}{\left(N \sum_j \bar{x}_j \left(1 - \bar{x}_j\right)\right)^2}$$

- $\mathrm{E_Y}\left(\mathrm{V}_\delta^1(\hat{\delta})\right) = V_Y^*(\hat{\delta})$
- Situations where Var(Y_i) depends on i:
  - Random treatment effect
  - Sample size varies between clusters

Robust Inference

1) Hypothesis Testing

Ho: $\delta = \delta_o$

Ha: $\delta \neq \delta_o$

$$Z = \frac{\hat{\delta} - \delta_o}{\sqrt{V_{\delta o}(\hat{\delta})}}$$

2) Confidence Intervals

$$\left\{ \delta : -Z_{1-\alpha/2} < \frac{\hat{\delta} - \delta}{\sqrt{V_{\delta}(\hat{\delta})}} < Z_{1-\alpha/2} \right\}$$

Alternative variance estimates

1. Consider using $V^{1}_{\hat{\delta}}(\hat{\boldsymbol{\delta}})$
   - Biased estimate of $V_Y\ (\hat{\delta})$
   - Bias correction complex, appears to depend (weakly) on true Var(Y), but multiplying by N/(N-1) seems to work pretty well
   - Reference distribution may be t-distribution?

2. $V^{2}(\hat{\delta})$
   - does not depend on δ; does not assume Var(Y_i) $\perp$ i
   - requires each intervention sequence occurs $\geq$2 times

Simulation Setting

H_o: $\delta = 0$

H_a: $\delta \neq 0$

- N = 12, 24, 36
- T = 5 time intervals
- β = -0.1 per time interval
- n = 10 (average) (constant, low var, high var)
- Error variance (σ^2) = 1
- Cluster variance (τ^2) = 0.2
- Intervention variance (η^2) = 0, 0.1, 0.4
- 10,000 replicates

Simulation Results – Type I error (δ = 0)

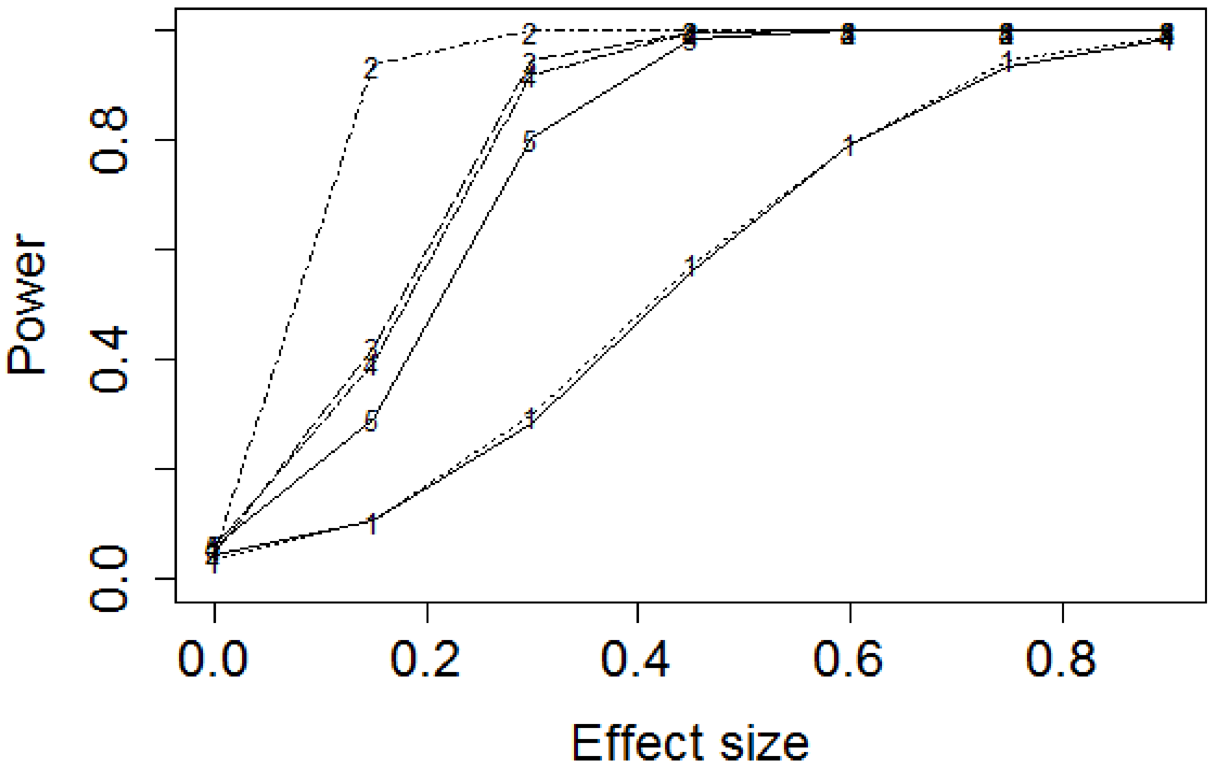
		$V^1_{\delta=0}(\hat{\delta})$			$V^1_{\delta=\hat{\delta}}(\hat{\delta})$			$V^2(\hat{\delta})$		
		Treatment variance (η^2)								
N	n	0	0.1	0.4	0	0.1	0.4	0	0.1	0.4
	constant	0.05	0.05	0.05	0.06	0.07	0.07	0.09	0.09	0.10
12	low var	0.05	0.05	0.05	0.06	0.07	0.07	0.10	0.09	0.09
	hi var	0.05	0.05	0.05	0.07	0.07	0.07	0.09	0.08	0.08
	constant	0.05	0.05	0.06	0.06	0.06	0.06	0.07	0.07	0.07
24	low var	0.05	0.05	0.06	0.06	0.06	0.06	0.07	0.07	0.07
	hi var	0.04	0.04	0.05	0.05	0.05	0.05	0.07	0.06	0.06
	constant	0.05	0.05	0.06	0.05	0.06	0.06	0.06	0.06	0.06
36	low var	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06
	hi var	0.05	0.06	0.06	0.06	0.06	0.07	0.06	0.06	0.06

Simulation Results – Power (δ = 1)

		$V^1_{\delta=0}(\hat{\delta})$			$V^1_{\delta=\hat{\delta}}(\hat{\delta})$			$V^2(\hat{\delta})$		
		Treatment variance (η^2)								
N	n	0	0.1	0.4	0	0.1	0.4	0	0.1	0.4
12	constant	0.59	0.57	0.51	0.64	0.61	0.55	0.65	0.62	0.55
	low var	0.59	0.56	0.49	0.63	0.60	0.53	0.64	0.60	0.54
	hi var	0.42	0.41	0.38	0.46	0.45	0.42	0.48	0.47	0.42
24	constant	0.90	0.87	0.81	0.90	0.88	0.82	0.90	0.88	0.80
	low var	0.89	0.87	0.81	0.90	0.88	0.82	0.90	0.88	0.81
	hi var	0.78	0.76	0.71	0.79	0.78	0.72	0.82	0.79	0.73
36	constant	0.97	0.97	0.94	0.98	0.97	0.94	0.97	0.97	0.93
	low var	0.97	0.96	0.93	0.97	0.96	0.94	0.97	0.96	0.93
	hi var	0.84	0.83	0.79	0.85	0.84	0.80	0.84	0.82	0.78

Simulation Results – Power

- Compare robust estimator to optimal estimate from mixed model



Solid = robust estimator
1 = independence
2 = random cluster
3 = random cluster + tx
4 = random cluster + time
5 = random cluster + tx + time

$ICC_{cluster}$ = .17
ICC_{tx} = .091
ICC_{time} = .038

Summary/Directions

- Gain in robustness; loss in efficiency
- Application/simulations for binary data
- Closed form estimates for more efficient model (e.g. exchangeable correlation) over permutation distribution?

Permutation tests – a cautionary tale

- N = 20, n constant
- 1000 replicates
- Permutation test based on Z-statistics

- S1 = Random intercept
- S2 = Random intercept + intervention

Asymptotic Type I error rate

	Indep	LMM intercept	LMM Intercept+ intervent	GEE indep	GEE exch
S1	.86	.04	.04	.10	.05
S2	.89	.85	.06	.12	.09

Permutation Type I error rate

	Indep	LMM intercept	LMM Intercept+ intervent	GEE indep	GEE exch
S1	.03	.05	.05	.05	.05
S2	.05	.14	.07	.04	.06

Permutation Variance of $\hat{\delta}$

$$V_P(\hat{\delta}) = \frac{\left\{ \begin{matrix} \sum_i \left[\sum_j Y_{ij}^2 \bar{x}_j \left(1-\bar{x}_j\right) + 2\sum_{j<j'} Y_{ij} Y_{ij'} \; \bar{x}_j \left(1-\bar{x}_j'\right)\right] - \\ \frac{2}{N-1} \sum_{i<i'} \sum_{i,j'} Y_{ij} Y_{i'j'} \bar{x}_{\min(j,j')} \left(1-\bar{x}_{\max(j,j')}\right) \end{matrix} \right\}}{\left(N \sum_j \bar{x}_j \left(1-\bar{x}_j\right)\right)^2}$$

- $\mathrm{E_Y}\left(\mathrm{V_P}\left(\hat{\delta}\right)\right) = V_Y^*\left(\hat{\delta}\right) + \delta\, C(x)$
- $V_P\left(\hat{\delta}\right)$ is a biased estimator of $V_Y\left(\hat{\delta}\right)$ unless δ = 0
- But bias does NOT depend on covariates that are constant within columns of SW design (e.g. time)

Estimate of $V_Y(\hat{\delta})$

- If every intervention sequence occurs $\geq$2 times then …

$$V^2(\hat{\delta}) = \sum_h \frac{\left\{ \begin{array}{c} \sum_{i=1}^{n(h)} \left[\sum_j {Y_{hij}}^2 \left(x_{hij} - \bar{x}_j\right)^2 + 2 \sum_{j<j'} Y_{hij} Y_{hij'} \left(x_{hij} - \bar{x}_j\right)\left(x_{hij'} - \bar{x}_{j'}\right) \right] - \\ \frac{2}{n(h)-1} \sum_{i<i'} \sum_{i,j'} Y_{hij} Y_{hi'j'} \left(x_{hij} - \bar{x}_j\right)\left(x_{hij'} - \bar{x}_{j'}\right) \end{array} \right\}}{\left(N \sum_j \bar{x}_j \left(1 - \bar{x}_j\right)\right)^2}$$

- $\mathrm{E}_{\mathrm{Y}}\left(\mathrm{V}^2\left(\hat{\delta}\right)\right) = V_Y \ \left(\hat{\delta}\right)$

Notes:

- One is tempted to try

$$V_\delta^c(\hat{\delta}) = \frac{\left\{\begin{array}{c}\sum_i \left[\sum_j Y_{ij}(\delta)^2 \left(x_{ij} - \bar{x}_j\right)^2 + 2\sum_{j<j'} Y_{ij}(\delta) Y_{ij'}(\delta) \left(x_{ij} - \bar{x}_j\right)\left(x_{ij'} - \bar{x}_{j'}\right)\right] - \\ \frac{2}{N-1}\sum_{i<i'}\sum_{j,j'} Y_{ij}(\delta) Y_{i'j'}(\delta)\, \bar{x}_{\min(j,j')}\left(1 - \bar{x}_{\max(j,j')}\right)\end{array}\right\}}{\left(N\sum_j \bar{x}_j\left(1 - \bar{x}_j\right)\right)^2}$$

- This estimate is unbiased for $V_Y(\hat{\delta})$ regardless of true covariance matrix and SW design …
- … but can be negative!