Randomization-Based Inference for Cluster Randomized Trials

Dustin J. Rabideau

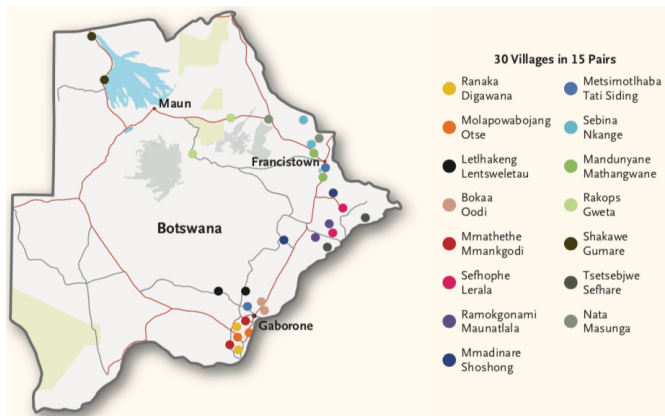November 20, 2023

Cluster Randomized Trial (CRT)

Individually-randomized trial: individuals are randomized

CRT: **groups/clusters of people** are randomized

Example: The Botswana Combination Prevention Project

30 Villages in 15 Pairs

Ranaka Digawana · Metsimotlhaba Tati Siding · Molapowabojang Otse · Sebina Nkange · Letlhakeng Lentsweletau · Mandunyane Mathangwane · Bokaa Oodi · Rakops Gweta · Mmathethe Mmankgodi · Shakawe Gumare · Sefhophe Lerala · Tsetsebjwe Sefhare · Ramokgonami Maunatlala · Nata Masunga · Mmadinare Shoshong

Makhema et al. (2019). Universal Testing, Expanded Treatment, and Incidence of HIV Infection in Botswana. *NEJM*.

Large HIV prevention CRT

- 30 communities, 8,551 individuals in HIV-incidence cohort
- Intervention: combination prevention package vs. standard-care
- Outcome: HIV infection, annual study visits (interval-censored)

Reasons to Conduct a CRT

1. Intervention is more naturally/feasibly applied at cluster level

2. Avoid treatment contamination

3. Capture population-level (indirect) effects of intervention

Statistical Analysis of a CRT

Statistical challenges in CRTs

- individual-level outcomes within a cluster are **correlated**
- small sample setting (only 30 clusters, 15/group)
- design features (pair-matched randomization)

Typical regression approaches for CRTs:

- **mixed effects model** (random effects/maximum likelihood)
- **marginal model via GEE** (generalized estimating equation)

Distributional assumptions not met or a small # of clusters

- **inaccurate** p-values and confidence intervals
- $\implies$ **Wrong conclusions could be drawn from the study!**

Randomization-Based Inference

(A.k.a. permutation methods, re-randomization tests)

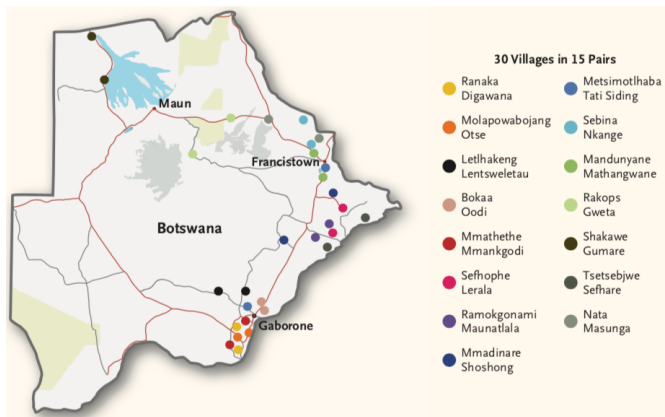Recent resurgence of interest in randomization-based inference for CRTs

Advantages:

- Distribution-free (outcome, correlation)
- Exact (small # clusters)

Challenges

- Less common, less familiar
- More computational time required
- Focus on tests/p-values; **confidence interval methods are limited**

Example: The Botswana Combination Prevention Project

Makhema et al. (2019). Universal Testing, Expanded Treatment, and Incidence of HIV Infection in Botswana. *NEJM*.

"The unadjusted HIV incidence ratio in the intervention group as compared with the standard-care group was 0.69 (**p=0.09**) by [randomization] test (95% confidence interval [CI], **0.46 to 0.90** by pair-stratified Cox model)."

Learning Objectives

1. Learn how to conduct a randomization test

2. Learn how to calculate a randomization-based confidence interval

3. Reinforce concepts by applying these methods to the Botswana Combination Prevention Project

Setting and Notation

Parallel CRT with 2 treatment groups

- # of individuals per cluster can vary
- Y_{ki} is outcome for the ith individual, kth cluster
- $X_k = 1$ (intervention group)
- $X_k = 0$ (control group)

Randomization Test

Quantity of interest: marginal effect of treatment (X) on outcome (Y)

$$\theta = g\{E(Y_{ki}|X_k = 1)\} - g\{E(Y_{ki}|X_k = 0)\}$$

- continuous Y (ignore g), θ represents **difference in means**
- binary Y and "logit" g, θ represents **(log) odds ratio**

To conduct a randomization test of no treatment effect $(H_0 : \theta = 0)$:

1. Fit regression model with **observed** data, get observed $\widehat{\theta}$
2. **Shuffle** treatment assignments, re-fit model, get $\widehat{\theta}^{(2)}$; Ditto $\widehat{\theta}^{(3)}$; ...
3. Calculate % of permuted estimates "as or more extreme" than $\widehat{\theta}$

Randomization Test

1. Fit regression model with **observed** treatment vector $\boldsymbol{x} = (x_1, \ldots, x_K)$ to get $\widehat{\theta}^{(1)} = \widehat{\theta}$

$$g\{E(Y_{ki}|X_k)\} = \mu + \theta x_k$$

```
> fit <- glm(y ~ x, data = ds)
> coef(fit)["x"] # thetaHat
```

⋆⋆ Note, we are fitting a model (GLM) typically used in the **independent data** (i.e., non-clustered) setting

Randomization Test

2. Shuffle treatment assignments and re-fit model with **permuted** treatment vector $\boldsymbol{X}^{(p)}$ to get $\widehat{\theta}^{(p)}$ (and repeat this step for $p = 2, \ldots, P$)
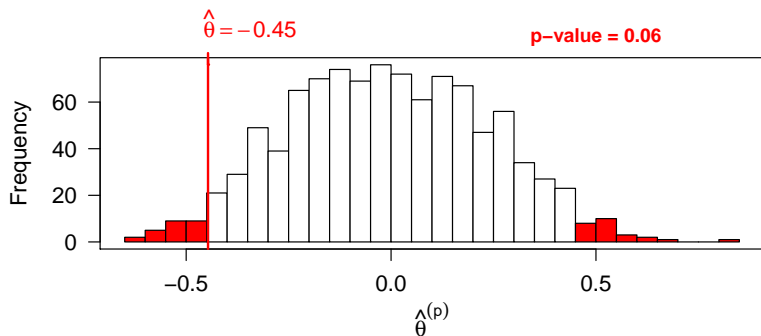
$$g\{E(Y_{ki}|X_k)\} = \mu + \theta X_k^{(p)}$$

```
> for (p in 2:P) {
>     ds$xp <- permute(ds$x) # special function
>     fit <- glm(y ~ xp, data = ds)
>     coef(fit)["xp"]} # thetaHat_p
```

| | Cluster | |
$\boldsymbol{X}^{(p)}$	1	2	3	4	5	6	$\widehat{\theta}^{(p)}$
$\boldsymbol{X}^{(1)} = \boldsymbol{x}$	1	1	1	0	0	0	$\widehat{\theta}^{(1)} = \widehat{\theta}$
$\boldsymbol{X}^{(2)}$	1	0	0	1	0	1	$\widehat{\theta}^{(2)}$
$\boldsymbol{X}^{(3)}$	0	1	1	1	0	0	$\widehat{\theta}^{(3)}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\boldsymbol{X}^{(P)}$	0	0	1	0	1	1	$\widehat{\theta}^{(P)}$

Randomization Test

3. Calculate p-value = proportion of $\{\widehat{\theta}^{(1)}, \widehat{\theta}^{(2)}, \ldots, \widehat{\theta}^{(P)}\}$ that are "as or more extreme" than (observed) $\widehat{\theta}^{(1)} = \widehat{\theta}$

Randomization-Based Confidence Interval

Need to "invert" randomization test to calculate a confidence interval (CI)

- Conduct **many** randomization tests to see which θs are "reasonable"

To calculate a randomization-based confidence interval for θ:

1. Conduct randomization test for a non-zero null value ($H_0 : \theta = \theta_0$)
2. Repeat across many different θ_0
3. Collect all θ_0 **not** rejected by this test; bounds form CI

Randomization-Based Confidence Interval

1. Conduct randomization test for a non-zero null value ($H_0 : \theta = \theta_0$)

Mathematically equivalent to $H_0 : \tau = (\theta - \theta_0) = 0$ (zero null)

$$g\{E(Y_{ki}|X_k^{(p)})\} = \mu + \theta_0 x_k + \tau X_k^{(p)}$$

- **Permuted** treatment $\boldsymbol{X}^{(p)}$ for offset-adjusted term $\tau X_k^{(p)}$
- **Observed** treatment $\boldsymbol{x}$ for **offset** $\theta_0 x_k$ (fixed across all permutations)

```
> fit <- glm(y ~offset(theta0 * x) + xp, data = ds)
> coef(fit)["xp"] # tauHat_p
```

⋆⋆ Note, this boils down to conducting a randomization test the same way as before, **but now with a fixed offset term in your model**
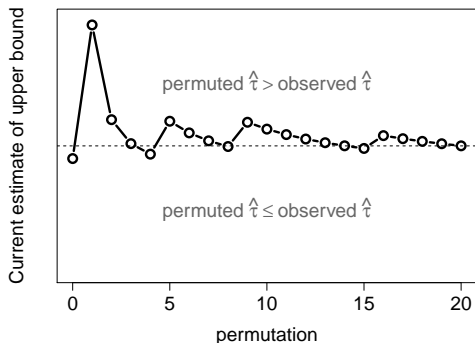
Randomization-Based Confidence Interval

2. Repeat across many different θ_0
3. Collect all θ_0 **not** rejected by this test; bounds form CI
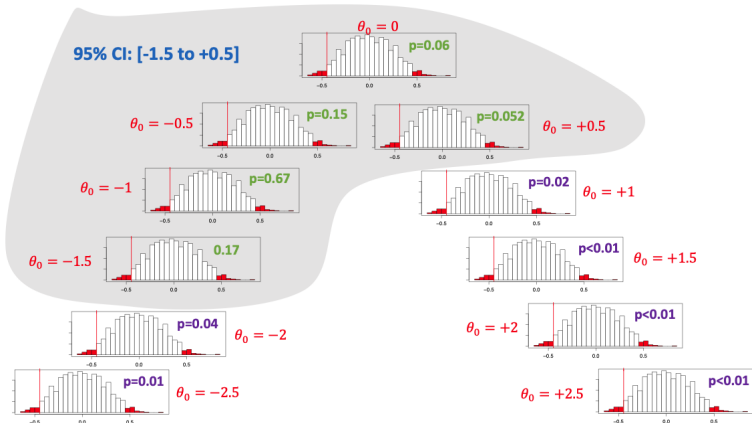
Fast CI Computation

Grid/binary search $= (P$ perms$) \times ($many $\theta_0) \quad \approx$ hours to days

Efficient search[1] $= (P$ perms$) \qquad\qquad\qquad \approx$ **seconds to minutes**

[1] Adapted from: (1) Garthwaite (1996). Confidence intervals from randomization tests. *Biometrics*. (2) Garthwaite and Jones (2009). A Stochastic Approximation Method and Its Application to Confidence Intervals. *J. Comput. Graph. Stat.*

Introduction   Test   **Confidence Interval**   Example   Appendix   drabideau@mgh.harvard.edu   17 / 27

Accounting for Study Design Features

E.g. stratified, pair-matched, restricted randomization

Typical solution: include additional term(s) in model
- can change value/interpretation of the targeted parameter
- can exacerbate GEE small-sample bias (or result in overcorrection)

Accounting for Study Design Features

E.g. stratified, pair-matched, restricted randomization

Typical solution: include additional term(s) in model
- can change value/interpretation of the targeted parameter
- can exacerbate GEE small-sample bias (or result in overcorrection)

Randomization-based solution: restrict $\boldsymbol{X}^{(p)}$ based on design
- E.g. BCPP: 30 communities, pair-matched randomization
  $\implies$ sample $\boldsymbol{X}^{(p)}$ from among $2^{15} \approx 33\text{K}$ (not $\frac{30!}{15!15!} \approx 155\text{M}$)
- maintain target of inference and parsimonious nonstratified model

R Package on Github: permuter

 My R package makes this all **very easy** to implement!

```
> devtools::install_github("djrabideau/permuter")

> fit <- glm(y ~ x, data = ds)
> permtest(fit, data = ds, ...)
> permci(fit, data = ds, ...)

> ?permtest
> ?permci
```
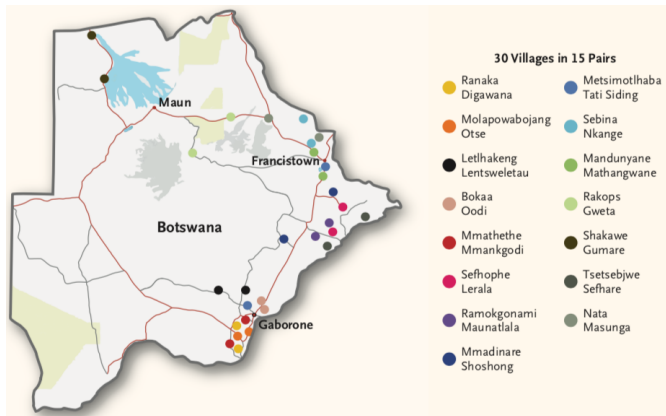
R package link: https://github.com/djrabideau/permuter

Example: The Botswana Combination Prevention Project

30 Villages in 15 Pairs

Ranaka Digawana
Molapowabojang Otse
Letlhakeng Lentsweletau
Bokaa Oodi
Mmathethe Mmankgodi
Sefhophe Lerala
Ramokgonami Maunatlala
Mmadinare Shoshong

Metsimotlhaba Tati Siding
Sebina Nkange
Mandunyane Mathangwane
Rakops Gweta
Shakawe Gumare
Tsetsebjwe Sefhare
Nata Masunga

Makhema et al. (2019). Universal Testing, Expanded Treatment, and Incidence of HIV Infection in Botswana. *NEJM*.

"The unadjusted HIV incidence ratio in the intervention group as compared with the standard-care group was 0.69 (**p=0.09**) by [randomization] test (95% confidence interval [CI], **0.46 to 0.90** by pair-stratified Cox model)."

BCPP Data

Primary outcome: HIV infection measured at annual study visits

- interval-censored time-to-event outcome

pid	community	pair	trt	hiv	left	right
01-001	1	1	Intervention	0	754	Inf
01-002	1	1	Intervention	1	681	765
01-003	1	1	Intervention	0	404	Inf
02-001	2	1	Standard	0	702	Inf
02-002	2	1	Standard	1	404	668
03-001	3	2	Intervention	0	354	Inf
⋮	⋮	⋮	⋮	⋮	⋮	
30-282	30	15	Intervention	0	689	Inf

BCPP randomization test

```
> fit <- survreg(Surv(left, right, type = "interval2") ~ trt)
> test <- permtest(fit, trtname = "trt", runit = "community",
                   strat = "pair", nperm = 5000)
> plot(ptest)
```


Randomization Distribution
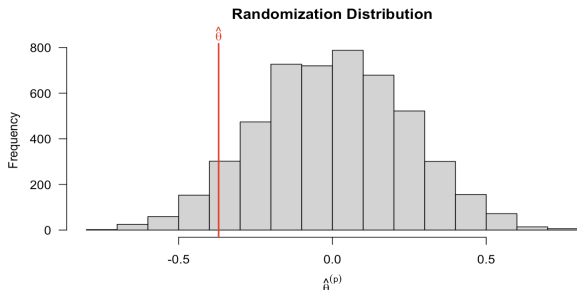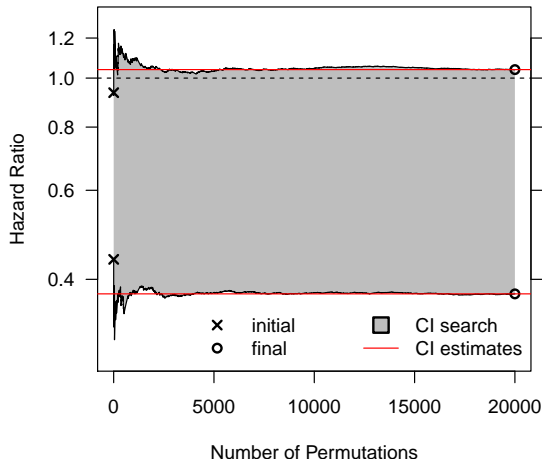
$\widehat{\theta} = -0.37$

Hazard Ratio $= 0.69$

Randomization **p=0.09**

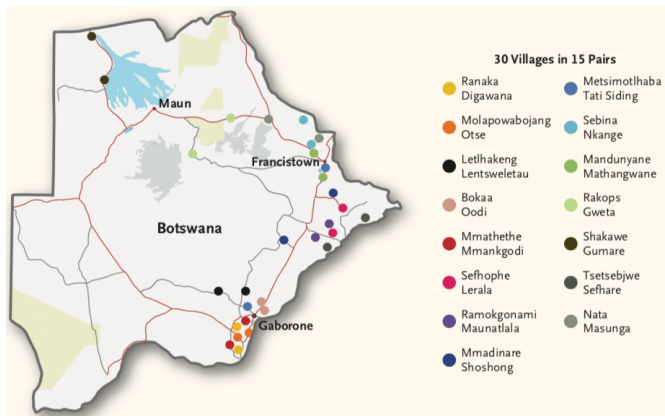⏱ 2 minutes on laptop

BCPP randomization-based confidence interval

```
> pci <- permci(fit, trtname = "trt", runit = "community",
                strat = "pair", nperm = 20000)
> plot(pci)
```


Randomization-based
95% CI: **0.37 to 1.04**

⏱ 40 minutes on laptop

Example: The Botswana Combination Prevention Project

30 Villages in 15 Pairs

- Ranaka Digawana
- Molapowabojang Otse
- Letlhakeng Lentsweletau
- Bokaa Oodi
- Mmathethe Mmankgodi
- Sefhophe Lerala
- Ramokgonami Maunatlala
- Mmadinare Shoshong
- Metsimotlhaba Tati Siding
- Sebina Nkange
- Mandunyane Mathangwane
- Rakops Gweta
- Shakawe Gumare
- Tsetsebjwe Sefhare
- Nata Masunga

Makhema et al. (2019). Universal Testing, Expanded Treatment, and Incidence of HIV Infection in Botswana. *NEJM*.

"The unadjusted HIV incidence ratio in the intervention group as compared with the standard-care group was 0.69 (**p=0.09**) by [randomization] test (95% confidence interval [CI], **0.46 to 0.90** by pair-stratified Cox model))."

Now, we get 0.69 (**p=0.09**) with 95% CI, **0.37 to 1.04** ✓

Summary

Randomization-based inference is a robust analysis strategy for CRTs

- Distribution-free (outcome, correlation)
- Exact (small # clusters)

To conduct a randomization **test** of no treatment effect ($H_0 : \theta = 0$):

1. Fit regression model with **observed** data, get observed $\widehat{\theta}$
2. **Shuffle** treatment assignments, re-fit model, get $\widehat{\theta}^{(2)}$; Ditto $\widehat{\theta}^{(3)}$; ...
3. Calculate % of permuted estimates "as or more extreme" than $\widehat{\theta}$

To calculate a randomization-based **confidence interval** for θ:

1. Conduct randomization test for non-zero null ($H_0 : \theta = \theta_0$) via **offset**
2. Repeat across many different θ_0
3. Collect all θ_0 **not** rejected by this test; bounds form CI

To Dive Deeper. . .

Recommended papers

- Rabideau and Wang (2021). Randomization-based confidence intervals for cluster randomized trials. *Biostatistics.*
- Rabideau and Wang (2021). Randomization-based inference for a marginal treatment effect in stepped wedge cluster randomized trials. *Stat. Med.*
- Ernst (2004). Permutation Methods: A Basis for Exact Inference. *Stat. Sci.*

Try out my R package

- https://github.com/djrabideau/permuter

Thanks!

| Method	HR	95% CI	p-value
Randomization, Marginal	0.640	[0.374, 1.039]	0.064
Randomization, Pair-Stratified	0.646	[0.369, 1.054]	0.068
Weibull, Frailty-Cluster	0.640	[0.432, 0.947]	0.025
Weibull, Frailty-Pair	0.641	[0.453, 0.905]	0.012
Weibull, Pair-Stratified	0.646	[0.457, 0.913]	0.013

The Botswana Combination Prevention Project
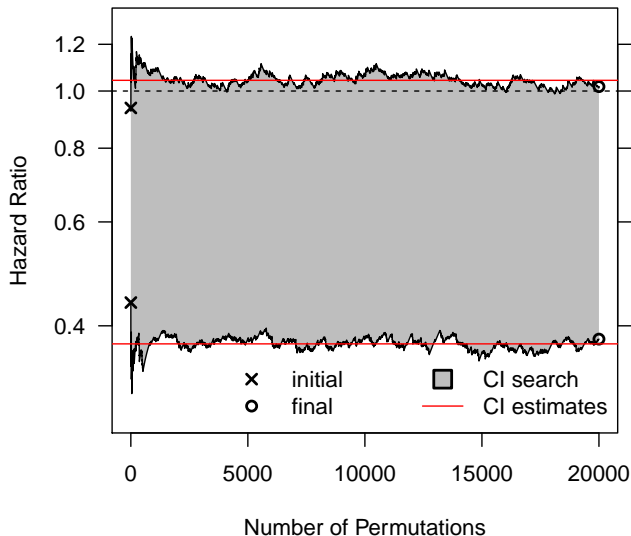
Monitoring 4 separate chains using different starting values
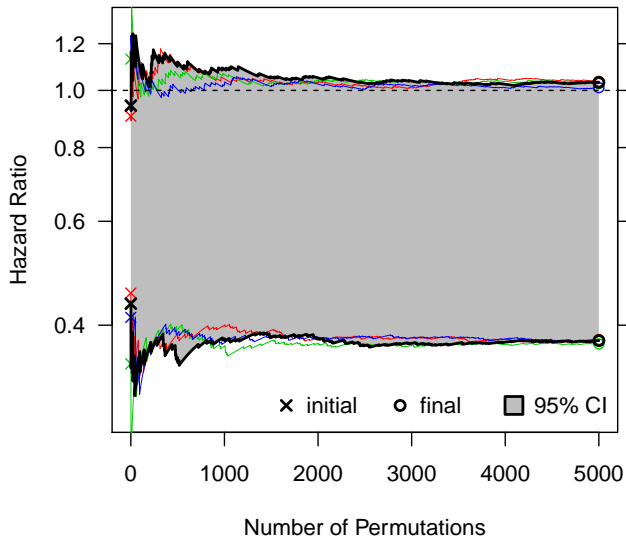
The Botswana Combination Prevention Project

Search procedure adapted from Garthwaite (1996)
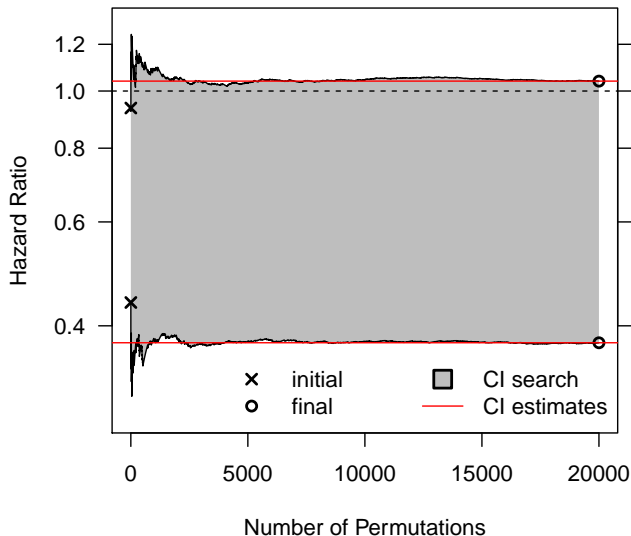
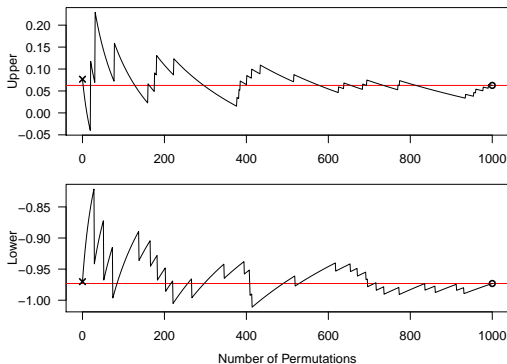# The Botswana Combination Prevention Project

Search procedure adapted from Garthwaite and Jones (2009)

R package
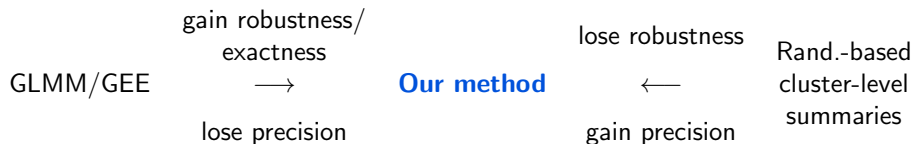
```
> m1 <- glm(bpepisodes ~ spnvac, family = poisson, data = pneumovac)
> ci <- permci(m1, trtname = "spnvac",
               runit = "randunit", data = pneumovac,
               nperm = 1000, ncores = 2, seed = 445, level = 0.95)
> print(ci$ci)
      lower       upper
-0.97314014   0.06265964
> plot(ci)
```

Our method **provides a compromise** between existing approaches:

	gain robustness/ exactness		lose robustness	
GLMM/GEE	$\longrightarrow$	**Our method**	$\longleftarrow$	Rand.-based cluster-level summaries
	lose precision		gain precision	

Randomization-Based CI for SWTs

Population model

$$(\boldsymbol{Y}_i | \boldsymbol{X}_i = \boldsymbol{x}_*) \sim F(\boldsymbol{\eta}_{\boldsymbol{x}_*}, \phi), \quad \boldsymbol{\eta}_{\boldsymbol{x}_*} = (\eta_{1x_1}, \ldots, \eta_{Jx_J})^T,$$
$$\eta_{jx} = g\{E(Y_{ijk} | X_{ij} = x)\} = \mu + \beta_j + \theta x$$

- constant treatment effect (θ) across clusters and time
- common **average** secular trend across clusters (i.e. same β_j for all i)
- correlation not impacted by treatment, but otherwise **unspecified**

Comply:
- Exchangeable correlation structure (Hussey and Hughes, 2007)
- Nested exchangeable (Hooper et al, 2015; Hemming et al, 2017)
- Exponential decay (Kasza et al, 2019)

Do not comply:
- Treatment heterogeneous correlation structure (Hughes et al, 2015)
- Treatment effect heterogeneity (models C-E in Hemming et al, 2017)

Simulations: Different Target Parameters

Marginal model: **cluster-average** treatment effect

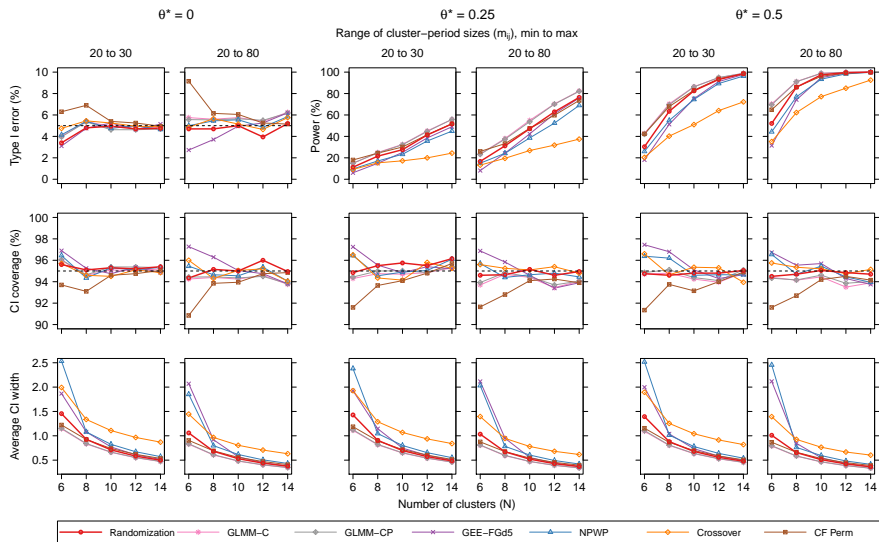- $\theta = g\{E(Y_{ki}|X_k = 1)\} - g\{E(Y_{ki}|X_k = 0)\}$

Mixed model: **cluster-specific** treatment effect

- $\theta^* = g\{E(Y_{ki}|X_k = 1, \gamma_k)\} - g\{E(Y_{ki}|X_k = 0, \gamma_k)\}$
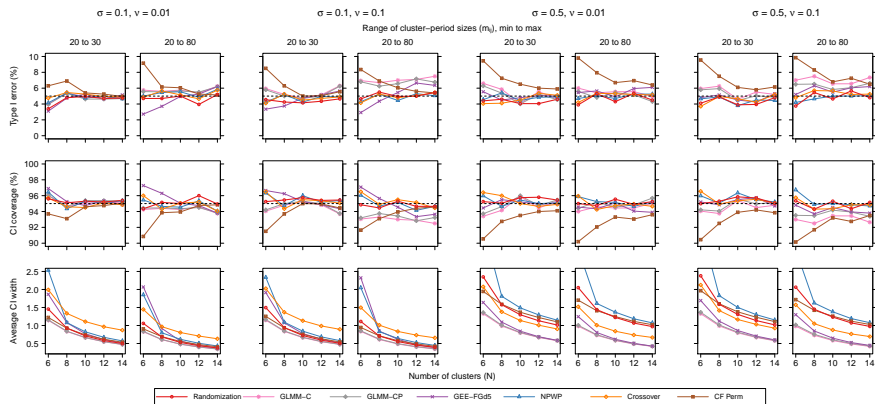- relation to marginal θ: integrate over random effects

Cluster-Level Analysis:

- (weighted) average of cluster(-period) summaries
- relation to marginal θ can be **complex**
  - contrast function (e.g. nonlinear)
  - weights
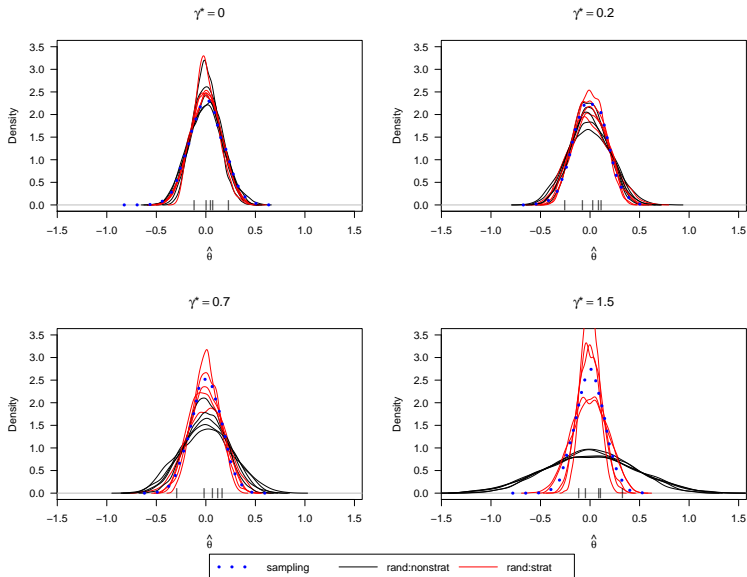 - cluster-period sizes
  - heuristic adjustments

# Simulations: Accounting for Study Design Features

Simulations: Stratified SWT with a Binary Outcome

- single binary stratification factor Z
- larger γ^* $\implies$ larger Y-Z association
- both nonstratified (-ns) and stratified analysis

Method	γ^*	CI coverage (%)			Average CI width		
				N			
		6	10	14	6	10	14
	0	94	95	96	1.67	0.72	0.50
Randomization	1.5	95	95	96	1.46	0.61	0.43
	1.5-ns	100	100	100	4.73	2.10	1.58
	0	94	93	95	1.12	0.65	0.47
GLMM-C	1.5	95	95	94	1.08	0.63	0.45
	1.5-ns	96	95	95	1.29	0.77	0.56
	0	98	94	95	2.11	0.77	0.51
GEE-FGd5	1.5	98	95	95	2.13	0.74	0.49
	1.5-ns	96	95	96	1.49	0.76	0.52
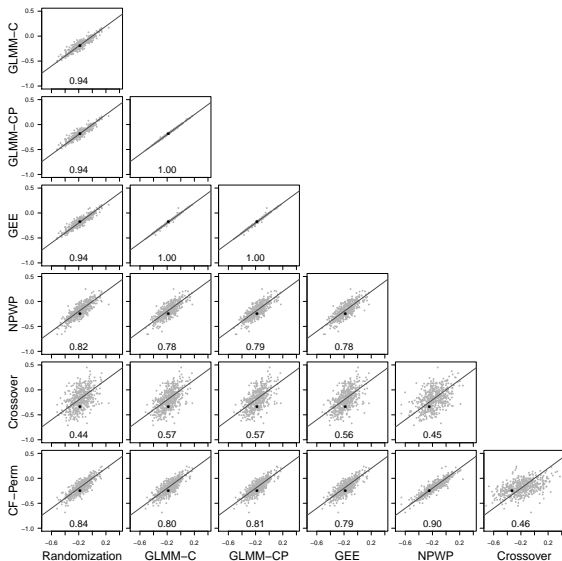
The XpertMTB/RIF Trial

SWT assessing 2 diagnostic tests of tuberculosis (TB)

- XpertMTB/RIF rapid test ($X = 1$) vs. smear microscopy ($X = 0$)
- 14 laboratories, 8 periods, 3,926 individuals diagnosed with TB
- composite binary outcome (death, dropout, drug failure/resistance)
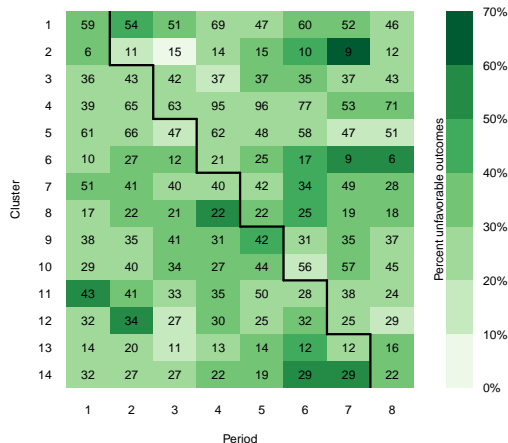
The XpertMTB/RIF Trial

Analysis type	Method	OR	95% CI	p-value
	Randomization	**0.84**	**[0.64, 1.07]**	**0.13**
Individual-level	GLMM-C	0.84	[0.68, 1.03]	0.09
	GEE-FGd5	0.83	[0.57, 1.21]	0.31
	NPWP	0.78	[0.61, 0.97]	0.02
Cluster-period	Crossover	0.72	[0.52, 1.01]	0.05
	CF-Perm	0.78	[0.61, 1.01]	0.06