Linking Design to Analysis of Cluster Randomized Trials: Covariate Balancing Strategies

Fan (Frank) Li

PhD Candidate in Biostatistics
Department of Biostatistics and Bioinformatics
Duke Clinical Research Institute
Duke University

NIH Collaboratory Grand Rounds on February 9, 2018

Acknowledgement

- NIH Collaboratory Biostatistics and Study Design Core Working Group
  - Elizabeth DeLong, PhD, David Murray, PhD, Patrick Heagerty, PhD, Elizabeth Turner, PhD, William Vollmer, PhD, Andrea Cook, PhD, Yuliya Lokhnygina, PhD
- Collaborators at Duke and Harvard
  - John Gallis, ScM, Melanie Prague, PhD, Hengshi Yu, MS
- Funding
  - This work was supported by the NIH Health Care Systems Research Collaboratory (U54 AT007748) from the NIH Common Fund

Outline

- 1. Introduction
- 2. Balancing strategies
  - 2.1 Stratification and pair matching
  - 2.2 Constrained randomization
- 3. Two lessons for statistical analysis
- 4. Summary

1. Introduction

Cluster (group) randomized trials

- Randomization at the cluster level (clinics, hospitals, etc.)
  - Intervention delivered at the cluster level
  - Outcome measured at the individual level
- Focus on parallel design
  - Intervention implemented simultaneously
- Limited number of clusters available
  - Most CRTs randomize ≤ 24 clusters [1]
- Chance imbalance is likely to occur after simple randomization (see an example that follows)

[1] Fiero MH, Huang S, Oren E, Bell ML (2016). Statistical analysis and handling of missing data in cluster randomized trials: a systematic review. *Trials*

An example trial

- Consider the reminder/recall (R/R) immunization study [2]
  - 2-arm parallel CRT with 16 counties (clusters)
  - to increase immunization rate in children 19-35 months
  - a population-based R/R approach (Trt)
  - a practice-based R/R approach (Ctr)
  - binary response variable, immunization status for children in contacted families
- Location known for all clusters (8 rural & 8 urban)

[2] Dickinson LM, Beaty B, Fox C, Pace W, Dickinson WP, Emsermann C, Kempe A (2015). Pragmatic cluster randomized trials using covariate constrained randomization: a method for practice-based research networks. *Journal of the American Board of Family Medicine*

Ideal scenario

- Symbolic representation

Location	# of counties	Symbols
Rural	8	●●●●●●●●
Urban	8	●●●●●●●●

- Assign 8 counties to each arm
- We wish to achieve "balance" after randomization

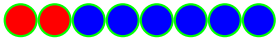
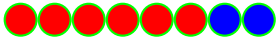
Arm	# of rural/urban counties	Symbols
Trt	4/4	●●●●●●●●
Ctr	4/4	●●●●●●●●

- Same number of urban (or rural) counties/arm $\Rightarrow$ balance

Chance imbalance

- Random allocation of 16 counties to two arms does not guarantee "balance"
    - balance defined by same number of urban counties/arm

- We may end up getting

Arm	# of rural/urban counties	Symbols
Trt	2/6	
Ctr	6/2	

- With a few clusters, the probability of getting an "imbalanced" random allocation is non-negligible ($\approx 1/8$)

- Chance imbalance becomes a bigger issue with more than one baseline variable

Why baseline balance

- Chance imbalance leads to [3]
  - poor internal validity
  - reduced study power/precision of estimates (issue magnified by small sample size)
- Need design-based adjustment of baseline covariates to avoid chance imbalance
- Design-based solution is possible since
  - all clusters are identified prior to randomization (baseline cluster characteristics specified)
  - unlike individually randomized trials with sequential enrollment

[3] Turner EL, Li F, Gallis JA, Prague M, Murray DM (2017). Review of recent methodological developments in group-randomized trials: Part 1–design. *Am J Public Health*

Baseline characteristics

- R/R immunization study
  - ① location (rural/urban)
  - ② % children with immunization record
  - ③ # children aged 15-35 months
  - ④ % up-to-date at baseline
  - ⑤ % Hispanic
  - ⑥ % African American
  - ⑦ average income
  - ⑧ pediatric-to-family medicine practices ratio
  - ⑨ # of community health centers
- Various types of covariates, most of which are continuous
- Goal: leverage design-based control of baseline covariates

2. Balancing strategies

Stratification

- Create distinct strata of clusters based on baseline covariates
  - straightforward with categorical variables
- Stratified randomization

	Location	Symbols	Randomization
Stratum 1	rural		$1:1$ to two arms
Stratum 2	urban		$1:1$ to two arms

- Balance is maintained within each stratum defined by location

Arm	# of rural/urban counties	Symbols
Trt	$4/4$	
Ctr	$4/4$	

Stratification

- Create distinct strata of clusters based on baseline covariates
  - continuous variables will be categorized (e.g. *high* versus *low*)

	Location	Avg income	# of counties	Randomization
Stratum 1	rural	low	○	1 : 1 to two arms?
Stratum 2	rural	medium	○○○	1 : 1 to two arms?
Stratum 3	rural	high	○○○○	1 : 1 to two arms?
Stratum 4	urban	low	none	none
Stratum 5	urban	medium	○○○○○	1 : 1 to two arms?
Stratum 6	urban	high	○○○	1 : 1 to two arms?

- Con: incomplete filling of strata with ↑ number of strata
  - unavoidable with a number of baseline covariates (R/R study)
  - sensitive to cutoff used in categorization
  - same drawback in individual RCTs

Pair matching

- Good matches $\Rightarrow$ an effective mechanism to create comparable groups
- Suppose location variable is of good prognostic values (the matching variable), can create eight pairs of clusters

	rural/urban counties	Symbols	Trt	Ctr
Pair 1	2/0	🔴🔴	🔴	🔴
Pair 2	2/0	🔴🔴	🔴	🔴
Pair 3	2/0	🔴🔴	🔴	🔴
Pair 4	2/0	🔴🔴	🔴	🔴
Pair 5	0/2	🔵🔵	🔵	🔵
Pair 6	0/2	🔵🔵	🔵	🔵
Pair 7	0/2	🔵🔵	🔵	🔵
Pair 8	0/2	🔵🔵	🔵	🔵

Pair matching

- Matching with multiple covariates relies on a multivariate distance metric
- Advantage[4]
    - allows for an efficient nonparametric design-based estimator
- Disadvantages [5]
    - loss of follow-up from one cluster removes its matches
    - difficult to properly calculate the intraclass correlation coefficient (ICC)
    - "break the matches"?

[4] Imai K, King G, Nall C (2009). The essential role of pair matching in cluster randomized experiments, with application to the Mexican universal health insurance evaluation. *Stat Sci*.

[5] Klar N, Donner A (1997). The merits of matching in community intervention trials: A cautionary tale. *Stat Med*.

Constrained randomization (CR)

- General idea
  - Specify the simple randomization space containing all possible allocation schemes
  - Assess "balance" for each possible allocation scheme
  - Randomize only within a constrained space with "balanced" allocation schemes
- Advantages[6]
  - accomondate a number of, and all types of covariates
  - does not complicate ICC calculation

[6] Li F, Turner E, Heagerty PJ, Murray DM, Vollmer W, Delong ER (2017). An evaluation of constrained randomization for the design and analysis of group-randomized trials with binary outcomes. *Stat Med*

Schematic illustration of constrained randomization

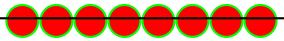
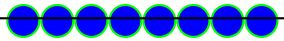
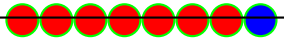
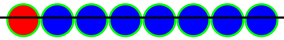
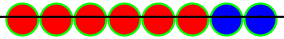
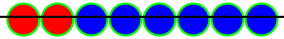
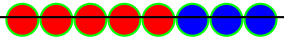
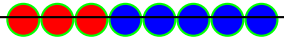
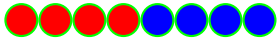
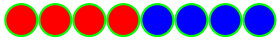
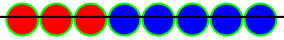
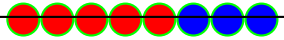
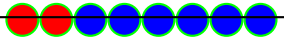
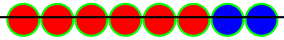
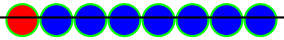
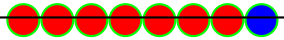
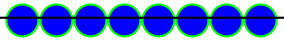
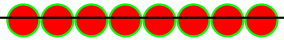
- R/R study with $n = 16$ clusters and 8 clusters/arm
- Simple randomization: 12,870 allocation schemes
- 9 allocation types of 8 rural (x=0) & 8 urban (x=1) clusters
- Balance score by a simple balance metric: $|\bar{x}_T - \bar{x}_C|$

# Rural in Arms	Treatment	Control	# of schemes	Balance
8/0	🔴🔴🔴🔴🔴🔴🔴🔴	🔵🔵🔵🔵🔵🔵🔵🔵	1	1.00
7/1	🔴🔴🔴🔴🔴🔴🔴🔵	🔴🔵🔵🔵🔵🔵🔵🔵	64	0.75
6/2	🔴🔴🔴🔴🔴🔴🔵🔵	🔴🔴🔵🔵🔵🔵🔵🔵	784	0.50
5/3	🔴🔴🔴🔴🔴🔵🔵🔵	🔴🔴🔴🔵🔵🔵🔵🔵	3136	0.25
4/4	🔴🔴🔴🔴🔵🔵🔵🔵	🔴🔴🔴🔴🔵🔵🔵🔵	4900	0.00
3/5	🔴🔴🔴🔵🔵🔵🔵🔵	🔴🔴🔴🔴🔴🔵🔵🔵	3136	0.25
2/6	🔴🔴🔵🔵🔵🔵🔵🔵	🔴🔴🔴🔴🔴🔴🔵🔵	784	0.50
1/7	🔴🔵🔵🔵🔵🔵🔵🔵	🔴🔴🔴🔴🔴🔴🔴🔵	64	0.75
0/8	🔵🔵🔵🔵🔵🔵🔵🔵	🔴🔴🔴🔴🔴🔴🔴🔴	1	1.00

Schematic illustration of constrained randomization

- Constrain to 4,900/12,870 allocations with most balance
  - Balance score = 0
  - 4 rural & 4 urban clusters/arm

- Randomize 16 clusters within the constrained subset of 4,900

Treatment	Control	# of schemes	Balance
		~~1~~	~~1.00~~
		~~64~~	~~0.75~~
		~~784~~	~~0.50~~
		~~3136~~	~~0.25~~
		4900	0.00
		~~3136~~	~~0.25~~
		~~784~~	~~0.50~~
		~~64~~	~~0.75~~
		~~1~~	~~1.00~~

Implementing covariate constrained randomization

- Step 1: Specify important baseline cluster-level covariates
- Step 2: Generate allocation schemes
  - Either enumerate all schemes (e.g. if $n \leq 18$)
  - Or simulate many schemes (e.g. 50,000) & remove duplicates
- Step 3: Select a **constrained randomization space** with sufficiently-balanced allocations according to **balance metric**
- Step 4: Randomly sample 1 scheme from constrained randomization space

Balance metrics

- Goal: balance K baseline cluster-level covariates
- Could consider any sensible balance metric (distance function)
- Class of balance metrics: $B = \sum_k \omega_k g(\bar{x}_{Tk} - \bar{x}_{Ck})$
- Two common balance metrics:

Balance metric	$g(t)$	Default weights (w_k)	Reference
$B_{(l2)}$	t^2	$1/s_k^2$	Raab and Butcher (2001) [7]
$B_{(l1)}$	$\|t\|$	$1/s_k$	Li et al (2017)[6]

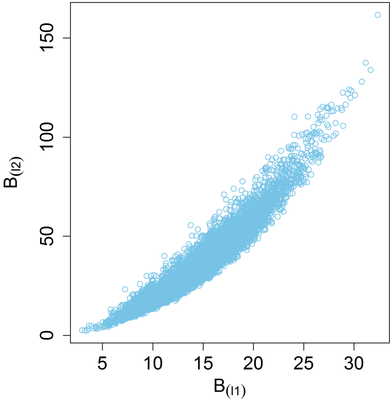
- Unitless metrics under default weights

[6] Li F, Turner E, Heagerty PJ, Murray DM, Vollmer W, Delong ER (2017). An evaluation of constrained randomization for the design and analysis of group-randomized trials with binary outcomes. *Stat Med*

[7] Raab GM, Butcher I (2001). Balance in cluster randomized trials. *Stat Med*

R/R Immunization Study: Two balance metrics

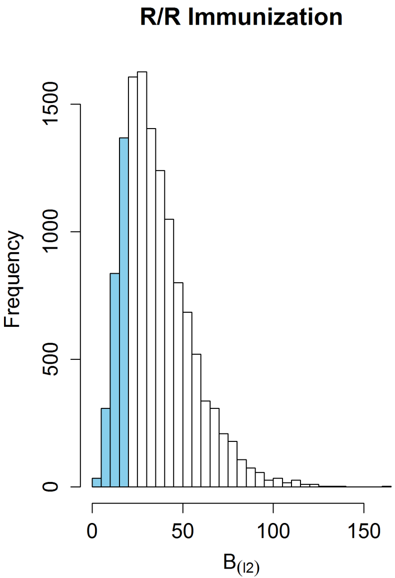
- Balance all 9 baseline covariates
- $l1$ and $l2$ metrics very similar: can use either one for constrained randomization



- Spearman rank correlation: $\lambda = 0.97$

Size of randomization space

- Balance all 9 baseline covariates
- $\binom{16}{8} = 12,870$ possible allocation schemes with equal-arm assignment
- Example: constrained randomization space 10% of simple randomization space



Size of constrained randomization space

- $q =$ size of the constrained randomization space as % of entire simple randomization space with lowest balance scores

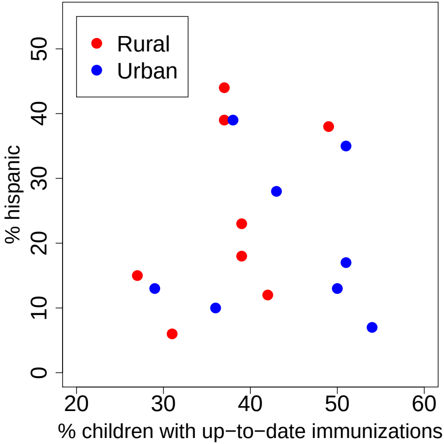
- q small but should not be too small
  - Risk deterministic allocation
  - May prohibit permutation test for a fixed α

- $q = 10\%$ works well in simulation experiments [6]
  - Power $\uparrow$ as $q \downarrow$ by balancing predictive covariates
  - Relationship not monotone, power may not $\uparrow$ if $q < 10\%$

[6] Li F, Turner E, Heagerty PJ, Murray DM, Vollmer W, Delong ER (2017). An evaluation of constrained randomization for the design and analysis of group-randomized trials with binary outcomes. *Stat Med*

Example: R/R Immunization study

- Goal: Increase child immunization rate
- Randomize 16 clusters (counties) to 'treatment' vs. 'control'
- Balance on 3 outcome-predictive baseline cluster covariates



Example: covariate CR in practice

- Allocate 16 clusters (8/arm) + balance on 3 covariates
- Use $B_{(l2)}$ metric + constrain at $q = 10\%$ of simple randomization (SR) space
- Compare mean covariate levels between arms under "best" balance, at "boundary" of 10% CR space & under "worst" balance (i.e. at worst SR allocation)

Covariate - Mean	"Best"		"CR Boundary"		"Worst"	
	Trt	Ctr	Trt	Ctr	Trt	Ctr
# of urban county	4	4	4	4	0	8
% hispanic	22.4	22.3	19.9	24.8	24.4	20.3
% up-to-date at baseline	40.8	40.9	40.3	41.4	37.6	44.0
balance score	$B_{(l2)} = 0.005$		$B_{(l2)} = 2.58$		$B_{(l2)} = 71.08$	

Example: covariate CR in practice

- Allocate 16 clusters (8/arm) + balance on 3 covariates
- Use $B_{(l2)}$ metric + constrain at $q = 10\%$ of simple randomization (SR) space
- Compare mean covariate levels between arms under "best" balance, at "boundary" of 10% CR space & under "worst" balance (i.e. at worst SR allocation)

Covariate - Mean	"Best"		"CR Boundary"		"Worst"	
	Trt	Ctr	Trt	Ctr	Trt	Ctr
# of urban county	4	4	4	4	0	8
% hispanic	22.4	22.3	19.9	24.8	24.4	20.3
% up-to-date at baseline	40.8	40.9	40.3	41.4	37.6	44.0
balance score	$B_{(l2)} = 0.005$		$B_{(l2)} = 2.58$		$B_{(l2)} = 71.08$	

Example: covariate CR in practice

- Allocate 16 clusters (8/arm) + balance on 3 covariates
- Use $B_{(l2)}$ metric + constrain at $q = 10\%$ of simple randomization (SR) space
- Compare mean covariate levels between arms under "best" balance, at "boundary" of 10% CR space & under "worst" balance (i.e. at worst SR allocation)

Covariate - Mean	"Best"		"CR Boundary"		"Worst"	
	Trt	Ctr	Trt	Ctr	Trt	Ctr
# of urban county	4	4	4	4	0	8
% hispanic	22.4	22.3	19.9	24.8	24.4	20.3
% up-to-date at baseline	40.8	40.9	40.3	41.4	37.6	44.0
balance score	$B_{(l2)} = 0.005$		$B_{(l2)} = 2.58$		$B_{(l2)} = 71.08$	

Example: covariate CR in practice

- Allocate 16 clusters (8/arm) + balance on 3 covariates
- Use $B_{(l2)}$ metric + constrain at $q = 10\%$ of simple randomization (SR) space
- Compare mean covariate levels between arms under "best" balance, at "boundary" of 10% CR space & under "worst" balance (i.e. at worst SR allocation)

Covariate - Mean	"Best"		"CR Boundary"		"Worst"	
	Trt	Ctr	Trt	Ctr	Trt	Ctr
# of urban county	4	4	4	4	0	8
% hispanic	22.4	22.3	19.9	24.8	24.4	20.3
% up-to-date at baseline	40.8	40.9	40.3	41.4	37.6	44.0
balance score	$B_{(l2)} = 0.005$		$B_{(l2)} = 2.58$		$B_{(l2)} = 71.08$	

Example: covariate CR in practice

- Allocate 16 clusters (8/arm) + balance on 3 covariates
- Use $B_{(l2)}$ metric + constrain at $q = 10\%$ of simple randomization (SR) space
- Compare mean covariate levels between arms under "best" balance, at "boundary" of 10% CR space & under "worst" balance (i.e. at worst SR allocation)

Covariate - Mean	"Best"		"CR Boundary"		"Worst"	
	Trt	Ctr	Trt	Ctr	Trt	Ctr
# of urban county	4	4	4	4	0	8
% hispanic	22.4	22.3	19.9	24.8	24.4	20.3
% up-to-date at baseline	40.8	40.9	40.3	41.4	37.6	44.0
balance score	$B_{(l2)} = 0.005$		$B_{(l2)} = 2.58$		$B_{(l2)} = 71.08$	

Example: covariate CR in practice

- Allocate 16 clusters (8/arm) + balance on 3 covariates
- Use $B_{(l2)}$ metric + constrain at $q = 10\%$ of simple randomization (SR) space
- Compare mean covariate levels between arms under "best" balance, at "boundary" of 10% CR space & under "worst" balance (i.e. at worst SR allocation)

Covariate - Mean	"Best"		"CR Boundary"		"Worst"	
	Trt	Ctr	Trt	Ctr	Trt	Ctr
# of urban county	4	4	4	4	0	8
% hispanic	22.4	22.3	19.9	24.8	24.4	20.3
% up-to-date at baseline	40.8	40.9	40.3	41.4	37.6	44.0
balance score	$B_{(l2)} = 0.005$		$B_{(l2)} = 2.58$		$B_{(l2)} = 71.08$	

Example: covariate CR in practice

- Allocate 16 clusters (8/arm) + balance on 3 covariates
- Use $B_{(l2)}$ metric + constrain at $q = 10\%$ of simple randomization (SR) space
- Compare mean covariate levels between arms under "best" balance, at "boundary" of 10% CR space & under "worst" balance (i.e. at worst SR allocation)

Covariate - Mean	"Best"		"CR Boundary"		"Worst"	
	Trt	Ctr	Trt	Ctr	Trt	Ctr
# of urban county	4	4	4	4	0	8
% hispanic	22.4	22.3	19.9	24.8	24.4	20.3
% up-to-date at baseline	40.8	40.9	40.3	41.4	37.6	44.0
balance score	$B_{(l2)} = 0.005$		$B_{(l2)} = 2.58$		$B_{(l2)} = 71.08$	

Application of CR to the R/R study

- Balance all of 9 baseline covariates
- Similar to Dickinson et al. (2015) [2], list the "best" and "worst" allocation schemes under CR with $q = 0.1$

Covariate - Mean (Sd)	"Best"		"CR Boundary"	
	Trt	Ctr	Trt	Ctr
# of urban county	4 (50)	4 (50)	4 (50)	4 (50)
% in CIIS	87.4 (7.5)	87.0 (8.4)	87.6 (5.4)	86.6 (9.9)
# of children	4172 (4465)	4221 (4707)	4068 (4640)	4325 (4530)
% up-to-date at baseline	41.4 (8.4)	40.3 (8.7)	42.1 (9.7)	39.5 (7.1)
% African American	3.3 (3.1)	2.5 (2.5)	3.5 (3.0)	2.3 (2.5)
% Hispanic	21.9 (12.1)	22.8 (14.5)	18.3 (11.8)	26.4 (13.5)
Average income ($1000/yr)	54.8 (19.0)	52.2 (13.1)	51.3 (12.0)	55.7 (19.5)
PM-to-FM ratio	0.26 (0.22)	0.30 (0.29)	0.23 (0.14)	0.33 (0.33)
# CHCs	4.4 (2.7)	4.4 (4.4)	4.5 (3.4)	4.3 (3.9)
balance score	$B_{(l2)} = 2.5$		$B_{(l2)} = 15.4$	

[2] Dickinson LM, Beaty B, Fox C, Pace W, Dickinson WP, Emsermann C, Kempe A (2015). Pragmatic cluster randomized trials using covariate constrained randomization: a method for practice-based research networks. *Journal of the American Board of Family Medicine*

Implementation: R and Stata packages

The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–23

cvcrand and cptest: Efficient design and analysis of cluster randomized trials

John A. Gallis
Duke University Department of Biostatistics
Duke Global Health Institute
Durham, NC
john.gallis@duke.edu

Fan Li
Duke University Department of Biostatistics
Durham, NC
frank.li@duke.edu

Hengshi Yu
University of Michigan Department of Biostatistics
Ann Arbor, MI
hengshi@umich.edu

Elizabeth L. Turner
Duke University Department of Biostatistics
Duke Global Health Institute
Durham, NC
liz.turner@duke.edu

cvcrand: Efficient Design and Analysis of Clust

Constrained randomization by Raab and Butcher (2001) <doi:10.1002/
suitable for cluster randomized trials (CRTs) with a small number of cl
based on the baseline values of some cluster-level covariates specified.
through clustered permutation test introduced by Gail, et al. (1996) <do
SIM220%3E3.0.CO;2-Q>. Motivated from Li, et al. (2016) <doi:10.10
baseline values of cluster-level covariates and cluster permutation test

Version:	0.0.1
Depends:	R (≥ 3.3.1)
Imports:	tableone
Suggests:	knitr, rmarkdown
Published:	2017-11-28
Author:	Hengshi Yu [aut, cre], John A. Gallis [aut], Fan Li
Maintainer:	Hengshi Yu <hengshi at umich.edu>
License:	GPL-2 | GPL-3 [expanded from: GPL (≥ 2)]

3. Two Lessons for Statistical Analysis

Lesson # 1: model-based inference

- Mixed-effects models
  - augment the linear model (or logistic model) with a random cluster effect
  - random effect terms describe the similarity between individual outcomes within a cluster (county)
- Should control for the prognostic baseline covariates balanced by constrained randomization (CR)[6]
  - model-based standard error ignores CR $\Rightarrow$ underpowered
- Basic principle: analysis should account for the design

[6] Li F, Turner E, Heagerty PJ, Murray DM, Vollmer W, Delong ER (2017). An evaluation of constrained randomization for the design and analysis of group-randomized trials with binary outcomes. *Stat Med*

Lesson # 2: permutation inference

- Basic idea:
  - Calculate a test statistic under the observed treatment assignment
  - Recompute the value of the test statistic under all other possible assignment $\Rightarrow$ null distribution
  - Compare the observed test statistic to the null distribution $\Rightarrow$ p-value
- Constrained randomization space should be used for valid inference [8]
- Basic principle: analysis should account for the design

[8] Li F, Lokhnygina Y, Murray DM, Heagerty PJ, Delong ER (2016). An evaluation of constrained randomization for the design and analysis of group-randomized trials. *Stat Med*.

4. Summary

Summary

- Constrained randomization is a powerful technique to balance multiple, possibly continuous baseline covariates in small cluster randomized trials
  - avoid categorization of continuous covariates (v.s. stratification)
  - randomization not based on pairs; ICC calculation unaffected (v.s. matching)
- Software to perform constrained randomization is made available in `Stata` and `R` by Duke group
  - `Stata` - `cvcrand` (CR) and `cptest` (permutation test)
  - `R` - `cvcrand` (CR) and `cptest` (permutation test)
  - documentations on `SSC` and `CRAN`
- Analysis of trial results should account for the design

Look forward

- Balance is an important consideration in pragmatic cluster randomized trials (with a limited number of clusters)
- Only considered parallel cluster randomized trials, where the interventions are implemented concurrently for all clusters
  - not always logistically feasible
  - stepped wedge designs
- Balance may benefit between-cluster comparisons
  - Invited session at Society of Clinical Trials (SCT), May 2018
- Lots of open statistical questions still need to be addressed

Welcome questions or comments for

- Fan (Frank) Li: `frank.li@duke.edu`
- Elizabeth Turner, PhD: `liz.turner@duke.edu`
- Elizabeth DeLong, PhD: `elizabeth.delong@duke.edu`

Thank you