Small-Sample Methods for Cluster-Robust Inference in School-Based Experiments

James E. Pustejovsky
UT Austin
Educational Psychology Department
Quantitative Methods Program
pusto@austin.utexas.edu

Elizabeth Tipton
Teachers College, Columbia University
Dept. of Human Development
tipton@tc.columbia.edu

March 2, 2017
Society for Research on Educational Effectiveness
Washington, DC

In brief…

- Analysis of social experiments often requires handling ***dependencies among outcomes*** using:
  - Multi-level modeling
  - Regression with cluster-robust variance estimation (CRVE)

- Conventional CRVE behave poorly when the number of clusters is small, and "small" depends on the model.

- McCaffrey, Bell, & Botts (2001; Bell & McCaffrey, 2002) proposed bias-reduced linearization variance estimator (BRL), Satterthwaite t-test

- Our work (Pustejovsky & Tipton, 2017) extends BRL
  - so that it works in panel models with fixed effects
  - F-test for multi-parameter hypothesis tests
  - software implementation in R and Stata (***clubSandwich*** package)

Model

- Main impacts model:

$$Y_{ij} = \beta_0 + \beta_1 T_{ij} + \boldsymbol{\beta}_2^t \mathbf{x}_{ij} + e_{ij}$$

- More generally,
  - Models with multiple treatment indicators
  - Treatment-by-covariate interactions

- In matrix form:

$$\mathbf{Y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{e}_i, \quad \mathrm{Var}\left(\mathbf{e}_i\right) = \boldsymbol{\Sigma}_i, \quad i = 1, \ldots, n$$

Estimation

- Estimate $\boldsymbol{\beta}$ by weighted least squares:

$$\hat{\boldsymbol{\beta}} = \mathbf{M}\left(\sum_{i=1}^{n} \mathbf{X}_i^t \mathbf{W}_i \mathbf{Y}_i\right), \quad \mathbf{M} = \left(\sum_{i=1}^{n} \mathbf{X}_i^t \mathbf{W}_i \mathbf{X}_i\right)^{-1}$$

- Standard CRVE:

$$\mathbf{V}^{CR} = \frac{n}{n-1} \times \mathbf{M}\left(\sum_{i=1}^{n} \mathbf{X}_i^t \mathbf{W}_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^t \mathbf{W}_i \mathbf{X}_i\right)\mathbf{M}$$

- Conventional to use $n - 1$ degrees of freedom for t-tests.

Bias-reduced linearization

- Corrects $\mathbf{V}^{CR}$ based on a ***working model*** for the error covariance structure:

$$\mathbf{V}^{BRL} = \mathbf{M}\left(\sum_{i=1}^{n} \mathbf{X}_i^t \mathbf{W}_i \mathbf{A}_i \hat{\mathbf{e}}_i \hat{\mathbf{e}}_i^t \mathbf{A}_i^t \mathbf{W}_i \mathbf{X}_i\right)\mathbf{M}$$

with adjustment matrices $\mathbf{A}_1,\ldots,\mathbf{A}_n$ chosen to satisfy

$$\mathrm{E}\left(\mathbf{V}^{BRL}\right) = \mathrm{Var}\left(\hat{\boldsymbol{\beta}}\right)$$

- Degrees of freedom corrections for hypothesis tests
  - Satterthwaite d.f. for t-tests (Bell & McCaffrey, 2002)
  - Approximate Hotelling's T^2 d.f. for F-test (Tipton & Pustejovsky, 2015; Pustejovsky & Tipton, 2017)

Approximate Hotelling Test

- We propose a generalization of the Satterthwaite approximation to the multi-dimensional case, with $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$
- Approximate the distribution of V^{BRL} using a Wishart distribution with degrees of freedom η.
- Estimate η by matching mean and **total variance** of V^{BRL}.

$$F_{AHT} = \frac{\eta - q + 1}{\eta q}\left(\mathbf{C}\hat{\boldsymbol{\beta}}\right)^t \left(\mathbf{C}\mathbf{V}^{BRL}\mathbf{C}\right)^{-1}\left(\mathbf{C}\hat{\boldsymbol{\beta}}\right)$$

$$F_{AHT} \dot{\sim} F\left(q, \eta - q + 1\right)$$

Effects of Tribes Learning Communities (Hanson et al., 2011)

- Social-Emotional Learning curriculum.
- Classroom-level randomization to TRIBES or BAU control.
- 10 participating schools in Grades 1-2.
- Original analysis used HLM with classroom level random effects, school fixed effects.

Effects of Tribes Learning Communities (Hanson et al., 2011)

- OLS estimation (seemingly unrelated regressions)
- Cluster SEs by school

Outcome	Impact Est. (ES units)	Conventional CRVE			Bias-Reduced Linearization		
		SE	df	p	SE	df	p
Aggressive behavior (T)	0.329	0.156	9	.065	0.173	7.0	.098
Rule-breaking (T)	0.312	0.157	9	.078	0.173	7.0	.114
Interpersonal strength (P)	0.209	0.079	9	.026	0.085	7.5	.041
Intrapersonal strength (P)	0.231	0.077	9	.015	0.081	7.4	.023

- Joint test of outcomes
  - Conventional: $F(4, 9) = 6.82$, $p = .008$
  - Bias-reduced linearization: $F(4, 4.3) = 3.70$, $p = .109$

Angrist & Lavy (2009)

- Cluster-randomized trial in 40 high schools in Israel.
- Tested effects of monetary incentives on post-secondary matriculation exam (Bagrut) completion rates.
- Longitudinal data, difference-in-differences specification.
- Focus on effects for higher-achieving girls

Hypothesis	Test	F	df	p-value
treatment effect (q = 1)	Standard	5.746	34.00	.022
	Satterthwaite	5.169	18.13	.035
Moderation by school sector (q = 2)	Standard	3.186	34.00	.054
	AHT	1.665	7.84	.250

Further considerations

- Magnitude of SE adjustment and degrees of freedom depend on:
  - Weighting
  - Cluster sizes
  - Balance
  - Covariate distribution

- Given these complexities, we recommend applying small-sample adjustment ***by default*** when using CRVE.

Software

- R package `clubSandwich`
  - Available on Comprehensive R Archive Network (v0.2.1)
  - Development version at https://github.com/jepusto/clubSandwich
  - Works with a wide variety of models (lm, lme, plm)

- Stata package `clubSandwich`
  - Available on Github: https://github.com/jepusto/clubSandwich-Stata
  - Wraps `reg` and `areg`

Future directions

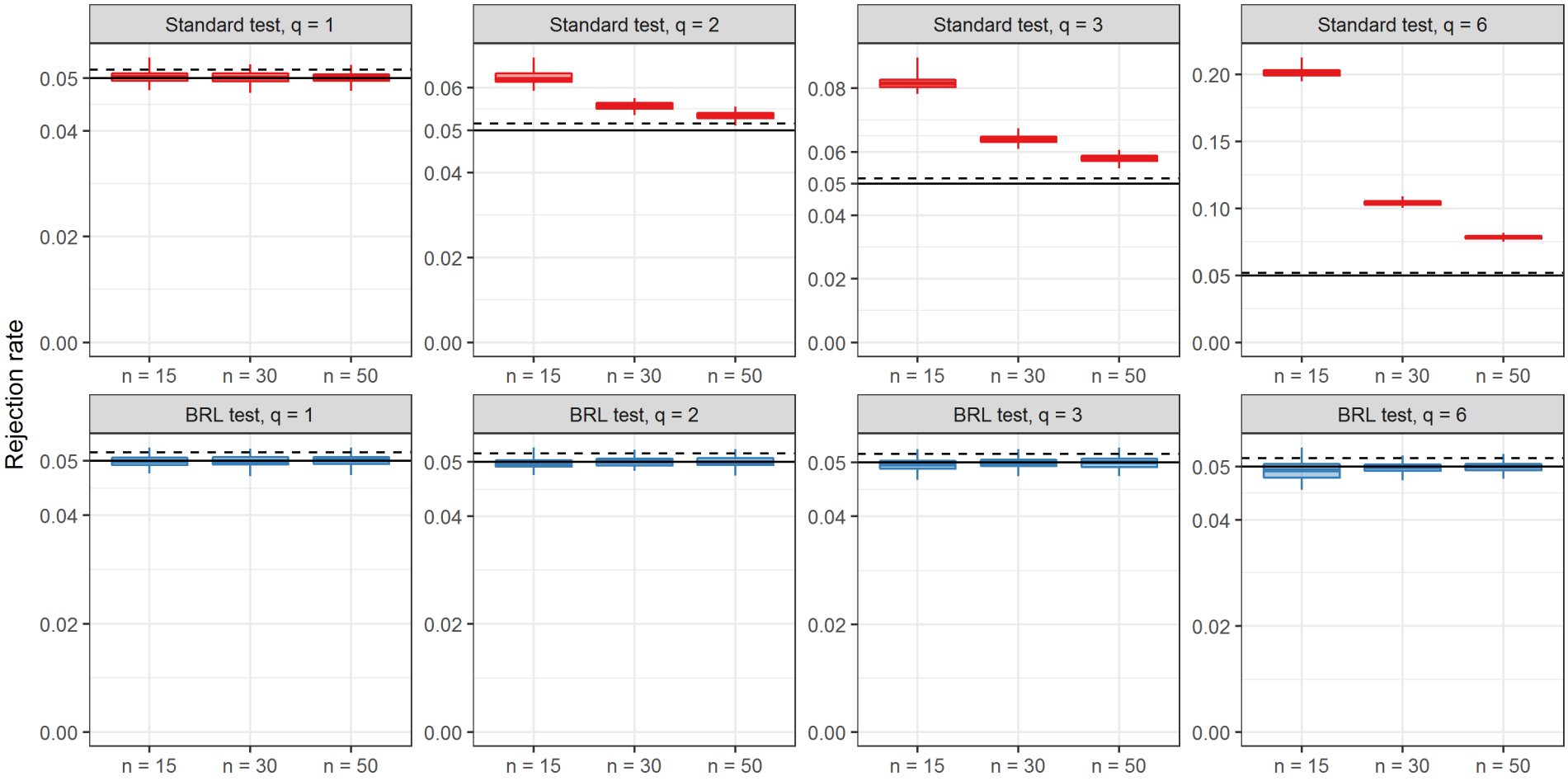
- Performance comparisons versus other small-sample corrections
  - Cluster-wild bootstrap (Cameron, Gelbach, & Miller, 2008; MacKinnon & Webb, 2016).
  - Randomization tests (Canay, Romano, & Shaikh, 2014).
  - Other degrees-of-freedom corrections from GEE literature (e.g., Fay & Graubard, 2001; Wang & Long, 2011).
  - Robust score (LM) tests.

- Extensions
  - Instrumental variables (2-stage least squares)
  - GEE models
  - Multi-way clustering (Cameron, Gelbach, & Miller, 2011)

References

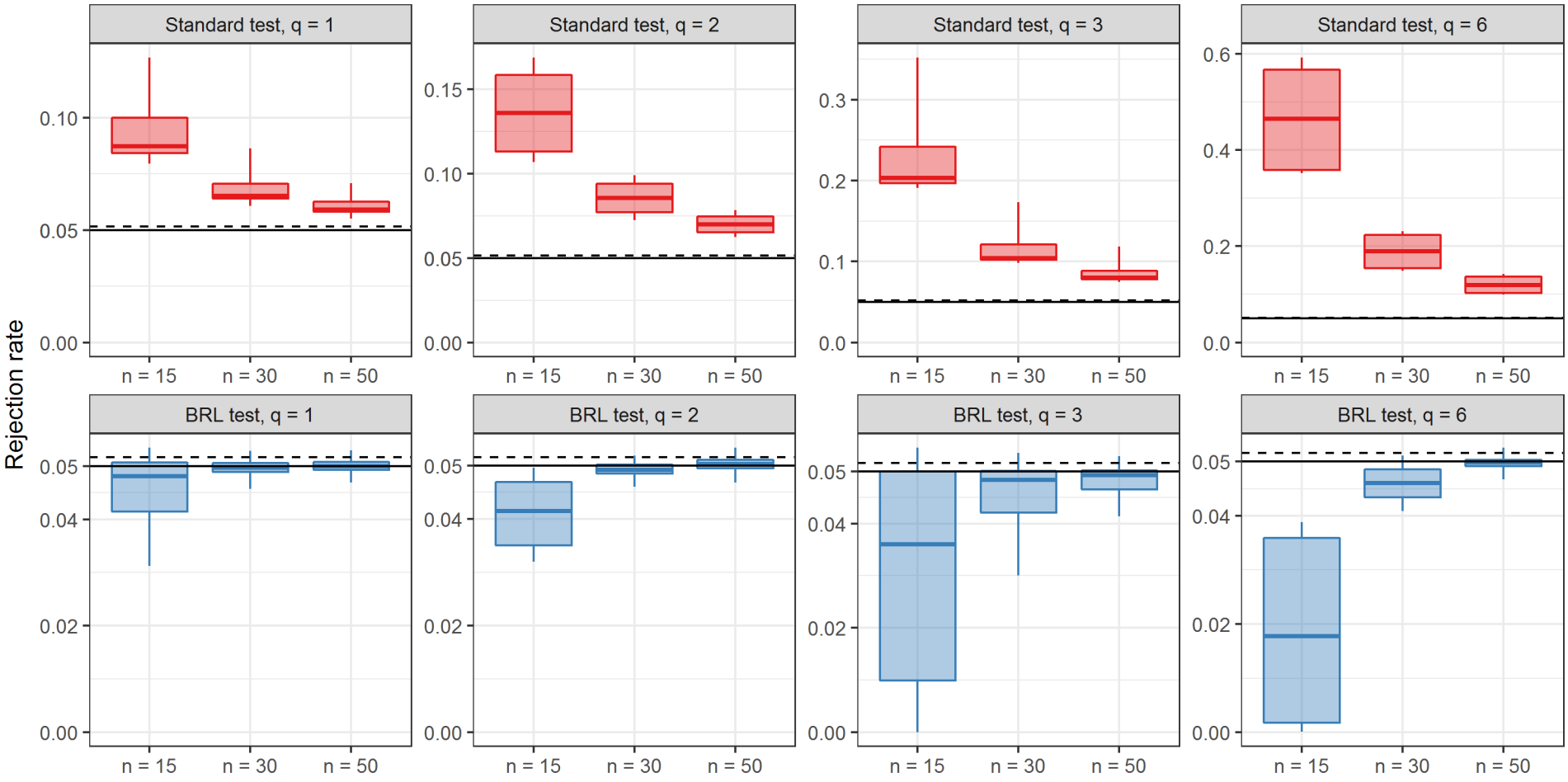
- Angrist, J. D., & Lavy, V. (2009). The effects of high stakes high school achievement awards : Evidence from a randomized trial. *American Economic Review*, *99*(4), 1384–1414.
- Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. *Survey Methodology*, *28*(2), 169–181.
- Cameron, A. C., Gelbach, J. B., and Miller, D. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427.
- Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, *29*(2), 238–249..
- Canay, I. A., Romano, J. P., & Shaikh, A. M. (2014). Randomization tests under an approximate symmetry assumption. Working paper.
- Fay MP and Graubard BI. Small-sample adjustments for Wald-type tests using sandwich estimators. Biometrics 2001;57: 1198-1206.
- Hanson, Thomas L., Jo Ann Izu, Anthony Petrosino, Bo Delong-Cotty, and Hong Zheng. Outcome Evaluation of Tribes Learning Communities in California, 2007-2010. ICPSR32821-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2012-12-20.
- Imbens, G. W. and Kolesar, M. (2016). Robust Standard Errors in Small Samples: Some Practical Advice. *Review of Economics and Statistics*, forthcoming.
- James-Burdurmy, Susanne. Randomized Experiment of Playworks Analytic Files for 2010-2011 and 2011-2012 Cohorts in Six United States Cities. ICPSR35638-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2016-09-20.
- Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. *Journal of Econometrics*, *142*(2), 655–674.
- MacKinnon, J. G. and Webb, M. D. (2016). Wild bootstrap inference for wildly different cluster sizes. *Journal of Applied Econometrics*, forthcoming.
- McCaffrey, D. F., Bell, R. M., & Botts, C. H. (2001). Generalizations of biased reduced linearization. In *Proceedings of the Annual Meeting of the American Statistical Association*.
- Pustejovsky, James E. & Elizabeth Tipton (2017). Small sample methods for cluster-robust variance estimation and hypothesis testing in fixed effects models. Journal of Business and Economic Statistics. In Press.
- Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics*, *40*(6), 604–634.
- Wang M and Long Q. Modified robust variance estimator for generalized estimating equations with improved small-sample performance. Statistics in Medicine 2011;30(11): 1278-1291.

Simulation results: Block-randomized trials



Note: q is the dimension of the hypothesis test.
Source: Pustejovsky & Tipton (2017).

Simulation results: Cluster-randomized trials



Note: q is the dimension of the hypothesis test.
Source: Pustejovsky & Tipton (2017).

Block-randomized/multi-site trials

- Model with block fixed effects:

$$Y_{ij} = \beta_i + \delta T_{ij} + e_{ij}$$

- Overall impact estimate:

$$\hat{\delta} = \frac{1}{W}\sum_{i=1}^{n} w_i \hat{\delta}_i, \quad W = \sum_{i=1}^{n} w_i$$

where $\hat{\delta}_1, \ldots, \hat{\delta}_n$ are treatment effect estimates from each block.

- Conventional CRVE (clustering by block):

$$\mathbf{V}^{CR} = \frac{1}{W^2}\sum_{i=1}^{n} w_i^2 \left(\hat{\delta}_i - \hat{\delta}\right)^2$$

Block-randomized/multi-site trials (cont.)

- BRL correction:

$$\mathbf{V}^{BRL} = \frac{1}{W^2}\sum_{i=1}^{n}\frac{w_i^2\left(\hat{\delta}_i - \hat{\delta}\right)^2}{\left(1 - w_i / W\right)}$$

- Satterthwaite df:

$$df = \left[\sum_{i=1}^{n}\frac{w_i^2}{\left(W - w_i\right)^2} - \frac{2}{W}\sum_{i=1}^{n}\frac{w_i^3}{\left(W - w_i\right)^2} + \frac{1}{W^2}\left(\sum_{i=1}^{n}\frac{w_i^2}{W - w_i}\right)^2\right]^{-1}$$

- Satterthwaite $df = n - 1$ if w_j are equal (otherwise $df < n - 1$).

Cluster-randomized trials

- Model (without covariates):

$$Y_{ij} = \beta_0 + \delta T_i + e_{ij}$$

- Overall impact estimate:

$$\hat{\delta} = \frac{1}{W_T}\sum_{i=1}^{n_T} w_i \hat{\mu}_i^T - \frac{1}{W_C}\sum_{i=1}^{n_C} w_i \hat{\mu}_i^C$$

where $\hat{\mu}_1^T, \ldots, \hat{\mu}_{n_T}^T$ and $\hat{\mu}_1^C, \ldots, \hat{\mu}_{n_C}^C$ are cluster-specific mean estimates.

Cluster-randomized trials (cont.)

- Conventional CRVE:

$$\mathbf{V}^{CR} = \frac{1}{W_T^2}\sum_{i=1}^{n_T} w_i^2\left(\hat{\mu}_i^T - \hat{\mu}_\bullet^T\right)^2 + \frac{1}{W_C^2}\sum_{j=1}^{n_C} w_i^2\left(\hat{\mu}_i^C - \hat{\mu}_\bullet^C\right)^2$$

- BRL correction:

$$\mathbf{V}^{BRL} = \frac{1}{W_T^2}\sum_{i=1}^{n_T} \frac{w_i^2\left(\hat{\mu}_i^T - \hat{\mu}_\bullet^T\right)^2}{1 - w_i / W_T} + \frac{1}{W_C^2}\sum_{j=1}^{n_C} \frac{w_i^2\left(\hat{\mu}_i^C - \hat{\mu}_\bullet^C\right)^2}{1 - w_i / W_C}$$

- If w_i are approximately equal (cf. Imbens & Kolesaar, 2016):

$$df \approx \frac{\left(n_T + n_C\right)^2\left(n_T - 1\right)\left(n_C - 1\right)}{n_T^2\left(n_T - 1\right) + n_C^2\left(n_C - 1\right)}$$

Effects of Playworks on school climate, student social skills and behavior (James-Burdurmy et al., 2013)

- Structured physical activity and recess coaching program.
- 29 participating schools, grouped in 9 blocks
- School-level block randomization to Playworks or BAU control.
  - 17 treatment schools
  - 12 control schools
- OLS estimation, including block fixed effects
- Cluster SEs by school

Effects of Playworks on school climate, student social skills and behavior (James-Burdurmy et al., 2013)

Outcome	Impact Est. (ES units)	Conventional CRVE			Bias-Reduced Linearization		
		SE	df	p	SE	Df	p
Teacher support for organized play	0.591	0.138	28	<.001	0.172	12.0	.005
Staff support for organized play	0.324	0.130	28	.019	0.156	12.2	.059
Student bullying/exclusion	-1.014	0.187	28	<.001	0.253	11.9	.002
Difficult transitioning to learning after recess	-0.840	0.112	28	<.001	0.143	11.8	<.001

Joint test of outcomes

- Conventional: $F(4, 28) = 23.5$, $p < .001$
- Bias-reduced linearization: $F(4, 9.0) = 10.6$, $p = .002$