# Robust and Pseudo-Robust Solutions to Lord’s Paradox

Robert E. Larzelere & Hua Lin
Oklahoma State Univ.

# Four Pseudo-Robust Solutions to Lord's Paradox

- #1: Adding pretest $Y_0$ to diffs-in-diffs model
- #2: Matching on pretest
- #3: Centering all data on pretest group means
- #4: Increasing variance of d.v.

# Robustness (Duncan +, 2014)

- Developmental science: 5-25% articles
  - Like econometrics: 66-68% of articles
    - Some check of replication or robustness
- Robustness
  - Across samples & sub-samples
  - Across alternative analyses
    - Esp. if contradictory biases
- Best solution to Lord's Paradox?
  - Not definitive

# Pseudo-Robustness #1

- Diffs-in-diffs controlling for pretest
  - Best of both?
  - Tx effect identical to ANCOVA

# Pseudo-Robustness #1

- $Y_2 - Y_1 = \gamma_0 + \gamma_1 X_1 + e$ (diffs-in-diffs)
- $Y_2 = b_0 + b_1 X_1 + b_2 Y_1 + e$ (ANCOVA)
- $Y_2 - Y_1 = \gamma_0 + \gamma_1 X_1 + \gamma_2 Y_1 + e$ (combined)
- $Y_2 = \gamma_0 + \gamma_1 X_1 + (1+\gamma_2) Y_1 + e$
  - Adding pretest makes $\gamma_1 = b_1$
    - $b2 = 1+\gamma 2$
- 2nd and 3rd equations above: equivalent by math

# Basis for other Pseudo-Robustness: Lin (2018)

- Lord's paradox: ANCOVA vs. difference-in-differences
  - Simulated
  - for corrective actions with a "known" result (Tx's for depression)

# Two Adjustment Methods

- **Difference scores: differences in differences / Change**

| | Pretest | Posttest | Change |
|---|---|---|---|
| Treatment | 9 | 5 | 4 |
| control | 2.5 | 2 | 0.5 |

- **Residualized score: ANCOVA/linear regression**

$$y_{ij1} = a + bx_j + cy_{ij0} + e_{ij}$$

# Lord's Paradox Simulated

## Lord's paradox

**Simulates the null hypothesis $H_0$ for difference score**

| | Pretest M | Posttest M | Change |
|---|---|---|---|
| Female | 130 | 130 | 0 |
| Male | 160 | 160 | 0 |

*r*(pre, post) = .48

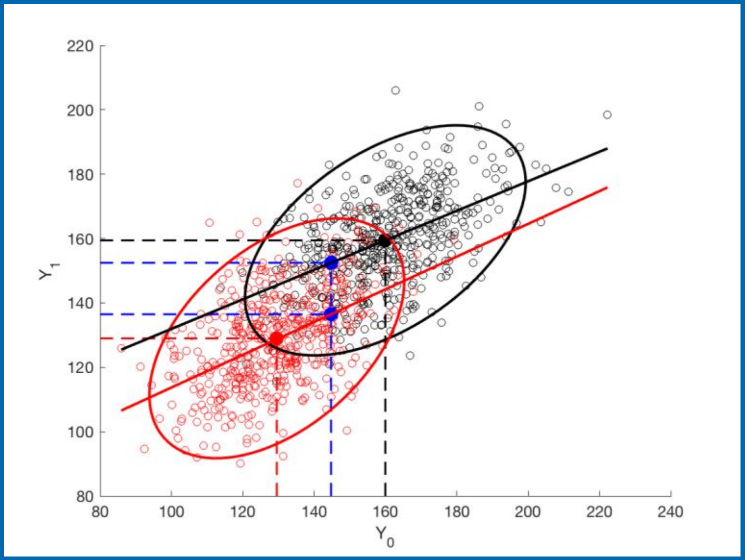
*SD* = 15

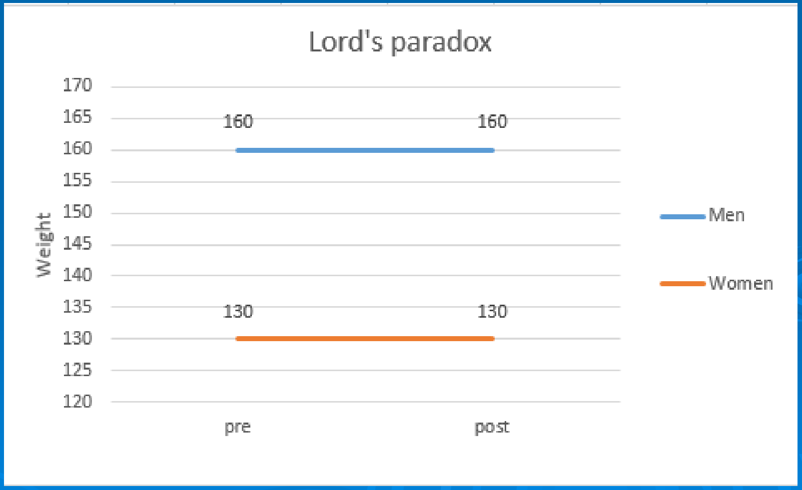
**Results:**

**Difference scores: *d* = .02 (n.s.)**

**Residualized scores: *b* = -15.60*****

***p < 0.001



# "Reversed" Paradox Simulated

## Reversed Lord's paradox

Simulates the null hypothesis $H_0$ for ANCOVA

| | Pretest | Posttest | Change |
|---|---|---|---|
| Women | 130 | 137.8 | - 7.8 |
| Men | 160 | 152.2 | 7.8 |

Mean pretest: $\overline{Y}_0 = 145$

Mean posttest: $\overline{Y}_1 = 145$

$$Y_{ij1} = a + b * girl_j + cY_{ij0} + e$$

$b = 0$

$c = 0.48$

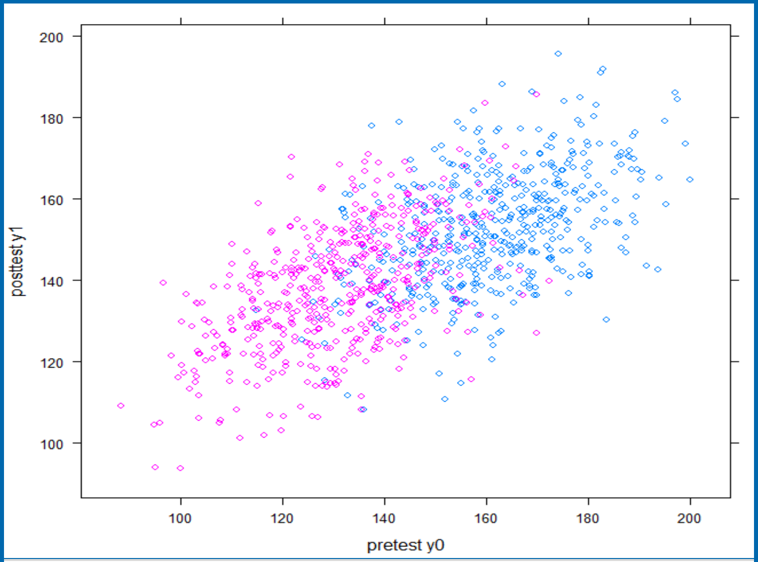
$a = \overline{Y}_1 - b \times \overline{Y}_0 = 75.4$

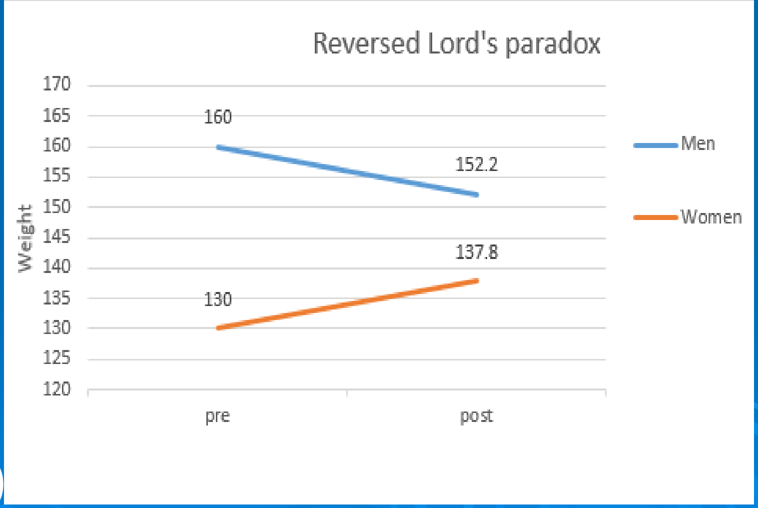
$SD(e) = 13.159$

**Results:**

Difference scores: $d = 15.61$***

Residualized scores: $b = 0.02$ (n.s.)



# Robust change-score estimates

Given equal variances across groups & times, etc.:

$d_1 - b_1 = (b_2 - 1)(M_{y20} - M_{y10})$

Thus, equating pretest group means makes the two change-score estimates equal (robust)

Equating pretest group means:

- Matching
- Centering all data on pretest group means

# Pseudo-Robustness #2: Matching: Simulations

Simulation Study

Differences-in-Diffs    Residualized Change

*Results from ANCOVA and Analyses of Simple Changes Scores for Two Simulated Datasets*

| Data | Pretest Difference | | Simple Change Score | | Residual Change Score | |
|---|---|---|---|---|---|---|
| | $d_0$ | $t(d_0)$ | $d_1$ | $t(d_1)$ | $b_1$ | $t(b_1)$ |
| Simulation of Lord's paradox data (to fit the null hypothesis for simple change scores) | | | | | | |
| Original data | –29.99*** | –31.68 | –0.02 | –0.002 | –15.60*** | 15.50 |
| Matched on pretest (1:1) | –0.09 | –0.08 | –15.53*** | –9.74 | –15.58*** | –10.55 |
| Simulation of reversed Lord's paradox (to fit the null hypothesis for ANCOVA) | | | | | | |
| Original data | –30.02*** | –31.68 | 15.61*** | 16.17 | 0.02 | 0.02 |
| Matched on pretest (1:1) | –0.09 | –0.08 | 0.11 | 0.07 | 0.06 | 0.04 |

Consistent, but biased

Consistent & unbiased?

# Pseudo-Robustness #2: Matching Pretests: Tx's for Depression

The FFCW data

*Results from ANCOVA and Analyses of Simple Changes Scores for Treatments for Depression in Mothers of the Fragile Families Data Set*

| | Pretest Difference | | Simple Change Score (Differences-in-Diffs) | | Residual Change Score (Residualized Change) | |
|---|---|---|---|---|---|---|
| Data | $d_0$ | $t(d_0)$ | $d_1$ | $t(d_1)$ | $b_1$ | $t(b_1)$ |
| *Psychological Treatment* | | | | | | |
| Original scale | 5.80*** | 19.44 | –2.13*** | –5.76 | 1.43*** | 6.12 |
| Matched on pretest (1:1) | 0 | 0 | 1.44** | 2.97 | 1.44*** | 3.51 |
| *Medication Treatment* | | | | | | |
| Original scale | 5.53*** | 16.41 | –1.89*** | –4.74 | 1.48*** | 5.78 |
| Matched on pretest (1:1) | 0 | 0 | 1.47** | 2.77 | 1.47*** | 3.28 |

Consistent, biased?

# Pseudo-Robustness #2: Matching: Kang (2022)

- Many steps to improve causal validity
  - Drop overly frequent spanking
  - Entropy balancing
    - Equated pretest means exactly

| | Effect of PP in Past Week | |
|---|---|---|
| | AR-1 | Diff-scores |
| Externalizing | .081*** | .081*** |
| Self-control | -.059*** | -.059** |
| Relational skills | -.059*** | -.059** |

# Pseudo-Robustness #2: Propensity-Score Matching

- Lin (2018) used Haviland et al. (2007) plan
  - Mixture modeling (3 depression trajectories)
  - Propensity-score matching within trajectories

| | **Effect on T5 Depression** | |
|---|---|---|
| | Residualized | Diff scores |
| Meds at T4 | | |
| Pretest matching | 1.49** | 1.38* |
| Propensity matching | 1.24* | 1.12 |
| Therapy at T4 | | |
| Pretest matching | 1.43** | 1.18* |
| Propensity matching | 1.24* | 1.02* |

# Do Propensity Scores Work for Corrective Actions?

- Home-visitations to reduce child abuse
- 3 studies found adverse effects with propensity- or entropy-score methods
  - Matone et al. (2012) injuries
  - Matone et al. (2018) severe injuries
    - Home visiting, Early Head Start, Parents as Teachers
  - Holland et al. (2022) investigated CPS reports
    - children removed from home, using diffs-in-diffs

# Are Propensity Scores Insuffient for Corrective Actions?

- Still biased like ANCOVA
  - Steiner et al. (2010)
  - Due to fallible covariates
- Has ANCOVA bias harmed parenting advice to at-risk parents?
  - Original Olds+ study: NFP’ers advised parents how to punish their preschoolers
  - Now: Positive parenting recommended

# Pseudo-Robustness #3: Dual-Centered Data

- **Dual-Centered ANCOVA**
- **Extension of Huitema's Quasi-ANCOVA**

$$Y_{1ij} - \overline{Y}_{0j} = \delta_0 + \delta_1 X_{ij} + \delta_2 (Y_{0ij} - \overline{Y}_{0j}) + \varepsilon$$

Center the posttest scores on pretest group means

Center the pretest scores on pretest group means

# Results Using Dual-Centered Data

| Data | Difference Scores $d_1$ | Difference Scores $t(d_1)$ | Residualized Change Score $b_1$ | Residualized Change Score $t(b_1)$ |
|---|---|---|---|---|
| Lord's example | -0.01 | -0.01 | -0.01 | -0.01 |
| Reversed | 15.61*** | 16.17 | 15.61*** | 18.76 |
| Sex costs talk | –0.08** | –2.77 | –0.08*** | –3.43 |
| Reasoning | –0.03* | –2.35 | –0.03* | –2.73 |
| Hospitalization | 0.16*** | 3.81 | 0.17*** | 4.91 |

More Power?
OR
Inflated $\alpha$

**Treatment** **Outcome**

Sex costs talk → Unprotected sex
Reasoning → Child aggression
Hospitalization → Physical health

# Pseudo-Robustness #4: Increasing Variance over Time

Given equal variances across groups & times, etc.:

$$d_1 - b_1 = (b_2 - 1)(M_{y20} - M_{y10})$$

Thus, the two change-score Tx estimates are equal (robust) when $b_2 = 1$

- Unrealistic limit when $\sigma^2$s of outcome are unchanged over time
- Possible with increasing $\sigma^2$s over time

# Pseudo-Robustness #4: $b_2$=1

- Example from Ding & Li (2019)
  - Article: 2 change-score analyses as brackets on true causal effect
- Example #2: Beneficial policy & electoral voting: flood disaster relief in Germany
  - $b_2$ = .997, Tx effects: 7.12 (ANCOVA) & 7.14 (diffs-in-diffs)

# Pseudo-Robustness #4: $b_2$=1

- Larzelere, Knowles et al. (2018)
  - Effects of 7 tactics by type of noncompliance
  - Robust results across change-score analyses
- $b_2 = 1$ when standardized $\beta_2 = SD_{Y0}/SD_{Y1}$
  - Externalizing: $r_2$ = .79 vs. 7.39/8.00 = .92
  - Internalizing: $r_2$ = .69 vs. 5.95/6.52 = .91
  - Total problems: $r_2$ = .75 vs. 16.87/17.69 = .94
- Increasing $\sigma^2$ of d.v. partially accounts for robustness

# Acknowledgements


- Funding from
  - NICHD grant #5 R03 HD107307
  - OK State U. Parenting Professorship