Reconsidering Lord’s Paradox: Implications for Valid Longitudinal Causal Inferences

Hua Lin & Robert E. Larzelere
Oklahoma State Univ.
Symposium Outline

- Intro: need to understand implications of Lord’s Paradox for causal inference
  - ANCOVA vs. Diffs-in-Diffs
- Pseudo-robustness across 2 change-score models
- Does group-mean centering bias Tx se’s?
- A test of Tx X Pretest interactions in diffs-in-diffs model
Outline to Intro to Symposium

- Foundational: Valid causal inferences of change
- Lord’s paradox: unresolved after 56 years
- Pros and cons
  - ANCOVA-type residualized change
  - Difference-score analyses
- Implications: analyzing corrective actions
Basic Human Devel. Questions

- Describe between-person diffs
- Explain in
- Optimize within-person change

But how should we analyze change?

- for valid causal explanations
  - Necessary for optimal applications
  - Difficult without randomization or equivalent
How to Analyze Change?

- **ANCOVA-type residualized change**
  - Predict $Y_2$ controlling for $Y_1$: $Y_2 / Y_1$

- **Difference-score analyses**
  - Predict $Y_2 – Y_1$

- **Lord’s (1967) Paradox:**
  - 2 change-score analyses often contradictory
Which is more causally valid?

- Residualized change
  - Cronbach & Furby (1970), Pearl (2016)

- Simple change
  - Allison (1990), Castro-Schilo & Grimm (2018)

- Both
  - Robustness: Duncan et al. (2014)
  - Bracketing: Angrist & Pischke (2009), Ding & Li (2019)
Which is more causally valid?

- Depends on assumptions:

- Assumptions (under null $H_0$)
  - Simple change: parallel slopes
    - Differential slopes undermine causal validity
    - e.g., regression toward the mean
  - Residualized change: ignorability (like RCT)
    - Covariates independent of Tx | stat model
    - Between-person diffs undermine causal validity
Null Hypothesis for Difference-Score Analysis

Lord's paradox

<table>
<thead>
<tr>
<th>Weight</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>120</td>
<td></td>
<td></td>
</tr>
<tr>
<td>130</td>
<td>130</td>
<td>130</td>
</tr>
<tr>
<td>140</td>
<td></td>
<td></td>
</tr>
<tr>
<td>150</td>
<td></td>
<td></td>
</tr>
<tr>
<td>160</td>
<td>160</td>
<td>160</td>
</tr>
<tr>
<td>170</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ANCOVA’s Null H₀

Null Hypothesis for ANCOVA

Reversed Lord's paradox

- Men: Decrease in weight from pre to post.
- Women: Increase in weight from pre to post.

Weights:
- Pre: Men = 160, Women = 130
- Post: Men = 152.2, Women = 137.8
Why Residualized Change?

- Greater statistical power
- Unbiased causal estimate IFF all confounders controlled for perfectly
- Doesn’t require Time-1 measure of d.v.
  - Propensity-score matching
Points Against Residualized Change

- Assumes covariates indpt of Tx
- Biased by between-person differences
  - Unadjustment bias – Campbell & Boruch (‘75)
  - Hamaker et al. (2015)
  - Berry & Willoughby (2017)
  - Hoffman (2015)
- Assumes regression toward a grand mean
  - Two equivalent distributions
Why Simple Change?

- Pure within-person changes
  - Not confounded with between-person diffs
- Overcomes underadjustment bias?
- Unbiased IFF parallel slopes assumption is correct
Points Against Simple Change

- Need equivalent measures @ Time 1 & 2
- Assumes interval scale without ceiling or floor
- Low reliability
- Difference score – $r$ with pretest score
- Cannot test Treatment X Pretest
Which is More Causally Valid?

- Kenny’s (2011) underappreciated work
  - “Change that we cannot believe in”
  - Kenny (1975): Contrast of two change-score analyses in NECG design
  - “It depends”
    - On trait vs. state components of pretest
  - Issue: fallibility of measures
My Interest in This Issue

- What alternative tactics > physical punish?
  - 45-year research program

- Two questions:
  - Are adverse r’s of PP causal or spurious?
  - What other tactics are more effective?
    - Esp. for difficult disciplinary situations
## 4 Corrective Actions by Type of Statistical Evidence (Externalizing)

<table>
<thead>
<tr>
<th>Evidence Type</th>
<th>Physical Punishment</th>
<th>Nonphysical Punishment</th>
<th>Therapy for Child</th>
<th>Ritalin</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cross-sectional $r$</td>
<td>.20***</td>
<td>.20**</td>
<td>.12**</td>
<td>.11**</td>
</tr>
<tr>
<td>Longitudinal $r$</td>
<td>.16***</td>
<td>.18**</td>
<td>.13**</td>
<td>.13**</td>
</tr>
<tr>
<td>$\beta$ (Y2</td>
<td>Y1)</td>
<td>.07***</td>
<td>.08^</td>
<td>.14^</td>
</tr>
<tr>
<td>$r$ with Y2 – Y1</td>
<td>-.04*</td>
<td>-.02^</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td>RCT</td>
<td>-.35*</td>
<td>-.64**</td>
<td>-.18**</td>
<td>--</td>
</tr>
</tbody>
</table>
Corrective Actions

- **Def:** An action selected to correct a perceived problem
  - Subsequent symptoms due to:
    - Poor prognosis of problem
    - Effect of action to modify that prognosis

- **ANCOVA biased against corrective actions**
  - Make them look less effective than they are
Intervention Selection Bias

Intervention Selection Bias

CORRECTIVE ACTION

PERFECT COVARIATE

$+r_1$

DETRIMENTAL OUTCOMES

$+r_2$

Unbiased $b_3$
Intervention Selection Bias

CORRECTIVE ACTION

FALLIBLE COVARIATE

+ \( r_1 \)

DETRIMENTAL OUTCOMES

Biased

+ \( b_3 \)

+ \( r_2 \)
Counterfactuals for Non-Physical Punish: 3 Analyses
ANCOVA Biased Against Most Corrective Actions

- By parents
  - All disciplinary responses to misbehavior
  - Helping with homework
  - Talks against deviant behaviors and peers

- By professionals
  - Therapy for kids and women
  - Medical Tx’s for kids and women
  - Out-of-home placements
  - Job training programs
Effects of Biases in Parenting Research?

- Evidence against denigrated corrective actions: Easy
  - spanking, harsh parenting

- Evidence for good corrective actions: Hard
  - APA & AAP: No cited evidence for tactics to replace spanking

- Weakens 1st step of translational research
  - Effectiveness of child Tx’s stagnant or down
    - e.g., Tx for conduct probs: $d = .76$ in 1963; $d = .36$ in 2017: MA by Weisz et al. (2019)
Lessons from Econometrics

- Robustness (Duncan et al., 2014)
  - Across analyses w contrasting biases
- More humility about causal evidence
- Generated regressors
Acknowledgement

➢ Consultants
  • Dave Kenny
  • Joshua Habiger & Tianyu Cao
  • B. Wade Brorsen

➢ Funding
  • NICHD grant #5 R03 HD107307
  • OK State Univ. Parenting Professorship
Extra, Unused Slides

- Means: 130 & 160; SD = 15
- Null H₀: No-Tx effect re simple gain scores
Null $H_0$: ANCOVA: Lin (2018)

- Ms: 130, 160, post: 137.8, 152.2; $SD = 15$
- Null $H_0$: No-Tx effect re ANCOVA
## Lord’s Paradox: Examples

<table>
<thead>
<tr>
<th>Example</th>
<th>Treatment (Time 1)</th>
<th>Outcome (Time 1 &amp; 2)</th>
<th>Sample Size</th>
<th>Data Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example 1</td>
<td>Parent-youth discussions about sexual risks</td>
<td>Subsequent unprotected sexual behaviors</td>
<td>$N = 4753$</td>
<td>Adolescent to Adult Health</td>
</tr>
<tr>
<td>Example 2</td>
<td>Disciplinary reasoning</td>
<td>Subsequent child aggression</td>
<td>$N = 2467$</td>
<td>Fragile Families Child Wellbeing</td>
</tr>
<tr>
<td>Example 3</td>
<td>Hospitalization</td>
<td>Subsequent physical health in mothers</td>
<td>$N = 3831$</td>
<td>Fragile Families Child Wellbeing</td>
</tr>
</tbody>
</table>
Lord’s Paradox: Results

<table>
<thead>
<tr>
<th>Data</th>
<th>Difference Scores</th>
<th>Residualized Change Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$d_1$</td>
<td>$t(d_1)$</td>
</tr>
<tr>
<td>Lord’s example</td>
<td>-0.02</td>
<td>-0.00</td>
</tr>
<tr>
<td>Reversed</td>
<td>15.61***</td>
<td>16.17</td>
</tr>
<tr>
<td>Sex costs talk</td>
<td>-0.08**</td>
<td>-2.77</td>
</tr>
<tr>
<td>Reasoning</td>
<td>-0.03*</td>
<td>-2.35</td>
</tr>
<tr>
<td>Hospitalization</td>
<td>0.16***</td>
<td>3.81</td>
</tr>
</tbody>
</table>

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