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
- Explain
- Optimize

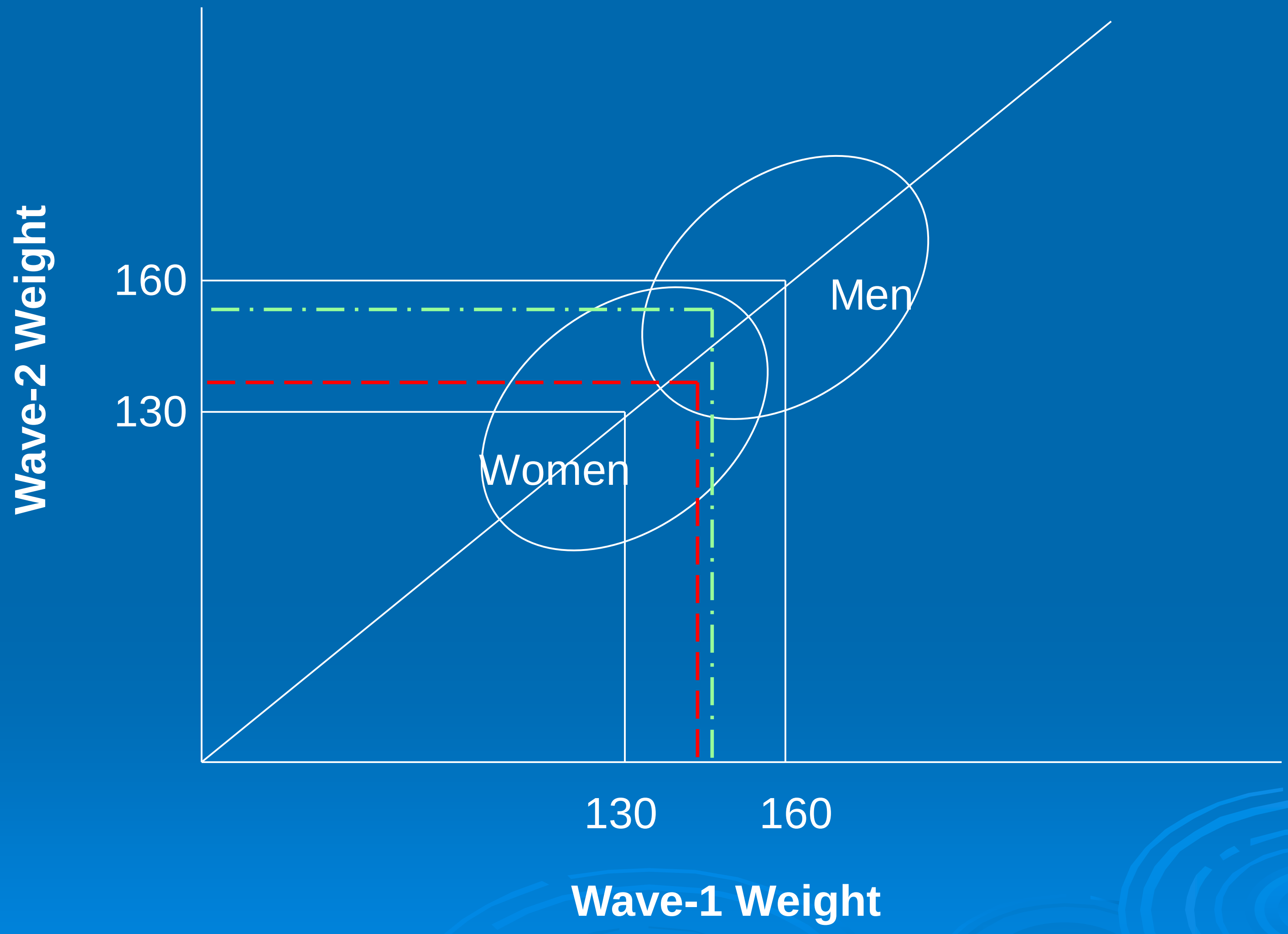
between-person diffs
in
within-person change

- 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 : $Y_2 \mid Y_1$
- Difference-score analyses
  - Predict $Y_2 – Y_1$
- Lord's (1967) Paradox:
  - 2 change-score analyses often contradictory

Lord's Paradox



Which is more causally valid?

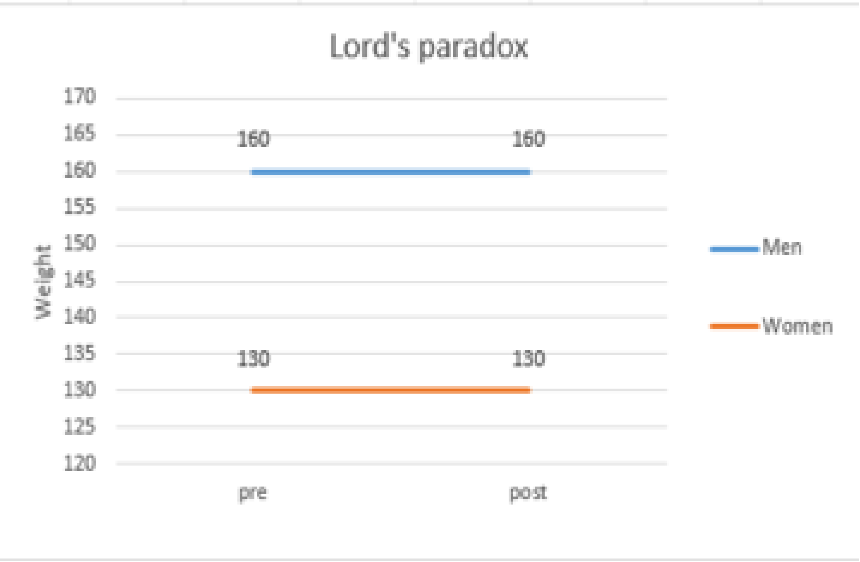
- Residualized change
  - Cronbach & Furby (1970), Pearl (2016)
- Simple change
  - Allison (1990), Castro-Schilo & Grimm (2018)
- Both
  - Robustness: Duncan et al. (2014)
  - Bracketing: Angrist & Pischke (2009), Ding & Li (2019)

Which is more causally valid?

- Depends on assumptions:
  - Holland & Rubin (1983), Wainer (1991)
- 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)
    - Coviariates independent of Tx | stat model
    - Between-person diffs undermine causal validity

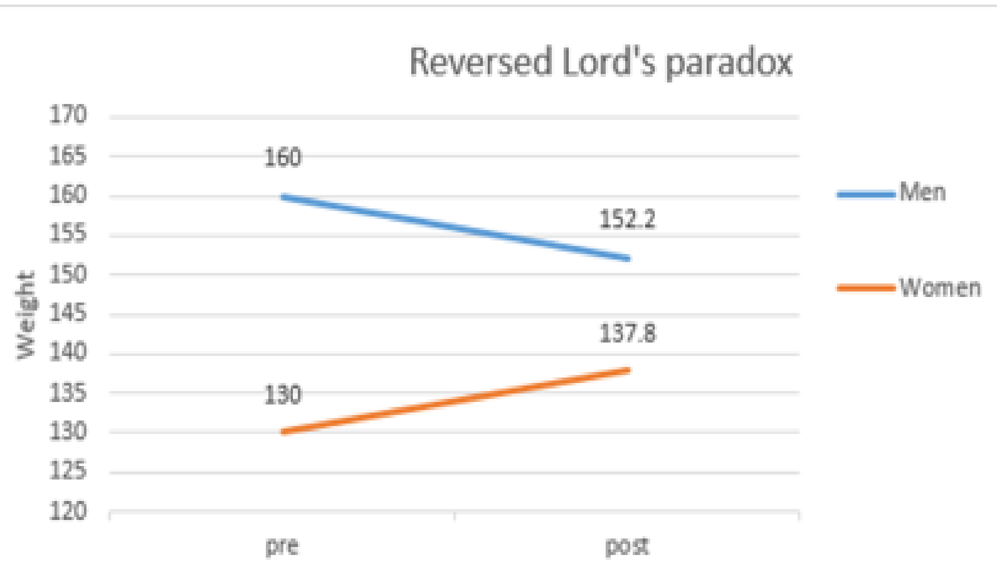
Difference-score's Null H_0

Null Hypothesis for Difference-Score Analysis



ANCOVA's Null H_0

Null Hypothesis for ANCOVA



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

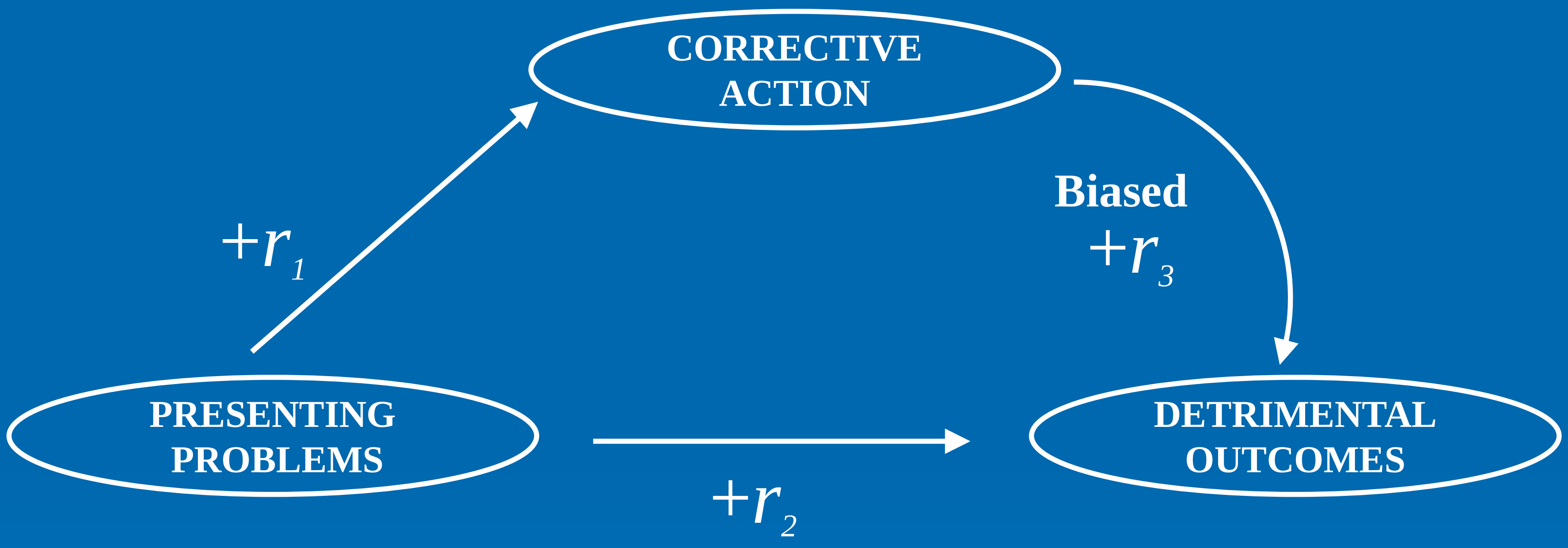
4 Corrective Actions by Type of Statistical Evidence (Externalizing)

Evidence Type	Physical Punishment	Nonphysical Punishment	Therapy for Child	*Ritalin*
Cross-sectional *r*	.20***	.20**	.12**	.11**
Longitudinal *r*	.16***	.18**	.13**	.13**
β (Y2 | Y1)	.07***	.08^	.14^	.06
r with Y2 – Y1	-.04*	-.02^	.01	.02
RCT	-.35*	-.64**	-.18**	--

Corrective Actions

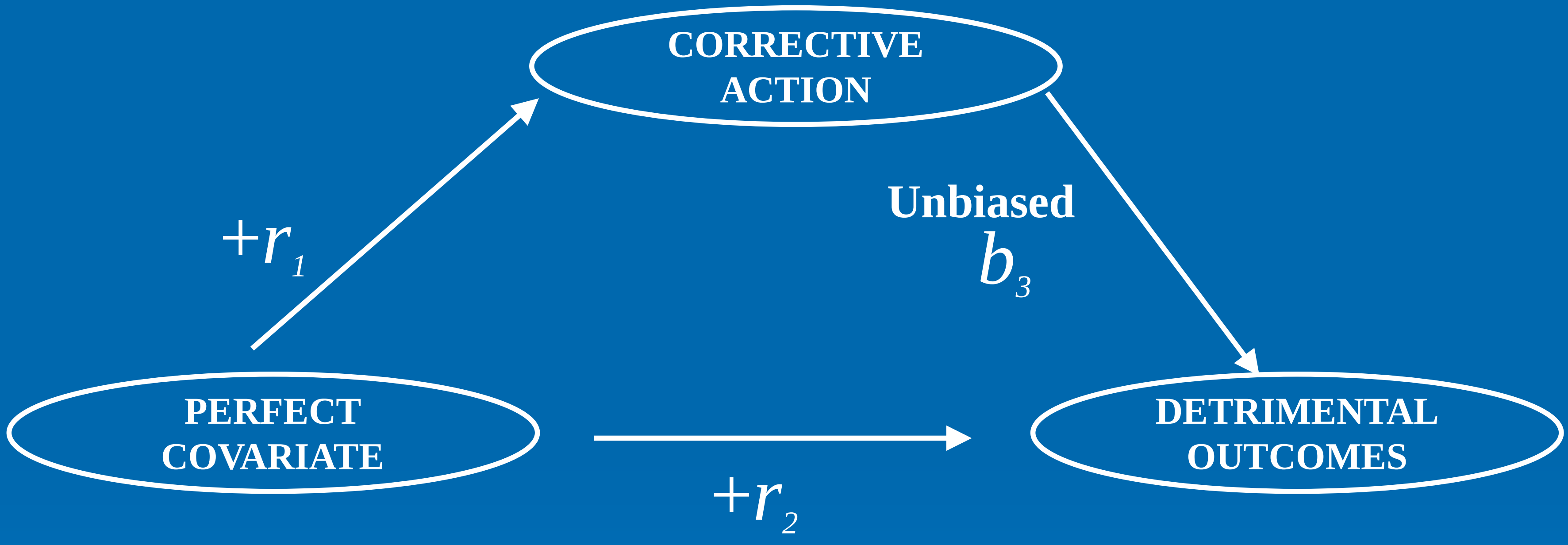
- Def: An action selected to correct a perceived problem
  - Subsequent symptoms due to:
    - Poor prognosis of problem
    - Effect of action to modify that prognosis
- ANCOVA biased against corrective actions
  - Make them look less effective than they are

Intervention Selection Bias

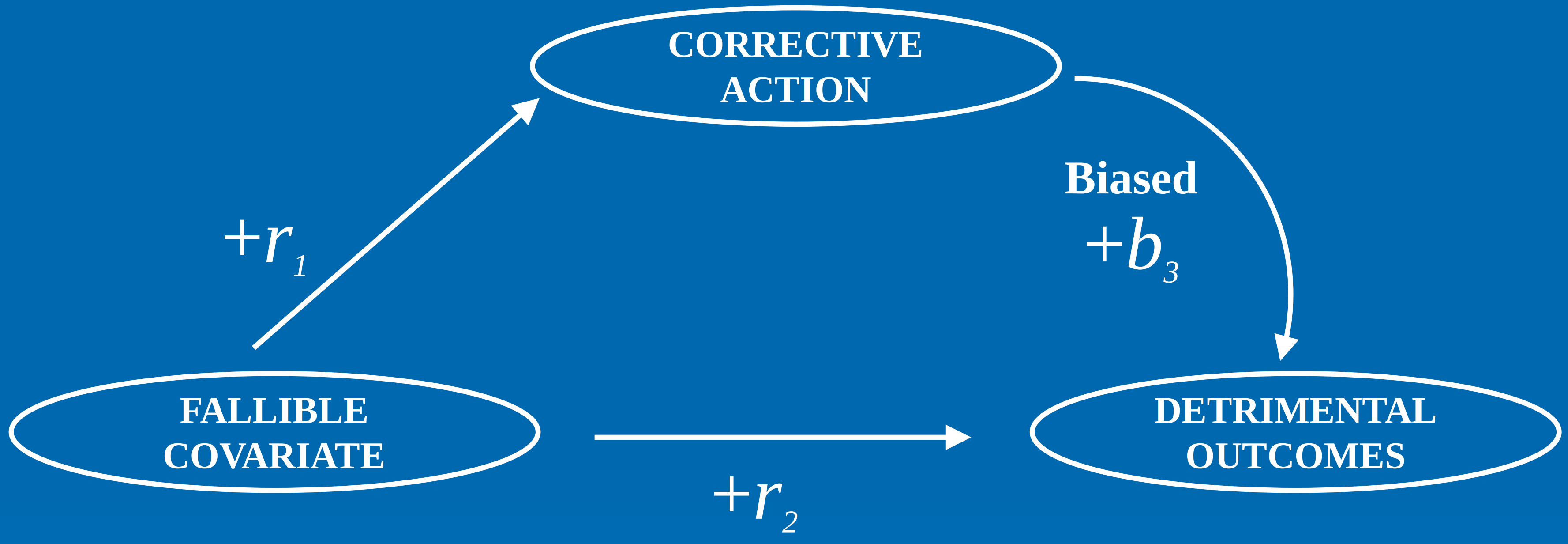


Larzelere et al. (2004). The intervention selection bias. *Psychol. Bull.*

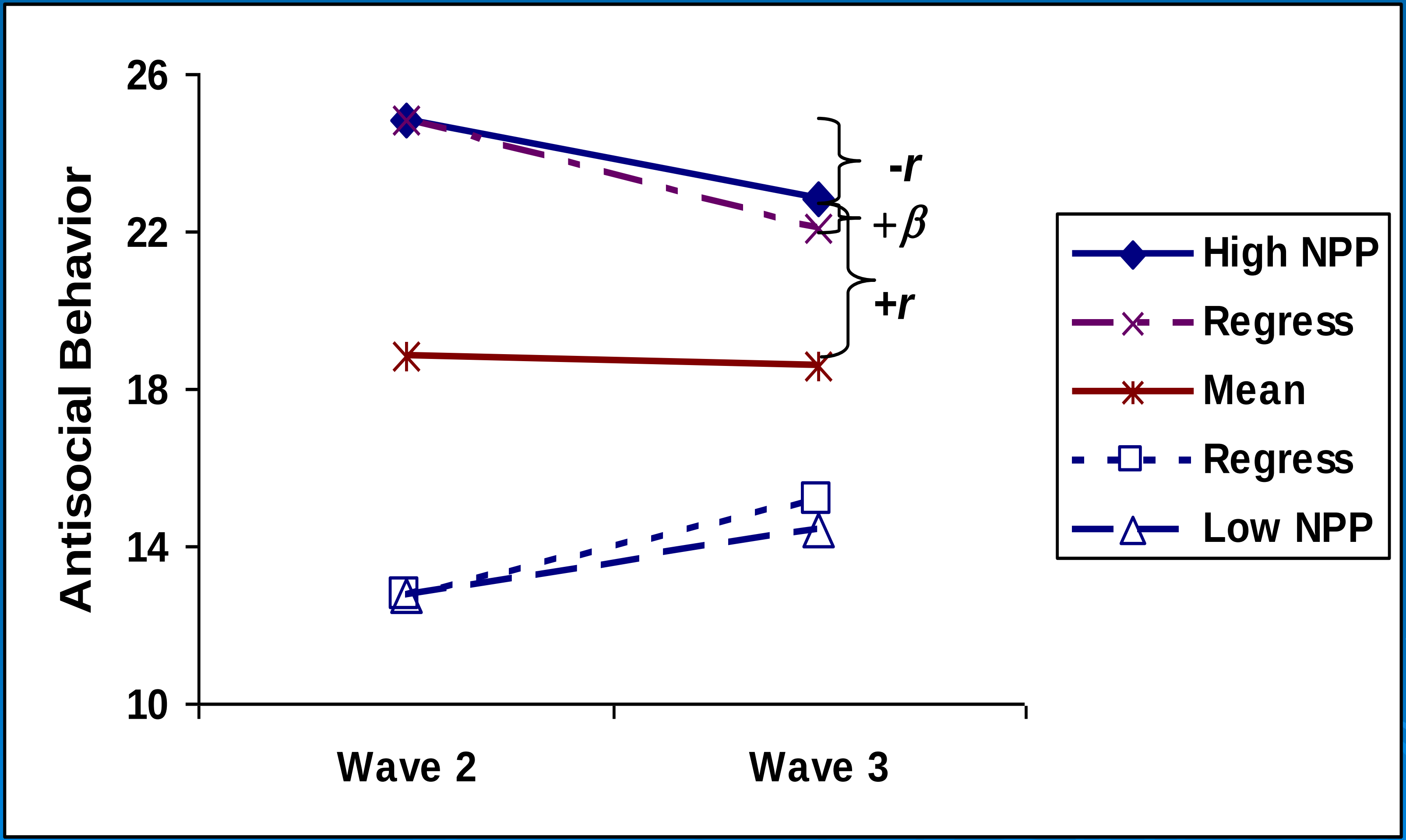
Intervention Selection Bias



Intervention Selection Bias



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

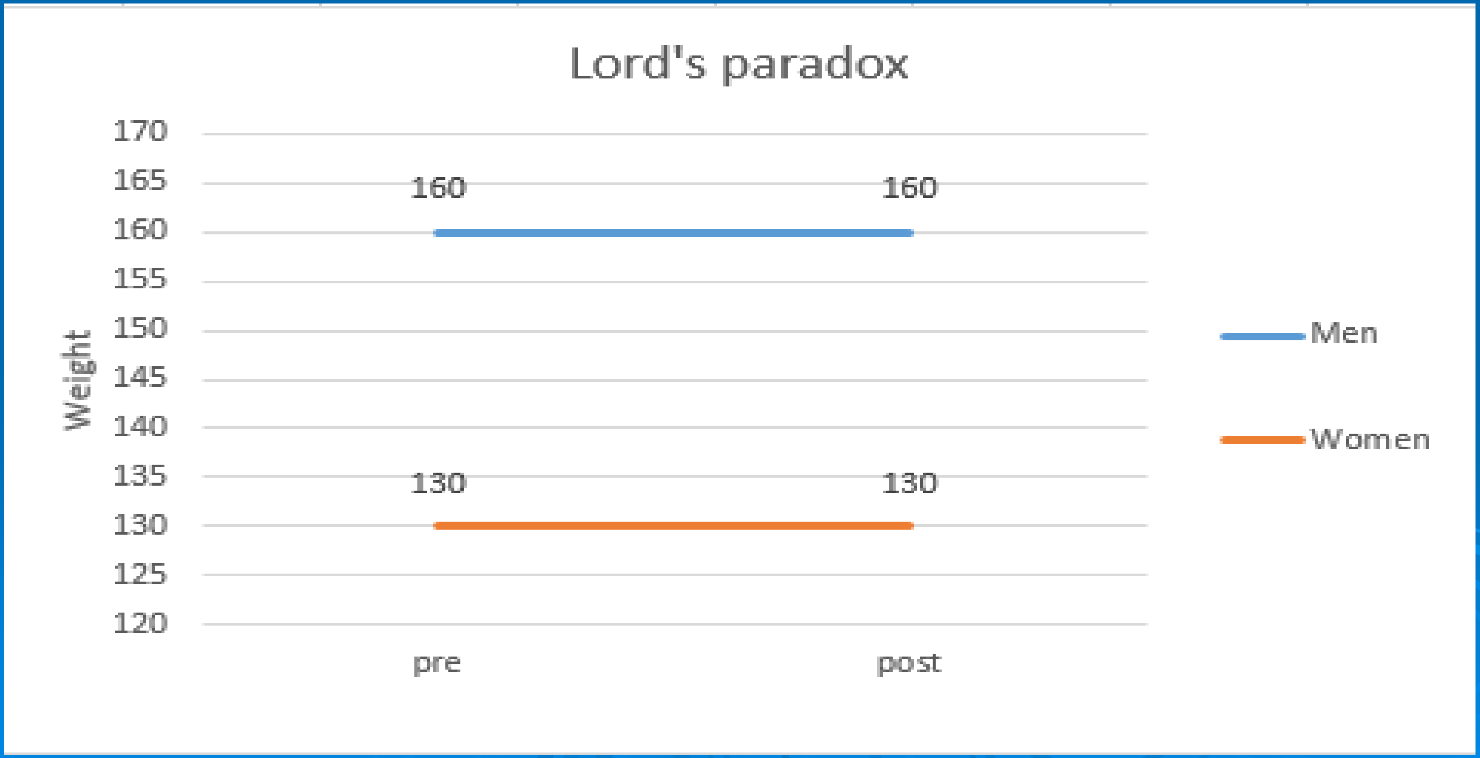
Acknowledgement

- Consultants
  - Dave Kenny
  - Joshua Habiger & Tianyu Cao
  - B. Wade Brorsen
- Funding
  - NICHD grant #5 R03 HD107307
  - OK State Univ. Parenting Professorship

Extra, Unused Slides

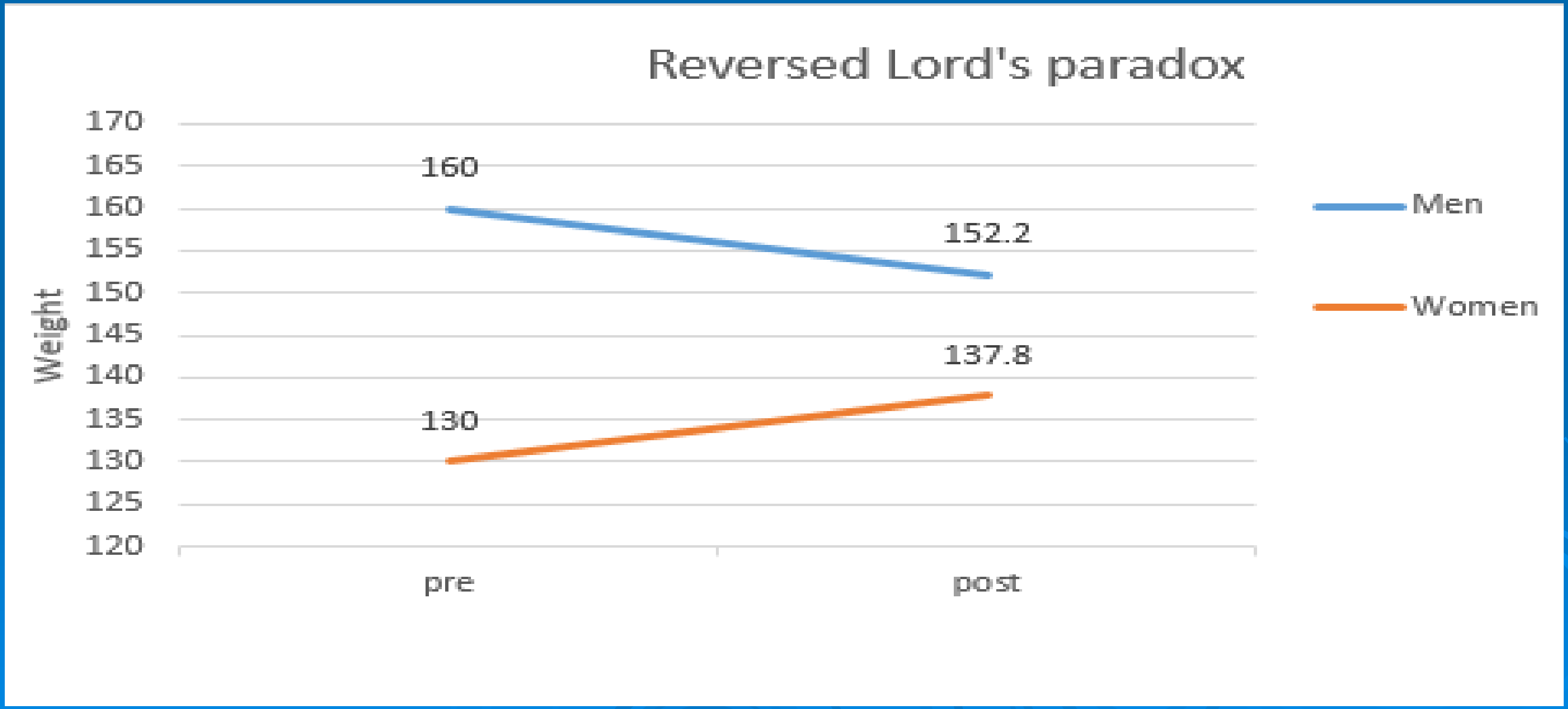
Null H_0: Difference-Scores: Lin (2018)

- Means: 130 & 160; *SD* = 15
- Null H_0: 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

	Treatment (Time 1)	Outcome (Time 1 & 2)	Sample Size	Data Set
Example 1	Parent-youth discussions about sexual risks	Subsequent unprotected sexual behaviors	$N = 4753$	Adolescent to Adult Health
Example 2	Disciplinary reasoning	Subsequent child aggression	$N = 2467$	Fragile Families Child Wellbeing
Example 3	Hospitalization	Subsequent physical health in mothers	$N = 3831$	Fragile Families Child Wellbeing

Lord's Paradox: Results

Data	Difference Scores		Residualized Change Score	
	d_1	$t(d_1)$	b_1	$t(b_1)$
Lord's example	-0.02	-0.00	-15.60***	15.50
Reversed	15.61***	16.17	0.02	0.02
Sex costs talk	–0.08**	–2.77	0.07**	2.78
Reasoning	–0.03*	–2.35	0.04**	3.51
Hospitalization	0.16***	3.81	–0.15***	–4.15

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