Computation of Regression-based Effect Size Measures

Tasha Beretvas,

Hyewon Chung, Wendy Machalicek, & Clare Riviello

University of Texas at Austin

Outline

- Large-*n* data and effect sizes (ESs)
- Single-*n* data and ESs
  - Problems
- Four-parameter reg'n models
  - Explanation
  - Adaptation
  - ESs (ΔR^2- and β-based)
- Comparison study of parametric ESs

Large-n Studies' Data

- Typically use standardized mean difference effect size (ES):

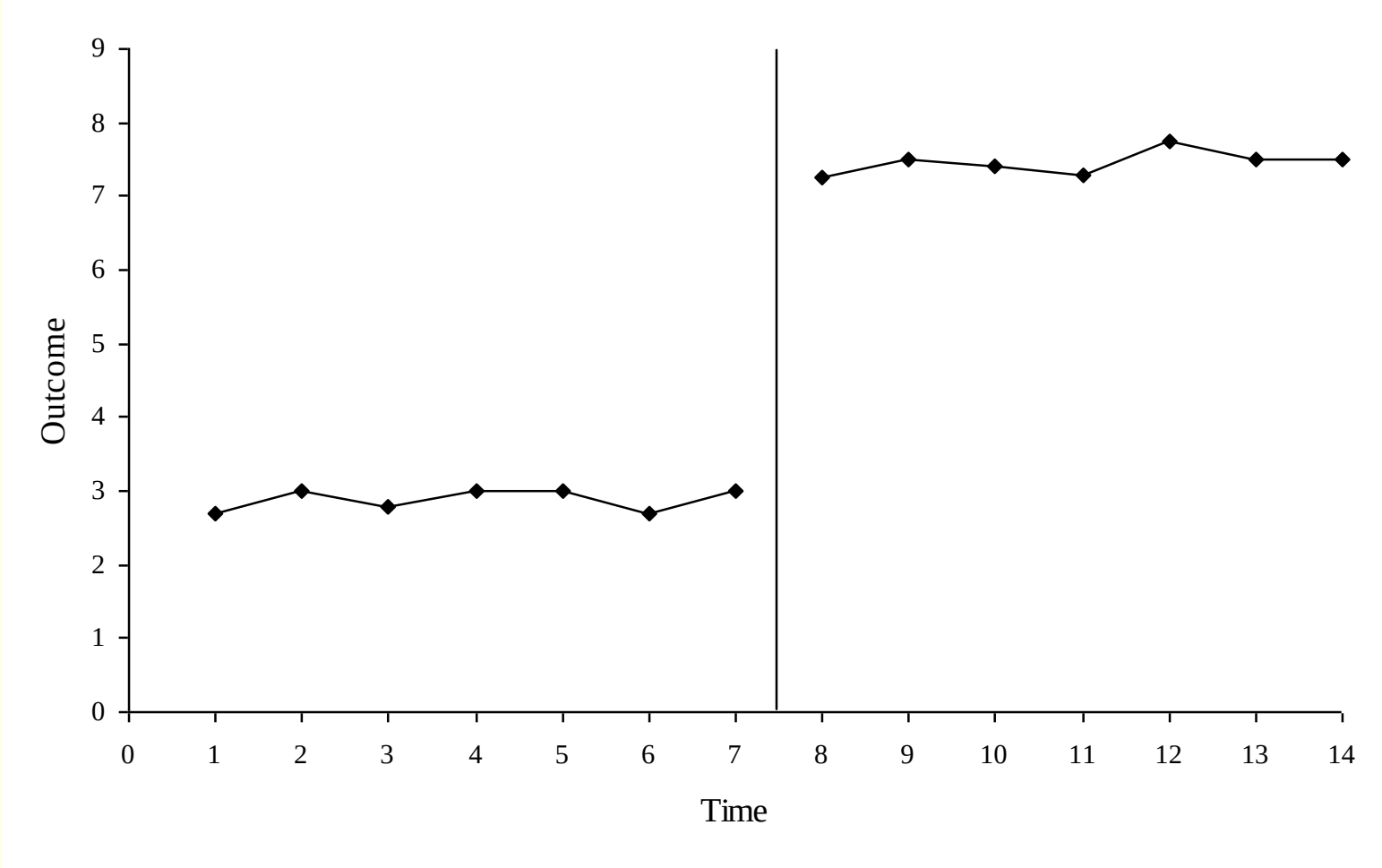
$$d = \frac{M_T - M_C}{s}$$

- Each participant is measured once on the outcome; data are independent.
- The standard deviation provides an estimate of the variability of these *independent* data points.

Single-n Studies' Data

- Most simply: repeated measures on an individual ***over time*** in two phases (time series data):
  - Baseline: phase A = "control"
  - Treatment: phase B = "treatment"
- Score at time point *t* is related to score at time (*t* – 1): i.e., <u>not independent</u>.

Single-n Studies' Data



Single-n Studies' Data Visual Analysis:

- Plots are evaluated for the presence of a treatment effect by simultaneously considering the following :
  - Sustainable <u>level</u> and/or <u>trend</u> changes
  - Baseline trends in expected direction
  - Overlapping data between phases
  - Variability changes within and across phases.

Single-n Effect Sizes

- PND, and other non-parametric ESs
- Seems reasonable that a standardized difference between scores in phase A and B could be used as an effect size (ES):

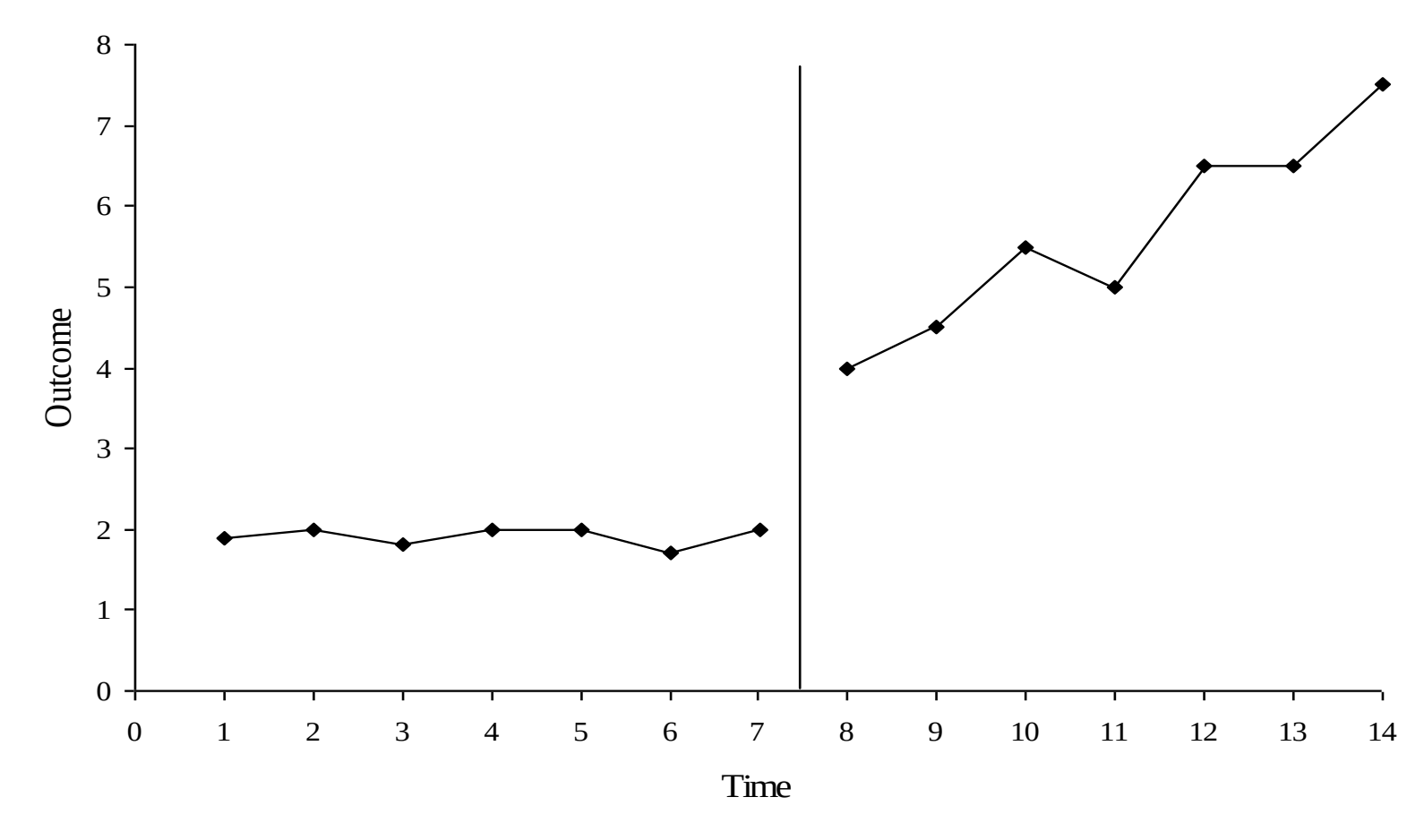
$$d = \frac{M_B - M_A}{s}$$

- Easy to interpret but problematic due to ***s***
  - Dependency (Autocorrelation, ρ_1) of data points ***within*** an individual.

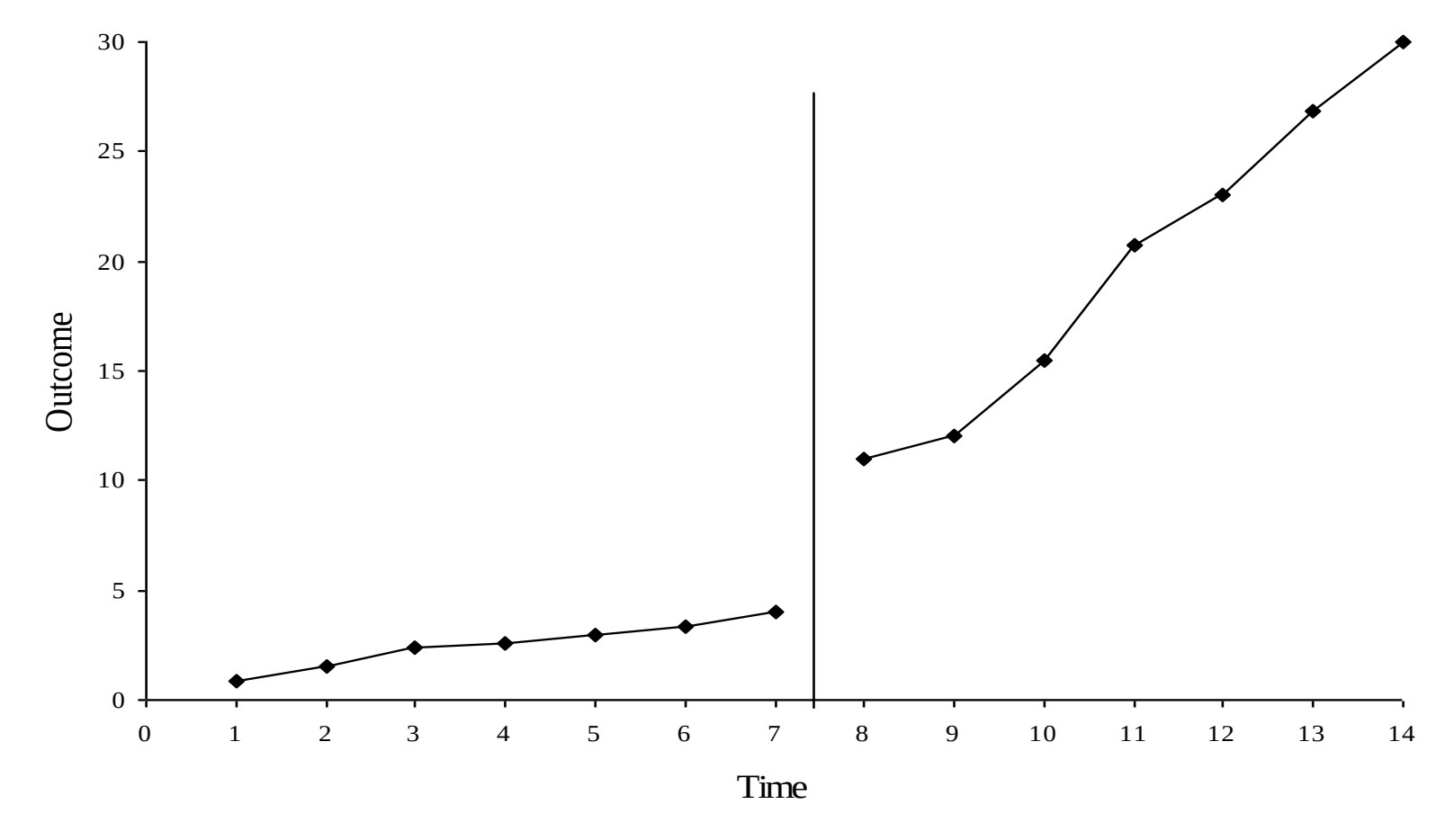
Problems with ***d*** for ***single-n*** designs

- In addition to potential autocorrelation problems, it would also ***only*** make sense to use ***d*** when there is ***no trend*** in the data.
- Regression equation for interrupted time series data should include a parameter for change in level and one to assess change in slope.

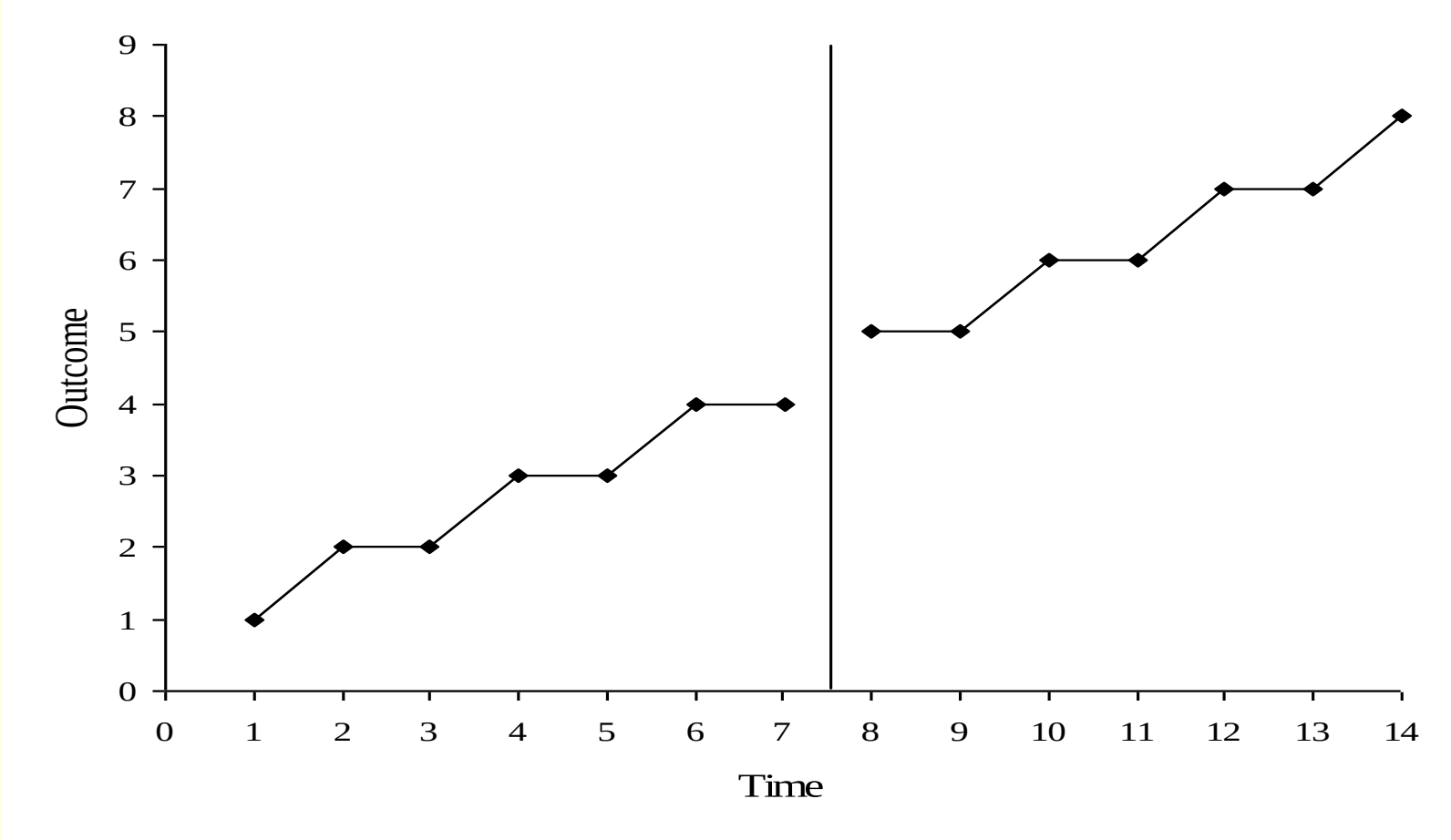
Trend in B phase, tx effect



Trend in A and B phases, tx effect



Trend in A and B phases, no Tx effect



Modeling change in level and slope

- Led to idea to reduce autocorrelation and permit assessment of treatment effect on both level and slope using an appropriate regression model specification.
- Huitema & McKean (2000) suggested parameterization of a piecewise regression eqn that included two coefficients that can be used to describe change in level and in slope from Baseline to Treatment phase.

4-parameter model (Huitema & McKean, 2000)

- The model:

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 D_t + \beta_3 [T_t - (n_1 + 1)] D_t + e_t$$

where

Y_t = outcome score at time t

T_t = time point

D = phase (A or B)

n_1 = # time points in phase A

4-parameter model - interpretation

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 D_t + \beta_3 [T_t - (n_1 + 1)] D_t + e_t$$

β_0 = baseline intercept (i.e. *Y* at time = 0)

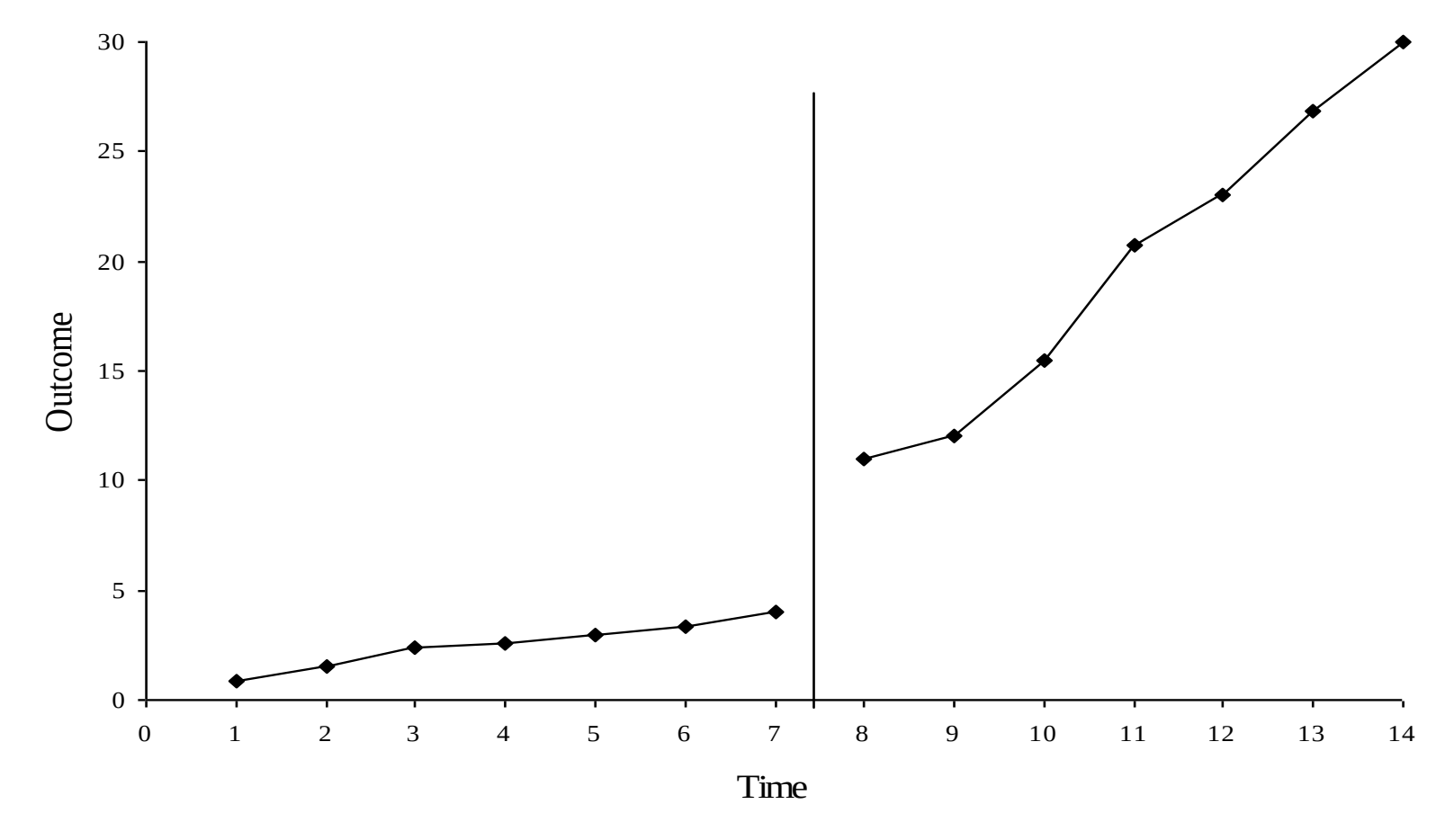
β_1 = baseline linear trend (slope over time)

β_2 = difference in intercept predicted from treatment phase data from that predicted for time = n_1+1 from baseline phase data

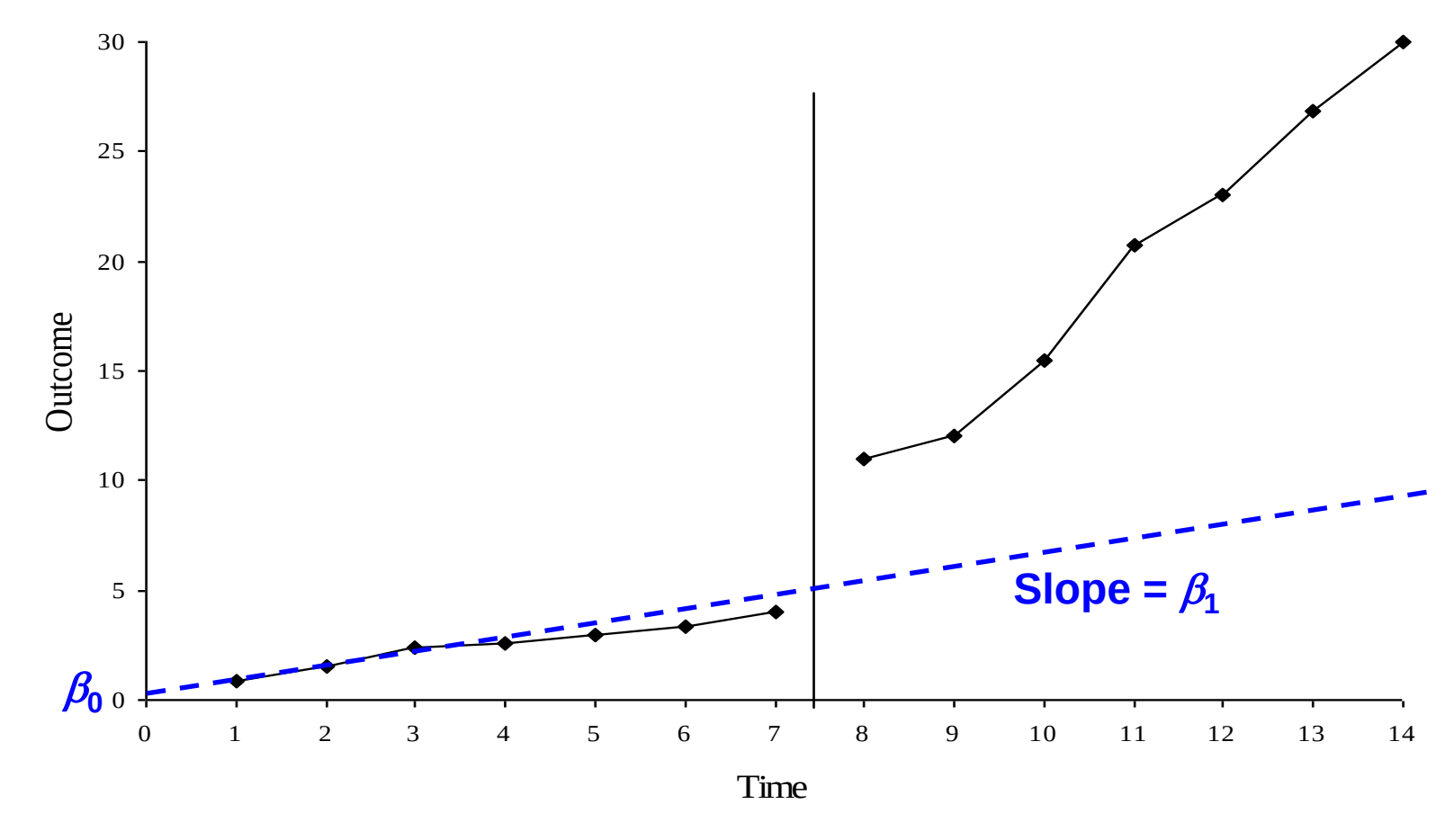
β_3 = difference in slope

- Thus $\boldsymbol{\beta_2}$ and $\boldsymbol{\beta_3}$ provide estimates of a treatment's effect on level and on slope, respectively.

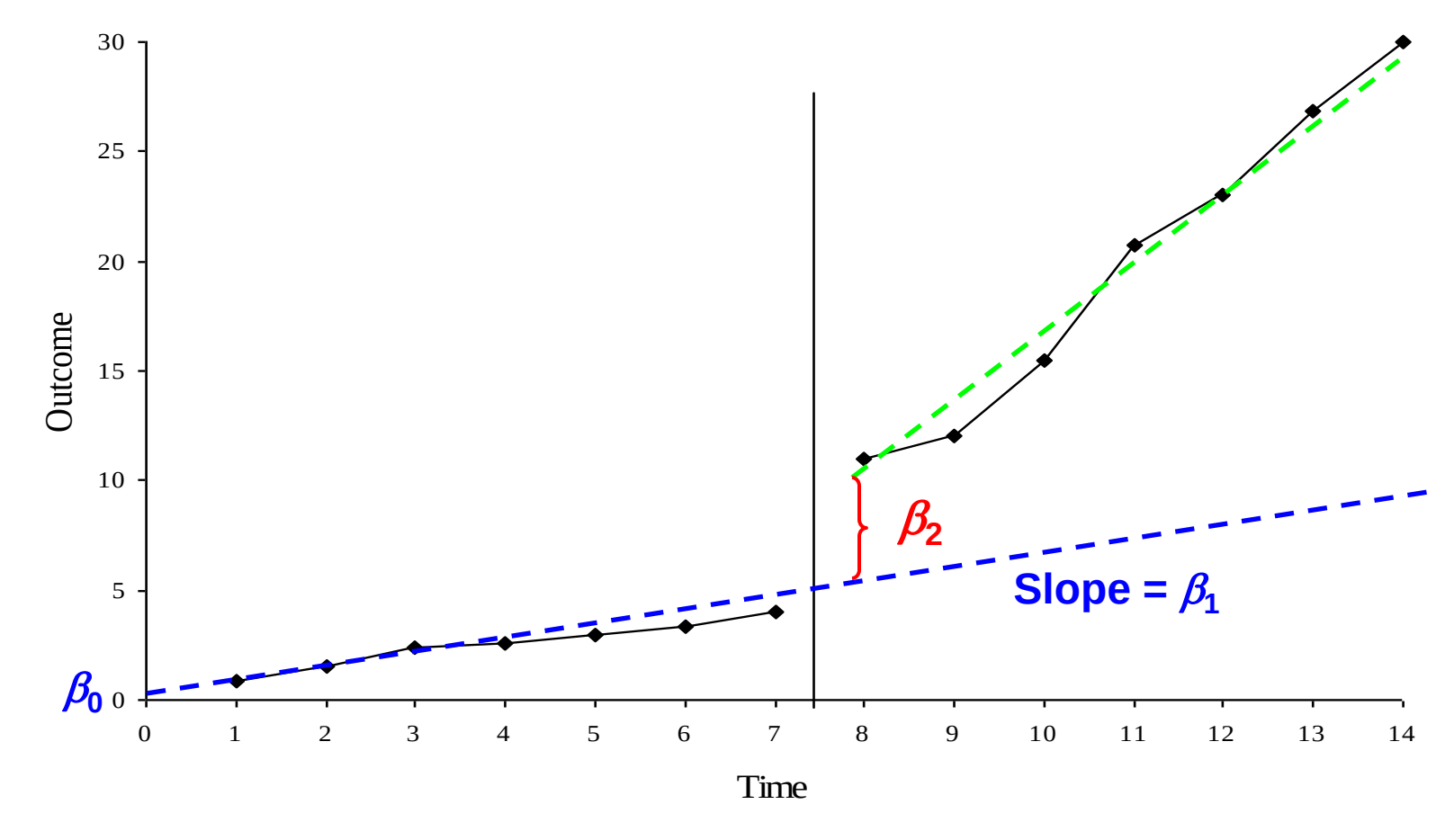
4-parameter model - interpretation



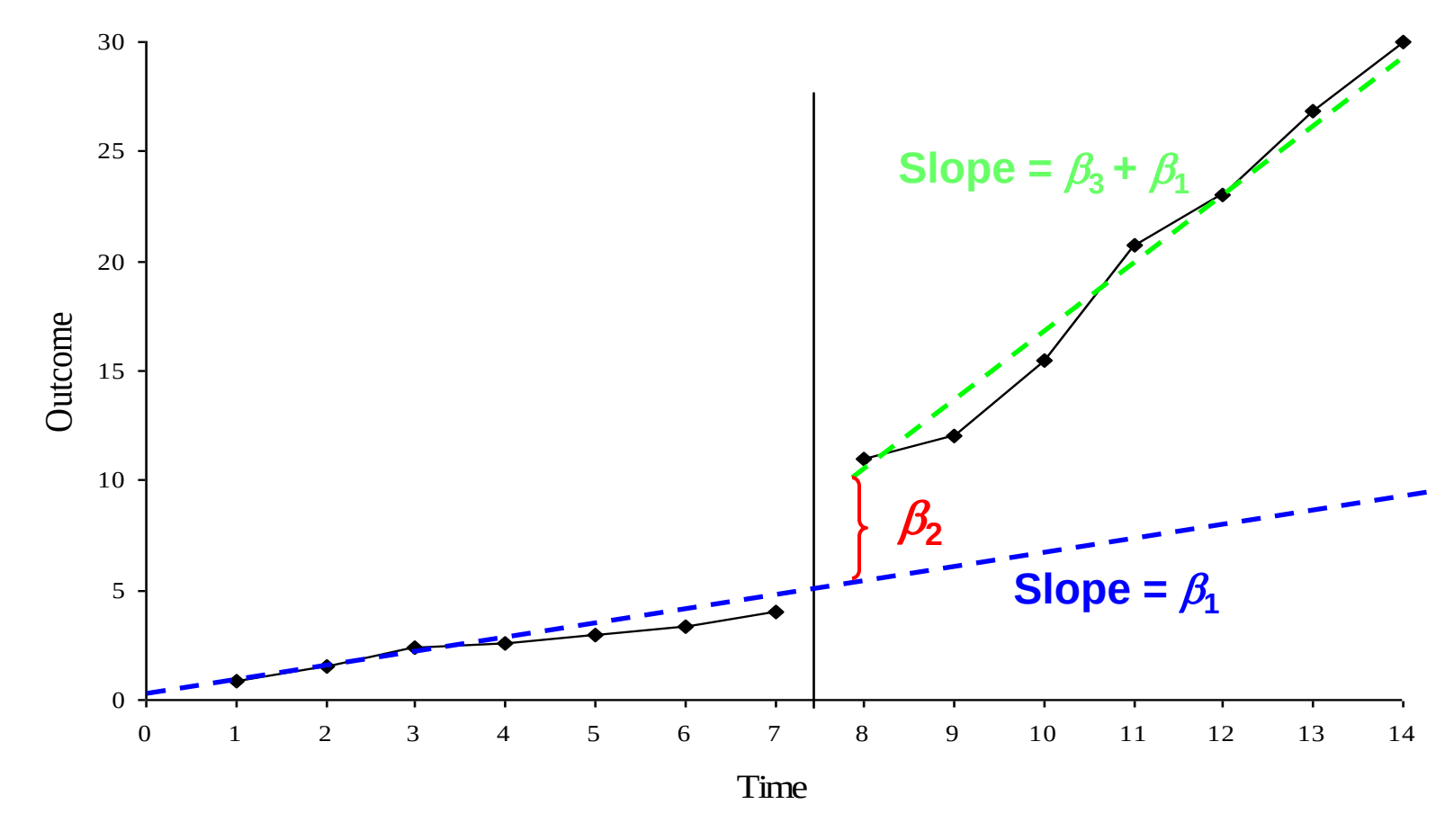
4-parameter model - interpretation



4-parameter model - interpretation



4-parameter model – interpretation



Real data example available upon request

δ_β Effect Sizes

- Van den Noortgate & Onghena (2003) suggested standardizing the β_2 and β_3 coefficients in the full model to provide standardized ESs:

$$\delta_{\beta_2} = \frac{B_2 - 0}{\sqrt{MSE}} \qquad \delta_{\beta_3} = \frac{B_3 - 0}{\sqrt{MSE}}$$

- They also suggested correcting the associated covariance (of the β_js):

$$\sigma_{\delta_{\beta_i}\delta_{\beta_j}} = \frac{\sigma_{B_i B_j}}{MSE}$$

- They recommended pooling the resulting β_js using multivariate synthesis methods

ΔR^2 Effect Sizes

Based on: $$\delta = \left(2t \Big/ \sqrt{df_E}\right)$$

and $$t(df_E) = \sqrt{F(1, df_E)}$$

An *F*-ratio can be used to test ΔR^2 for a Restricted versus a Full model:

$$F = \frac{\left(R^2_{Full} - R^2_{Rst}\right) \Big/ \left(\nu_{Rst} - \nu_{Full}\right)}{\left(1 - R^2_{Full}\right) \Big/ \left(n - \nu_{Full} - 1\right)}$$

where $\nu_{Rst} - \nu_{Full} = 1$

ΔR^2 Effect Sizes

- Models estimated to calculate $\delta_{\Delta R_S^2}$

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 D_t + e_t$$

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 D_t + \beta_3 [T_t - (n_1 + 1)] D_t + e_t$$

- Models estimated to calculate $\delta_{\Delta R_I^2}$

$$Y_t = \beta_0 + \beta_1 T_t + \beta_3 [T_t - (n_1 + 1)] D_t + e_t$$

$$Y_t = \beta_0 + \beta_1 T_t + \beta_2 D_t + \beta_3 [T_t - (n_1 + 1)] D_t + e_t$$

ΔR^2 Effect Sizes

- Type I error rates of $\delta_{\Delta R_I^2}$ and $\delta_{\Delta R_S^2}$ were acceptable in conditions where $\rho_1 = 0$ but were inflated, otherwise.
- When incorrect autoregressive model estimated, (i.e., when $\rho_1 = 0$) Type I error rates were inflated.
- When $\rho_1 \neq 0$, type I error rate inflated regardless of model assumed.
  - However, unclear about extent of autocorrelation in residuals when 4-parameter model used to explain pattern in AB designs.

Note on Autocorrelation Estimation

- With small *n*, commonly used estimates of ρ_1 are biased.
- A modified test statistic (Riviello & Beretvas, 2007) was found to preserve Type I error rates at the nominal level:

$$z^*_{HM} = \frac{r_1 + \frac{P}{n}}{\{1-[E(r_1)]^2\}\sqrt{\frac{(n-2)^2}{n^2(n-1)}}}$$

Research Question

- How well do the following parametric ESs correspond?
  - Standardized mean difference, δ,
  - Van den Noortgate & Onghena's, δ_{β_2} and δ_{β_3}
  - Beretvas & Chung's $\delta_{\Delta R_I^2}$ and $\delta_{\Delta R_S^2}$
- How autocorrelated are residuals of 4-parm model in multiple baseline data?

Method - Data

- Multiple baseline Communication Skills data from study designed to synthesize results of school-based interventions for students with ASD (Machalicek et al., 2007).
- 9 studies, 29 participants, 84 effect sizes.
- Effect sizes for outcomes for which a negative effect meant that the treatment worked were reverse-coded.

Method - Analyses

- Three kinds of effect sizes, (5 ESs) were calculated for <u>each baseline</u>.
- Each effect size was pooled together for <u>each participant</u> (univariate pooling for single δ; otherwise, GLS was used)
- Each effect size was pooled together for <u>each study</u>.
- Correlations calculated between five effect sizes ($\delta, \delta_{\beta_2}, \delta_{\beta_3}, \delta_{\Delta R_I^2}, \delta_{\Delta R_S^2}$).

List of Studies

Study	Reference
1	Johnson et al. (2004)
2	Buffington, et. al. (1998)
3	Hetzroni & Tannous (2004)
4	Schepis, Reid, & Sutton (1998)
5	Hetzroni and Shalem (2005)
6	Sigafoos et al. (2004)
7	Keen, Sigafoos, & Woodyatt (2001)
8	Smith and Camarata (1999)
9	Kravits et al. (2002)

Results

Pooled estimates of δ by Study

Study	β_2	ΔR^2_I	β_3	ΔR^2_S	δ
1	2.570	1.194	0.765	1.347	2.844
2	5.172	1.112	0.368	0.895	3.222
3	0.544	0.666	0.196	0.504	1.041
4	2.481	0.582	0.396	0.595	1.402
5	0.469	0.386	0.589	1.388	1.728
6	3.200	0.371	0.089	1.432	3.188
7	-0.133	-0.713	0.019	0.355	0.079
8	1.141	0.338	0.194	0.599	1.120
9	2.535	0.838	0.082	0.251	2.900

Results

Correlations between Studies' δs (k = 9)

	β_2	ΔR^2_I	δ	β_3	ΔR^2_S
β_2	1				
ΔR^2_I	.486	1			
δ	.841	.399	1		
β_3	.166	.370	.275	1	
ΔR^2_S	.265	-.048	.506	.616	1

Results

Correlations between Participants' δs (k = 29)

	β_2	ΔR^2_I	δ	β_3	ΔR^2_S
β_2	1				
ΔR^2_I	.471	1			
δ	.778	.417	1		
β_3	.332	-.205	.288	1	
ΔR^2_S	.094	-.197	.216	.421	1

Results

Correlations between Baselines' δs (k = 84)

	β_2	ΔR^2_I	δ	β_3	ΔR^2_S
β_2	1				
ΔR^2_I	.136	1			
δ	.811	.152	1		
β_3	-.005	.013	.193	1	
ΔR^2_S	-.069	.077	.211	.753	1

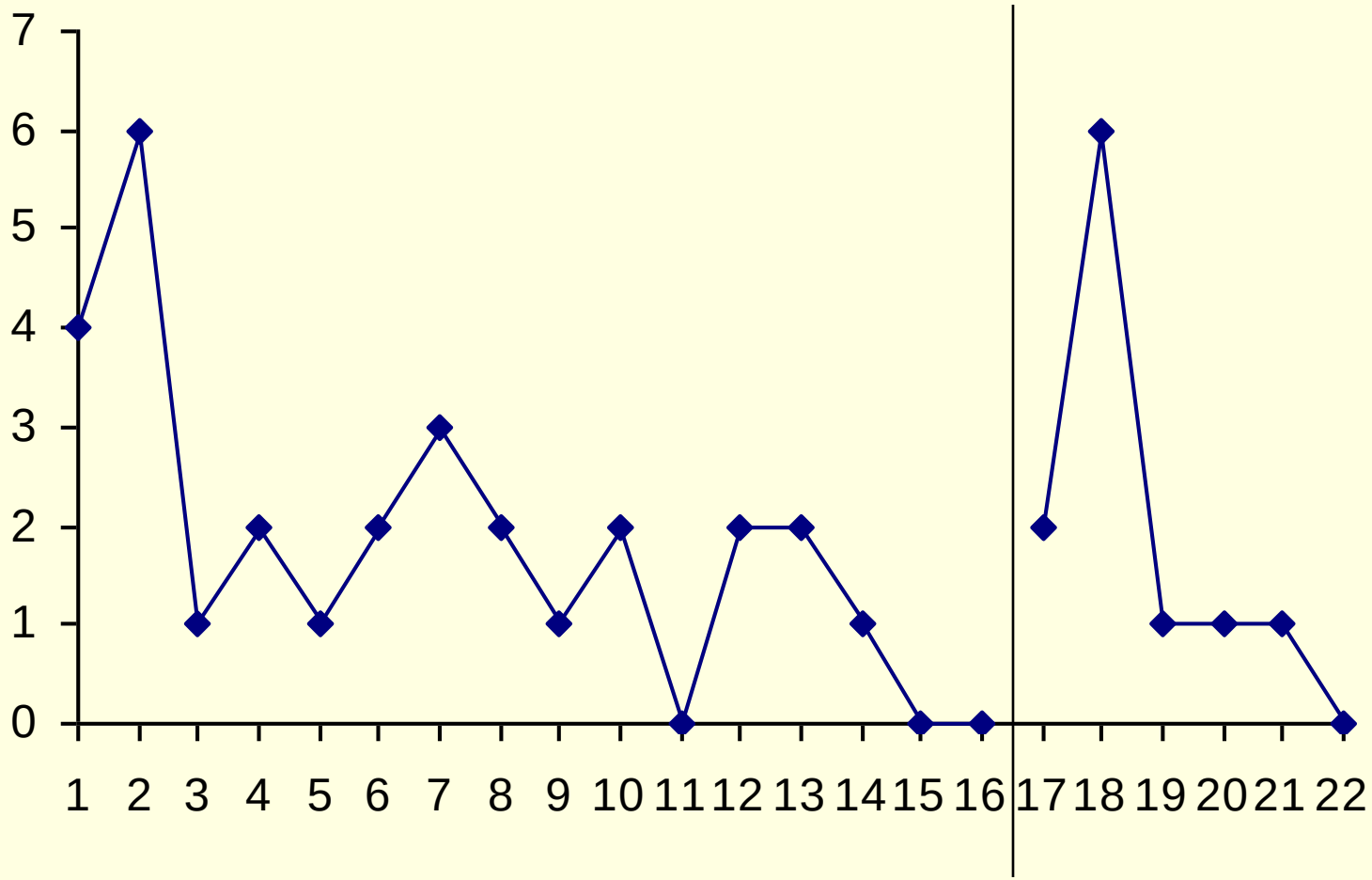
Results

Correlations between Baselines' δs (k = 84)

	β_2	ΔR^2_I	δ	β_3	ΔR^2_S	PND
β_2	1					
ΔR^2_I	.136	1				
δ	.811	.152	1			
β_3	-.005	.013	.193	1		
ΔR^2_S	-.069	.077	.211	.753	1	
PND	.497	-.029	.674	.145	.151	1

Biggest disparity in ranks (β_2:ΔR^2_I)

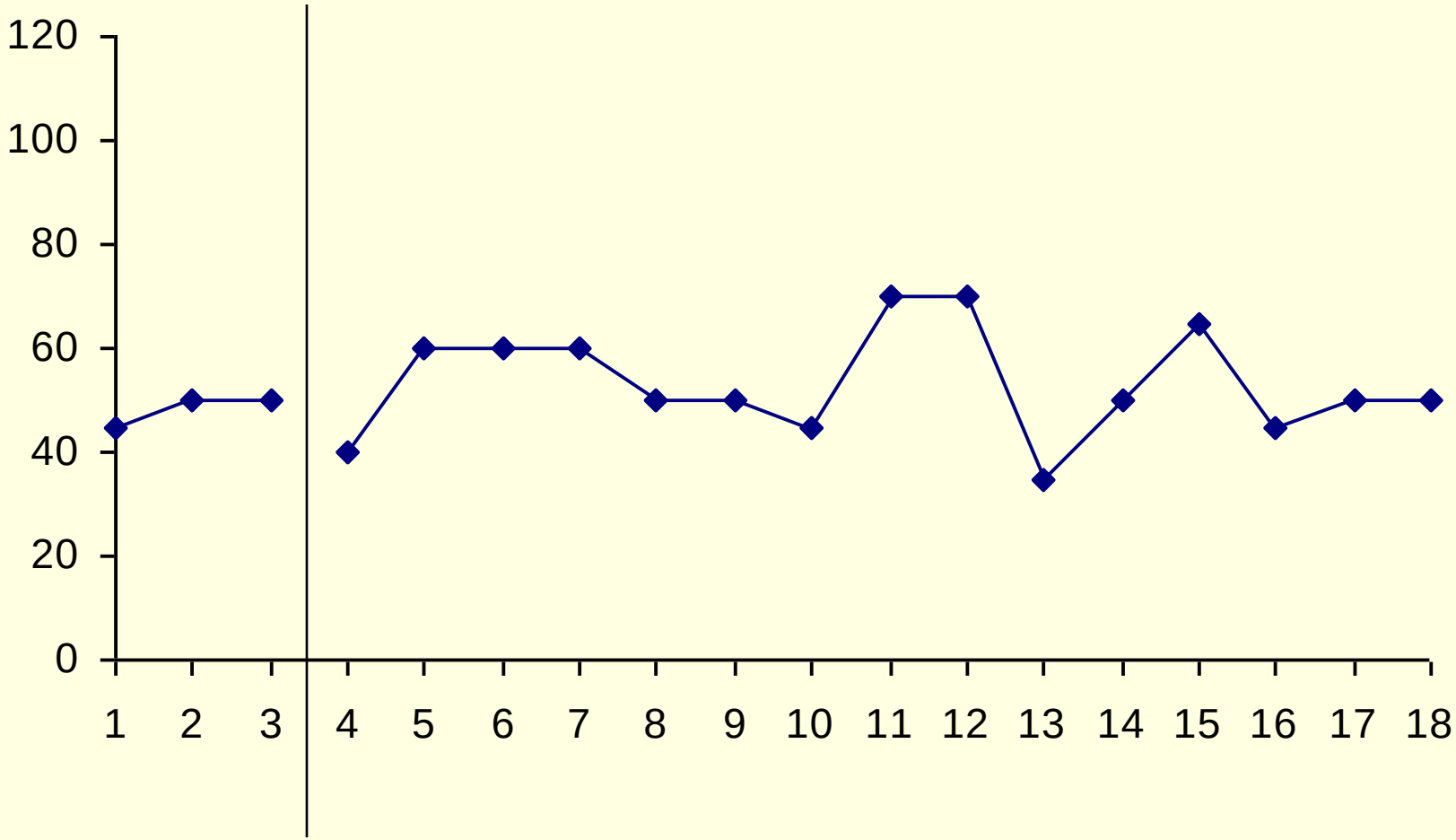
(-2.60: -1.32)



R^2_{Full}=0.411; R^2_{Rst}=0.135; B_2=3.62; B_0=3.63

Biggest disparity in ranks (β_3:ΔR^2_S)

(-0.26:2.88!)



R^2_{Rst}=0.043; R^2_{Full}=0.052; B_3=-2.679; B_1=2.5

Discussion

- Great disparities in effect size metrics' results.
  - Source of disparities requires further research.
- 64.29% of baseline data points had significantly autocorrelated residuals ($p < .05$) so the corresponding ΔR^2 indices would be biased.
  - Unclear ρ_1 would affect other ESs.
  - Estimation of auto-regressive equations needs to incorporate corrected estimate of ρ_1.
- Large degree of heterogeneity.
- Asympotically these ESs should work.
- Need 2 ESs to assess change in level *and* slope.
- More work needed!