BEING IDIOGRAPHIC WITH GROUP DATA: SEEING IS BELIEVING, WITHOUT *P*

Neville M Blampied
Department of Psychology
University of Canterbury
Neville.blampied@canterbury.ac.nz

New Zealand Applied Psychology Conference
Waikato University, 23-24 October, 2014

Abstract

Methodological reform in psychology calls for research to be more idiographic and less dependent on group statistical inference using null-hypothesis significance testing. Recommended alternatives include the use of *the new statistics*; attention to measurement error, reliable change, Effect Size and clinical/practical significance; more extensive use of graphs and visual analysis; and abandonment of over-reliance on *p* (e.g., Association for Psychological Science; Cumming; Klein; Task Force on Statistical Inference). This has major implications for applied psychology, given that the application of knowledge is almost always idiographic (i.e., to the single case) while applied research has overwhelmingly been done within the nomothetic, group statistical tradition. This paper describes a synthesis of these alternative approaches to data analysis that presents data on change over time visually for each participant, while presenting group statistics in a way consistent with the new statistics approach. This is done using Modified Brinley Plots, scatter-plots that compare individual scores at time 1 (normally pre-treatment) with scores at various times post-treatment. If the origin and axis scales are the same no or little change is shown by data points clustering on or about the 45^{o} diagonal line. Change associated with treatment (improvement or deterioration) is shown by shifts away from the diagonal. Interpretation is enhanced by the addition of clinical cut-offs, and indicators of means, variances, confidence intervals, measurement error, reliable change, and effect sizes. Both between-group and within-group data may be presented and analysed in this way.

Detecting therapeutic change – conventional approaches

Step 1

	Time 1			Time 2		
	1			1		
	2			2		
	3			3		
	.			.		
	.			.		
	.			.		
	.			.		
	n			n		
		MEAN t1			MEAN t2	

The essence of an RCT

		TREATMENT GROUP				
	Time 1			Time 2		
	1			1		
	2			2		
	3			3		
	.			.		
	.			.		
	.			.		
	.			.		
	n			n		
		MEAN t1	≠		MEAN t2	
		CONTROL GROUP				
	Time 1			Time 2		
	1			1		
	2			2		
	3			3		
	.			.		
	.			.		
	.			.		
	.			.		
	n			n		
		MEAN t1	=		MEAN t2	
	Treatment effect = significant Group x Time Interaction					

Problem

Excellent INTERNAL validity

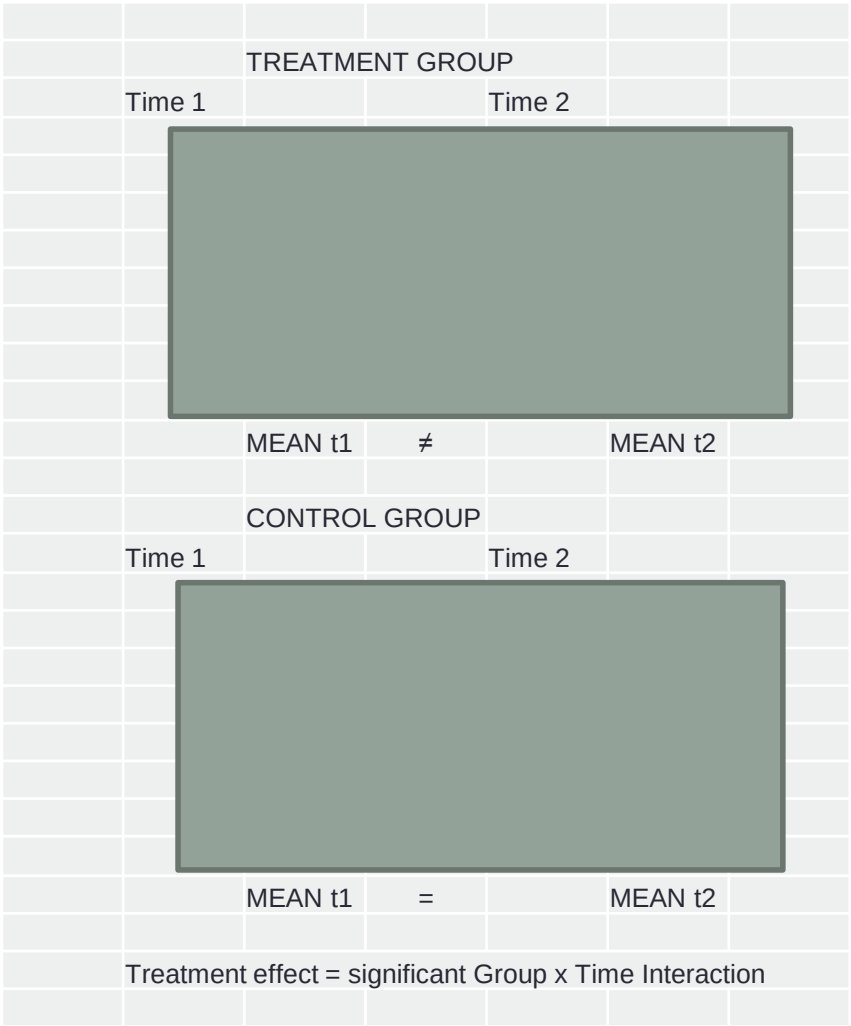
Poor EXTERNAL validity

Excessive INDUCTIVE GENERALIZATION

… we … investigate 150 [undergraduates]… for 20 minutes of their lifetimes, and think of them as interchangeable physical objects that do not change over time – which allows us to present our significant result as if it were about all mankind [sic] and all time.

Gigerenzer (1987). Probabilistic thinking and the fight against subjectivity. In *The probabilistic revolution.* Vol 2. p 13 - 33.

Individuals become invisible



Throughout its history as a science, psychology has been plagued by a double standard in its treatment of the individual …

In psychological discourse (both scientific and applied) the individual … is constantly given high relevance. In contrast, the individual case is usually forgotten in the practice of psychological research because it is replaced by samples of subjects ***that are assumed to represent some general population****.*

(Valsiner, 1996, p2)

RCTs & the *uniformity myth*

(Kiesler, 1966)!

To find out what people do in general, we must first discover what each person does in particular, then determine what, if anything, these particulars have in common. … the former [Nomothetic laws] can be discovered only after we find the latter [ideographic laws].

(Thorngate, 1996, pp 75-76)

… by attempting to describe only the average, one runs the risk of describing nobody in particular.

(Molden & Dweck (2006; p192-203)

Application is always to the single case

(Allport, 1942)

Statistical group outcome reports convey very little about what types of individual change are typical.

(Sobell, et al, 1995; p 658 – 59)

Information regarding within-treatment variability of outcome is of the utmost importance to clinicians.

(Jacobson & Truax; 1991, p 12)

… there is a strong perception that problems exist in generalizing a nomothetic result to an idiographic situation.

(Barlow & Nock, 2009; p 20)

So - Can we be more idiographic in applied research?

My trick ….

- Use visual analysis
- Show data for all participants
- Show change directly

And

- Show group/phase means
- Show variance or Confidence Intervals (95% CIs)
- Show Effect Sizes (ES)
- Show CIs for ES
- Show Clinical cut-offs
- Show Reliable Change
- Show % with reliable change

But

- No mention of *p* values < 0.05!

Core element - scatterplot

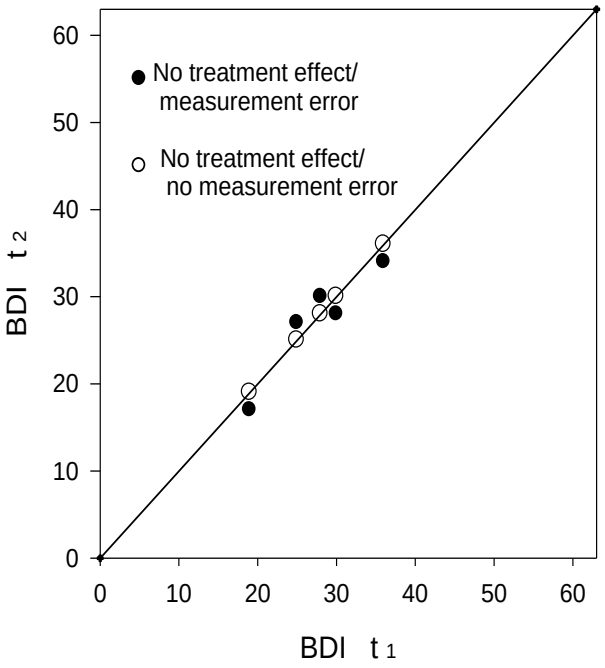
In a scatterplot

IF

- X & Y axes have same start
- X & Y axes have same scale
- Plot same individual's data on same measure @ t_1 (X) against t_2(Y)

Then

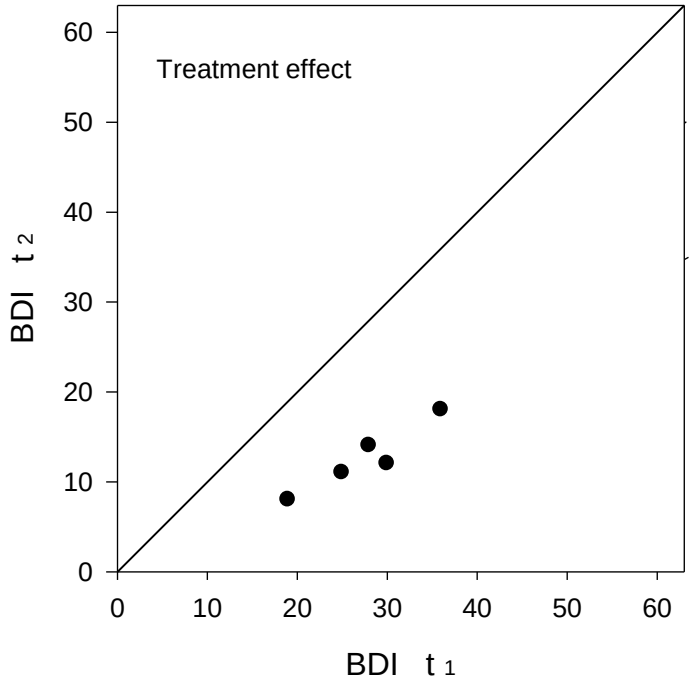
- Diagonal = no change
- X = Y
- Unsystematic variation = measurement error



Scatterplot …

BUT

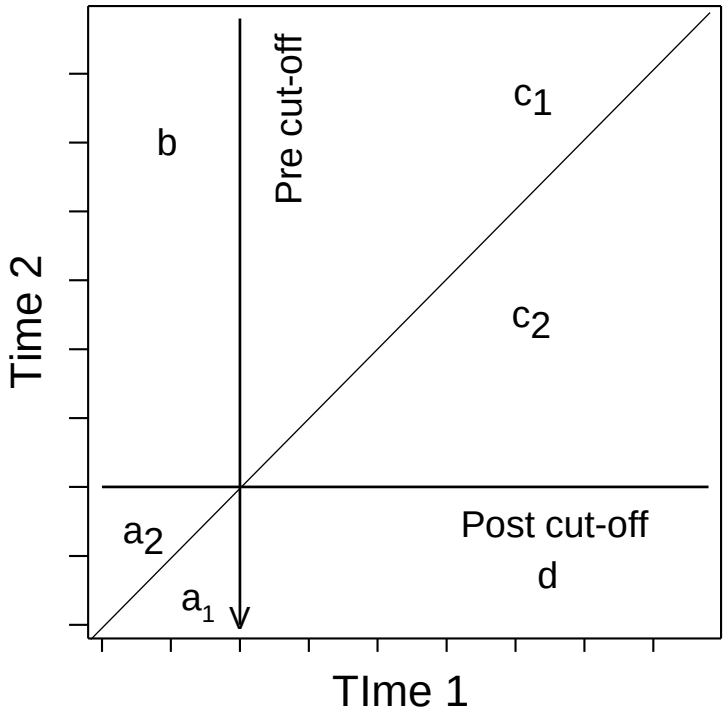
Systematic treatment effects show as deviations from diagonal



Scatterplot features

- Clinical cut-offs
- Arrow indicates improvement

Interpretation		
	t1	t2
a1	non clinical	non-clinical
a2	non-clinical	non-clinical
b	non-clinical	clinical
c1	clinical	clinical- worse
c2	clinical	clinical-better
d	clinical	non-clinical

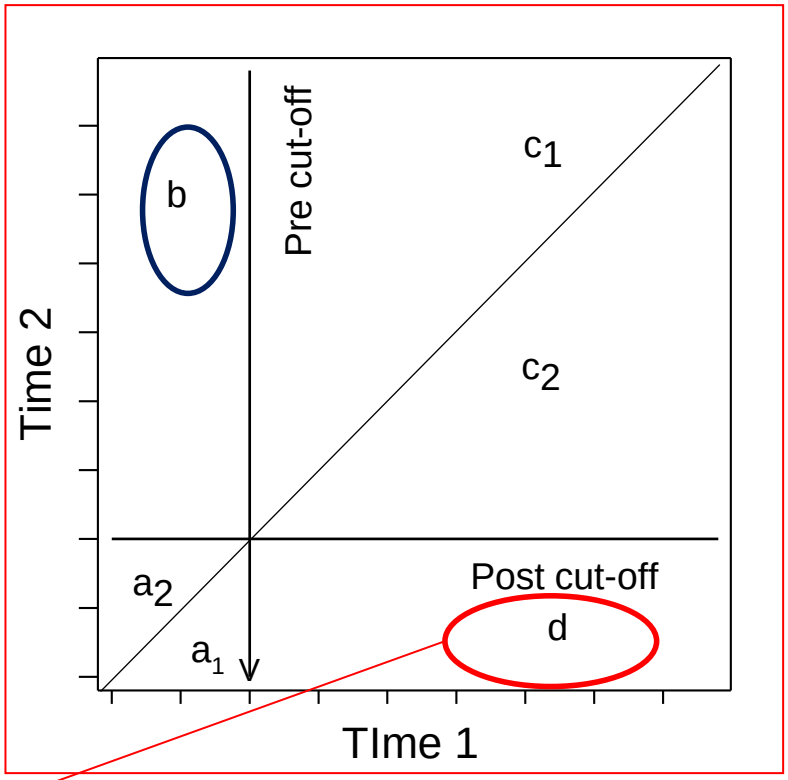


Scatterplot features

- Clinical cut-offs
- Arrow indicates improvement

Interpretation		
	t1	t2
a1	non clinical	non-clinical
a2	non-clinical	non-clinical
b	non-clinical	clinical
c1	clinical	clinical- worse
c2	clinical	clinical-better
d	clinical	non-clinical

Shame!

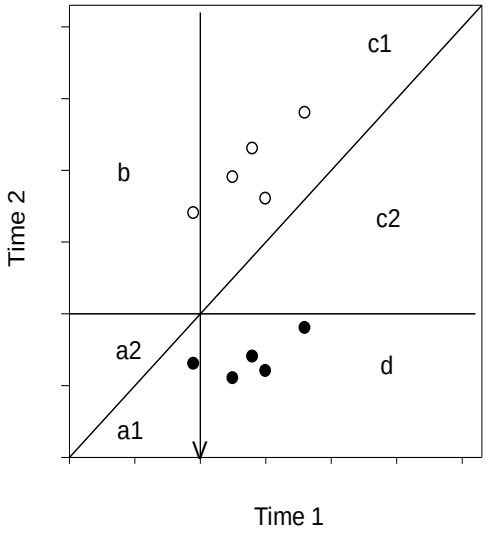


bingo!!

(after
Jacobson et al., 1984)

Scatterplot interpretation

You can classify individual outcomes



Further aids to interpretation – Reliable Change

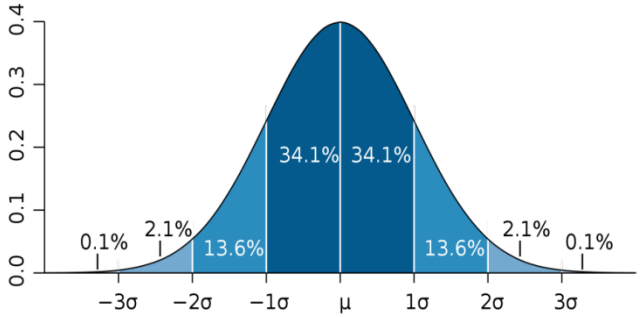
- Introduced by Jacobson et al (1984) & Jacobson & Truax (1991)
- Uses elementary psychometrics
- Based on Standard Error of Measurement

  (SE_M)

$$RC = x1 - x2/S_{DIFF}$$

S_{DIFF} = Change Score SE_M

- RC > 1.96 is unlikely to be error –
  - score has to lie in 5% tail of measurement error distribution



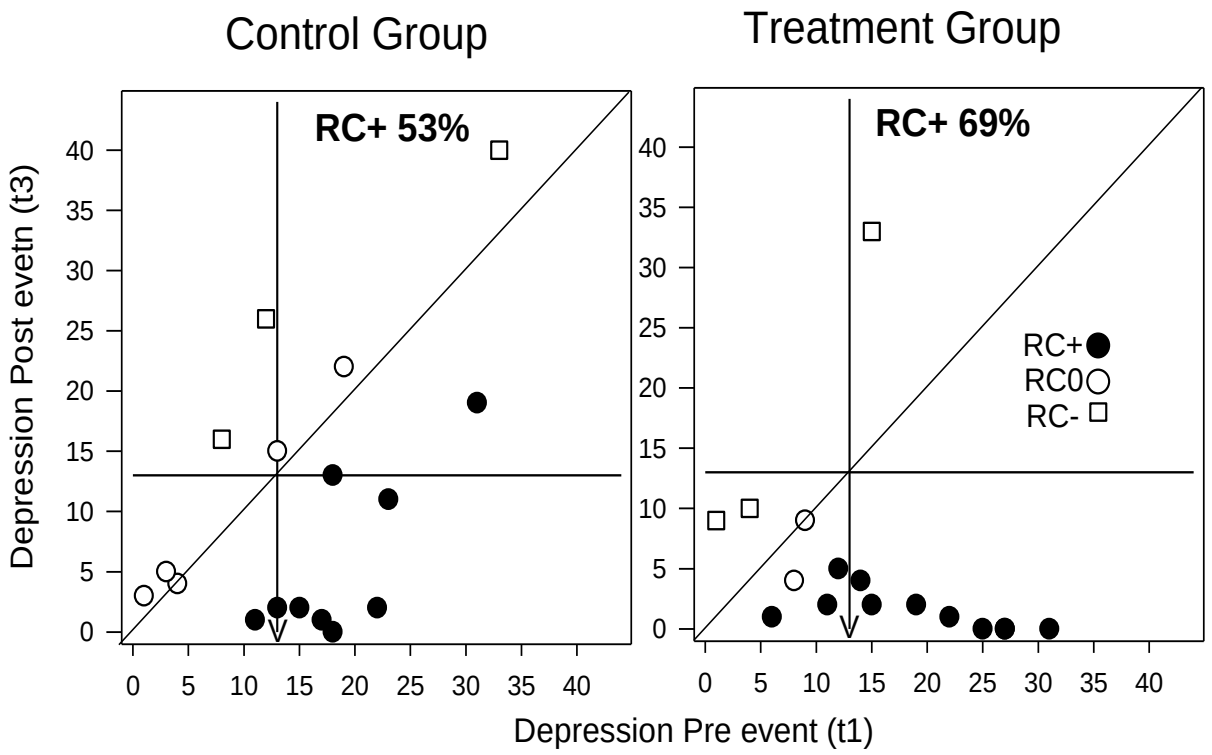
- Use RCI to classify each person
  - Reliably improved +
  - Reliably deteriorated -
  - Indeterminate 0

Example of modified Brinley Plot with RCI

Classifies each person RC+, RC-, or RC0

Shows % of group showing positive Reliable Change (RC+)

From Rucklidge & Blampied (2011)



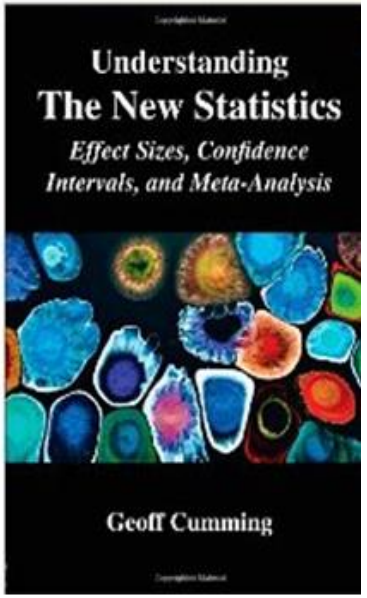
The new statistics

- Estimation }
- Precision – }*
  - Confidence intervals
- Effect sizes
- Meta-analysis
  - (best evidence synthesis)
- Replication

*Both lead to concern for

Measurement: validity/reliability/error

- Does not use NHST or p <?

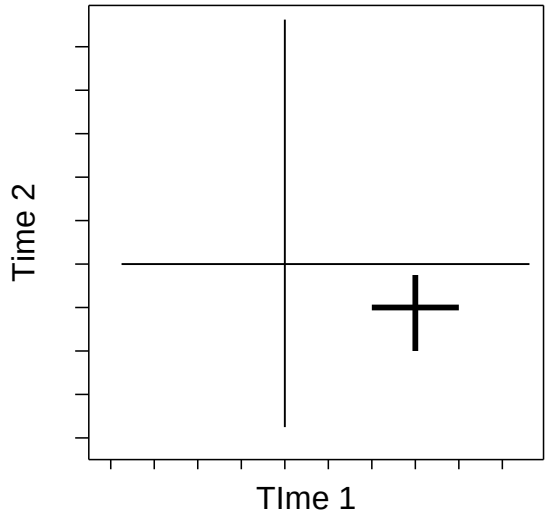


. . . *friends do not let friends compute p* [quoted in Klein, 2013).

I conclude from the arguments and evidence I have reviewed that best research practice is not to use NHST at all [Cumming, 2012]

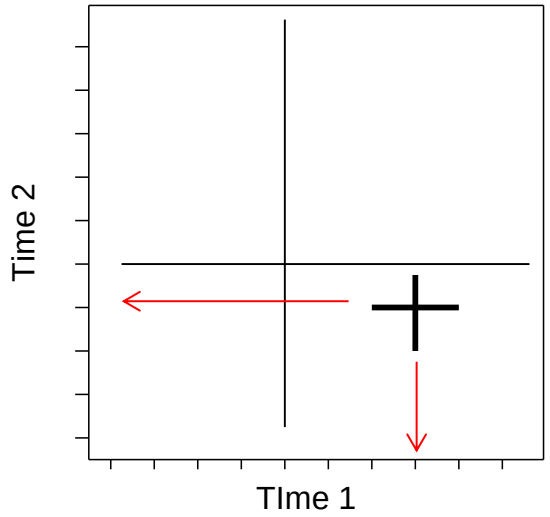
Display means and 95% CIs

- Intersection of bold lines = t_1 & t_2 means
- Length of line about middle = +/- 95% CI



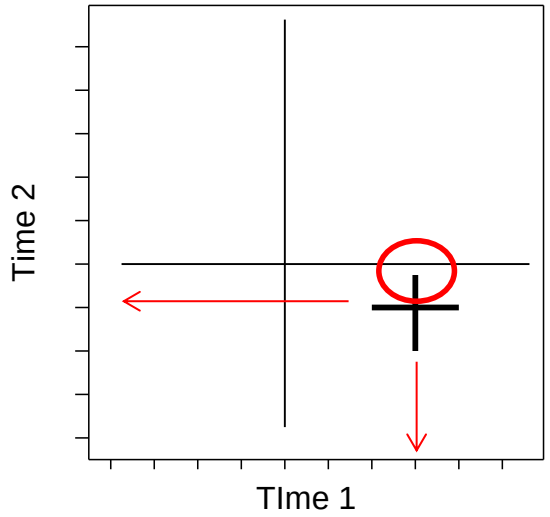
Display means and 95% CIs

- Intersection of bold lines = t_1 & t_2 means
- Length of line about middle = +/- 95% CI



Display means and 95% CIs

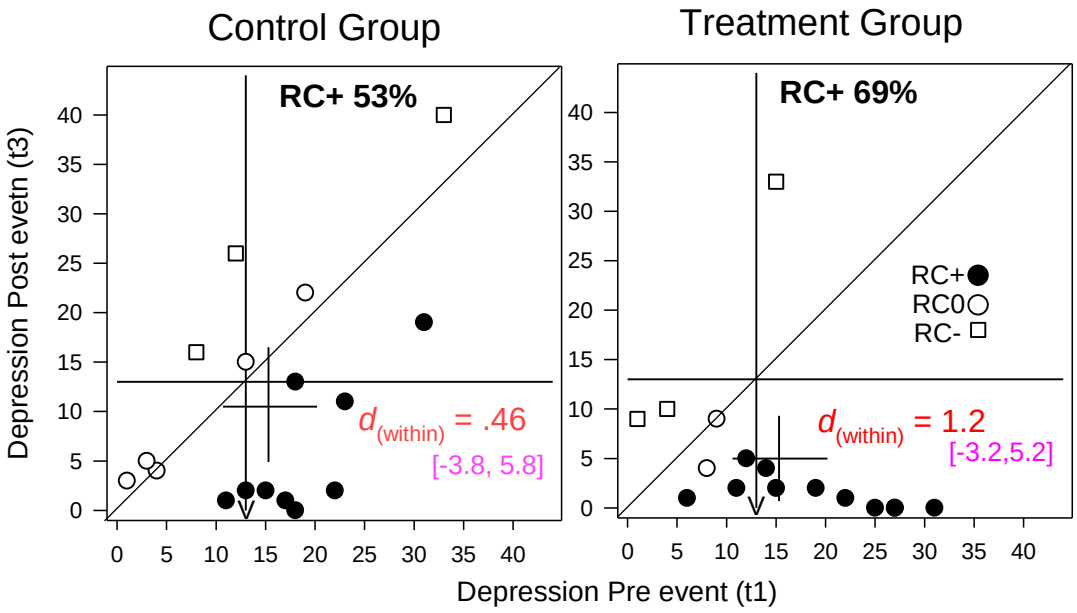
- Intersection of bold lines = t_1 & t_2 means
- Length of line about middle = +/- 95% CI
- And, CI does not cut t_2 clinical cut-off



Adding Effect Size information

Shows Cohen's *d (within)* ES

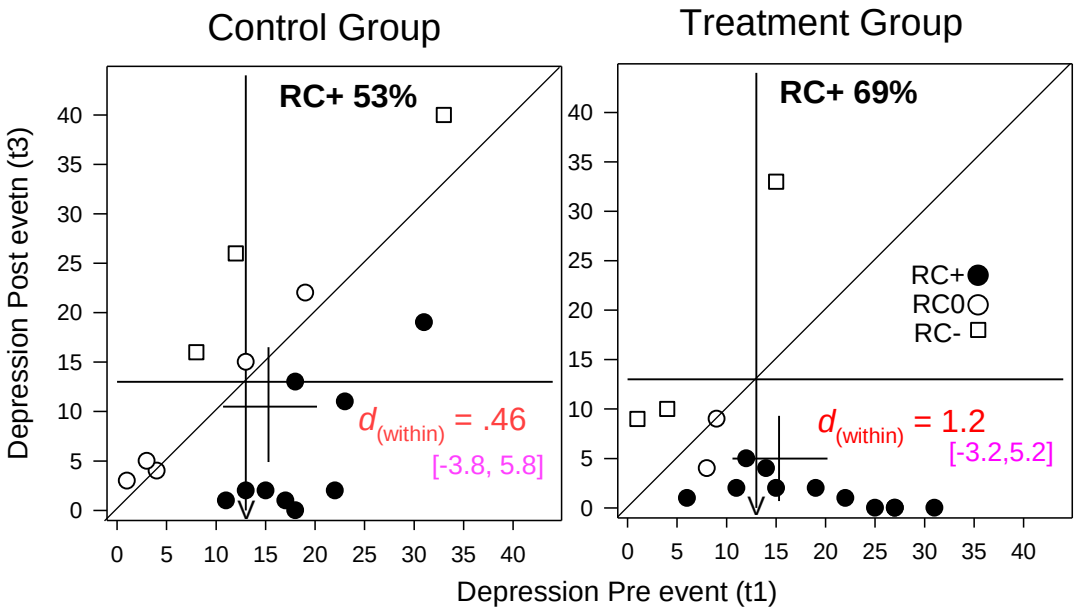
Shows 95% CI for *d*



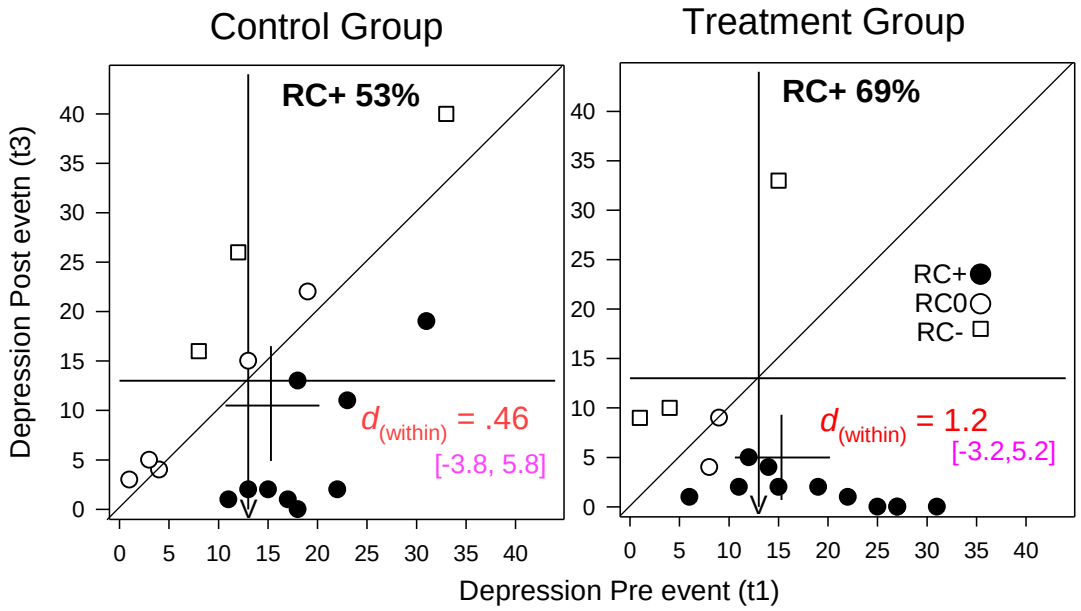
Putting it all together

In the ideal case we want to see a large ES with high precision of estimation. Additionally, it is useful to integrate information from the RC since for any given value of an ES, one that is based on a large proportion of participants demonstrating reliable change is more credible than one where few do.

(Blampied, 2014)



Seeing is believing, without *p*

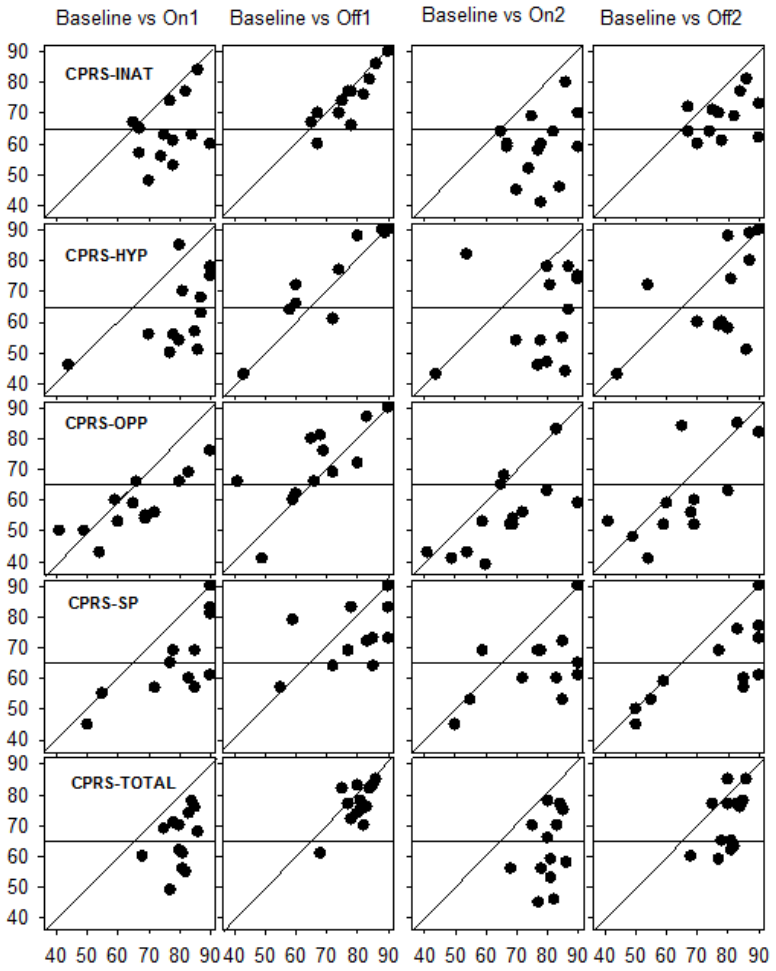


- Visual analysis √
- Groups√
  - Individuals √
  - Means √
    - 95% Confidence intervals √
  - Reliable Change √
  - Effect size √
    - 95% CI on ES √
- % with reliable change √
- Clinical significance√

& plots efficiently display a lot of data

- 15 participants
- 5 study phases
- 5 Measures

[from Gordon, 2014].



So - Can we be more idiographic in applied research?

YES

And, modified Brinley Plots form the basis for generating a whole new set of single case research designs, but with groups rather than single cases going through treatment!

References

For background, see

Blampied, N.M. (1999) A legacy neglected: Restating the case for single-case research in cognitive-behaviour therapy. *Behaviour Change, 16,* 89 - 104.

Blampied, N.M (1999). Murray Sidman – An appreciative introduction. *Behaviour Change, 16,* 75 - 78.

Blampied, N.M. (2000). Single-case research designs - A neglected alternative*. American Psychologist, 55,* 960.

Blampied, N.M. (2001). The third way - Single-case research, training and practice in clinical psychology. *Australian Psychologist.* 36, 157 – 163.

Blampied, N.M. (2013). Single-case research and the scientist-practitioner ideal in applied psychology. In G. Madden, et al. *Handbook of Behavior Analysis*, Vol 1. Washington, DC: APA.

Blampied, N.M. (2014). Using modified Brinley plots to analyse behaviour change in individuals within groups. Under review.

RC – what you need to know to compute

Info about the measure

- s = SD of reference data-set
- r_{xx} = Test-retest reliability of measure (Chronbach's alpha)

Used to compute

1. SE_M
2. S_{DIFF}

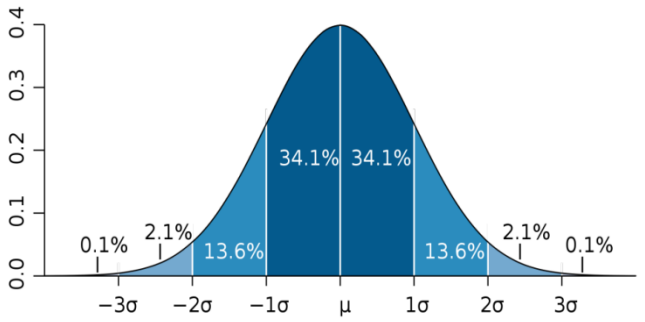
Both are a form of Standard Deviation

S_{DIFF} is S_{EM} of the Error Distribution of the Difference Scores

Info about measurement error distribution

- Is a Normal distribution
- SE_M is the Standard Deviation of the error distribution
- 95% of errors lie within

  +/- 1.96 SE_M [1.96~2SDs]



RC computation

Steps & formulae

1. Compute Standard Error of Measurement

$SE_M = s\sqrt{1-r_{xx}}$

2. Compute S_{DIFF}

$S_{DIFF} = \sqrt{2(SE_M^2)}$

3. Compute the difference score for each individual

Diff = x1-x2

4. Compute RC = x1 – x2/S_{DIFF}

Example

1. If s = 7.5

Test-retest alpha, r = .80

$SE_M = 7.5\sqrt{1-.8} = 3.35$

2. $S_{DIFF} = \sqrt{2(3.35*3.35)} = 4.74$

3. So if

x1 = 47.75

x2 = 32.5 Diff = 15

4. RC= 15/4.74 = 3.16

3.16 > 1.96, so difference that large not likely due to measurement error – in 5% tail of error distribution