

Evidence fit for evidence- based practice: Implications for the curriculum & new ways of looking at data

Neville M Blampied,
University of Canterbury
Christchurch, New Zealand
ICOPE 06 Flagstaff AZ
August 2014

Neville.blampied@canterbury.ac.nz

Supported by the Erskine Fund
University of Canterbury



Challenges for 21st C Psychology

... in the final analysis, the progress of ... psychology, as of every other science, will be determined by the value and account of its contributions to the advancement of the human race. (Witmer, 1907)

- The paradox of a commitment to evidence-based practice but a refusal to act on criticisms of the standard model of research
- The enduring mismatch between the needs of the scientist-practitioner and the standard model of research

We need evidence-based research practice
AND
evidence-based professional practice

21st C Challenges

We need evidence-based research practice

AND

evidence-based professional practice

AND

We need a methods curriculum that fits our graduates for
both roles

Evidence-based

- Researcher
- Practitioner

Current curriculum does not deliver this

The inference revolution in Psychology

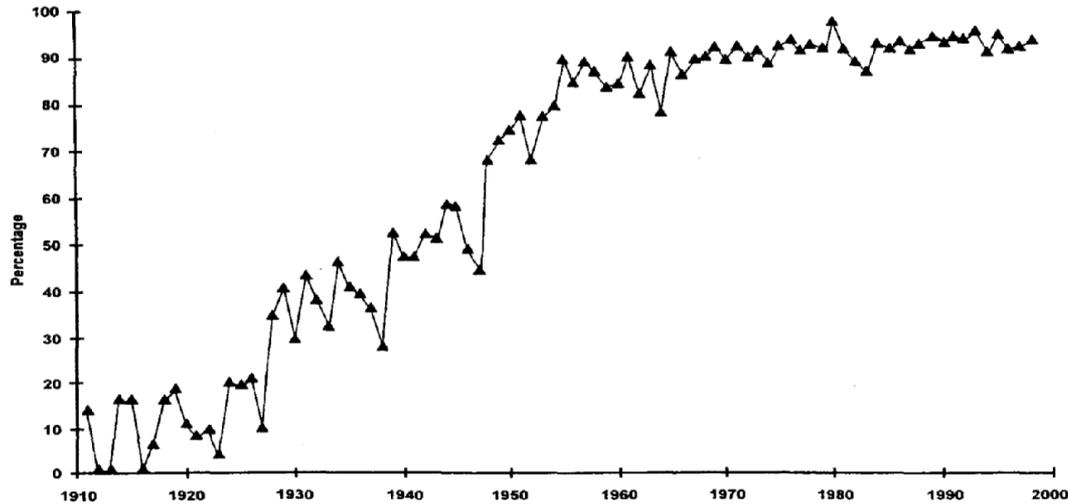


Figure 1.1. Percentage of articles reporting results of statistical tests in 12 journals of the American Psychological Association from 1911 to 1998. From "The Historical Growth of Statistical Significance Testing in Psychology—And Its Future Prospects," by R. Hubbard and P. A. Ryan, 2000, *Educational and Psychological Measurement*, 60, p. 665. Copyright 2001 by Sage Publications. Reprinted with permission.

Although their initial understanding of these techniques was often defective, psychologists embraced them with considerable enthusiasm. [Danziger, 1987].

Fisher's contributions to statistics

Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis. (Fisher, 1935).

- Devised (or improved on)
 - Control groups
 - Randomization
 - Factorial designs, & Analysis of Variance
 - Null Hypothesis Tests (NHST)

The methods ... of Fisher are beautifully suited and highly effective [for] ... the purposes for which they were developed – population genetics, agricultural research, and industrial quality control.

Johnston & Pennypacker (1980), p 91

Science according to Fisher

- Postulate a null hypothesis about population
 - Recruit as large a sample* as possible.
 - Randomly allocate participants to treatment conditions.
 - Aggregate individual data into group averages (sample statistics).
 - Draw inferences about the population from the sample statistics.
 - Use Null-hypothesis significance tests [NHST] to assess the data assuming H_0
 - reject H_0 if $p < \alpha$ (typically, α set at $p < 0.05$).
 - If H_0 rejected, accept H_A
 - Worry about Type 1 error (false rejection of H_0)
 - Don't worry about Type 2 error (false rejection of H_A)
- * From a known population

Criticisms of this *Standard Model*

Criticisms of the Standard Model

- Criticism has increased substantially from the 1990's
 - Harlow, L.L., et al. (1997) *What if there were no significance tests?*
 - APA Taskforce on Statistical Inference (1999)
 - Nickerson (2000). Null-hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5, 241-301
 - Ziliak & McCloskey (2008). *The cult of statistical significance.*

Criticism has focussed on

- use of NHST as the only decision rule
- Lack of power of much research

Criticisms of NHST

NHST has been called:-

- a *religion* (Cohen, 1990)
 - an *addiction* (Schmidt & Hunter, 1997)
 - an *enduring tyranny* (Loftus, 1991)
 - a *justification for betraying the evidence of the raw data* (Faverau, 1993)
 - *An empty and damaging ceremony* (McCloskey & Ziliack, 2008)
 - a ... *kind of sorcery, a form of statistical shamanism* (Lambdin, 2012)
 - *is among the most questionable things we do* (Trerweiler & Stricker, 1998)
 - *Is the most bone-headedly misguided procedure ever institutionalized in the rote training of science students* (Rozenbaum, 1997)
- one of the worst things that ever happened in the history of psychology.*** (Meehl, 1978).

Responses to criticism

Responses to criticism

There is no comprehensive, definitive refutation of the criticism of NHST!

Generally speaking, most defenders of NHST believe that it has been misinterpreted and badly used for decades.

[Balluerka, et al., 2005]

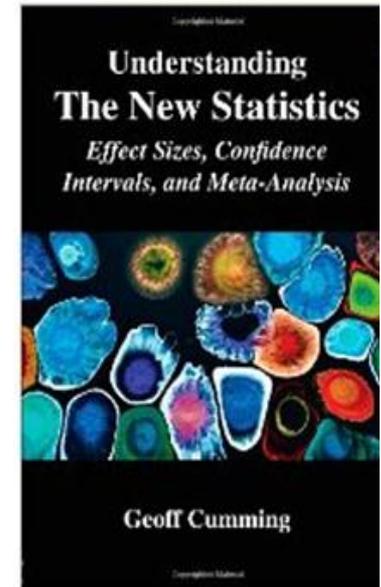
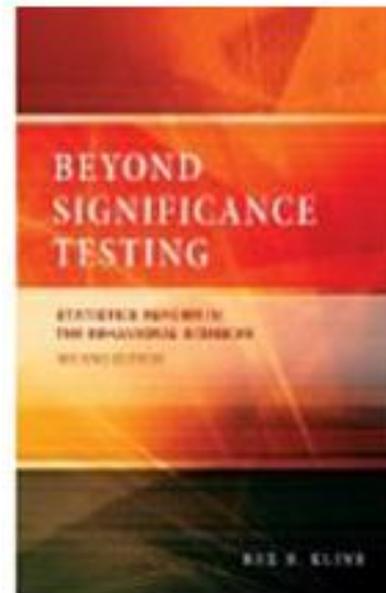
Conclusion

We need to reform the methods curriculum in psychology!

New ways of doing science (1)

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, 2013]

New ways of doing science (2)

The new statistics - Replication

November 2012; 7 (6)



» Index By Author

» Cover Photo

Special Section on
Replicability in Psychological
Science: A Crisis of
Confidence?

Special Section on Research
Practices

Replication draws for its inductive force on
the principle of unlikely successive coincidences

[Barlow et al., 1984]

*An ounce of replication is worth a ton of inferential
statistics.*

[Steiger, 1990]

*... we must finally rely, as have the older sciences,
on replication.*

[Cohen, 1994]

replication is at 'the heart of any science'...

[Schmidt, 2009]

New ways of doing science (3)

Re-discovering the individual

The double standard

Throughout its history as a science, psychology has been plagued by a double standard in its treatment of the individual subject... 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, 1986]

Idiographic science

Barlow & Nock (2009). Why can't we be more idiographic in our research? *Perspectives on Psychological Science*, 4, 19 – 21

Molenaar (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, 2, 201-218

The application of knowledge is always to the single case [Allport, 1942]

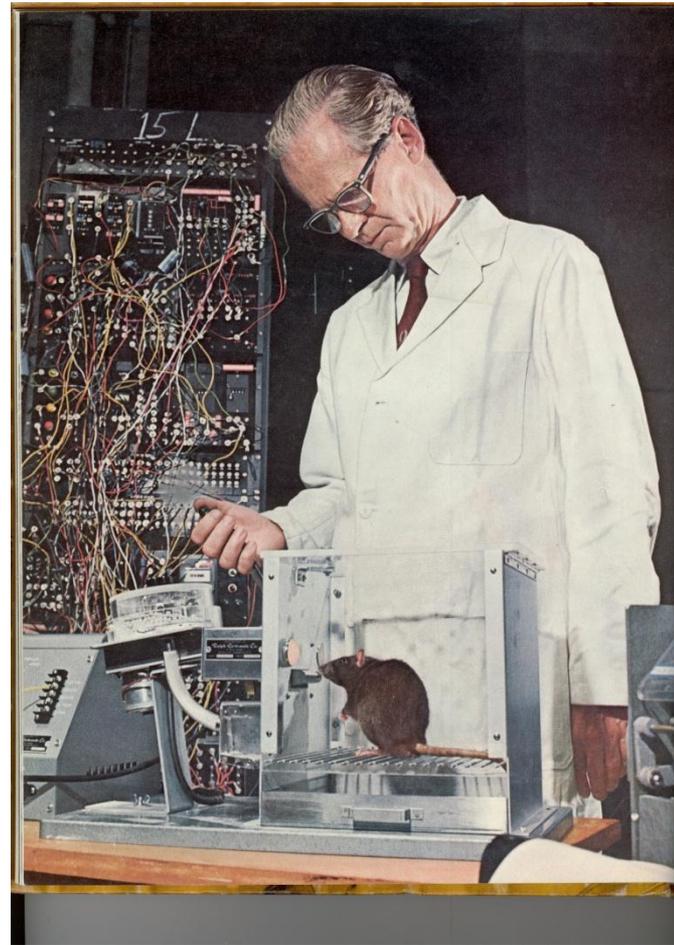
You can't treat a population [Grice, 2012]

New ways of doing science (4)

Single-case research – individuals + replication

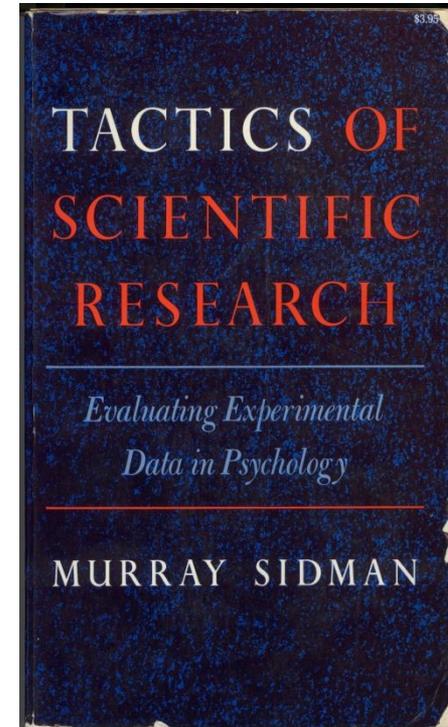
the kind of science proposed here belongs on the non-statistical side. ... In placing itself in that position it gains the advantage of a kind of prediction concerning the individual that is necessarily lacking in a statistical science. ... the student of behavior [cannot] predict what a single organism will do if his laws apply only to groups. Individual prediction is of tremendous importance, so long as the individual is to be treated scientifically.

[Skinner, 1938]



Single-case research ...

The soundest empirical test of the reliability of data is provided by replication ... intersubject replication is a more powerful tool than intergroup replication. ... With intersubject replication ... each additional experiment increases the representativeness of the findings. Indeed, replication of an experiment with two subjects establishes greater generality ... than does replication with two groups of subjects whose individual data have been combined. [Sidman, 1960]



Single-case research...

What it does not do:

- Rejects sampling theory.
- Does not average over participants.
- Does not use NHST-based inference.

What it does do:

- Maintains a commitment to quantification
- Uses rigorous experimental designs that permit causal inferences
- Uses Visual Analysis as its key data analysis technique
- Uses Replication as its key justificatory procedure

New ways of doing Science (4)

Visual Analysis

Recent advocacy -

Loftus, G.R. (1993). A picture is worth a thousand p values: On the irrelevance of hypothesis testing in the microcomputer age. *Behavior Research Methods, Instruments, and Computers*, 25, 250 – 256.

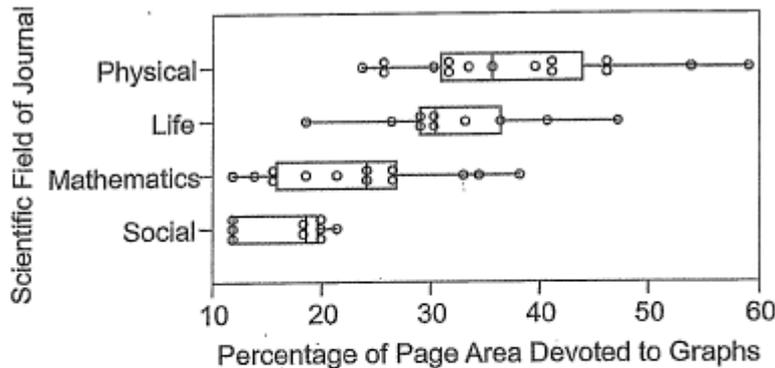
Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594 – 604.

*As soon as you have collected your data, before you compute any statistics, **look at your data.***

[Wilkinson & Task Force, 1999, emphasis in original]

Visual Analysis ...

you can learn a lot by looking. Yogi Berra



*Visual analysis ... can simultaneously detect curvilinear trends, repeating patterns or cycles in data, delayed or lagged responses following intervention onset, and within-phase changes in variability. In addition it can detect changes in mean level and trend slope across phases ... **That breath is not equalled by any other analytic technique.***

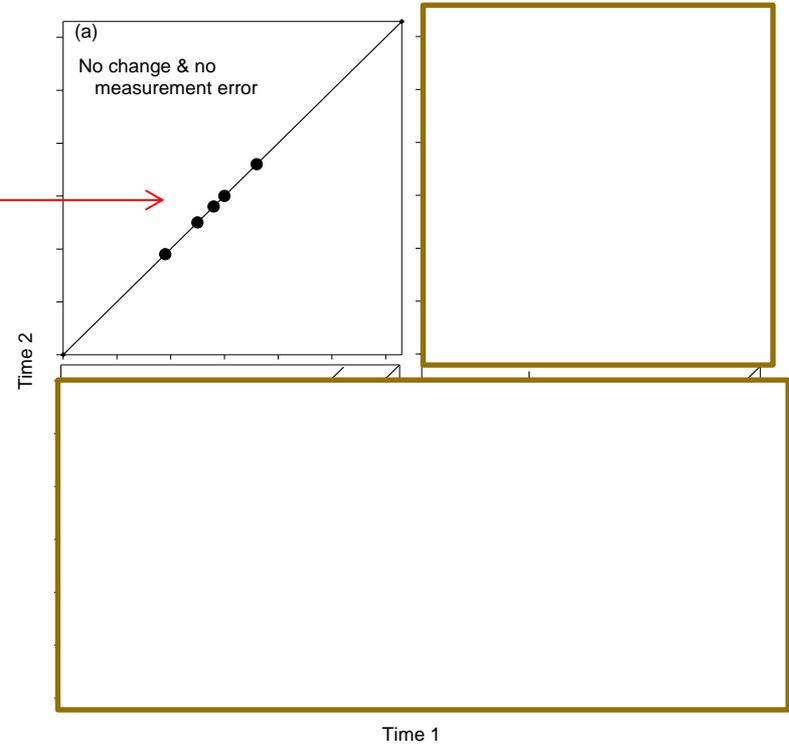
[Parker & Hagen-Burke, 2007]

[Wilkinson, I. (2012). Graphic displays of data. In H. Copper [Editor-in-chief]. *APA Handbook of Research Methods in Psychology: Vol 3* (pp73 – 100). Washington, DC:APA]

An example of innovation in visual analysis

Example

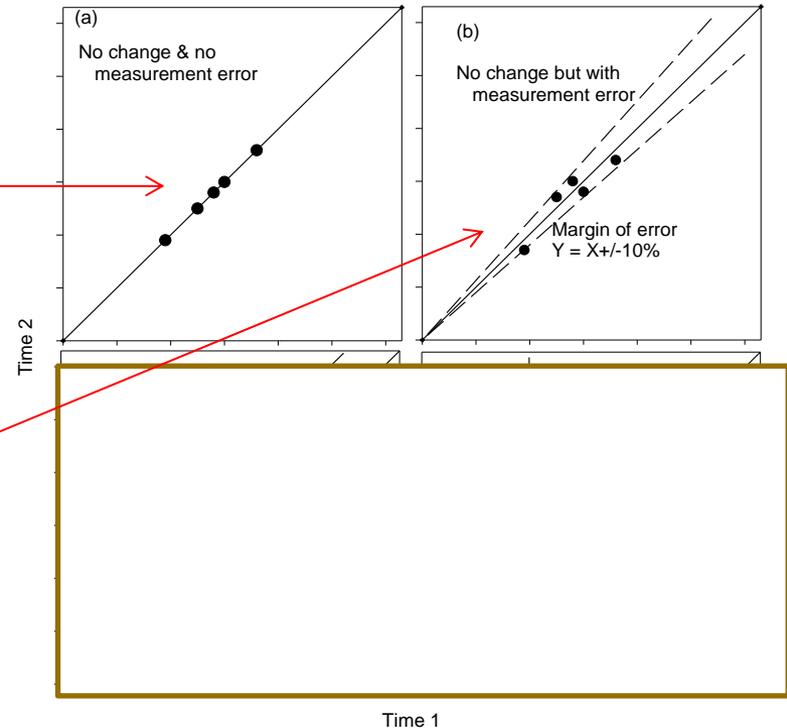
On a scatter-plot
The diagonal is
the line of no effect



Example

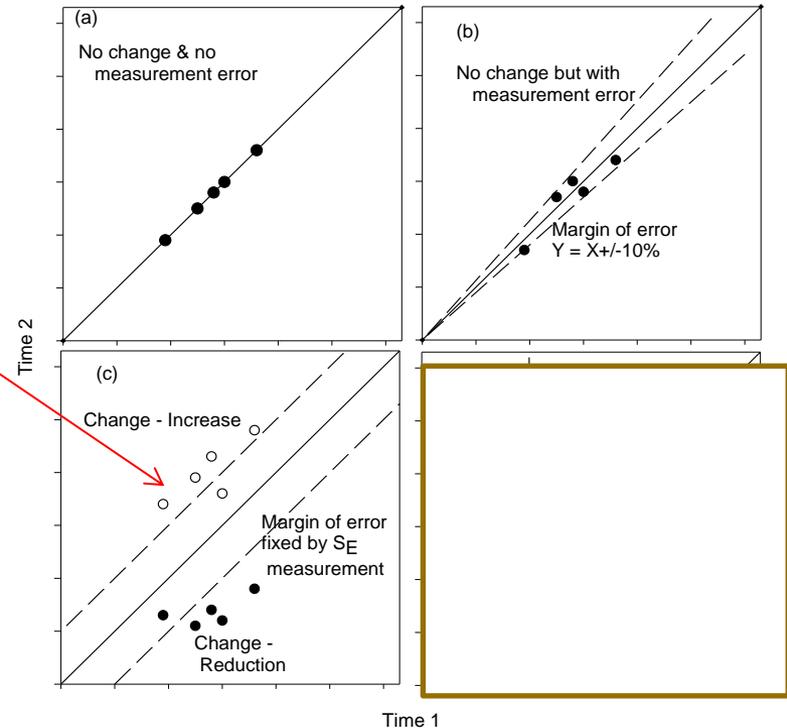
On a scatter-plot
The diagonal is
the line of no effect

Measurement error/
unsystematic variability
shows as variability round
the line



Example

On a scatter-plot
Systematic effects show as
displacements up
or down

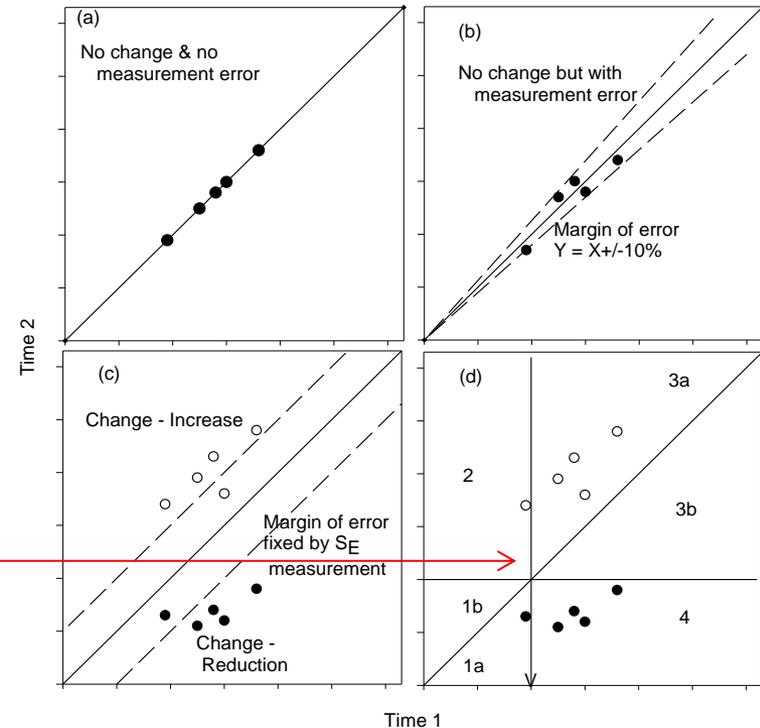


Example

On a scatter-plot
Systematic effects show as
displacements up
or down

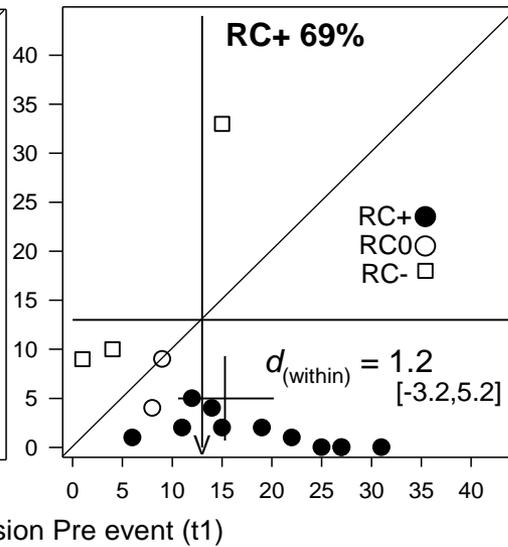
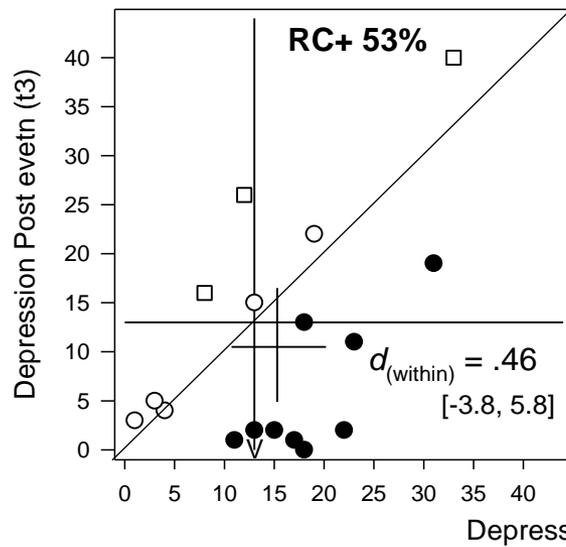
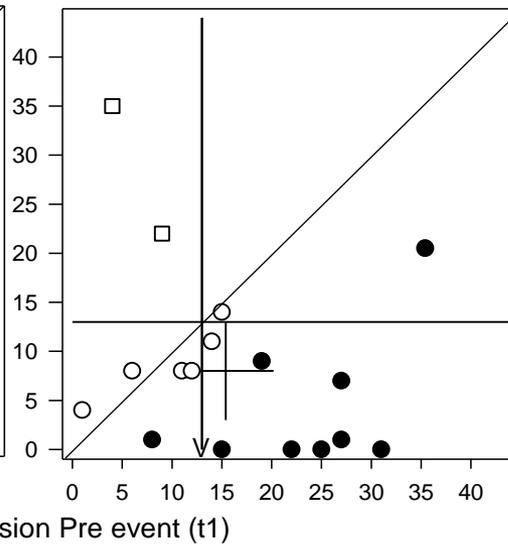
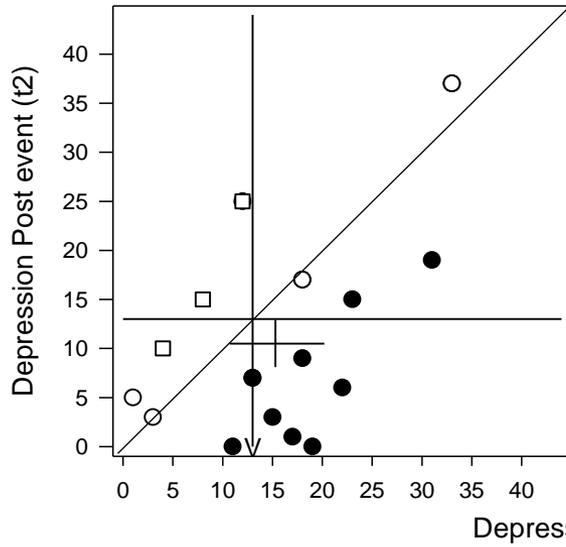
Clinical cut-offs can be
added

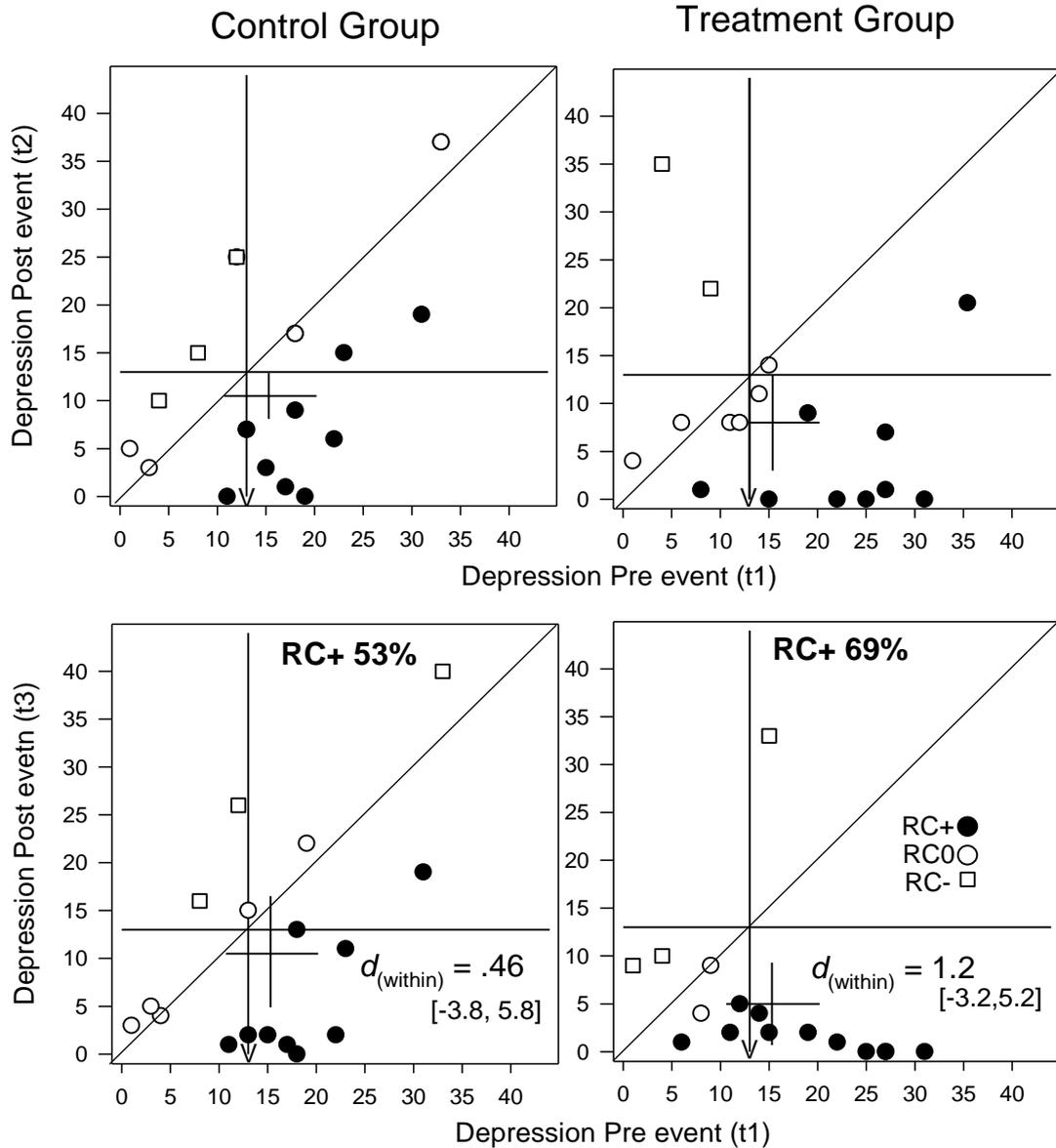
to help interpretation



Control Group

Treatment Group





- Visual analysis ✓
- Groups ✓
- Individuals ✓
- Means ✓
 - 95% Confidence intervals ✓
- Reliable Change ✓
- Effect size ✓
 - 95% CI on ES ✓
- % with reliable change ✓
- Clinical significance ✓
- [Blampied, 2014]

Implications

We need to take responsibility for changing the methods curriculum to achieve evidence-based research practitioners/graduates

- Personal
- Collaborative
- Institutional
- Organizational

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.

Abstract

Current proposals for methodological reform in psychology call for research to be more idiographic and less dependent on group averaging and null-hypothesis statistical testing. This paper reviews this development in the context of evidence-based practice and considers several major changes in the methodology curriculum required if reform is to occur and if evidence is to be fit for use in evidence-based practice. These changes include the teaching of the new statistics (estimation, confidence intervals, effect sizes, and meta-analysis); visual analysis techniques for the display of individual data in group contexts; replication; single-case research designs; and more sophisticated statistical tools (e.g., P-factor analysis). The presentation will focus on the construction and interpretation of modified Brinley plots, a technique for analysing change over time that is particularly suitable for idiographic analysis of outcome research in behavioural and cognitive-behavioral therapies. Modified Brinley plots are 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° diagonal line. Change associated with treatment (improvement or deterioration) is shown by shifts away from the diagonal. Interpretation is aided 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 with these plots, and they form the basis of new visual displays for group research using single-case research designs.