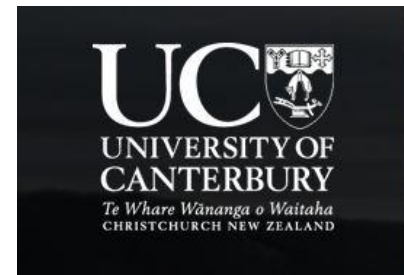# The Scientist-Practitioner Ideal in the 21ˢᵗ C

Responding to evidence that the evidence-base for practice is flawed.

Neville M Blampied

Department of Psychology

University of Canterbury

Department Seminar,

Department of Psychology

University of Auckland

21 August, 2013.

Development #1
The Scientist-Practitioner ideal

- 1907: *The methods of clinical psychology are … invoked wherever the status of an individual mind is determined by* *observation and experiment* (Witmer, p 251).

- 1947: A clinical psychologist has *applied and theoretical knowledge in three major areas:* *diagnosis, therapy, and research.* (Committee on Training in Clinical Psychology, p540)

- 1949: The Boulder Model – The Boulder Conference on Graduate Education in Clinical Psychology (Raimy, 1950): Training to include training in diagnosis, therapy, research methods, statistics, and a PhD dissertation.

The Scientist-practitioner ideal

… describes three primary and inter-related activities or roles. …[1] a consumer of new research findings … usually new assessment or treatment techniques …[2] an evaluator of his or her own interventions using empirical methods [to] increase accountability. [3] the practitioner as a researcher, producing new data from his or her own setting and reporting these data to the scientific community.
(Barlow, Hayes, & Nelson, 1984, p 4)

Development #2
The *Inference Revolution* in Psychology

The prophet of the inference revolution

Sir Ronald Aylmer Fisher

- *Statistical Methods for Research Workers* (1925).
- *The Design of Experiments* (1935)
- Devised (or improved on)
  - Control groups
  - Randomization
  - Factorial designs, &
  - Analysis of Variance & Null Hypothesis Tests (NHST)

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

Fisher's Influence on Psychology

- First use of ANOVA in psychology in 1935
  - 16 more examples by 1940

- Psychology ultimately adopted an "uneasy synthesis" of Fisher's statistics with those of J Neyman and E. Pearson (Gigerenzer, 1987).
  - Hypothesis testing
  - Power
  - Type 2 error
  - *Alpha* as decision criterion

  Although their initial understanding of these techniques was often defective, psychologists embraced them with considerable enthusiasm. (Danziger, 1987, p 35).

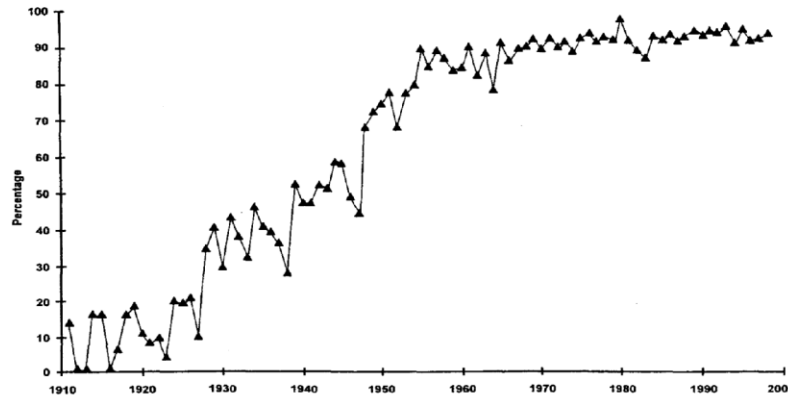# The inference revolution: 1911 - 1998

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.

By the early 1950's, half of the psychology departments in leading American universities offered courses on Fisherian methods. … By 1955, more than 80% of experimental articles in leading journals used inferential statistics to justify conclusions … Editors … made significance testing a requirement … and used the level of significance as a yardstick for evaluating quality.

(Gigerenzer,1991, p 255.)

From 1950 onwards:
The *Standard Model* of Research

- Postulate hypotheses about populations

- Recruit as large a sample as possible.

- Randomly allocate participants to treatment conditions.

- Aggregate individual data into group averages and generate sample statistics.

- Draw inferences about the population from the sample statistics.

- Use inferential statistics to separate out experimental effects from error.

- Use Null-hypothesis significance tests [NHST] to determine the scientific legitimacy of the results (typically, requiring $p < 0.05$)

A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give … significance

(Fisher, 1926, p 504).

The two developments converge

Science = Fisherian experimentation & statistical analysis

Therefore, science in Scientist-practitioner = Standard Model science

And, therefore, clinical/applied psychology research overwhelmingly done within Standard Model

With unfortunate consequences

With unfortunate consequences

Misuse of Statistical Tests in Three Decades of Psychotherapy Research

This article reviews the misuse of statistical tests in psychotherapy research studies published in the *Journal of' Consulting and Clinical Psychology'* in the years 1967-1968, 1977-1978, and 1987-1988.

We found that p values received an absurdly central position in many studies… often at the expense of descriptive statistics. …In general, null hypothesis tests have been misinterpreted and misapplied all along; researchers give an impression of playing the binary decisional game by using terms such as tests and significance level, but in actuality, they disregard its rules by overemphasizing and misinterpreting p values, using multiple pseudo-alpha levels, flirting with borderline p values, and inflating Type I error rates in the great majority of studies.

(Dar, Serlin, & Omer, 1994, p79).

And not much has changed since

Dar et al (1994) found that 30% of articles reported clinical significance …our coding in Period 4 [1993 – 2001] shows … only 40% of articles made any attempt to discuss clinical significance. This is serious. In a major journal dedicated to the research of psychotherapy and other interventions, clinical significance should be relevant to more than 40%...

(Fidler, Cumming,Thompson et al.,2005. Towards improved statistical reporting in the Journal of Consulting & Clinical Psychology. *JCCP, 73,* 136 -143.)

Development #3
Evidence-based practice

Evidence-based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients.

(Sackett, et al., BMJ, 312, 1996, p 72)

Evidence-based practice in psychology is the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences. … [It] starts with the patient and asks what research evidence … will assist the psychologist in achieving the best outcome … [and] articulates a decision-making process for integrating multiple streams of research evidence…into the intervention process.

(APA Presidential Taskforce on Evidence-based Practice, 2006, p273)

EPB is *a transdisciplinary, ideographic approach*

(Spring, 2007, p 611)

BUT

The standard model does not suit the needs of much applied research

This is not a recent discovery

…there is a growing disaffection from traditional experimental designs and statistical procedures which are held inappropriate to the subject matter… [particularly] in the area of therapeutic change…

(Bergin & Strupp, 1972. *Changing frontiers in the science of psychotherapy,* p 440).

Mismatch is intrinsic to the Standard Model

Hypotheses are about populations

Not about
- The sample
- The individual cases in the sample

Yet, applied psychology needs to apply knowledge at the individual level – EPP *starts with the patient/ is*

ideographic

But the individual disappears in our research

Psychology without individuals

The idea of eliminating the individual … was an innovation [in scientific method] unique to psychology.

… we … investigate 150 freshman from Michigan 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 and all time.

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

The "double standard" in Psychology

Throughout its history as a science, psychology has ben 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.

How do we get from the population to the individual?

Well – we can't, legitimately

So

- We infer from the sample to the population (legitimately – except that psychologists rarely specify the population they have sampled from)

- We attribute generic, prototypic, ideal status to the population average (a step of inductive generalization) – e.g., gender generalizations

…characteristics of that abstracted individual may easily become attributed to particular concrete individuals with whom psychologists … work

(Valsiner, 1986).

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

(Molden & Dweck, 2006, 192-203).

This is a major problem

The inductive inference from samples … to the abstract individual, and … deductively back to the multitude of concrete human beings, is guided by a number of implicit assumptions…. *that obscure insight into the science and hamper its applications*.

(Valsiner, 1986. *The individual subject and scientific psychology, p2).*

Development # 4
Criticisms of the Standard Model

- Criticism began early – in 1940's
- Criticism has increased substantially from 1990's
- Criticism has focussed on use of NHST ($p < .05$) as the only decision rule and the lack of power in much psychological research.

  - Harlow et al. (1997) *What if there were no significance tests?*
  - APA Taskforce on Statistical Inference (1999)

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)

- *the most bone-headedly misguided procedure ever institutionalized in the rote training of science students* (Rozenbaum, 1997)

- *Sorcery* (Lambdin, 2012)

Paul Meehl on NHST

Paul Meehl on NHST

*… Sir Ronald has befuddled us, mesmerized us, and led us down the primrose path. I believe that the almost universal reliance on merely refuting the null hypothesis as the standard method … is a terrible mistake, is basically unsound, poor scientific strategy, and **one of the worst things that every happened in the history of psychology.***

[Is] a potent but sterile intellectual rake who leaves in his merry path a long train of ravished maidens but no viable scientific offspring.

(Meehl (1978). p 817)

The latest broadside

Significance tests as sorcery: Science is empirical—significance tests are not

Since the 1930s, many of our top methodologists have argued that significance tests are not conducive to science. Bakan (1966) believed that "everyone knows this" and that we slavishly lean on the crutch of significance testing because, if we didn't, much of psychology would simply fall apart. If he was right, then significance testing is tantamount to psychology's "dirty little secret." This paper will revisit and summarize the arguments of those who have been trying to tell us—for more than 70 years—that p values are not empirical. If these arguments are sound, then the continuing popularity of significance tests in our peer-reviewed journals is at best embarrassing and at worst intellectually dishonest.

Lambdin, 2012, *Theory & Psychology, 22,* 67–90

False assumptions about *p*

- = probability of replication

- A p from a large-n study is better than small n

- = the confidence in our results

- Automates induction

- Permits inference from the population to our research hypothesis

- Estimates the veracity of our research hypothesis

- Permits inferences about the sample/individuals

- Measures the validity of inductions from results

- Gives the probability of Ho being true/false

- Gives the probability that H_A is true

- Is the probability the results happened by chance

(after Lambdin, 2012, p 73)

Psychology -Addicted to p

"I'm right there in the room, and no one even acknowledges me."

The New Yorker, 9/18/06
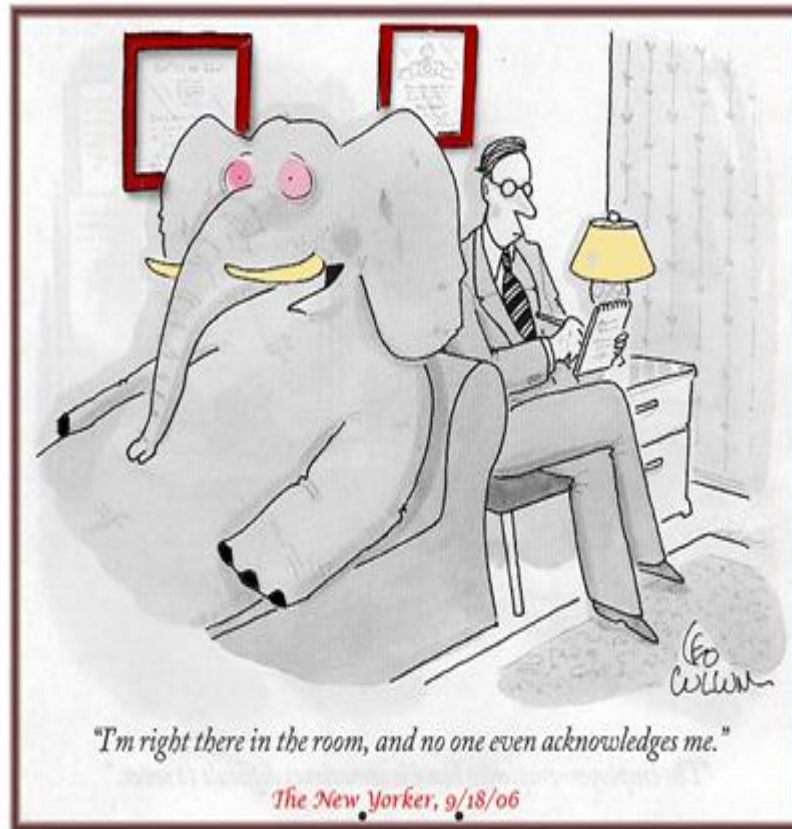
Alternatives?

The APA Taskforce on Statistical Inference

Alternatives?

Doubling one's thinking is likely to be much more productive than doubling one's sample size
(McClelland, 2000, p964)

The APA Taskforce on Statistical Inference

- Confidence intervals
- Effect sizes

See Geoff Cumming (2012). *Understanding the new statistics* [www.thenewstatistics.com]

Replication – coming to a journal near you?

November 2012; 7 (6)

Perspectives on
Psychological
SCIENCE

» Index By Author
» Cover Photo

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

Special Section on Research
Practices
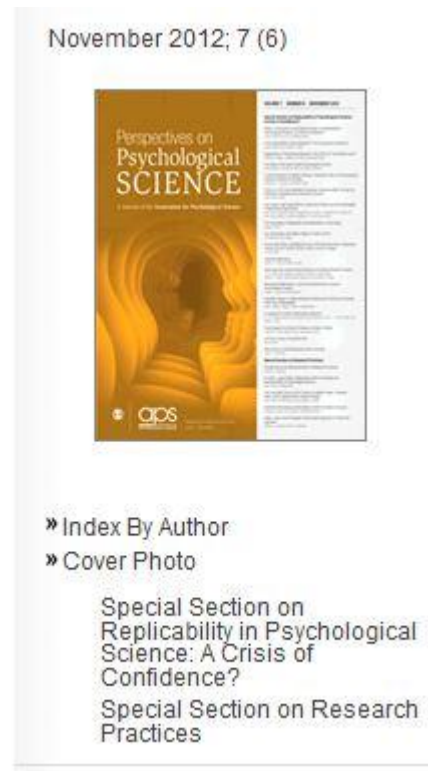
A science based on replication
What would it be like?

A science based on replication
What would it be like?
Single-case Research

Replication draws for its inductive force on

the principle of unlikely successive coincidences

(Barlow et al., 1984).

Single-case research - Origins

Individual prediction is of tremendous importance, so long as the individual is to be treated scientifically.

Skinner,1936. *The behaviour of organisms, p444).*
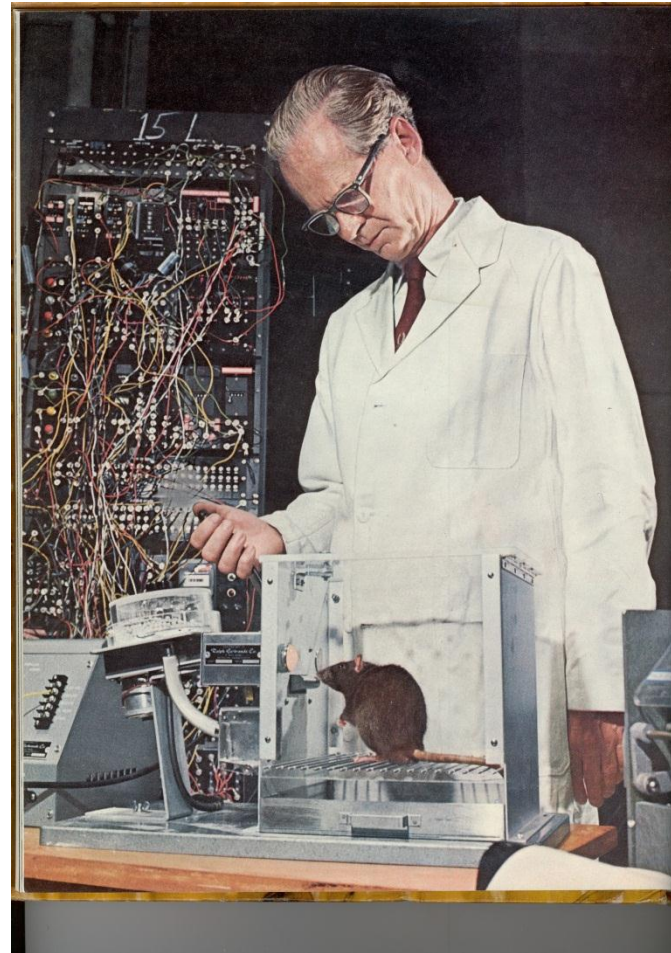
Formal exposition of Single-case designs Murray Sidman (1960)

In psychological experiments that employ large populations and group-data techniques, repetition is rare. … such repetition would help to establish the reliability of the central tendency, but would bear little, if any, relevance to the question of generality or representativeness with respect to individuals.

TACTICS OF SCIENTIFIC RESEARCH

Evaluating Experimental Data in Psychology

MURRAY SIDMAN

[Sidman, 1960, p74)
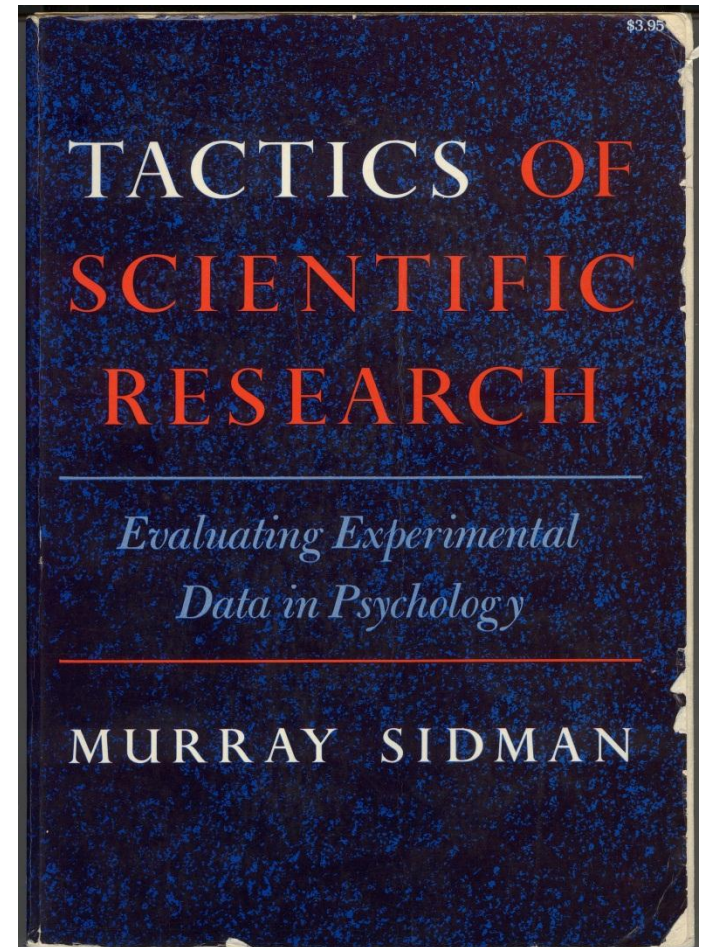
Single-case research designs in applied settings: Hersen & Barlow (1976)

Replication is at the heart of any science. In all sciences it serves … two purposes. First, to establish the reliability of previous findings and, second, to determine the generality of these findings under differing conditions.

(Hersen & Barlow, 1976, p317)
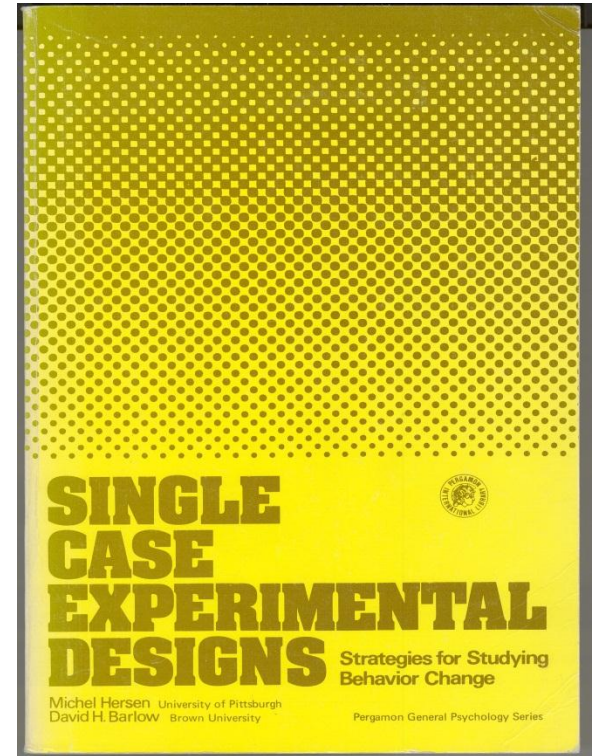
Replication – thinking about the individual

Replication – thinking about the individual

- The core unit for the visual analysis is a scatter-plot of repeated measures data

  (after Brinley, 1965)

- Typically, baseline/t_1 on the X-axis and

- Data from another phase (t_2) on the Y-axis
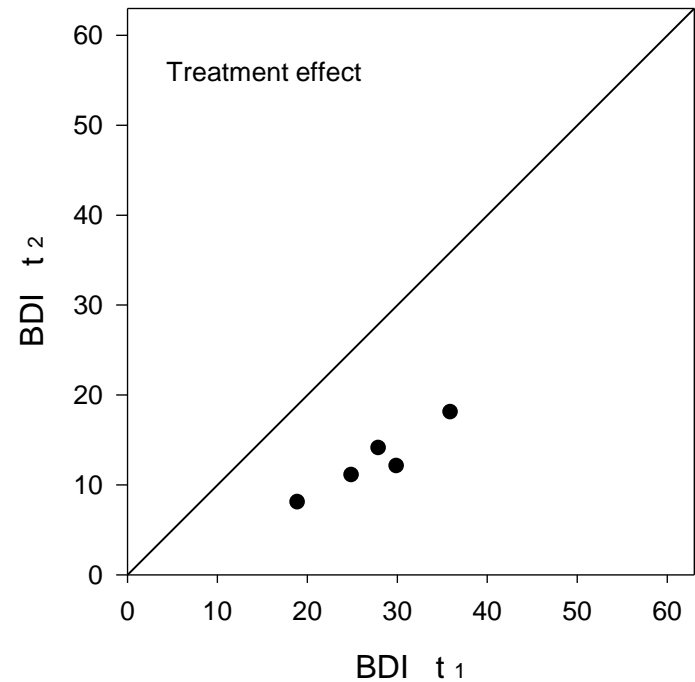
An alternative visual analysis

- The diagonal is the line of no effect/no change

- Filled points indicate no treatment effect but "measurement error"
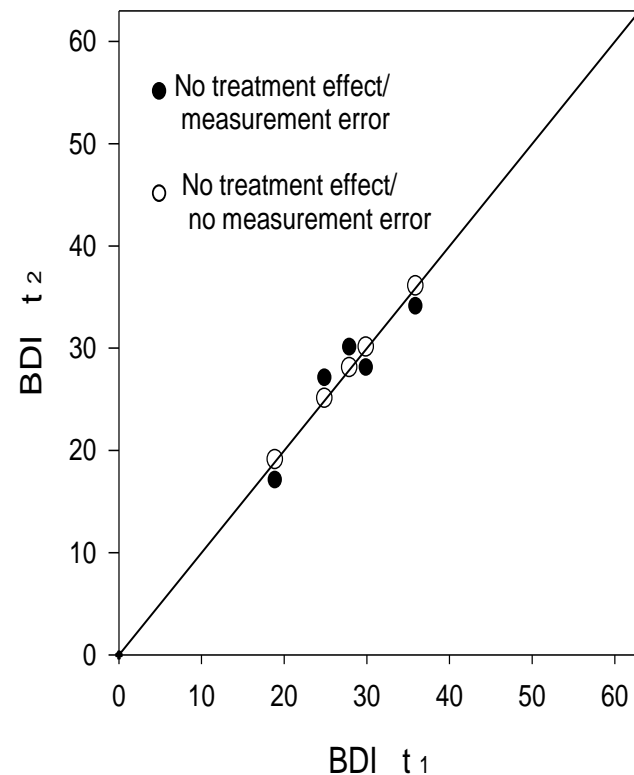
Using the scatter plot

- If there is an intervention effect between t_1 and t_2, then the data points will shift systematically away from the diagonal
  - Here, down-shift indicates improvement
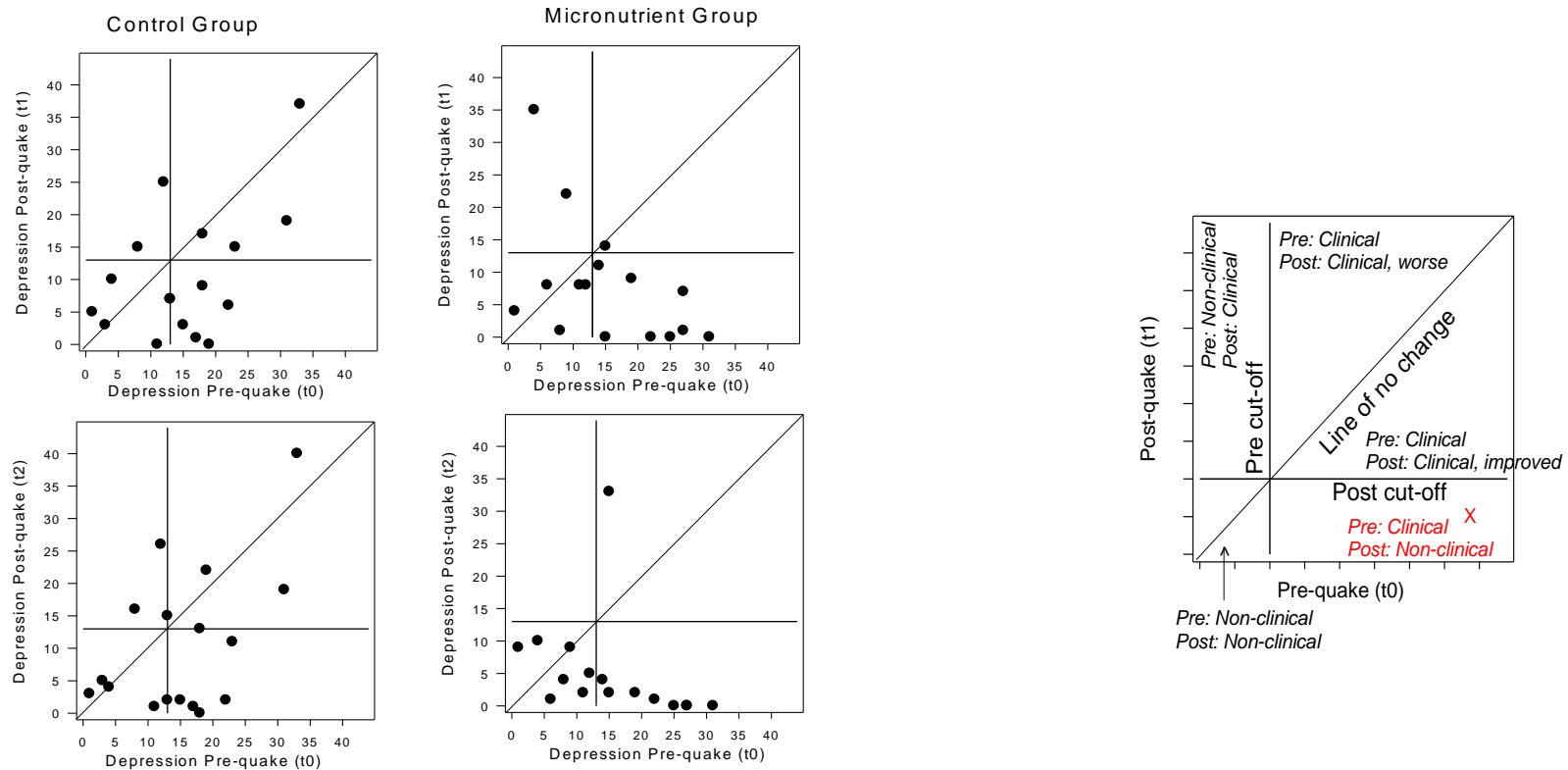
Using the scatter plot

■ You can add a line indicating a clinical cut-off score (or any other expected/predicted value)
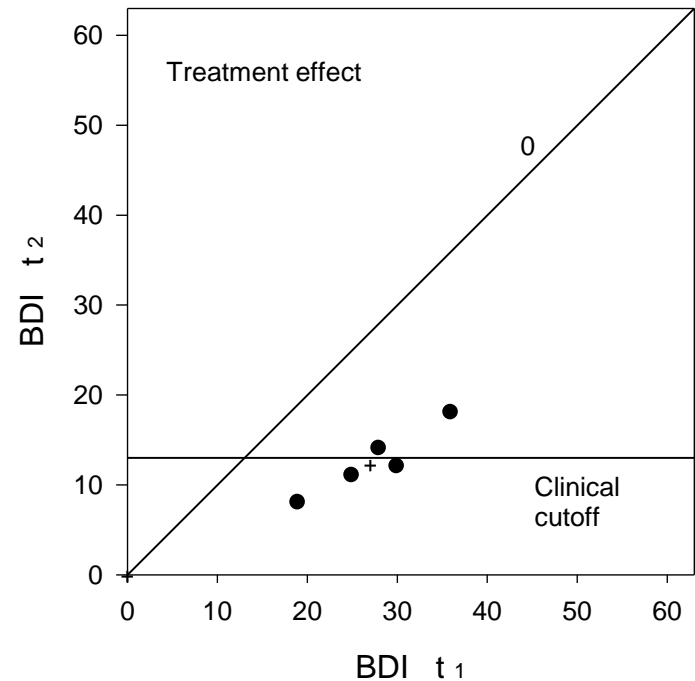
Example: Depression – pre & post earthquake

Example: Depression – pre & post earthquake

Scatter-plots & Single-case designs: Adaptations to group research

Why investigate groups?

- If the natural unit of analysis is a group
  - Family
  - School class
  - Work group
  - Community
- If the treatment is designed as a group application
  - "group therapy".

- If there are pragmatic reasons of efficiency and cost-effectiveness to work with a group.

  … the ideal intervention should also be available in group format in order to reach more clients.
 (King & Olendick,2006, p164).

So

- Groups are formed in some way

- Groups move through phases of standard single-case design

- Data are acquired as 1-point per person per phase, or aggregated to 1-point per person per phase (or sub-phase).

A reversal design

- The group moves from baseline to intervention and back to baseline conditions

- A single composite score is presented for each participant in each phase
    - The example is from a study of drug effects on infant sleep problems

[Based on France, Blampied & Wilkinson, 1999]
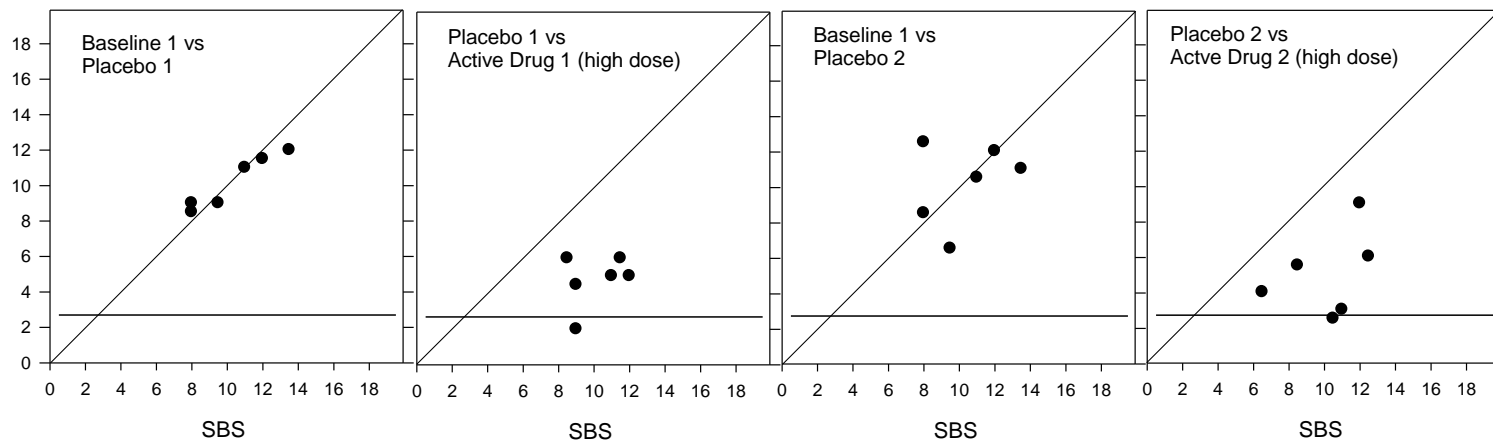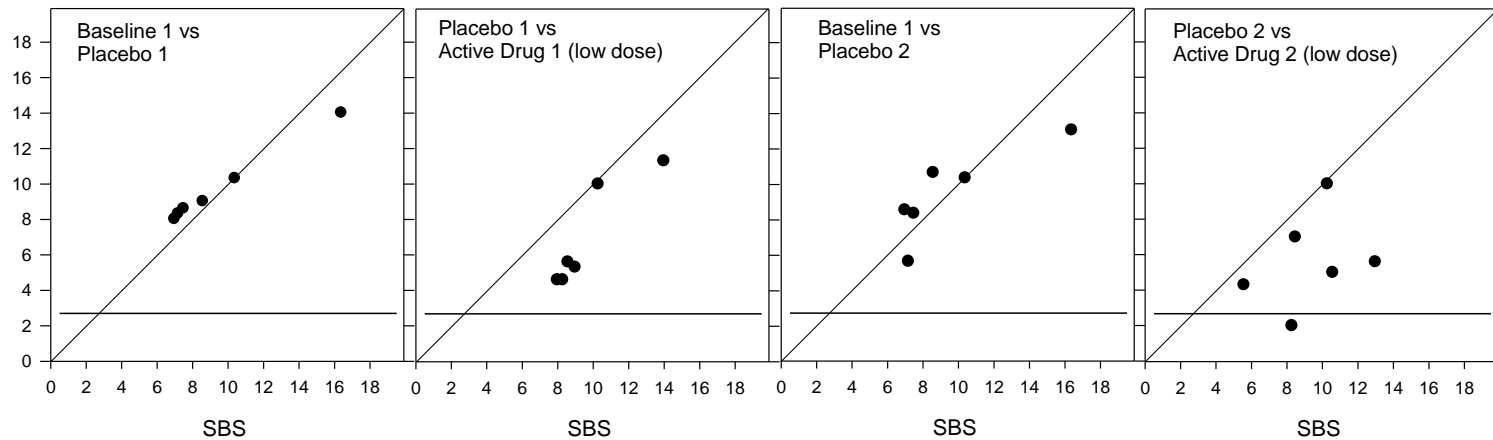
A Multiple-baseline design – across participants

- A multiple baseline across groups of participants
- Energy use by apartment-dwellers across three apartment blocks when daily consumption feedback supplied

  (invented data)

A Multiple-baseline design – across participants

- A multiple baseline across groups of participants
- Energy use by apartment-dwellers across three apartment blocks when daily consumption feedback supplied
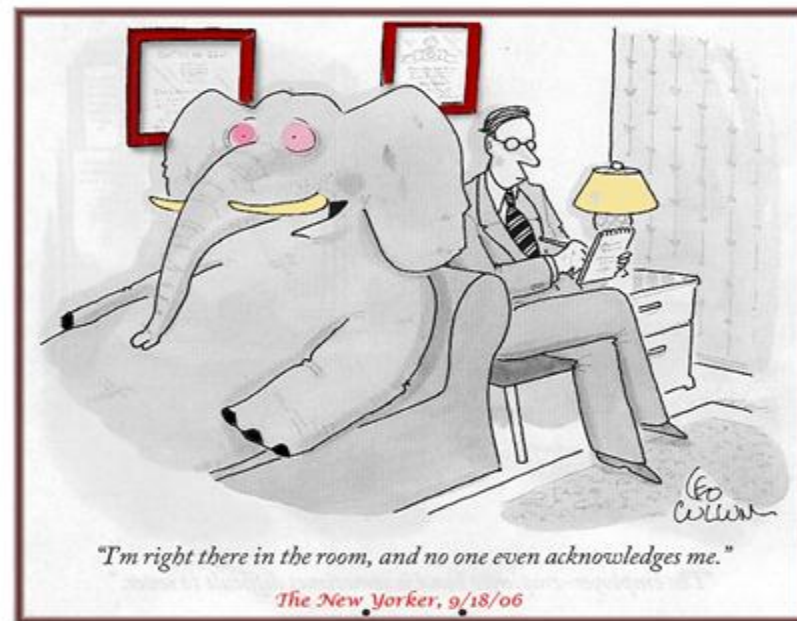
    (invented data)
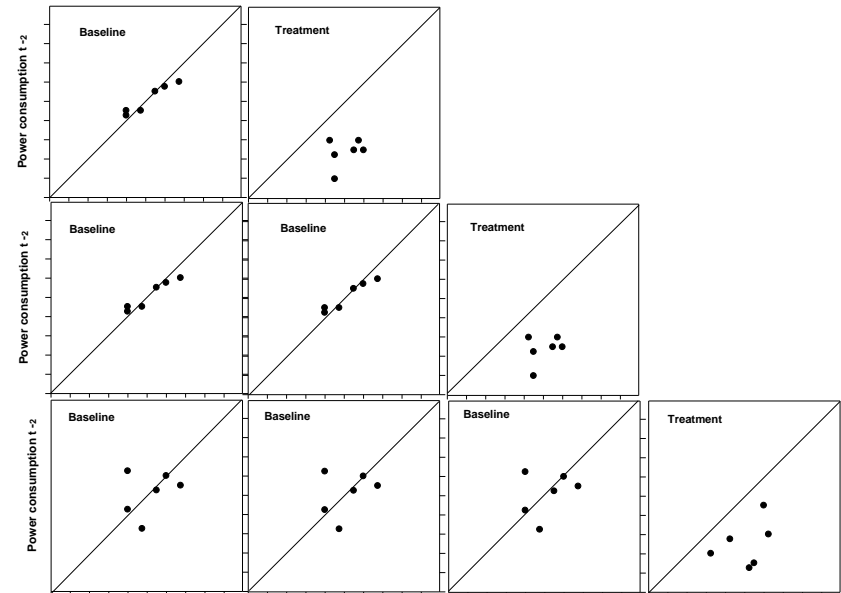
Conclusion
There are alternatives

…it is time to put our p values away and get around to the business of science.

(Lambdin, 2012, p85)

"I'm right there in the room, and no one even acknowledges me."

The New Yorker, 9/18/06

Friends don't let friends use p

(quoted in Kein (2013), *Beyond Significance Testing*)

Abstract

The ideal of the scientist-practitioner as the basis for applied psychology was one of the great achievements of 20th C psychology. Unfortunately, the idea became inextricably linked to the adoption by Psychology of a research methodology (I call it the Standard Model) based on a fusion of the ideas of Fisher and of Neyman and Pearson that are internally incompatible, and poorly adapted to the needs of applied research. Methodological criticism of the Standard Model has grown in intensity and comprehensiveness in the past 50 years. It has been almost completely ignored by researchers. We now have the paradox of applied psychologists, as scientist-practitioners, being expected to conduct evidence-based practice, while researchers themselves persistently ignore the evidence that their methods are flawed. I will review some of this history, and consider some of the ways that we might change our methods to better meet the needs of scientist-practitioners and evidence-based practice. I will particularly discuss the utility of single-case research approaches to applied practice.

Blampied, N.M. (2013). Single-case research and the scientist-practitioner ideal in applied psychology. In G. Madden (Ed.). *Handbook of Behavior Analysis* Vol 1. (pp 177 – 197). Washington, DC: American Psychological Association.