



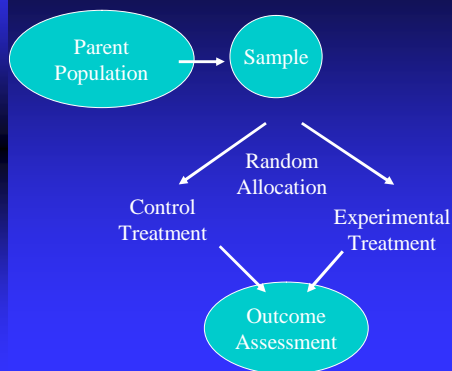
Recovery after brain damage and response to rehabilitation therapy: experimental design

Alan Sunderland
School of Psychology
University of Nottingham

Overview

- Behavioural methods –
 - ◆ Randomised controlled trials
 - ◆ Single case methods
 - Cross-sectional studies
 - Time-series designs
- Functional imaging studies of response to rehabilitation therapy

The Randomised Controlled Trial (RCT)



The Randomised Controlled Trial Why is it the gold standard?

- Can have high **Internal Validity**
= High confidence in inference of causal relationship (that it was the treatment that did it).
- Can have high **External Validity**
= High confidence that the results will generalise to other members of the population.

The Randomised Controlled Trial

- The accepted method in medical clinical trials e.g. testing new drugs.
- But it is difficult to apply where there are individual differences in problems, individually determined treatments, and complex outcomes.
- It has been argued that RCTs are premature in cognitive neuroscience until individual differences and treatment effects are better understood e.g. therapy for visual neglect (Barrett et al., J Cog Neuroscience, 2006, 1223-1236).

Single Case Experiments My Definitions

- “Case study” – a report of unplanned observations on a single case. Open to bias.
- “Single Case Experiment” – a planned study using specified measures and an experimental design:
 - ◆ cross-sectional design – different conditions at one moment in time e.g. neuropsychological investigation.
 - ◆ time-series experiment (TSE) – change over time under different conditions e.g. intervention effect.

Single-case Investigations vs Group Studies

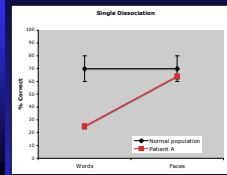
Single Case

- What is the pattern of performance for this individual?
- This person may be unique (an important insight or a misleading one-off)
- Statistical inference with single case measures can be difficult - can you be sure that the "pattern of performance" is more than error variation?

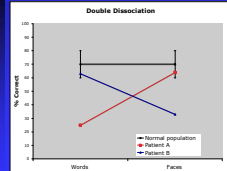
Group Studies

- What characterises human performance?
- Tend to consider average effects only (ignores outliers and dissociations)
- Group studies can use well developed, robust statistical methods.

Dissociations

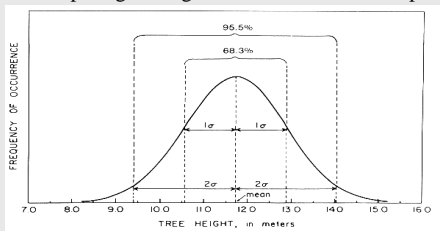


Indicates a selective impairment. But not necessarily separate processing systems. (The better task may make smaller demands on a shared resource e.g. sustained attention)



Suggests modularity - some processes are unique to each task and selectively impaired.

Comparing A Single Case to the Normal Population

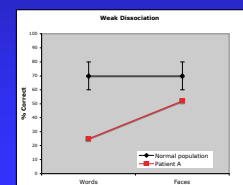
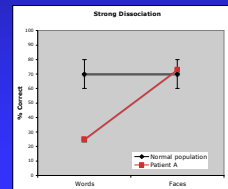


Normal Mean (M)	Normal SD	Score (x)	Standard scores (z scores)	Wechsler scaled scores (ss)	Percentile
11.7	1.15	9.4	-2.00	4	2

A common convention is that a z score < -2 is evidence of clear impairment (= 2SDs below the norm, p<.05, 2-tailed)

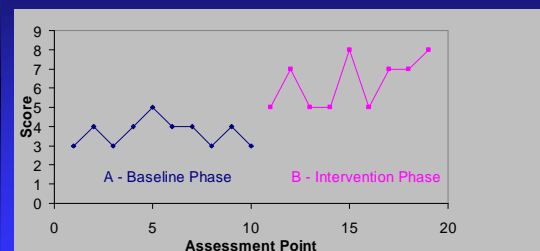
Some complications -

- Z scores should not be used with small control groups. Use a modified t-test (Crawford et al., 2005).
- Dissociations are only solid evidence of selective impairment when other scores are well within the normal range:-



Time-Series Experiments

AB – The simplest TSE



Single-case Time-series vs Clinical trials

- | | |
|---|--|
| <p>Single Case</p> <ul style="list-style-type: none"> ■ Was there a treatment effect for this case? ■ But may be true only of this case ■ Cause-effect inference <u>may</u> be weak ■ Useful where there are major individual differences, or as pilot work | <p>Randomised Controlled Trial</p> <ul style="list-style-type: none"> ■ Is this treatment generally effective? (External validity) ■ But considers the average effect only ■ The RCT offers high internal validity. ■ Useful for a standard treatment for a homogeneous population |
|---|--|

Design Issues

Selecting a measure – need for repeatability.

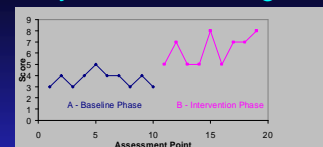
- Standard assessments (e.g. questionnaires; cognitive tests) are seldom suitable because of practice effects or familiarity.
- Most TSEs use simple rating scales or behavioural measures designed for the experiment e.g. time to complete task; number of events per day etc.
- One limitation of TSEs are situations where there is no suitable measure.
- Standard assessments may be used before & after to compare with a reference population.

Time-series assessment



- The standard assessments do not allow us to see the detail of change over time and the relationship to onset of treatment.
- The time-series measures show this but are not standardised against any reference population.

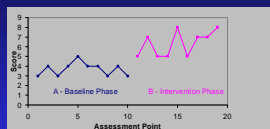
Weak Internal Validity of the AB design.



- Poor control over effects of external events. By coincidence something else may change at the AB boundary.
- Limited opportunities for statistical testing of treatment effects.

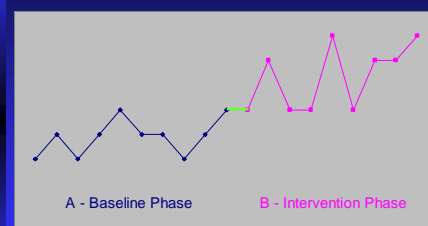
Problems with conventional statistics with all time-series designs

- The statistical tests we use in group studies rely on independent observation points e.g. 2 groups of 10 people, or one group of 10 tested twice.
- Successive points from a single case are not independent (=autocorrelated).
- Use of standard t-tests, ANOVAs or the nonparametric equivalents are therefore prone to Type 1 errors.
- Suitable statistical techniques (e.g. time-series analysis which models change over time) need many data points.



Design Issues Visual analysis or statistical analysis?

Visual judgement - Is there a treatment effect ? -



YES?

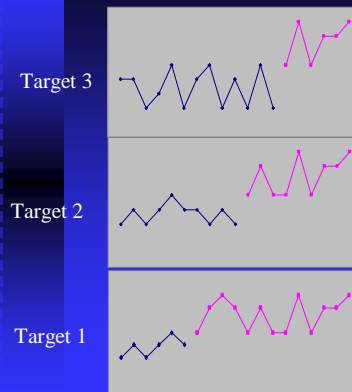
No?

Raters tend to make Type 1 errors (Franklin et al., 1996)
But most published studies still rely on visual judgement.
Justified only where consistency is high and the effect is huge.

Summary on AB Design

- The simplest TSE
- Visual analysis is unreliable
- Appropriate statistical tests have low power
- Useful as an initial pilot study

The Multiple Baseline Design



Several targets are identified for intervention.

Intervention is staged in pre-planned or randomly assigned manner.

If the dependent variables follow the same staging, then this a powerful demonstration of causation.

Multiple Baseline Design.

- Has higher internal validity than the AB design but does not require reversal of treatment effects like ABAB. It therefore has a wider range of clinical applicability.
- By randomly assigning the order of targets for change or the length of baselines, results can be statistically evaluated (randomisation test).
- A limitation is that treatments which show generalisation across targets cannot be evaluated within a single case (but multiple baseline across subjects is a feasible small N design).

Functional imaging studies of response to rehabilitation therapy

- A major challenge for imaging research requiring a longitudinal study (cf Ward et al., 2003) and needs to distinguish between effects of the passage of time versus treatment effects.
- Some studies have attempted to avoid this problem by looking at short-term effects in chronic-stage patients e.g. Musso et al:-

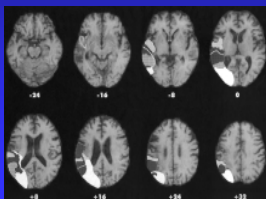
Musso et al.

"Training induced plasticity in aphasia."

Brain, 1999, 122, 1781-1790.

PET study of training effects (N=4).

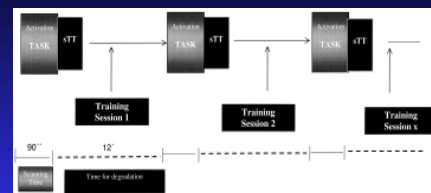
All patients were right-handed and had suffered from a stroke in the temporoparietal regions of the left hemisphere, including the presumed Wernicke's area. Time since stroke > 8 years. All had moderate-severe impairment on the Token Test.



MRI showing lesion sites in the 4 patients.

23

Scanning Paradigm



During each of 12 PET scans the subjects underwent the same activation task, which required them to follow one of two different auditory commands: 'take' or 'point to' an everyday object. The oral commands were presented every 10 s from prerecorded tapes.

In the interval between the scans (~12 min), while waiting for the degradation of radioactivity, the subjects were exposed first to the shortened Token Test, which was followed by one of the 11 training sessions.

24

Intensive Training - 11 sessions of ~8 min each.

Tactile-Verbal matching. An everyday object was given to the subject along with different oral verbal commands.

Visual-Verbal matching task. Point to named object.

Semantic Decision Making . Evaluate the semantic correctness of the sentences.

Phonological Decision Making. Judge the phonological accuracy of the sentences.

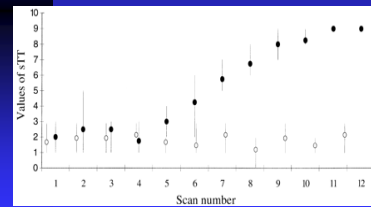
Picture-Sentence Matching. Pictures were shown to the patient with three descriptive oral sentences.

25

Effect of Training

“To assess the specificity of the training, the same four patients who underwent the PET scanning were tested off-line (i.e. outside the camera) using the same setup. However, instead of the language comprehension training described previously, we exposed the patients to non-specific stimuli such as writing, watching television and speaking. This stimulation was also organized in 11 sessions of 8 min each. “

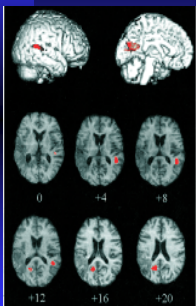
Presumably immediately before scanning?



Open circles = baseline.
Black circles = scanning session.

26

PET results



Averaged results show significant correlations in between language improvement and activation in right superior temporal (Wernicke's homologue) and left precuneus.

But individual results show a scatter of different areas:-

- Patient 1 – Left frontal & temporal.
Right middle temporal.
- Patient 2 – Left occipital.
Right frontal, temporal & parietal.
- Patient 3 – Left parietal (perilesional).
Right postcentral.
- Patient 4 – Left inferior frontal

27

Critical Evaluation.

“Our study shows for the first time that short-term clinical recovery of language performance correlated with fast modification of the activation pattern in a bilateral cortical network comprising brain regions directly or indirectly related to language.”

BUT

- Did training have an effect? (AB design)
- Were the brain activation changes caused by training? (or just time)
- Is this at all analogous to language rehabilitation?
- Individual results show no consistent pattern.

So, probably not a very informative study!

28

Additional References

Todman, J. B. and P. Dugard (2001). Single-case and small-n experimental designs. London, Lawrence Erlbaum.

Crawford JR, Garthwaite PH. 2005. Testing for suspected impairments and dissociations in single-case studies in neuropsychology: Evaluation of alternatives using Monte Carlo simulations and revised tests for dissociations. *Neuropsychology* 19:318-31

Dunn JC, Kirsner K. 2003. What can we infer from double dissociations? *Cortex* 39:1-7