

It's a Difference, but Is It Significant?

Alistair Campbell

Clinical significance, statistical significance, the CBT monoculture, the politics of service provision, and the search for additional ways to be effective.

Over the past weeks I have facilitated three workshops for psychologists. I have a standard introduction these days which alerts participants that I am not going to teach cognitive behavioural techniques. It is *not* an apology but a conscious political statement to explicitly identify that therapy can be practised in more than one modality. Interestingly, the majority of participants nod and smile on hearing this introduction. Many later say what a relief it is to hear about something other than CBT. These days, it certainly seems that nearly every workshop is about CBT for one condition or another. I think that people are starting to tire of the monoculture.

One of the reasons for this is that therapists are, I think, very dedicated to helping their clients. They want to use treatments that work, but no technique is going to work for all clients all of the time. To paraphrase Barnum, you can only help some of the people some of the time. Currently, though, therapists are given little variety in their diet as they seek additional ways to be effective. It is mainly the same mantra of 'CBT for this' and 'CBT for that'. We face the reality that what we are told is effective is actually what we have been doing, and it certainly does not work all of the time. So, there is a dissonance between the dominant story, of CBT as the only ethically valid treatment, and our clinical experience. I hope this dissonance can lead to a more intelligent understanding of how therapy is effective in the real world.

The evidence base on which proponents try to base CBT is not in fact the granite mountain they would like to believe. It is more like a pile of gravel. There are foundational elements which are simply not secure. The term 'evidence based' is used as a political tool to garner resources and legitimacy but it relies on an acceptance of narrow frameworks for doing and understanding science. I have made no secret of my belief about the inherent failure of randomised controlled trials as a methodology for understanding therapeutic interventions. There are, however, additional reasons to challenge the orthodoxy which maintains that the best evidence is that provided in support of CBT. One of the most fundamental of these is basically statistical and that is what I want to introduce in this article.

Most of us, whether we are psychologists or not, have been raised with the idea that the importance and meaning-

fulness of an experiment's findings can be measured by the alpha level (α). That is, if something is said to be significant at less than .05 we often assume that this means something. If it is less than .01 or .001 we have been trained to think that this means something even better. How many of you, then, would be surprised to know that this core statistical concept is basically considered to be invalid? That many authors have argued that the reporting of alpha should be banned from scholarly journals (Denis, 2003)? Not many, I expect. The failure of alpha is one of those family skeletons for anyone who relies on standard (Fisherian) statistics.

Statistical significance is too often confused with the idea of the size or power of an effect. Over time, the technical meaning of 'significance' has become confused with the term 'important', and the way it is used often reflects a basic misunderstanding of probability theory (Matthey, 1998). When something is reported to be *statistically* 'significant' it is often understood to mean that this is an indication of importance when often the true difference between means or amount of variance accounted for is of minimal *practical* significance (McCartney & Rosenthal, 2000). Thus the problem with null hypothesis significance testing is that something can be significant statistically yet be quite trivial in absolute terms. The probability value is crucially affected by the size of the sample and even very small differences between means can be statistically significant if enough data is gathered.

For example, consider a hypothetical weight reduction program. With 20 clients in both a control and treatment group, a mean difference in weight between the groups of 0.1kg is both statistically nonsignificant and obviously too small a difference to support the treatment program. But, if you keep the mean difference the same (0.1kg) and increase the numbers per group to 200 the difference becomes statistically significant at $p < .05$. With $N = 2000$ the difference becomes significant with $p < .0001$. The mean difference between the groups is still only .1 units but it is clear that by simply increasing N it is possible to create statistical significance.



Alistair Campbell, Co-ordinator, Research into Practice. E-mail: alistair.campbell@jcu.edu.au

The problem for real-world clinical data is that samples tend typically to be small whereas experimental studies often recruit large numbers of subjects. Therefore, the analysis of real world data is more likely to lead to non-significant differences whereas controlled studies are more likely to show statistical significance.

With this in mind it becomes important to be able to report data in such a way that both power and practical significance is taken into account. Calculation of effect size has been identified as one solution to the issue (Fan, 2001; Zakzanis, 2001) but, as these approaches are also susceptible to the size of N , they do not really address the fundamental issue of determining what differences are meaningful (Denis, 2003). Another suggestion has been to use statistical analyses that are more relevant to clinically derived data and to identify criteria of significance that are statistically and heuristically meaningful (Kazdin, 1999). The best-known approach of this kind is that suggested by Jacobson and a range of collaborators (Jacobson & Truax, 1991). The basic principle underlying Jacobson's work is the assumption that a change which is 'clinically significant' will have something to do with a return to normal levels of functioning. This leads to an assumption that people seeking therapy fall within a distinct, 'dysfunctional', population as regards some domain of measures of symptomatology, functioning, quality of life, and so on. From this it can be argued that there are three possible ways to operationalise clinically significant change following treatment; (1) that functioning should lie outside the dysfunctional range, where 'range' is defined as two standard deviations beyond the mean (depending on the score direction of the scale) for that population; or (2) that functioning should fall within the range of the functional, or normal, population within an agreed number of standard deviations; or (3) that the level of functioning after therapy is closer to the mean of the functional population than to the mean of the dysfunctional population (Jacobson & Revenstorf, 1988).

Because of the technical and practical complexity of some of the criteria for establishing clinical significance, it is not a well-known procedure, particularly amongst clinicians. But it has been argued that these analyses should be part of any standard approach to evaluating the effectiveness of clinical interventions and that editorial guidelines for publication should require the reporting of these statistics (Thompson, 2001, 2002). How often, though, do you see clinical significance reported in the treatment literature? The answer is, hardly ever.

One of the problems for researchers is that the criteria for clinical significance make this a conservative approach. Data that looks very good using statistical significance is frequently much less so with clinical significance. High rates of improvement using the statistical significance benchmark are often reduced to rates of between 30% and 40% (Kazdin, 1999). However, these percentages are more in line with clinicians' experience of changes for their clients in the real world and are probably more realistic. Using clinical significance it would be much harder to support the mono-

culture of CBT because it is a safe bet that most interventions could be shown to be effective between a third and a half the time.

Let me recapitulate. Therapists are interested in doing what works. We are told that it has been 'scientifically proven' that CBT is the only therapy that works. It is touted as the only approach that has an 'evidence base'. Yet, that is not the experience of clinicians working in the real world. What we know is that a range of things work. That some things work for some people but not for others and that sometimes an approach that worked for a person before does not work now. The rhetoric about CBT that abounds today is part of a general social conservatism and it needs to be understood as a political statement more than a science. If anything, much of the hype about CBT and evidence-based practice smacks of scientism more than it does of science.

But the evidence-base lobby has a powerful position and clearly dominates the therapy scene at a range of levels — governmental, educational and agency. It is important then — at least I certainly think so (Campbell, 2005) — for clinicians to be able to arm themselves to challenge the dogma. Clinical significance is not the best argument against the current orthodoxy but it forms part of a chain of understandings which enables the development of a sophisticated attack on the gravel foundations of the evidence base position.

It is not enough to say that something is statistically significant. Statistical significance does not provide a metric of meaningfulness or importance. Therefore, if someone wants to support CBT as proven to be the predominant intervention, the yardstick cannot be a significant difference. It must make a meaningful difference defined in terms that are relevant to the real world and based on a current understanding of the most appropriate approach to statistics. I would argue that proponents of CBT as the one true therapy have no such evidence base.

References

- Campbell, A., 2005. Evidence Based Practice in Australia: Where is the Debate? Paper presented at the Inaugural Australian Counselling and Supervision Conference: Integrating research, practice and training, Brisbane, Qld.
- Denis, D., 2003. Alternatives to Null Hypothesis Significance Testing, *Theory & Science*, 4, 1. Retrieved 2003, from <http://theoryandscience.icaap.org/>
- Fan, X., 2001. Statistical Significance and Effect Size in Education Research: Two Sides of a Coin, *The Journal of Educational Research*, 94, 5: 275–282.
- Jacobson, N. S. & Revenstorf, D., 1988. Statistics for Assessing the Clinical Significance of Psychotherapy Techniques: Issues, Problems, and New Developments, *Behavioral Assessment*, 10, 2: 133–145.
- Jacobson, N. S. & Truax, P., 1991. Clinical Significance: A Statistical Approach to Defining Meaningful Change in Psychotherapy Research, *Journal of Consulting and Clinical Psychology*, 59, 1: 12–19.

- Kazdin, A. E., 1999. The Meanings and Measurement of Clinical Significance, *Journal of Consulting and Clinical Psychology*, 67, 3: 332–339.
- Matthey, S., 1998. $p < .05$. But Is It Clinically Significant? Practical Examples for Clinicians. *Behaviour Change*, 15, 2: 140–146.
- McCartney, K. & Rosenthal, R., 2000. Effect Size, Practical Importance, and Social Policy for Children, *Child Development*, 71, 1: 173–180.
- Thompson, B., 2001. Significance, Effect Sizes, Stepwise Methods, and Other Issues: Strong Arguments Move the Field, *The Journal of Experimental Education*, 70, 1: 80–93.
- Thompson, B., 2002. 'Statistical', 'Practical', and 'Clinical': How Many Kinds of Significance do Counselors Need to Consider? *Journal of Counselling and Development*, 80: 64–71.
- Zakzanis, K. K., 2001. Statistics to Tell the Truth, the Whole Truth, and Nothing but the Truth: Formulae, Illustrative Numerical Examples, and Heuristic Interpretation of Effect Size Analyses for Neuropsychological Researchers, *Archives of Clinical Neuropsychology*, 16: 652–667. ©

Relationships Australia

GRADUATE CERTIFICATE IN CLINICAL SUPERVISION AND CONSULTATION

commencing
3 March 2006

The Australian Institute for Relationship Studies is offering this training to experienced practitioners in the counselling, health, welfare and related areas to enable them to undertake supervisory roles within their organisations or offer clinical supervision and consultation as a private practitioner.

This course draws on a range of supervision models and frameworks informed by systems, psychodynamics, adult learning, management and professional ethics.

The Australian Institute for Relationship Studies is the professional training division of Relationships Australia

For more information or an application form please contact us on 9806 3288 or visit our website www.relationships.com.au

