

## **Two Babies in Two Bathtubs -- Don't Throw Out Either, But Rather Advance Both: Discussion of Edwards, Eells, and Messer Papers**

**FRANZ CASPAR**<sup>a,b</sup>

<sup>a</sup> Psychologie Clinique, Université de Genève, Unimail, 40 Blv. du Pont d'Arve, CH 1211 Genève 4

<sup>b</sup> Correspondence concerning this article should be addressed to Franz Caspar, Psychologie Clinique, Université de Genève, Unimail, 40 Blv. du Pont d'Arve, CH 1211 Genève 4

Email : [Franz.Caspar@pse.unige.ch](mailto:Franz.Caspar@pse.unige.ch)

### **ABSTRACT**

This article discusses three contributions by Edwards, Eells, and Messer to an article series on "seeking an equal place at the therapy research table" for the pragmatic case study method represented in this PCSP journal. The discussion focuses on a single but complex theme that these three authors as a group directly address. This theme is the importance of recognizing that both randomized clinical trials (RCTs) and case studies are methods that have both strengths and limitations with regards to improving the efficiency and effectiveness of psychotherapy, and that actually these strengths and limitations are complementary. In reaching our ultimate goal – improving the future practice of therapy with individual clients – we thus need both RCTs and case studies. Subsequently, one of the challenges for the psychotherapy research field is to establish a constructive, working relationship between these two paradigms.

*Key words:* case studies; randomized clinical trial (RCTs); generalization from research; politics of research

---

### **INTRODUCTION**

It is my pleasure to discuss three thoughtful articles and their introduction written by intelligent and knowledgeable colleagues (Edwards, 2007; Eells, 2007; Fishman, 2007; Messer, 2007), who have given much thought on the issue. These contributions are so rich in thoughts and information that it is not easy to go beyond just repeating, "Everything has been said, but not yet by me!" I will nevertheless try to further raise awareness by referring to some of the points made in these contributions, and by adding some further ideas.

First of all, I would like to commend Daniel Fishman not only on founding an exciting journal, but also in bringing together people who have contributed to single case research in the past and engaging them in further efforts to not only make the possibilities and advantages of single case research known, but also to contribute to the development of the methodology for reporting and studying single cases. It is easy to connect to stereotypes of what single case

research is, and get applause for using witty sayings such as "the plural of anecdote is not data," or "the plural of anecdote is not evidence." It is also relatively easy to criticize randomized clinical trials (RCTs), focusing on their weaknesses. It is much more demanding to criticize RCTs to a point where they are removed from an unjustified pedestal and brought into a rational comparison with other approaches, and then to demonstrate the contrasting and complementary strengths of systematic case study approaches. The articles in this module do this and can be considered of very important value in the advancement of methods contributing to improved clinical practice. The articles realize the importance of not throwing out the baby with the bathwater – neither for RCTs nor for case studies – but of rather advancing them both.

## WHAT IS THIS ALL ABOUT?

Our whole discussion is ultimately -- to put it simply -- about improving the quality of future therapies based on what we researchers have been scientifically learning about our present therapies, and knowing how to best avoid errors in this process. Unfortunately, none involved in the field can deny the influence of non-scientific, personal, political, and financial factors going beyond these (in principle) simple questions. These factors include processes such as (a) researchers' adherence to a particular therapeutic approach, for which they may have invested tons of time and money; and (b) researchers' preferences for particular basic models, because the more people share them, the more attention will be given to their own work and career, and the easier it will be to get it financed and published. Were that it were possible for researchers to be successful just based on their intelligence, creativity, and industriousness! Unfortunately, there are real-world limitations in implementing some logically elegant research designs. For example, a waiting-list control group is a good idea for satisfying criteria of the internal validity of a study. However, in a world which holds (fortunately) possibilities for alternative treatment in most cases, "real," seriously suffering patients will not simply wait for treatment if randomly assigned to the waiting-list condition, but will do something else. Remaining in the waiting-list group may thus indicate low clinical validity. Criteria for the design quality of studies can thus be in unavoidable conflict with one another.

When I was a child, I loved to read books about gardening. I used to plan ideal gardens, which have continued to be beyond the time and money available to me. However, I also dealt with concrete problems such as fertilization. There was a figure with a barrel (see below) to illustrate a basic principle of fertilization.



*Figure 1: The Barrel Model of Fertilization: The Weakest Component Determines the Situation*

This basic principle is that the liquid can be only as high as the shortest stave reaches. Thus, for example, if we have too little nitrogen, a sufficient amount of phosphate is of no use. Applying this logic to psychotherapy research: it is of no use to emphasize internal validity if clinical validity is low. It also makes no sense to be proud of a beautiful experimental design if it has little relation to everyday practice. In the absence of clear rules of how low one criterion (like the “clinical validity” stave in the above figure) may be without jeopardizing an entire study, or how and to what extent one can be compensated by the other, methodological discussions in psychotherapy research often resemble irrational arguments (Caspar, 2006). For example, arguing why phosphate is less important than nitrogen, or why a satisfactory level of nitrogen is inevitably linked to a lack of phosphate, or why certain fertilizers should be used even if they have negative side effects (I remember, for example, that pork manure contains poisonous chloride) does not make any real sense. In sum, global discrediting of a particular methodology is not constructive. Rather, what is rationally helpful is to systematically consider the specific strengths and weaknesses of any particular method. The papers in this module reflect this view in that they focus on articulating and exploring the strengths and weaknesses of the case study method.

One might add that new knowledge from therapy research is useless (apart from advancing the careers of researchers) unless it is used by practitioners to improve their practice. It also seems useless for researchers to complain that practitioners are not doing what they are supposed to do. It seems much more fruitful to develop ways to intrigue clinicians and interest them in therapy research. With case studies we have a vehicle that is appealing to many practitioners, because their work is always directly with individual cases. It would seem, then, that for researchers, appreciating the distinctive knowledge value of individual case studies and

databases of them (Fishman, 2005) leads to a path for developing genuine collaboration between researchers and practitioners.

Eells (2007) does an excellent job in pointing out and illustrating that when deciding what method to use, it depends on the research question. In this context, it is well to remember the old joke: A drunk looks for his keys in the light of a lantern. When asked whether he has lost his keys there, he says no, but that the light is better for searching. Using wrong methods for researching good questions because those methods are popular and available is as wrong as limiting us to certain questions just because they fit the available methodology well.

## GENERALIZABILITY

An obvious but central question when deriving conclusions from past cases -- be they from individual case studies or from clients in RCTs -- is: "Does this apply to the concrete case I want to treat?" In group designs and in the attempt to find more general laws, the question comes in the form of the generalizability question. As generalizability can hardly be expected for patients in general, the usual strategy is to define homogeneous diagnostic groups for whom the claim of validity is made. It is relatively easy to show that in everyday practice, only a small percentage of clients correspond to such homogenous groups. Beutler, Malik, Alimohamed, Harwood, Talebi, and Noble (2003) calculate that out of 397 diagnostic groups, only for 51 have manualized treatments been developed and evaluated. Certainly, one has to be aware that for the epidemiologically more important groups, empirically supported treatments are available; but the situation remains unsatisfactory if one thinks of the patients who do not have the "luck" of being in one of the more common groups, uncomplicated by comorbidity. It seems obvious that what is derived from a group of depressed patients in general cannot easily be applied to depressed patients with borderline personality disorder.

Why is this such a problem? One reason is in the numbers: If the percentage of patients not corresponding to a diagnostic category served by a treatment is large, we are empty-handed as far as empirically supported treatments are concerned. For as RCTs are time and money consuming, and as literally millions of combinations of patient properties have to be served, there is little hope that we will be able to develop sufficient information applicable to all patients if remaining within the RCT paradigm. An escape may be the development of treatment principles. Castonguay and Beutler (2005) have decided to develop them in a rather traditional way, based on group research rather than single case studies. They did this for good reasons (acceptance by the "RCT establishment" for the idea of principles instead of or complementing standardized treatments), but this way they are not using information residing in single cases. Such case studies could be a complementary approach for developing and verifying principles .

Edwards (2007), above all, deals with the question of applicability, in an excellent and relevant way. Nevertheless, he leaves a number of questions open. For example, how can it be definitively ascertained as to which characteristics of past cases are relevant to a practitioner's

present case? Salience of characteristics (to what extent they “stick out” for a number of possible reasons) may, for example, be a property of patient characteristics that facilitate making a link for better or worse. I have repeatedly presented cases and was -- although one hopes, of course, that others can relate them to their own cases -- not convinced that all the links that were made across cases should have been made. Representativeness is another property that, as shown in the work of Kahnemann and Tversky (1982), may also lead to errors.

The applicability problem cannot be solved by following simple algorithms. For example, following the rule of applying gained knowledge only to patients who fully satisfy the criteria of the same disorder neither excludes (a) the overinclusion error (e.g., the inappropriate application to a patient of the same diagnostic group who is different, for example from an interpersonal perspective); nor (b) the underinclusion error (e.g., renouncing the application of useful, borderline-related insights to a patient with depression who, while without borderline diagnosis at the same time shows some borderline patterns and for whom borderline-specific techniques could have been used). Discussions, such as the one by Eells (2007) on how causality conclusions are formed, are extremely important in informing us about how errors based on case studies might be limited.

A word related to Edward's vacuum tube (2007, p. 12), a fascinating example which merits consideration in this context. I would like to emphasize, however, that vacuum tubes (representing the possibility of radically excluding disturbing variables) do not exist in the world of psychotherapy. We have to live with various factors polluting our data, and this is one of the reasons for using many cases in the hope that undesired sources of variance will neutralize each other in our cross-case analyses.

## **MERITS AND LIMITS OF RCTs**

The main argument in favor of experimental research – all of us have learnt this in methodology classes and it is still true – is that this is the clearest avenue to causal argumentation. Although there are alternative methodologies such as path analyses or, for single cases, multiple baseline designs, critical discussions of experimental designs and RCTs which do not acknowledge this are under suspicion of being biased or demagogic. While it is also clear that experimental research can never prove anything but only disconfirm null hypotheses, its principal advantage in causal argumentation has to be acknowledged. Whether it is worth paying the numerous prices related to such research is a different question. The quote from the Edwards (2007) paper in this module, "Paradoxically, statistical analysis has been misused in the service of claims for absolute certainty – despite the fact that the main aim of statistical decision theory is to provide an analysis of how to make decisions rationally in the absence of such certainty" (p. 12), deserves being read twice!

In RCTs we need both a manualization of the therapeutic procedures and a homogenization of the patients. It seems that sometimes in the discussion, manualization and

homogenization take on an importance in themselves, while the reason why they are important is lost. Remembering them may increase our degrees of freedom. Manualization (along with ratings of therapist theoretical adherence) are a means for knowing with sufficient precision how patients were treated. This is clear progress if compared to traditional descriptions of procedures as, for example, "psychodynamic" without specification. It should not be forgotten, however, that there are alternatives for providing information about the actual procedure. Grawe, Caspar, and Ambühl (1990), for example, left open what precise procedure therapists would use as a consequence of different types of case conceptualization. To know how the actual procedure was, it was recorded and described by neutral observers. This is demanding, but the procedure is then (at least) equally well known as it would be when manualized and controlled for adherence. Analogously, groups of patients are commonly described with a primacy of ICD/DSM diagnoses, which are not always actually that relevant. Also for RCTs, other properties, for example interpersonal style, can be equally appropriate. To reflect on what conditions are being defined as obligatory for RCTs to serve what purpose, and to think about alternatives, is an emancipation from the dictate of being unnecessarily locked into one research methodology.

A misconception about the value of experimental data is that they would exist or be of value without interpretation, which is normally concept driven. Interpretations are already built into the design of a study, and continue when evaluating and interpreting the results. This is nicely illustrated by the lab rat who says to his cage compatriot, "We have succeeded training this human. Every time we press the lever, he gives us something to eat!" This is not to criticize experimental research, but to emphasize that as far as dependence on interpretations is concerned, the difference with single case studies is not as big as some may claim or implicitly assume.

An important point that is often neglected in the discussion, but convincingly emphasized by Edwards (2007), is the limited amount of variance (which may be expressed as effect sizes) determined by techniques. This leaves open a possibility that Stiles, Barkham, Twigg, Mullin, Mellor-Clark, and Cooper (2006) have shown – namely, that in the transition from RCT to practice-based research, the superiority of an approach can be lost. This is not a criticism of RCTs in general, but an argument in favor of more humility. It should also be noted that advocates of RCTs have shown some humility when replacing the phrase, "empirically *validated* treatments," by the more modest phrase, "empirically *supported* treatment." Good RCTs are extremely demanding on those who conduct them. It is understandable that some colleagues increase their motivation as well as chances of getting the money they need by discounting competing (and seemingly easier) methodology. From a longtime perspective this is, however, is likely to be highly counterproductive.

When talking about merits of RCTs and experimental research in general, it should also be mentioned that some crucial methodological aspects have been discussed relatively late in the development of the approach. For example, allegiance effects (that is, the observation that researchers typically find a superiority of the approach they initially prefer) has been discussed only relatively recently. Although suggestions for how to solve the problem have been made, it is

actually uncommon for researchers to pay attention to it in current experimental studies. Or consider another problem. Publication bias (that there is a tendency to report and publish only significant results while the non-significant disappear), as discussed by Bakan (1966) already over 30 years ago, has only recently had binding consequences, as the good journals accept articles only if they are related to studies registered already before it was known whether the author would find significant results. This suggests that it takes a relatively long time of maturation of a methodological approach before some important questions are discussed and solutions are proposed and respected on a large scale. Although single case studies of one kind or another have a long history in psychotherapy, systematic and methodologically rigorous therapy case studies of the type aspired to by the PCSP journal are quite new; and thus it can be assumed that with a relatively short time so far for maturation, much development for these case studies is still ahead to overcome some of the weaknesses that make them less favorable in the eyes of many colleagues today.

## MERITS AND LIMITS OF CASE STUDIES

A virtue of the articles in this module is that they help to clarify the distinction between false stereotypes of case studies and what they actually are and can be. A point, which is so crucial that I want to repeat it, is the likelihood that a false concept is disconfirmed. Consider = lobotom. This procedure of destroying a part of the brain was publicly demonstrated in so many cases (so much about the criteria of publication and the literal meaning of anecdotes!) that the Portuguese inventor Antônio Egas Moniz received the Nobel price for it in 1949. Would randomized clinical trials have revealed the uselessness of the method? Yes, in particular if they had included side effects. Would single case studies have revealed it? Traditional case presentations did obviously not before thousands of patients (and people who would not receive an ICD diagnosis) had been "treated." Single case research following the suggestions made in this module and in PCSP generally would – hopefully! Clearly, the discussion of the power to disconfirm and how it can be maximized must continue! The point made by Messer (2007) that it is not satisfactory that we have access to original data only via the author of a case report is certainly one of the crucial points. His discussion of how individual data should be obtained and recorded deserves due attention, and he describes and discusses the problems in his usual, penetrating manner.

One of the limitations of case studies that I find relevant is their claimed advantage to deal with complexity. However, in a limited number of cases, some properties co-vary unavoidably, whether clinically meaningfully or by chance. It is a rule of thumb, for example in factor analysis, that we need about five times as many objects/subjects as variables if we want to establish consistent and valid links. If we look at an individual case or a few cases, only our concepts can tell us which of the covariations or coincidences we find are random and deceiving, or meaningful. Yet if we want to be empirical, it is problematic to depend to a high degree on preexisting concepts. For this problem of forced covariation by a high number of variables with a

low number of cases – which is, to be clear, not a problem of thinking in group statistics – no solution is known to me.

To place RCTs in opposition to single case studies also makes me think of an old distinction made by action theory (Volpert, 1974): the distinction between global signs and super signs. A global sign would, in our context, for example, be "transference interpretation" or "anxiety exposure," as long as this is not filled with detailed ideas of what this actually is. The same terms could be super signs if they were related to such detailed concepts. With both, one could operate in an abstract way, for example to compare effect sizes and to discuss for which disorders they help. But only if they are super signs – as exposure should be for a behavior therapist and transference interpretation for a dynamically oriented therapist – do they give relevant access to action. The obvious way for RCTs to create super signs and thus to enable action is to describe and train procedures based on manuals. Another approach would be to transform global signs into super signs by case reports. While it is pretty obvious that therapists cannot act only on the base of global signs, as far as I know, nobody so far has studied what can be contributed by manuals, on the one hand, and what by case reports, on the other.

I would like to add a perspective that I find particularly important. If we think of all the possibilities of increasing our knowledge about psychotherapy with limited financial resources, a great potential lies in using psychotherapies that naturalistically take place *outside* the relatively few taking place *within* an RCT design. The very fact that furthering single case research would enlarge the data base that can be used for advancing our knowledge is in itself a strong argument in favor of this type of research.

As a basis for believing in a concept or principle and for giving it a heavy weight when treating our patients, we should ideally expect that it be able to pass *both tests*, the experimental as well as the single case approach. This is one way of seeing the approaches as complementary rather than exclusive of one another. Another way is well described by Eells (2007, p. 48) in his concept of "analyze then aggregate" (rather than only "aggregate then analyze"). This method has been repeatedly fruitful in my research (e.g. Caspar, Grossmann, Unmüssig & Schramm, 2005), and I keep using it to illustrate how noxious an unjustified exclusion of one or the other approach would be. I particularly agree with Eells as far as the importance of single case research for systematically investigating treatment failures.

A last word of advice for those who want to advance single case research: Consider doing at least one good RCT! This will increase your credibility and show that you are advancing single case studies not because you are unable to do experimental research, but because you are deeply convinced that for some or even many questions and concerns, it is simply the superior approach.



## REFERENCES

- Bakan, David. (1966). The test of significance in psychological research. *Psychological Bulletin*, 77, 423-437.
- Beutler, L. E., Malik, M., Alimohamed, S., Harwood, T. M., Talebi, H., Noble, S., et al. (2003). Therapist variables. In M. J. Lambert (Ed.), *Bergin & Garfield's handbook of psychotherapy and behavior change*. (5th Ed.), pp. 227-306. New York: Wiley.
- Caspar, F. (2006). Forschungsdesigns in der Psychotherapieforschung: Die Diskussion um Randomisierte Klinische Studien. In A. Brüggemann & R. Bromme (Hrsg.), *Entwicklung und Bewertung von anwendungsorientierter Grundlagenforschung in der Psychologie* (S. 38-46). Berlin: Akademie-Verlag.
- Caspar, F., Grossmann, C., Unmüssig, C., & Schramm, E. (2005). Complementary therapeutic relationship: Therapist behavior, interpersonal patterns, and therapeutic effects. *Psychotherapy Research*, 15, 91-102.
- Castonguay, L., & Beutler, L. (Eds.). (2005). *Principles of therapeutic change that work*. New York: Oxford University Press.
- Edwards, D.J.A. (2007). Collaborative versus adversarial stances in scientific discourse: Implications for the role of systematic case studies in the development of evidence-based practice in psychotherapy. *Pragmatic Case Studies in Psychotherapy*, 3, Module 1, Article 2, 6-34. Available: <http://pcsp.libraries.rutgers.edu>
- Eells, T.D. (2007). Generating and generalizing knowledge about psychotherapy from pragmatic case studies. *Pragmatic Case Studies in Psychotherapy*, 3, Module 1, Article 3, 35-54. Available: <http://pcsp.libraries.rutgers.edu>
- Fishman, D.B. (2005). Editor's introduction to PCSP – From single case to database: A new method for enhancing psychotherapy practice. *Pragmatic Case Studies in Psychotherapy*, 1, Module 1, Article 2, 1-50. Available: <http://pcsp.libraries.rutgers.edu>
- Fishman, D.B. (2007). Seeking an equal place at the therapy research table: An introduction to a series on the pragmatic case study method. *Pragmatic Case Studies in Psychotherapy*, 3, Module 1, Article 1, 1-5. Available: <http://pcsp.libraries.rutgers.edu>
- Grawe, K., Caspar, F., & Ambühl, H. R. (1990). Differentielle Psychotherapieforschung: Vier Therapieformen im Vergleich: Die Berner Therapievergleichsstudie. *Zeitschrift für Klinische Psychologie*, 19(4), 294-376.
- Kahnemann, D., & Tversky, A. (1982). Intuitive prediction: Biases and corrective procedures. In D. Kahnemann, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases*, pp. 414-421. Cambridge: Cambridge University Press.
- Messer, S.B. (2007). Psychoanalytic case studies and the pragmatic case study method. *Pragmatic Case Studies in Psychotherapy*, 3, Module 1, Article 4, 55-58. Available: <http://pcsp.libraries.rutgers.edu>
- Stiles, W. B., Barkham, M., Twigg, E., Mullin, T., Mellor-Clark, J., & Cooper, M. (2006). Effectiveness of cognitive-behavioural, person-centred and psychodynamic therapies as practiced in United Kingdom mental health settings. Paper presented at the 37th Annual Conference of the Society for Psychotherapy Research, University of Edinburgh, June.
- Volpert, W. (1974). *Handlungsstrukturanalysen*. Köln: Pahl-Rugenstein.