

Commentary on: Affect Phobia Therapy for Mild to Moderate Alcohol Use Disorder: The Cases of “Carey,” “Michelle,” and “Mary”

The Quest for Causality in Psychotherapy Research

BJÖRN PHILIPS ^{a,b}

^a Department of Psychology, Stockholm University, Stockholm, Sweden

^b Correspondence regarding this article should be sent to Björn Philips, Department of Psychology, Stockholm University, SE-106 91 Stockholm, Sweden
Email: bjorn.philips@psychology.su.se

Editor’s Note: Björn Philips is one of the secondary authors of the target article describing the cases of “Carey,” “Michelle,” and “Mary” (Frankl, Wennberg, Berggraf, & Philips, 2020). His role in these cases includes being the main research supervisor of My Frankl, the first author of the target article and the therapist for the three cases. In addition, Dr. Philips has had an impressive psychotherapy research career, conducting randomized controlled trials as well as qualitative research, naturalistic outcome studies, and process studies. With this background as context, I asked Dr. Philips to reflect on his experience with the case studies of Carey, Michelle, and Mary, and to write a commentary on methodological aspects of case studies versus group designs.

ABSTRACT

This commentary on the article by Frankl, Wennberg, Berggraf and Philips (2020) focuses on methodological aspects of case studies versus group designs in psychotherapy research. Experimental case study designs such as ABAB design and multiple baseline design have a long tradition within behavior therapy. These research designs are especially useful for testing newly developed therapy methods and investigating the effectiveness for treatment of rare disorders. However, experimental case study design is most appropriate for single-component treatments for patients with one circumscribed problem. Randomized controlled trials (RCTs) are considered as the gold standard for testing and establishing the efficacy of a particular therapy method for a particular problem. However, the RCT design also bears some methodological shortcomings, such as low external and construct validity, simplistic epistemological assumptions, and only being able to establish average causal effect (thus not giving the clinician clear guidelines on how to work with individual patients). Rigorous process research is useful for identifying change mechanisms in psychotherapy. Finally, pragmatic case studies have a great potential of increasing our knowledge about psychotherapy and its effectiveness. This potential could be increased even further if pragmatic case studies integrated some methods from process research and if the results from multiple case studies were analyzed together in meta-syntheses.

Key Words: research design; experimental case study; randomized controlled trial (RCT); psychotherapy process research; pragmatic case study; clinical case study; case study

This commentary is in response to an invitation from Daniel Fishman, PCSP's editor, explained in his note above. My goal is to comment on the usefulness of various research designs and how they help us answer some important questions upon which psychotherapy researchers focus: Does this particular psychotherapy work for this particular patient population, and if so, what causes it to work? The Frankl et al. article provides a very rich description of the psychotherapy process and outcome in three cases of Affect Phobia Therapy (APT). But to what extent and how do these three case studies bring clarity to whether APT is effective for women with mild to moderate Alcohol Use Disorder (AUD) and explain how the treatment works? However, before answering this question, I first want to step back and comment generally on different research designs in psychotherapy research.

EXPERIMENTAL N=1 CASE STUDY DESIGNS

Experimental N=1 case studies (also called "N=1 design," "single case studies," "experimental one-person designs," or "intensive designs") have a long tradition within behavior therapy. In these studies, experimental manipulations are conducted in the treatment of one single patient, to enable causal inferences whether the treatment is effective. The experimental manipulation consists of alternating between periods of treatment and baseline periods, following a pre-determined schedule, and using continuous measurements of the primary outcome variable.

The classical experimental N=1 case study design is the ABAB design. Such a design starts with a baseline period (A), followed by a treatment period (B), after which treatment is withdrawn during a new baseline period (A), and finally treatment is reinstated (B). The baseline periods should be long enough to show stability in the outcome measure, i.e., it should allow for temporary fluctuations until the person's baseline level on the measure is determined. With this design, a causal effect of the treatment is inferred when the target outcome is stable during the first A, improvement is visible during the first B, improvement ceases and/or the problem increases during the second A, and finally, new improvement takes place during the second B (Barlow, Nock & Hersen, 2009).

Experimental N=1 case study designs were recommended and could lead to a treatment being considered as empirically supported in the early definition of "empirically supported treatment" by Division 12 of the American Psychological Association (Chambless and Hollon, 1998; also see <https://www.div12.org/psychological-treatments/frequently-asked-questions/>).

The ABAB design is best suited for a single-component treatment being tested on a patient with a circumscribed problem. It is not suitable for a patient with multifaceted problems being in a complex form of psychotherapy, for which therapeutic action to a large extent is based on the therapeutic relationship, such as in psychodynamic therapy (PDT). The results of an ABAB study using PDT could be obscured by many things, as predicted by the psychodynamic theory of change. For example: (1) symptom improvement might not begin immediately as treatment sets in, since it can take time for a productive therapeutic relationship to develop; (2) when treatment is withdrawn and the second baseline period begins, you might see a "sleeper effect," in which the patient still remains on an improved plateau without deteriorating; or (3) you might see the opposite, a patient with insecure attachment who might react drastically to the withdrawal of treatment, and hence feel abandoned and deteriorate rapidly to functioning worse

than at initial baseline. Also, the trauma of abandonment might be so powerful that the patient does not improve again when treatment is reinstated.

Another type of experimental case study design may be more adequate for patients with vulnerability to separation—the multiple baseline design, which is a small N design. A multiple baseline design study should include at least three patients, each in an AB design, i.e., first a baseline period and subsequently a treatment period, all with continuous measurements. In order to draw causal inferences about the effectiveness of the treatment, the baseline periods must vary in length among the included patients, based on when their therapy begins. The shortest baseline period in the study must still be long enough to allow for stability in the outcome measure.

In a concurrent multiple baseline study (i.e., all patients are included at the same time point), the procedure is: when a clear treatment effect is visible in patient 1, then patient 2 begins treatment; and subsequently, when a clear treatment effect is visible in patient 2, then patient 3 begins treatment. Thus, the hypothesis of treatment effectiveness is supported if the study shows that all included patients start improving and keep improving only after the start of treatment. The different lengths of the baseline periods rule out the alternative explanation that improvement always appears after a certain time period, perhaps due to spontaneous remission or the positive attention from being in a research study (Barlow, Nock & Hersen, 2009).

A multiple baseline design was what we tried to implement in the study reported by Frankl et al. (2020). For practical reasons, we used a non-concurrent multiple baseline design. The lengths of the baseline periods were pre-set to four, six, and eight weeks, and the patients were randomized with regard to baseline duration. In hindsight, we realized some shortcomings in our implementation: 1) we should have included more than three patients, in order to detect trends even if not all the individual psychotherapy cases were successful; (2) we should have chosen longer baseline periods, to ensure stability of the outcome measurement; (3) we should not have included a patient whose result on the outcome measure was zero during the entire baseline period, because then it was impossible to show any improvement; and (4) in order to more effectively measure change, we should have chosen a more sensitive, fine-grained primary outcome measure than days of heavy episodic drinking, such as total alcohol consumption in grams.

Just as in the ABAB design, the multiple baseline design might be better suited for single-component treatments being tested on patients with one circumscribed problem. As can be seen in the case of Carey in our study, she benefited greatly in psychotherapy with regard to emotional and relational functioning. But unfortunately, as she had ceased drinking already before the baseline period, we could not observe any improvement in our pre-determined primary outcome variable – days of heavy episodic drinking.

Experimental case studies are especially useful for generating hypotheses on what is effective in psychotherapy, for developing new psychotherapeutic methods, and to study therapy effectiveness for rare disorders. For if a certain mental disorder is very rare, it might be impossible to test a treatment in an experimental group design—because you could not reach a sample size large enough for statistical power.

EXPERIMENTAL GROUP DESIGNS (RCTS)

The experimental group design used in psychotherapy research is the randomized controlled trial (RCT; also called an “efficacy trial”). The RCT is typically considered as the “gold standard” for testing the efficacy of a particular treatment for a particular disorder and thus the foundation for establishing whether that treatment is “empirically supported” (Chambless and Hollon, 1998).

In a classical RCT, the included patients are randomized into two arms—either the treatment group or a control condition (which could be a waiting list, a placebo intervention, or “treatment as usual”). A primary outcome measure is chosen, which should be a valid and reliable measure of the target problem for the patient population. For example, in an RCT of a treatment for major depressive disorder, the primary outcome measure should be a valid and reliable depression scale. The RCT is thus an example of experimental methodology, in which the intervention (treatment or control) is the independent variable and change in the outcome measure is the dependent variable. The rationale for the RCT is that it should rule out alternative explanations for differences in outcome between the two study arms—if patients are randomized to either treatment or control, and all other conditions are the same for the two groups, then difference in efficacy would be the only logical explanation for a difference in outcome.

There are variants of RCT design besides the classical one of treatment versus control. Examples are an RCT with a new treatment versus an already established evidence-based treatment, or an RCT with more than two arms (e.g., psychotherapy + medication versus psychotherapy alone versus medication alone versus waiting list control).

The logic behind the RCT design is to maximize internal validity, i.e., to use an experimental methodology in which alternative explanations are eliminated in order to establish a causal connection between the independent variable (treatment) and dependent variable (outcome). Many authors have pointed out how these procedures have the disadvantage of leading to poor external validity, i.e., to poor generalizability of the findings to settings outside the experimental situation.

Lambert (2013) describes how naturalistic outcome studies carried out in an ordinary clinical setting (also called “effectiveness studies”) have strong external validity by using procedures opposite to those in many RCTs, for example: (a) by not using strict criteria for selecting patients; (b) by not having a pre-determined length and “dose” of therapy; (c) by not monitoring therapists’ fidelity to the treatment method in a meticulous way and making sure it adheres to the prescribed manual; and (d) by including therapists who work at the clinic and who are not, as in RCT studies, specially skilled therapists who have undergone extensive special training and supervision. In sum, in psychotherapy research there seems to be an inverse relationship between internal and external validity.

A number of years ago, Seligman (1995) pointed out that randomization in itself is a procedure that contradicts a beneficial situation often found in typical clinical contexts. In the non-experimental world, many patients actively seek a particular therapy method and sometimes a particular therapist, and this leads to stronger motivation and facilitative treatment expectations.

Another weakness of the RCT design is that it does not fit psychotherapy as well as pharmacological treatment. In clinical trials testing a medication, the patient, the doctor, and the nurse are unaware of which patients receive the actual medication and which patients receive sugar pills (placebo)—that is, the trials are double-blinded in order to eliminate the influence of placebo effects (expectations of positive outcome). Such double-blinding is practically impossible in psychotherapy research.

An additional problem is that RCTs most often test the efficacy of an entire treatment package (e.g., CBT versus waiting list) and hence, even if the trial shows that the treatment is efficacious, we still do not know what made the treatment work. What were the crucial therapeutic interventions and change mechanisms that helped the patients to improve? Kazdin (2013) refers to this as a lack of construct validity in RCTs of treatment packages. Hence, just as experimental case studies, RCTs are more appropriate for single-component treatments for patients with one circumscribed problem.

A major critique of experimental designs in psychotherapy research is that they might be based on incorrect epistemological assumptions. Both experimental case studies and RCTs are based on the assumption of linear causality, i.e., such an experiment tests hypotheses of the type “A causes B” (equivalent to the scientific laws of Newtonian mechanics). However, much evidence suggest that the human mind and the practice of psychotherapy would better be characterized as complex systems (equivalent to quantum physics, meteorology, or economics), which indicates that psychotherapy research should rather use models like non-linear dynamics and chaos theory. To complicate things further, the human mind contains elements that are not present in even the most complex natural sciences—elements that are likely to influence the process and outcome of psychotherapy, such as consciousness, intentionality, subjectivity, and emotions. Hence, quantitative research on psychotherapy needs to analyze the complex interplay of multiple factors by using advanced statistical methods such as multilevel modelling.

A particularly important point about psychotherapy RCTs is that not one that I know of has found that every patient who received the experimental treatment improved, while all patients in the control condition deteriorated or remained unchanged. Thus, experimental group designs do not establish absolute natural laws of psychology. Instead, they establish the average causal effect for the average patient. Here we have a problem of generalizability: a clinician who is about to choose a treatment for a new patient might not know whether this is an average patient or an atypical patient.

PARTICULAR RESEARCH CHALLENGES FOR PSYCHOTHERAPY WITH SEVERELY DISTURBED CLIENTS

My own experience of managing an RCT on psychotherapy for patients with concurrent borderline personality disorder (BPD) and substance use disorder (SUD) increased my conviction that the RCT is an inappropriate research design for patients who are severely disturbed, often traumatized, and lacking trust and hope that treatment providers can help them. It often takes a long time to build a secure relationship with such individuals and to strengthen their motivation to dare begin psychotherapy and work with their problems. At that moment, randomization is detrimental to the patients who end up in the control condition and do not receive the treatment they are eager for. Our study was aiming for the inclusion of 80 patients

(based on a statistical power calculation), but after five years of hard work and consuming all of the external research funding, we had to stop at $N = 46$ (Philips, Wennberg, Konradsson & Franck, 2018). In fact, previous RCTs of psychotherapy for concurrent BPD and SUD had even smaller sample sizes (Lee, Cameron & Jenner, 2015).

In my view, the just-mentioned small sample sizes are an indication that psychotherapy with fragile patients, such as dual-diagnosis patients, should be studied with research designs other than a typical RCT. An exception might be an RCT comparing two bona fide psychotherapy methods so that none of the patients received lesser therapy. But such a comparative RCT would be extremely expensive if it were to meet design needs for sufficient statistical power and adequate treatment duration (probably more than one year).

In our study of mentalization-based treatment (MBT) versus standard SUD treatment for dual diagnosis patients, we found no significant effect of MBT, partly because the staff taking care of the control group helped many of those patients by referring them to some sort of psychotherapy (Philips, Wennberg, Konradsson & Franck, 2018). That was good news for the patients, but it was bad news for our research study and for the RCT design!

THE USEFULNESS OF PROCESS RESEARCH

Since RCT designs leave a knowledge gap about the change mechanisms in psychotherapy, this gap has to be filled by research focusing directly on this question. Partly this can be done by using the experimental component designs with dismantling and constructive strategies (Kazdin, 2013). However, these questions can also be addressed using rigorous psychotherapy process research.

An example of such an approach took place in analyzing data from the above-mentioned trial on MBT for comorbid BPD and SUD. In the study, we used a microanalytic sequential process design, showing that within a session therapist interventions directed at exploring mental processes were connected with a subsequent higher patient level of reflective function—a study using longitudinal multilevel modelling that strengthens the possibility of concluding that this connection was causal. In other words, therapist interventions guiding the patient to explore mental processes lead to increased patient reflective functioning and accordingly, this is a change mechanism in MBT (Möller, Karlgren, Sandell, Falkenström & Philips, 2017). Such rigorous process research could be based on psychotherapy sessions from RCTs or naturalistic studies. An advantage of using data from RCTs is that they have more well-defined patient samples and treatment methods; hence, it is easier to know which generalizations of the findings are adequate.

PRAGMATIC CASE STUDIES

Pragmatic case studies following the guidelines of this journal have a great potential for increasing our knowledge about psychotherapy and what makes it work (or not work) for particular patients. In comparison with the more anecdotal case studies that are common in the psychotherapy literature, pragmatic case studies have substantially higher scientific value because they are comprehensive and based on rigorous methodological strategies: the use of thorough session notes from the therapist, the use of video- or audio-taped sessions, and the use

of standardized, quantitative patient self-rating scales that constitute a triangulation of the qualitative analysis employing a “mixed methods” model (Fishman, Messer, Edwards & Dattilio, 2017).

Using this methodology, the present case studies by Frankl et al. (2020) provide some important qualitative and quantitative evidence that brief APT can be helpful for at least some women—like Carey—who have AUD but who also have basically benign and trustful expectations of other people, such as the therapist, and who have not developed a severe physical alcohol dependence. Likewise, the case studies provide some parallel evidence that brief APT is insufficient for some women—like Michelle and Mary—who have AUD together with substantial insecurity and mistrust in their relational expectations and/or a severe physical alcohol dependence. The nature of this evidence that the APT therapy had a causal impact in effecting substantial change for Carey but not for Michelle and Mary is based on a combination of the logic of systematic observation and analysis of clinical narrative integrated with concurrent quantitative case analysis (e.g., Fishman et al., 2017), an approach illustrated in the Carey, Michelle, and Mary case studies (Frankl et al., 2020).

What could add even more scientific rigor to pragmatic case studies would be if the clinical narrative analysis were even more systematized by having the taped therapy sessions observed by independent assessors using established psychotherapy process measures. As illustration, I would like to mention the groundbreaking work of Enrico Jones (2000) in which he analyzed cases of psychoanalytic psychotherapy session-by-session using the Psychotherapy Process Q-set (PQS). The PQS is designed to be used for audio- or videotaped records or verbatim transcripts of psychotherapy sessions and can be used for any psychotherapy theoretical approach. The instrument consists of 100 items covering a wide range of attitudes, experiences, or behaviors of the patient, his/her interaction with the therapist, and the therapist’s actions and attitudes. The 100 PQS items are arranged in an ipsative, forced normal distribution by judges. In the Jones method, the researcher first makes PQS ratings of ongoing sessions in a particular individual therapy case. The researcher then factor analyzes those ratings and identifies what Jones calls, within the context of psychoanalytic therapy, “interaction structures,” that is, factors that capture specific patterns of interaction between the patient and the therapist developing during therapy and, in successful cases, worked through and dissolved due to the therapist identifying the relational pattern and intervening.

As an example of Jones’ (2000) method, he presents the therapy of Ms. B, a young female patient with depression in 126 sessions over 21 months. A factor analysis of the PQS ratings of every other session resulted in a Q-sort factor labeled “Provoking Rescue.” This factor revealed an interaction structure in which Ms. B was passive and silent with strong feelings of inadequacy and inferiority, and in which the therapist often felt ineffective and prompted to rescue them both from the painful silence by becoming more active. Analysis of the therapy process, in the context of the Provoking Rescue factor, showed that the therapist was gradually able to interpret these interactions as, among other meanings, a way for the patient to avoid thinking and reflecting upon her painful memories and to avoid making herself vulnerable. Slowly this Provoking Rescue interaction structure decreased in frequency; and, using time-series analysis (Gottman, 1981), Jones was able to show that the decrease of the Provoking Rescue factor predicted improvement in Ms. B’s depressive symptoms and daily functioning, as assessed by standardized self-report measures. The PQS has also been used in process studies

with group designs involving various psychotherapy methods, e.g., an early study by Jones and Pulos (1993) and more recent studies in which I have been involved, such as Philips et al. (2018) and Lilliengren et al. (2019). In the latter study, the PQS was employed to study, in part, psychotherapy process in Affect Phobia Therapy.

Besides adding process assessment by independent observers, another further step that would increase the value of pragmatic case studies immensely would be systematic meta-syntheses, following procedures resembling those of quantitative meta-analyses: clear definition of the target question (which patient population and therapy method), thorough search for publications, well-defined criteria for the selection of articles, and stringent methods for analyzing the data. Such meta-syntheses of case studies would be extremely useful in advancing our knowledge about psychotherapy.

In sum, the psychotherapy research designs that I have discussed in this commentary—experimental case studies, randomized controlled trials, process research, and pragmatic case studies—are all useful, but they differ with regard to what research questions they are designed to answer and the context for which they are appropriate. Thus, these different psychotherapy research designs complement each other in our research endeavour to provide knowledge about psychotherapy and its effectiveness.

REFERENCES

- Barlow, D. H., Nock, M. K., & Hersen, M. (2009). *Single case experimental designs: Strategies for studying behavior change, 3rd edition*. Boston, MA: Pearson.
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7–18.
- Fishman, D. B., Messer, S. B., Edwards, D. J. A., & Dattilio, F. M. (2017). *Case studies within psychotherapy trials: Integrating qualitative and quantitative methods*. New York: Oxford.
- Frankl, M., Wennberg, P., Berggraf, L., & Philips, B. (2020). Affect Phobia Therapy for mild to moderate Alcohol Use Disorder: The Cases of “Carey,” “Michelle,” and “Mary.” *Pragmatic Case Studies in Psychotherapy 15*(3), Article 1, 214-257. Available: <http://pcsp.libraries.rutgers.edu/>
- Gottman, J.M. (1981). *Time series analysis*. New York: Cambridge University Press.
- Jones, E. E. (2000). *Therapeutic action: A guide to psychoanalytic therapy*. Northvale, NJ: Jason Aronson.
- Jones, E.E., & Pulos, S.M. (1993). Comparing the process in psychodynamic and cognitive-behavioral therapies. *Journal of Consulting and Clinical Psychology, 61*, 1993, 306-316.
- Kazdin, A. (2013). *Research design in clinical psychology, 4th ed*. Boston, MA: Pearson.
- Lambert, M. J (2013). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change, 6th edition* (pp. 169-218). Hoboken, NJ: Wiley.
- Lee, N. K., Cameron, J., & Jenner, L. (2015). A systematic review of interventions for co-occurring substance use and borderline personality disorders. *Drug and Alcohol Review, 34*, 663–672.

B. Philips

Pragmatic Case Studies in Psychotherapy, <http://pcsp.libraries.rutgers.edu>

Volume 15, Module 3, Article 3, pp. 272-280, 01-09-20 [copyright by author]

- Lilliengren, P., Philips, B., Falkenström, F., Bergquist, M., Ulvenes, P. & Wampold, B. (2019). Comparing the treatment process in successful and unsuccessful cases in two forms of psychotherapy for cluster C personality disorders. *Psychotherapy*, *56*, 285-296.
- Möller, C., Karlgren, L., Sandell, A., Falkenström, F. & Philips, B. (2017). Mentalization-based therapy adherence and competence stimulates in-session mentalization in psychotherapy for borderline personality disorder with co-morbid substance dependence. *Psychotherapy Research*, *27*(6), 749-765.
- Philips, B., Karlsson, R., Nygren, R., Rother-Schirren, A., & Werbart, A. (2018). Early therapeutic process related to dropout in mentalization-based treatment with dual diagnosis patients. *Psychoanalytic Psychology*, *35*, 205-216.
- Philips, B., Wennberg, P., Konradsson, P. & Franck, J. (2018). Mentalization-based treatment for concurrent borderline personality disorder and substance use disorder: A randomized controlled feasibility study. *European Addiction Research*, *24*, 1-8.
- Seligman, M. E. P. (1995). The effectiveness of psychotherapy. The consumer reports study. *American Psychologist*, *50*, 965-974.