Psychoanalytische Perspectieven, 2013, 31, 1: 59-78

EXPERIMENTAL VERSUS NATURALISTIC PSYCHOTHERAPY RESEARCH: CONSEQUENCES FOR RESEARCHERS, CLINICIANS, POLICY MAKERS AND PATIENTS

Mattias Desmet

Vakgroep voor Psychoanalyse en Raadplegingspsychologie
H. Dunantlaan, 2, B-9000 Gent
Tel.: ++/32/(0)9 264 63 58, Mattias.Desmet@UGent.be

Summary: During the first half of the twentieth century, psychotherapy research was synonymous to single case research. Research and practice were highly integrated in this era, but to be considered full-fledged scientific research, the case descriptions lacked methodological rigor. From 1970 onwards, the experimental Randomized Controlled Trial (RCT) design gained momentum in the field of psychotherapy research and the single case paradigm was marginalised. In this article, it is argued that the classical RCT design is ethically troublesome, created a dramatic gap between research and practice, fails to yield the promised *objective* evaluation of the efficacy of psychotherapy, and systematically disadvantages the types of therapy that prove to be most effective in everyday clinical practice. For several reasons, returning to the classical single case paradigm, however, is not an option. As an alternative, a research program based on naturalistic effectiveness studies and empirical single case studies is put forward. It is argued that, compared to RCT research, this type of research: (1) is methodologically superior; (2) is more informative towards clinicians; (3) is a more reliable basis for anticipating cost-effectiveness of psychotherapy.

Key words: RCT Design, Experimental Psychotherapy Research, Naturalistic Psychotherapy Research.

Received: January 14th 2013; ***Accepted:*** February 4th 2013.

Breuer and Freud's *Studies on Hysteria* (1895*d*) has often been considered the historical starting point of psychotherapy research. This publication comprises five detailed case studies of hysterical women, the psychotherapeutic techniques used to treat them, and a comprehensive theoretical framework underlining the mechanisms at work in the illness (Breuer & Freud, 1895*d*: 185-251). In the subsequent decades, Freud published five further case studies that became fundamental to the worldwide dissemination of

psychoanalysis and its hegemony in the field of psychotherapy in the first half of the twentieth century (Freud, 1905*e* [1901]; 1909*b*; 1909*d*; 1911*c* [1910]; 1918*b* [1914]). Until today, Freud's case studies continue to hold importance in psychoanalytic training institutes (Kächele, Schachter, & Thomä, 2009: 106-130) and, following Freud, psychoanalysts continue to produce case studies in order to share and develop their conceptual understanding of clinical material. While no exact figures are available for the first half of the twentieth century, the total number of psychoanalytic case studies published in academic, peer-reviewed journals since that time could easily amount to several thousand (see Desmet, Meganck, Seybert, Willemsen, Geerardyn, Van Camp et al., 2013). Clinical research, practice and training is highly integrated in this respect: research is published in the form of clinical case studies, which, in turn, reflect and inform clinical practice, and provide a foundation for the training of new practitioners.

The classical single case study design, however, is frequently discredited in the academic world, particularly as psychology aspires to be a science, and science aspires towards a positivistic ideal of the empirical and experimental paradigm (Kächele et al., 2009). Case studies tend to focus on the subjective experiences articulated by patients, which some might find more reminiscent of fiction than science. Moreover, the treatment process described in case studies does not take place within the parameters of an experimentally controlled design, nor is it based on systematic qualitative or quantitative assessment procedures. Finally as such studies are always concerned with the single case, there is disagreement among authors as to whether the findings can be generalised the broader population.

In the second half of the twentieth century, the dominance of psychoanalysis gradually decreased and other therapeutic orientations – such as behavioural therapy, systemic therapy, experiential therapy, gestalt therapy, cognitive behavioural therapy (CBT), etc. – gained momentum. Randomized Controlled Trials (RCT) on the efficacy of pharmacological interventions saw their first publications at the beginning of the 1950s, and under the impetus of cognitive-behavioural theorists, psychotherapy research adopted this design as the new ideal for clinical research. In 1966, the first RCT study on the efficacy of behavioural therapy for agoraphobia was published (Gelder & Marks, 1966) and by 1980, the first RCT meta-analysis was published (Smith, Glass, & Miller, 1980). The rise of the RCT paradigm was proportionate to the decline of the classical single case study paradigm and psychotherapy almost seamlessly became, above

all, an evidence-based practice. In this study, I will highlight some of the consequences of this turning point in the history of psychotherapy research and argue that it was a mistake. It is not my belief that returning to the classical single case study paradigm is the solution. However, a naturalistic research program that combines the strengths of the single case study design with a more methodologically rigorous approach may be progressive.

Problems with RCT designs

An RCT design evaluates the efficacy of a therapy in the following way: two or more patient groups are randomly selected; in group one (the experimental group) the therapy of interest is administered; in group two (the control group) a placebo is administered. In pharmacological double-blind studies, neither the researchers nor the patients know to which treatment condition they are assigned. When the clinical trial is complete, researchers carry out statistical analyses on the data obtained to compare the effects of the treatment group with the no-treatment group.

Though the rationale of an RCT design may be appropriate for medical and pharmacological research, virtually every aspect of it (the randomization of patients across groups, double-blind administration of a specific treatment or a placebo) is highly problematic for psychotherapy research. First, a double-blind administration of psychotherapy or a control condition is impossible. The therapist obviously knows whether a *real* psychotherapy or a control treatment is being administered, which implies different expectations with respect to the treatment outcome. Different expectations, however, interfere with a measurement of efficacy that is free of bias.

Second, the use of a control group is problematic. In pharmacological research, it's clear what is being controlled for, namely the psychological effects of the intervention. In a double-blind study the administration of the placebo has the same psychological effects as those associated with the administration of the real medication, but no chemical effects. Thus, by comparing the effects of the real medication with the effects of the placebo, one can assess the net chemical effect of the medication. In psychotherapy research it's more difficult to determine what is to be controlled for. In many respects, the basic function of the control group in psychotherapy research is completely different from the function of the placebo group in pharmacological research. In psychotherapy research, it makes no sense to control for

possible psychological effects of a treatment, since all effects are psychological. The first possibility, then, is that one controls for whether the treatment outcome is better than no treatment at all. In this case, one compares a group undergoing the treatment of interest with a group that remains on the waiting list (i.e., the so-called waiting list condition). The problem, of course, is that patients in this group not only receive no treatment, but also become frustrated in terms of their demand for therapy and their belief that they will not be cured. Such a scenario is not only ethically problematic, it also makes it difficult to determine whether any eventual observation of differences between the control group and the experimental group are due to improvement in the experimental group or deterioration (due to frustration) in the control group (see Rifkin, 2007).

A possible solution for this problem is the use of a so-called *non-specific treatment condition*, which is usually considered to be the equivalent of the placebo condition in pharmacological research. Researchers believe that, apart from the specific therapeutic technique being administered, a substantial portion of the effect of psychotherapy is due to non-specific aspects of the therapeutic contact. These non-specific aspects include the attention received or the amount of treatment contact, human interaction variables such as warmth, empathy, and social support, and positive outcome expectations (Mohr, Spring, Freedland, Beckner, Arean, Hollon et al., 2009). Therefore, the *non-specific treatment condition* aims to provide an equal share of these non-specific factors as the experimental condition, i.e., the group that receives the therapy of interest. Practically speaking, this condition is often conceived as a type of non-directive speech between the therapist and patient, in which no specific therapeutic techniques are used. The problem here is that this kind of interaction is quite close to what takes place in psychodynamic and experiential therapies. In this context, however, it would concern a psychodynamic or experiential therapist that hadn't received any training and has the intention of making the therapy fail rather than to make it work. Thus, the use of such control conditions is not so much a comparison of the effects of a specific psychotherapy with a placebo, but rather a comparison of one type of psychotherapy with a caricature of another.

Other problematic aspects of RCT research become apparent when one compares it with naturalistic research (i.e., research that investigates everyday psychotherapeutic practice, discussed more thoroughly below). In an excellent review article, Westen, Novotny, and Thomspon-Brenner (2004) thus identified several problematic

aspects of RCT research worthy of discussion, three of which we summarise below.

First, RCT designs compare the efficacy of different treatments for similar patients. This requires homogeneous groups of patients at the level of complaints and symptoms. In other words, patients are selected with isolated symptoms; co-morbidity is avoided as much as possible. In everyday practice, such patients are the exception rather than the rule. In a meta-analysis focusing on exclusion rates in RCT designs, Westen and Morrison (2001) found that approximately 65% of the patient group was excluded due to Axis I co-morbidity. Furthermore, 60 to 80% of the group had one or more personality related problem (problems with intimacy, anger, self-assertion, etc.) that also led to an increased chance of being excluded from RCT studies (Morrison, Bradley, & Westen, 2003). The crux of their argument is to illustrate that patient groups included in RCT studies are not representative of the patients treated in everyday practice, which obviously limits the generalizability of their findings. One solution to this problem could be to include a more varied sample of patients in RCT studies. For example, by administering quantitative measures of co-morbidity and personality-related problems, such variables could be kept under control. However, this is highly problematic from a psychometric point of view as it assumes personality variables can be measured in an accurate way, which they cannot. We will discuss in more detail below that including such variables in an RCT design leads to a dramatic loss of power in the statistical analyses.

Second, in an ideal RCT design, treatments are as short as possible. The longer the time span, the higher the chance that confounding variables will influence the patient (e.g., spontaneous recovery from symptoms) and thus reduce the reliability of the findings. This methodological restriction means that, ideally, the psychotherapy treatment should be as short as possible in order to optimise the internal validity of the RCT design. Long-term therapies, consequently, lose scientific credibility, not because empirical research proves that they are less effective, but because they do not fit the research design as well as short therapies. Besides the fact that it is absurd to adjust the therapy to the research design instead of tuning the research design to the therapy, this practice is even more problematic when considered in the light of naturalistic psychotherapy research, which demonstrates that across all psychotherapeutic orientations, both the size and durability of the effects of psychotherapy are positively correlated with the length of the treatment (e.g., Howard, Kopta, Krause, & Orlinski,

1986; Seligman, 1995: 968). Evidently, the better a therapy works in real practice, the less it fits an RCT design.

How then can we explain the significant positive effects of short-term therapies observed in RCT designs? Evidence from naturalistic studies suggests that such observations are highly misleading. Indeed, the bulk of naturalistic studies confirm that, in terms of symptomatic relief, most patients tend to improve significantly in the first 15 sessions (Howard et al., 1986; Kopta, Howard, Lowry, & Beutler, 1994; Seligman, 1995 in Westen et al., 2004). However, Ilardi and Craighead (1994) show that a closer examination of these data reveals that, in fact, most of the improvement occurs in the first five sessions, that is, before the intervention that was proposed to yield the therapeutic effects has been administered. Other researchers even show that 15% significantly improve after making the initial phone call to the therapist's office (Kopta et al., 1994 in Westen et al., 2004). These improvements, however, are usually short term and have been interpreted as being the effect of the restoration of hope in the patient (Howard, Lueger, Maling, & Martinovich, 1993). According to Westen et al. (2004), follow-up studies show that up to 88% of the patients that terminate therapy after this initial improvement seek further treatment within the next two years. After long-term therapy, however, patients tend to maintain the progress made in the therapy and even continue to improve after the treatment (Shedler, 2010).

Third, within an RCT design, a treatment is considered an experimental manipulation, which should be as similar as possible across all patients in the same group. This requirement results in an increased pressure to manualize and protocolize treatments. However, naturalistic research of manualized treatments concludes that in everyday practice, the less a therapist sticks to the manual, the more effective he/she is (Westen et al., 2004). This again suggests that the better a therapy works in real practice, the less it fits an RCT design.

Over and above the problems associated with the RCT design itself, a more fundamental problem exists at the psychometric level. As mentioned above, RCT designs use statistical tests to formulate conclusions concerning the effects of the treatment administered. However, statistical tests always take place under the assumption that the variables under investigation are measured without error. Empirical research shows that this assumption is violated in virtually all psychological research carried out in group designs. Studies on cross-method agreement (i.e., studies that assess convergence between methodologically independent measurements of the same

psychological variable) are illustrative of this. Review studies show that cross-method agreement ranges somewhere between .00 and .45 (expressed in Pearson Correlations): correlations between self-reports and reports of spouses, partners, peers and clinicians range from .14 to .44 (Meyer, Finn, Eyde, Kay, Moreland, Dries et al., 2001); correlations between self-reports and projective tests range from .04 to .13 (*Ibid.*); correlations between self-reports and observer ratings range from .15 to .32 (*Ibid.*); correlations between explicit measures and implicit measures range from -.22 to .33 (Cunningham, Preacher, & Banaji, 2001; Bosson, Swann, & Pennebaker, 2000).

To make the size of these correlations more tangible for those not familiar with statistical analyses, Table 1 presents data of a thought experiment in which a carpenter measures the same windows by means of three different measurement methods, and in which the three resulting series of measurements correlate .45 (which equals the upper limit of cross-method agreement in psychological research).

	Measurement obtained by Method 1	Measurement obtained by Method 2	Measurement obtained by Method 3
Window 1	71 inch	51 inch	24 inch
Window 2	39 inch	79 inch	59 inch
Window 3	63 inch	87 inch	51 inch
Window 4	39 inch	67 inch	83 inch
Window 5	12 inch	39 inch	8 inch
Window 6	47 inch	31 inch	63 inch
Window 7	43 inch	59 inch	24 inch
Window 8	12 inch	35 inch	4 inch

TABLE 1: THREE SERIES OF A CARPENTER'S MEASUREMENTS OF THE SAME WINDOW THAT CORRELATE .45

It is clear from these data that convergence between psychological measurements is low. Very often, researchers do not interpret these observations as indications of measurement error, but rather conclude that distinct methods assess distinct aspects of the variables under investigation (e.g., Meyer et al., 2001). For example, lack of convergence between a questionnaire measure and an implicit measure of the personality trait *dependency* might suggest that the questionnaire taps into dependency insofar as subjects can and want to report it (i.e.,

explicit dependency), while the implicit measure would capture aspects of dependency that operate outside of the subject's awareness (i.e., implicit dependency). While this might seem plausible, empirical research challenges this interpretation. If it is correct, cross-method agreement should increase if the assessment focuses on specific aspects of psychological variables. For example, the results of methodologically independent measurements of implicit aspects of a variable (e.g., Implicit Association Test, Subliminal attitude-prime Task, Implicit Self-Evaluation Survey, etc.) should correlate strongly, and similarly, results of methodologically independent measurements of explicit aspects of a variable (e.g., self-reports, peer reports, reports of spouses, etc.) should also correlate strongly. Empirical research shows, however, that convergence between different implicit measures ranges from -.14 to .34 (e.g., Cunningham et al., 2001; Bosson et al., 2000); convergence between different explicit measures ranges from .00 to .45 (Meyer et al., 2001). This low cross-method agreement is thus observed at all levels of specificity of psychological assessment, which indicates that it should be interpreted in terms of measurement error, rather than in terms of different aspects of variables being assessed.

In an earlier study (Desmet, Roelstraete, Meganck, Inslegers, & Vanheule, 2013), we performed a mathematical simulation study to assess the impact of such portions of measurement error on statistical hypotheses testing. The findings were dramatic. As a consequence of measurement error, a real correlation of .60 – which is a very strong correlation, comparable to the correlation between the geographic variables *distance of a place to the equator* and *temperature in a place* (National Oceanic and Atmospheric Administration, 1999 in Meyer et al., 2001) – would be observed as a correlation of only .12. Our results support the argument put forward by Kalton and Schuman (1982: 43) over 25 years ago, namely that "a failure of data to fit a theory is usually as likely to cast doubt on the theory as on the measuring instruments".

In contrast to personality variables, which tend to produce very low measurement precision (Meyer et al., 2001), psychological variables that can be measured in the most precise way include *general distress* and directly observable symptoms. We argue that if psychotherapy research limits itself to the measurement of these variables, it could maintain a certain degree of reliability. However, this would mean that the most important variables for psychotherapy research (namely personality and character related variables) should be left out of

consideration. Thus, a researcher using an RCT design is in a catch-22 situation: either he excludes arguably the most important variables from the research design because they cannot be measured accurately, or he includes them, but uses measurements that are so imprecise that all ensuing statistical tests are worthless.

The lack of precision at the psychometric level leads up to profound ambiguity in research practice. For example, because outcome measures (e.g., Hamilton Rating Scale for depression or Beck Depression Inventory) often yield divergent results, several measures are often included in the same study and the researcher must choose which one to include/exclude in his data-analysis. He must also choose between different scoring methods for these measures (subscale scores or total scores, composite scores of several measures or scores for all measures separately, factor scores or raw scores, etc.). These choices are usually made on a subjective basis and have major implications for the conclusions drawn from the data. Other examples of the choices that must be made by researchers concern which variables to control for in the data-analysis, the inclusion/exclusion criteria for participants, how to handle missing values and outliers, etc. These issues open up a space in which the subjectivity of the researcher has free play.

The so-called allegiance effect (Luborsky, Singer, & Luborsky, 1975) is illustrative with respect to this issue. The allegiance effect refers to the fact that the major part of the variance (up to 73% of the total variance, i.e., a Pearson correlation of .85 between outcome and researcher allegiance, Luborsky, Diguer, Seligman, Rosenthal, Johnson, Halperin et al., 1999) in the findings of RCT studies can be explained by the theoretical orientation of the researcher. In other words, if a researcher with a primary allegiance to CBT therapy investigates the efficacy of psychotherapy, he/she will usually observe that CBT is the most effective therapy; if a psychoanalyst does the same, he/she will usually find that psychoanalysis is most effective. Thus, while the RCT design was introduced to exclude such subjective influences, it actually confirms it. Its claim of *objectivity* and scientific superiority thus appears to be unfounded.

There is other evidence that the results of RCT studies may be seriously biased. By means of the RCT design, the evidence-based practice paradigm aims to introduce mathematical and experimental rigor into the investigation of the process and outcome of psychotherapy. If this attempt is to be considered successful, the findings of different RCT studies should converge. However, the

results of meta-analyses indicate that the findings of different RCT studies do not converge. Below we will take as an example efficacy research on psychotherapy for depression, as this is studied most extensively.

Imagine the scenario: a clinician needs to decide which type of therapy he/she should recommend to a depressed patient and consults the five most recent meta-analyses that compare the efficacy of different types of psychotherapy for depression. For the sake of reliability, he consults only meta-analyses that have been published in high ranking journals. First, he finds the meta-analysis of Cuijpers, van Straten, Andersson, and van Oppen (2008*a*), which integrates results of 53 RCT studies. This meta-analysis concludes that CBT, non-directive supportive psychotherapy, behavioural activation, psychodynamic therapy, problem focused therapy, and social skills training all produced significant and equally strong effect sizes, measured immediately after the treatments ended; non-directive therapy showed significantly lower effect sizes; interpersonal therapy showed significantly higher effect sizes. Based on these findings, the clinician is inclined to recommend interpersonal psychotherapy.

However, to avoid basing his decision on only one meta-analysis, the clinician consults another study, one of Driessen, Cuijpers, de Maat, Abbass, de Jonghe, and Dekker (2010) which presents the findings of 13 RCT studies. This meta-analysis concludes that CBT, behavioural therapy, supportive therapy, non-directive counselling, and art therapy all produced significant and equally strong effect sizes immediately after the treatment; psychodynamic therapy and interpersonal therapy also produced significant effect sizes, but these effects were significantly lower than the other treatments. The clinician is now confused; the first meta-analysis indicates that interpersonal therapy is more effective than the other treatments; the second meta-analysis indicates that it is less effective.

He thus decides to consult a third meta-analysis, one published in 2010 by Tolin. This study concludes that all therapies are equally effective, with exception of psychodynamic therapy, which proved to be slightly less effective than the others. The clinician starts to become desperate. He decides to stop looking at the effects measured immediately after the treatment, as these effects don't appear to be consistent. Instead, he decides to look at the effects observed at one-year follow-up. After all, this is what is most important for his patient, whether the effects prove durable or not. In the meta-analysis of Cuijpers et al. (2008*a*), he finds that at one-year follow-up, all

differences between the different therapies disappeared. In the meta-analyses of de Maat, Dekker, Schoevers, and de Jonghe (2009) and Leichsenring and Rabung (2008), however, he finds that the results of psychodynamic therapy are not only lasting, but even improve over time.

The clinician is confused. He decides to investigate whether antidepressants are a better solution. He finds three meta-analyses published in the last decade in high-ranking journals that compare the efficacy of psychotherapy and antidepressant medication. The meta-analyses of de Maat, Dekker, Schoevers, and de Jonghe (2006) and Imel, Malterer, McKay, and Wampold (2008), report on 10 and 28 RCTs, respectively, and find no difference between antidepressant medication and psychotherapy, measured immediately after the treatment. However, a meta-analysis of Cuijpers, van Straten, Andersson, and van Oppen (2008*b*), reporting on 30 RCTs, finds that SSRIs produce significantly better results compared to psychotherapy, again measured immediately after treatment. Looking at long-term follow-up, the clinician finds that psychotherapy is superior to antidepressants in all meta-analyses.

The clinician thus decides to tell his patient that the findings from the most recent research concerning which therapy would be best for him are mixed. The only thing that is clear is that antidepressant medication doesn't appear to lead to long-term improvement and may have some negative side-effects, but could help him improve in the short term without having to make too much effort. However, the clinician must add that a placebo would probably be equally as effective, as that is what the meta-analysis of Fournier, DeRubeis, and Hollon (2010) indicates in the *Journal of the American Medical Association*, one of the most prestigious journals in the medical field. "To put it simply", the clinician concludes, "just select the therapy you feel most attracted too, but heed caution with antidepressants". Here, the evidence-based clinician has probably given the most objective and informed advice that he can, under the circumstances...

This example illustrates that, at present, the conclusions offered by RCT research with regard to the efficacy of psychotherapies must be considered questionable by clinicians and policy makers. Most clinicians already know this. According to Abma (2011) 90% of clinicians do not read RCT studies, even if they are published in the highest-ranking journals. This is not surprising; even if the findings of RCT studies did converge, one is hard pressed to know what exactly a clinician can do with them. RCT studies provide a purely statistical

description of the process and outcome of psychotherapy. Such a description does not allow the clinician relate the phenomena under investigation to the phenomena of everyday practice. In order to relate research findings to clinical practice, statistical descriptions have to be complemented with narrative accounts of the clinical processes under investigation.

Policy makers often restrict funding for clinical training and/or treatment to cognitive-behavioural methods. This decision seems to be based on the high number of RCT studies that (appear to) demonstrate the efficacy of CBT, rather than on the actual results of these studies. Not only do policy makers fail to recognise the fundamental mismatch between classical RCT designs and the very nature of psychotherapy, they also fail to recognise that, in the end, all therapies demonstrate equally strong effect sizes in RCT research (Luborsky et al., 1975). Whether this phenomenon – usually referred to as the Dodo Bird Effect – is due to uniformity in the efficacy of therapies or to the allegiance effect mentioned above doesn't actually make a difference: either way, the decision of policy makers to favour CBT is not based on sound evidence.

A promising shift in this state of affairs appears to be taking place in Sweden. Similar to other Western countries, Sweden joined the trend towards evidence-based psychotherapy and invested two billion Swedish crowns into short, manualized CBT treatments for patients with depression and anxiety. The results of this project were reported in *Socionomen*, a Swedish journal for social workers. Miller (2012) summarised the findings in the following way: "The widespread adoption of the method [of CBT] has had no effect whatsoever on the outcome of people disabled by depression and anxiety. Moreover, a significant number of people who were not disabled at the time they were treated with CBT became disabled, costing the government an additional one billion Swedish crowns. Finally, nearly a quarter of those who started treatment, dropped out, costing an additional 340 million".

Thus far, I have argued that incorporating the classical RCT design into psychotherapy research entails several ethical problems, creates a dramatic gap between research and practice, does not live up to its promise of controlling for the subjectivity of the researcher, and finally, does not lead to an efficient investment of government funding.

It would be an erroneous conclusion, however, to interpret this criticism of contemporary psychotherapy research as an argument in

favour of pharmacological treatments. Research into the efficacy of pharmacological interventions unmistakably shows pseudoscientific characteristics as well. It is partially subject to the same methodological problems as those mentioned above (e.g., problems associated with the validity of psychometric assessments) in addition to problems specific to pharmacological research itself, the most obvious of which concerns commercial interests (i.e., not publishing RCT studies which have negative outcomes for pharmaceutical companies). This has been extensively documented (e.g., Dickersin, 1990). The total publication bias is radical: meta-analyses which are exclusively based on published studies frequently find antidepressants to be significantly more effective than placebos, a finding which disappears when one includes un-published studies into the analyses (e.g., Fournier et al., 2010).

Pharmacological treatments are usually situated in a materialistic theoretical framework. From a materialist point of view, by definition, the ultimate cause is always a material cause, regardless of whether the phenomenon to be explained is situated at the material level or at the level of human subjectivity. In materialism, the totality of subjective experience is considered a side-effect, a not-intended consequence of complex material interactions. Human suffering, in this context, is a mechanical problem, and the ultimate treatment should be a surgical or biochemical (pharmacological) intervention, as a means to reverse the mechanical defect. In spite of being reductionist, or maybe because of being reductionist, this concept of man and of human suffering is dominant in contemporary scientific discourse and, in a more latent way, in mainstream discourse as well.

However, empirical research into the efficacy of pharmacological interventions – situated at the heart of the materialistic discourse – yields the most amazing demonstrations of a phenomenon that resists every explanation in materialistic terms. In such research, the observations of the placebo effect are ubiquitous to the extent that every study not controlling for it is considered inferior. Apparently, the refusal of the materialistic discourse to recognise psychological causality does not prevent it from existing. The placebo effect demonstrates the ubiquity and the impact of suggestion in the treatment of disease. Every therapeutic act, whether biological or psychological, evokes the representation of being cured, and that in itself appears to be responsible for a substantial part of its effects. Most medical interventions are in the first place psychotherapeutic interventions. These observations reveal the significance of psychotherapy and justify the investment of time and money in it.

Psychotherapy researchers, however, should use this time and money appropriately. Contemporary RCT research predominantly serves to persuade rather than to reveal. Revelation, however, is what a researcher should aim for, and honesty is the core virtue in science. When a researcher doesn't live up to that virtue but strives for other ends, such as persuasion, his work should not be qualified as science. Moreover, contemporary research should not only be rejected on the basis of methodological and ethical principles (as argued above); it is also problematic from a strategic/tactical point of view. While it might effectuate success in the short run because it meets criteria of naïve experimental science, it won't defy the ravages of time. Sooner or later, but probably sooner, the dramatic incapacity of this paradigm to lead to any valid or clinically useful conclusions will be undeniable, and this house of cards will collapse. Until then, it risks discrediting the field of clinical practice, as it systematically devalues the therapies that prove fruitful there.

Alternatives to the RCT design: Naturalistic Psychotherapy Research

The demand for evidence-based research is nevertheless justified, both from the perspective of the patient and policy maker. The development of a new evidence-based paradigm, suitable for clinical practice and psychotherapy research, is therefore imperative. Such research must be ethically sound, reintegrate research and practice, and avoid the use of pseudoscientific quantification procedures. It should also provide adequate control for the subjectivity of researcher or therapist, and must be a reliable basis for policy makers.

Below I present a concise description of a naturalistic research program that potentially meets these criteria. Naturalistic psychotherapy research is grounded in a fundamentally different epistemological conception than experimental research. In experimental designs, the researcher tries to infer the laws that determine the behaviour of the object under study by bringing it into the experimental environment and subjecting it to experimental conditions. Conversely, in naturalistic research, the researcher studies the object in its natural habitat, without disturbing it, and aims to reduce the effect of his presence as researcher to an absolute minimum.

This program comprises a combination of two strands of naturalistic research – naturalistic effectiveness research (as opposed to the experimental efficacy study, see for example Seligman, 1995:

965) and naturalistic single case study research – which are not new, but often wrongly considered inferior to the RCT design.

Effectiveness studies are important for policy makers, as an alternative to efficacy studies. In effectiveness studies, symptoms, complaints and other relevant outcome variables of a large group of patients are registered throughout the psychotherapeutic process, as it happens in everyday practice. Formal characteristics of the psychotherapeutic treatment (i.e., the type of treatment, duration of therapy, frequency of sessions, etc.) are registered. In some studies, researchers include a group of patients (i.e., a naturalistic control group) that had signalled mental health problems to a professional but did not enter therapy. Importantly, patients included in effectiveness studies are not randomized and no attempts are made to standardise the treatment given. After gathering the data, comparisons of the effects of different types of treatments are made in a way that is similar to RCT studies. Effectiveness studies are thus also limited by the use of psychometric measuring instruments, but not by the methodological shortcomings outlined above. On the condition that clinicians/researchers limit these measurements to *general distress* and *symptoms*, they can be considered methodologically sound, and as such, provide policy makers more reliable data than RCT studies.

Naturalistic single case study research, on the other hand, is indispensable for clinicians. In naturalistic case study research the therapeutic process is documented and studied in detail, and in such a way that it interferes minimally with daily practice. This, of course, does not mean that the act of investigation leaves no traces: such research is suitable only for certain patients, where the therapist can reasonably believe that the research component will not interfere with the therapeutic process. In permissible cases, the therapy sessions are recorded with the informed consent of the patient. Throughout the treatment, psychological and biological variables are measured, medication use and health care expenses are documented. Contrary to what is often believed, recording sessions proves to have little or no inhibiting effect on patients' speech (Kächele et al., 2009). Psychometric evaluations are limited to self-report questionnaires, assessing *general distress* and the patient's specific symptoms. Few studies on the validity of psychometric measurements in single case study designs have been carried out until now. However, it can be anticipated that for most variables included in single case study designs, psychometric measurements will be far more precise compared to group designs. Important sources of measurement error,

such as differential interpretations of scales between subjects and spontaneous fluctuations in the variables under investigation, are absent in single case study research or can be controlled for. Future psychometric research should nevertheless address this issue. In addition to the therapy recordings and the self-report data, saliva samples can be used to register evolutions in the concentrations of hormones, such as cortisol, testosterone, and androsterone, and to detect changes at the level of RNA and DNA mechanisms throughout the therapeutic process. Gathering saliva samples is a non-intrusive and reliable method of investigating the biological effects of psychotherapy. In this respect, it is preferable to brain imaging techniques which interfere substantially with the naturalistic course of the therapy, and the results of which are often hard to interpret (Vull, Harris, Winkielman, & Pashler, 2009).

When the therapy is complete, the therapy recordings are transcribed verbatim and the researcher can carry out an analysis of the data. There are numerous ways in which the analysis can be carried out, depending on the theory the researcher uses. In the first stage of the analysis a clinical description – a classical single case study – of the therapeutic process is written by the therapist. The primary goal of the subsequent data analysis is to challenge this clinical description. The core variables put forward in the clinical description can be systematically analysed in the audiotapes transcriptions to study the extent to which it matches or diverges from the data. In this way the initial clinical case description can be reviewed until the researchers have a comprehensive qualitative case study which captures the main aspects of the clinical process. This case description gives an outline of important turning points that took place during the therapy and, more specifically, at which stages of the therapy these turning points occurred. It also presents an extensive qualitative description of such turning points. In the final phase, statistical analyses of the data can tap into the degree to which these turning points are predictive of evolutions in the scores on the self-report measures of *general distress* and symptoms, as well as evolutions in biological variables. The final report should present both a classical clinical case description and an empirical analysis of the specific aspects of this clinical description. The clinical description will permit clinicians to relate the findings to their own practice; the empirical data, on the other hand, will control for any subjective effects of the researcher and participating therapist.

Each case must be conceived as a piece of craftsmanship. From the moment one tries to force all clients into exactly the same procedure, one crosses an ethical boundary. The craft of psychotherapy has to be cultivated through conceptually guided, careful study of series of single cases, in ones own clinical practice and in that of others. Academic research should provide clinicians with a database of detailed case studies that are accessible and useful for clinical practice. Such case studies will not provide clinicians with straightforward advice on how to treat patients with similar complaints; they only provide a systematised clinical database which can help clinicians to see patterns in the complexity of psychotherapeutic processes. Every clinician has his/her own particular way of practicing psychotherapy. In generalising findings across cases, statistical aggregation of findings is of minor importance. As soon as one begins to aggregate quantitative observations across cases, one is subject to the same psychometric limitations as group designs.

The research strategy outlined above implies turning away from a naïve, positivist conception of evidence-based practice. In such a (utopic) conception, academic researchers develop psychotherapeutic procedures through experimental research, and clinicians apply these procedures, blindly trusting the *empirical evidence*. Such an approach is centred on the empirically validated procedure itself, which ought to be applicable across patient groups. The subjectivity of the patient is considered of secondary importance, something that should have as low an impact as possible on the therapy. From an ethical perspective, such a model risks repeating what is probably the major cause of psychic suffering: namely the denial of one's subjectivity by someone else.

Conclusion

In this study we argued that the contemporary use of the RCT research design in psychotherapy research is ethically problematic, it entails a dramatic gap between research and practice, it does not control for the subjective effect of the researcher, and leads to flawed conclusions. In this respect, it systematically discredits the types of therapy that lead to lasting effects, while promoting therapies that usually do not work in real practice. In their struggle to survive, long term and non-manualized treatments have attempted to force themselves into the straightjacket of the RCT design. By doing so, long-term treatments have succeeded in gaining some credibility in

the field of evidence-based practice, and as such are still on the map today. Instead of putting time and money into this questionable enterprise, more energy should be invested in naturalistic effectiveness research and in naturalistic single case study research. Naturalistic effectiveness research is important to show policy makers the beneficial effects of psychotherapy. As such, it is a more reliable basis to evaluate the cost-effectiveness of psychotherapy than RCT research. Naturalistic single case study research, on the other hand, is particularly important for clinicians, as it has the potential to be psychometrically sound, to close the gap between research and clinical practice, and to open up a new and powerful scientific approach to psychological causality and the interplay between psychic and physical processes.

References

R. Abma (2010), *Evidentie en existentie: Evidence-based behandelen en verder*, Tilburg, Kenniscentrum voor Levensbeschouwing en Geestelijke Volksgezondheid.

J.K. Bosson, W.B. Swann, & J.W. Pennebaker (2000), "Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited?", *Journal of Personality and Social Psychology*, vol. 79, pp. 631-643.

J. Breuer, & S. Freud (1895*d*), *Studies on hysteria*, *S.E.*, II.

P. Cuijpers, A. van Straten, G. Andersson, & P. van Oppen (2008*a*), "Psychotherapy for Depression in Adults: A Meta-Analysis of Comparative Outcome Studies", *Journal of Consulting and Clinical Psychology*, vol. 76, pp. 909-922.

P. Cuijpers, A. van Straten, G. Andersson, & P. van Oppen (2008*b*), "Are psychological and pharmacological interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies", *Journal of Consulting and Clinical Psychology*, vol. 69, pp. 1675-1685.

W.A. Cunningham, J.K. Preacher, & M.R. Banaji (2001), "Implicit attitude measures: Consistency, stability, and convergent validity", *Psychological Science*, vol. 12, pp. 163-170.

S.M. de Maat, J. Dekker, R.A. Schoevers, & F. de Jonghe (2006), "Relative efficacy of psychotherapy and combined therapy in the treatment of depression: A meta-analysis", *Psychotherapy Research*, vol. 16, pp. 566-578.

S.M. de Maat, J. Dekker, R.A. Schoevers, & F. de Jonghe (2009), "Relative efficacy of long-term psychoanalytic therapy: A systematic review of empirical studies", *Harvard Review of Psychiatry*, vol. 17, pp. 1-23.

M. Desmet, R. Meganck, C. Seybert, J. Willemsen, F. Geerardyn, I. Van Camp, F. Declercq, R. Inslegers, E. Trenson, S. Vanheule, L. Kirshner, I. Schindler, & H. Kächele (2013), "Psychoanalytic single cases published in ISI-ranked journals: The construction of an online archive", *Psychotherapy and Psychosomatics*, vol. 82, pp. 120-121.

M. Desmet, B. Roelstraete, R. Meganck, R. Inslegers, & S. Vanheule (2013), "Measurement error and its consequences for psychological research: Problems and solutions", under review.

K. Dickersin (1990), "The existence of publication bias and risk factors for its occurrence", *Journal of the American Medical Association*, vol. 263, 1385-1359.

E. Driessen, P. Cuijpers, S.C.M. de Maat, A.A. Abbass, F. de Jonghe, & J.J.M. Dekker (2010), "The efficacy of short-term psychodynamic psychotherapy for depression: A meta-analysis", *Clinical Psychology Review*, vol. 30, pp. 25-36.

J.C. Fournier, R.J. DeRubeis, & S.D. Hollon (2010), "Antidepressant drug effects and depression severity: A patient-level meta-analysis", *Journal of the American Medical Association*, vol. 303, pp. 47-53.

S. Freud (1905*e* [1901]), "Fragment of an analysis of a case of hysteria", *S.E.*, VII, pp. 7-122.

S. Freud (1909*b*), "Analysis of a phobia in a five-year-old boy", *S.E.*, X, pp. 5-149.

S. Freud (1909*d*), "Notes upon a case of obsessional neurosis", *S.E.*, X, pp. 151-249.

S. Freud (1911*c* [1910]), "Psychoanalytic notes upon an autobiographical account of a case of paranoia (dementia paranoides)", *S.E.*, XII, pp. 9-82.

S. Freud (1918*b* [1914]), "From the history of an infantile neurosis", *S.E.*, XVII, pp. 1-122.

M.G. Gelder, & I.M. Marks (1966), "Severe agoraphobia: A controlled prospective trial of behavior therapy", *British Journal of Psychiatry*, vol. 112, pp. 309-319.

K.I. Howard, S.M. Kopta, M.S. Krause, & D.E. Orlinsky (1986), "The dose-effect relationship in psychotherapy", *American Psychologist*, vol. 41, pp. 159-164.

K.I. Howard, R. Lueger, M. Maling, & Z. Martinovich (1993), "A phase model of psychotherapy: Causal mediation of outcome", *Journal of Consulting and Clinical Psychology*, vol. 54, pp. 106-110.

S.S. Ilardi, & W.E. Graighaid (1994), "The role of nonspecific factors in cognitive-behavior therapy for depression", *Clinical Psychology: Science and Practice*, vol. 1, pp. 138-156.

Z.E. Imel, M.B. Malterer, K.M. McKay, & B.E. Wampold (2008), "A meta-analysis of psychotherapy and medication in unipolar depression and dysthymia", *Journal of Affective Disorders*, vol. 110, pp. 197-206.

H. Kächele, J. Schachter, & H. Thomä (2009), *From psychoanalytic narrative to empirical single case research*, New York, Routledge.

G. Kalton, & H. Schuman (1982), "The effect of the question on survey responses: A review", *Journal of the Royal Statistical Society Series A-statistics in Society*, vol. 145, pp. 42-73.

S. Kopta, K. Howard, J. Lowry, & L. Beutler (1994), "Patterns of symptomatic recovery in psychotherapy", *Journal of Consulting and Clinical Psychology*, vol. 62, pp. 1009-1016.

F. Leichsenring, & S. Rabung (2008), "Effectiveness of long-term psychodynamic therapy: A meta-analysis", *Journal of the American Medical Association*, vol. 300, pp. 1551-1565.

L. Luborsky, L. Diguer, D.A. Seligman, R. Rosenthal, S. Johnson, G. Halperin, M. Bishop, & E. Schweizer (1999), "The researcher's own therapeutic allegiances: A 'wild card' in comparisons of treatment efficacy", *Clinical Psychology: Science and Practice*, vol. 6, pp. 95-132.

L. Luborsky, B. Singer, & L. Luborsky (1975), "Comparative studies of psychotherapy", *Archives of General Psychiatry*, vol. 32, pp. 995-1008.

G.J. Meyer, S.E. Finn, L.D. Eyde, G.G. Kay, K.L. Moreland, R.R. Dries et al. (2001), "Psychological testing and psychological assessment", *American Psychologist*, vol. 56, pp. 128-165.

S. Miller (2012), http://scottdmiller.com/icce/revolution-in-swedish-mental-health-practice-the-cognitive-behavioral-therapy-monopoly-gives-way/

D.C. Mohr, B. Spring, K.E. Freedland, V. Beckner, P. Arean, S.D. Hollon, J. Ockene, & R. Kaplan (2009), "The selection and design of control conditions in randomized controlled trials of psychological interventions", *Psychotherapy and Psychosomatics*, vol. 78, pp. 275-284.

C. Morrison, R. Bradley, & D. Westen (2003), "The external validity of efficacy trials for depression and anxiety: A naturalistic study", *Psychology and Psychotherapy: Theory, Research and Practice*, vol. 76, pp. 109-132.

National Oceanic and Atmospheric Administration (1999), "Climate research data: The daily historical climatology network", Raw data retrieved December 9, 1999, from the World Wide Web: http://www.ncdc.noaa.gov/ol/climate/research/ushen/daily.html.

A. Rifkin (2007), "Randomized controlled trials and psychotherapy research", *American Journal of Psychiatry*, vol. 164, pp. 7-8.

M.E.P. Seligman (1995), "The effectiveness of psychotherapy", *American Psychologist*, vol. 50, pp. 965-974.

J. Shedler (2010), "The efficacy of psychodynamic therapy", *American psychologist*, vol. 65, pp. 98-108.

M.L. Smith, G.V. Glass, & T.I. Miller (1980), *The benefits of psychotherapy*, Baltimore, MD, Johns Hopkins University Press.

D.F. Tolin (2010), "Is cognitive-behavioral therapy more effective than other therapies? A meta-analytic review", *Clinical Psychology Review*, vol. 30, pp. 710-720.

E. Vull, C. Harris, P. Winkielman, & H. Pashler (2009), "Puzzlingly high correlations in fMRI studies on emotion, personality, and social cognition", *Perspectives on Psychological Science*, vol. 4, pp. 274-290.

D. Westen, & K. Morisson (2001), "A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: an empirical examination of the status of empirically supported therapies", *Journal of Consulting and Clinical Psychology*, vol. 69, pp. 875-899.

D. Westen, C.M. Novotny, & H. Thomspon-Brenner (2004), "The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials", *Psychological Bulletin*, vol. 130, pp. 631-663.