PSYCHOLOGY MISCELLANY

No.225 - August 2025

Evaluating Therapy and Treatment for Mental Disorders

Kevin Brewer

ISSN: 1754-2200

orsettpsychologicalservices@phonecoop.coop

This document is produced under two principles:

1. All work is sourced to the original authors. The images are all available in the public domain (most from http://commons.wikimedia.org/wiki/Main_Page). You are free to use this document, but, please, quote the source (Kevin Brewer 2025) and do not claim it as you own work.

2. Details of the author are included so that the level of expertise of the writer can be assessed. This compares to documents which are not named and it is not possible to tell if the writer has any knowledge about their subject.

This document is presented for human readers.

Kevin Brewer BSocSc, MSc

An independent academic psychologist, based in England, who has written extensively on different areas of psychology with an emphasis on the critical stance towards traditional ideas.

A complete listing of his writings at http://psychologywritings.synthasite.com/. See also material at https://archive.org/details/orsett-psych.

CONTENTS

1. FACTORS IN THE EVALUATION

1.1. INTRODUCTION

In relation to mental health issues [1], two questions stand out:

1. Does therapy/treatment work? [2]

2. Is one particular therapy/treatment better than another? [3]

Definitive answers are sought, but rarely found. In terms of generalisations, it is probably fair to say, "it depends", and "it varies". Or put another way, certain therapies/treatments work for certain individuals with certain problems in certain situations. This is obviously not very satisfactory. There are studies (and of good

[1] Other terms include mental health problems, mental disorders, mental illness, psychopathology, psychological problems, and psychiatric problems. They all refer in different ways to problems that an individual experiences in relation to their mental health. This is to distinguish it from physical health, though the two are interlinked, and such distinctions are often false.

[2] "Treatment" is used to refer to those that induce a physical changes (eg: medication; surgery) while "therapy" focuses upon producing psychological change. This distinction is not without overlap and problems, but it is commonly used and convenient.

[3] These two questions can be applied to physical health problems, and many of the same issues addressed will arise.

methodological quality) that show success, but the point is about generalisations. What follows is an outline of factors, issues, and problems that make answering the two questions above in a definitive manner difficult.

Ephron (2023) commented: "People and their problems are, by definition, individual, so it has proven difficult to compare like with like. And with estimates suggesting there are more than 500 types of therapy on offer, most yet to be tested in rigorous trials, linking outcomes to particular treatments is difficult" (p39).

McLeod (1997) observed that "a single piece of research, no matter how well thought-out, can in itself make any only a limited contribution to understanding. Research is like a huge jigsaw or mosaic, with individual studies serving to 'fill in' a bigger pattern. Each study must be read within the context of the global picture that is the research 'literature'" (p490).

The most straightforward way is to compare a group of people in therapy with those not to see who improves. This has been done many times, but there are problems with such studies - some are methodologically sounder than others. So not all studies are equal (table 1.1).

- Defining and measuring recovery.
- Different between therapists who use same technique (ie: "good"/"bad" therapists).
- Difference between clients (ie: some individuals more motivated than others).
- Spontaneous recovery (some individuals get better without therapy).
- What is happening outside the therapy to the client.

Table 1.1 - Main problems of trying to establish if therapy works compared to no therapy.

1.2. GETTING BETTER WITHOUT TREATMENT

1.2.1. Spontaneous Recovery or Remission

Some conditions wax and wane over time, and symptoms come and go in intensity. So, some individuals left without treatment would have improved anyway. Sometimes

people change their behaviour, not necessarily through conscious planning, but in response to their symptoms, and this leads to the improvement. This could be one reason why individuals in the placebo or control group improve, and thus the importance of having such groups as comparisons for the treatment group.

External factors like life events could have an impact here.

When individuals improve without help, there is a temptation to suggest that the condition was not that bad. This could be linked to the "hello-goodbye effect" (Cronbach 1990). The idea is that individuals exaggerate the severity of their symptoms when first seeking help ("hello"), but in time play down or report the true severity, particularly when ending treatment ("goodbye"). The success of treatment may also be exaggerated to "please" health professionals. Cronbach's (1990) original use was in reference to participants wanting to create "an overly positive image" of the success of a psychotherapeutic change programme for offenders which they completed (Walters et al 2002).

So, "when patients first present for treatment, they need to justify their request for help, so there is a sub-conscious tendency to exaggerate symptoms. After patients have received treatment, they may want to please the therapist, or at least not hurt the therapist's feelings, so there is a tendency to minimise symptoms" (Choi and Pak 2007 p430).

Choi and Pak (2007) sub-divided the hello-goodbye effect into:

a) "Faking bad bias" - "whereby participants try to appear sick to qualify for support" (p430).

b) "Faking good bias" (or social desirability or obsequiousness bias) - "whereby participants may systematically respond in the direction they perceive to be desired by the investigator" (p430).

Example: Post-Traumatic Stress Disorder

After experiencing a traumatic event, sone individuals will develop symptoms described as Post-Traumatic Stress Disorder (PTSD), and these may continue over years, while others will show spontaneous remission. There is great interest in predicting which trajectory an individual will take.

One example of a study is by Kolassa et al (2010), who investigated 444 refugees, from the Rwandan conflicts in the early 1990s, living in Uganda in 2006-07. A checklist of thirty war- and non-war-related traumatic event types were used, along with the seventeen-item "Post-traumatic Diagnostic Scale" (PDS) (Foa et al 1997).

Kolassa et al (2010) summed up: "The results show a clear dose-response effect of traumatic load on the probability of long-term spontaneous remission from PTSD: Higher cumulative exposure to traumatic events is associated with a lower probability of remission, with an additional traumatic event type experienced associated with an 8% lower chance of spontaneous remission from PTSD. In addition, this study confirmed previous findings of dose-response effects, with higher numbers of traumatic event types experienced associated with higher probabilities of developing lifetime and current PTSD as well as higher severity of lifetime and current PTSD symptoms" (pp171-172).

1.2.2. Changes in the Control Group

There are a number of other possible reasons why individuals not receiving the treatment or therapy may change, including:

1. Placebo (appendix 1A), expectations, or self-fulfilling prophecy (or nocebo – getting worse due to expectations).

2. Improvements occur simply due to length of time, while some treatments may take some time to work.

3. Being part of a study/"Hawthorne effect".

4. Cognitive dissonance and treatment.

1.3. METHODOLOGICAL ISSUES IN STUDY DESIGN

1.3.1. Control Group

A good study should have a control (or placebo) group. This is a group that does not receive the therapy or treatment, and ideally the members do not know this (known as "blinding"). In the case of medication, the control group will receive a placebo (a "dummy pill"). There is an ethical problem with a group who deliberately

has help withheld. One way around this concern is to use individuals on a waiting-list for therapy or treatment.

The control group shows the number of individuals who improve without help, and so the treatment can be compared to this, not to zero improvement. A certain number of individuals improve because of the "placebo effect".

Another issue with placebo medication is that individuals may know because they do not experience the side effects of the treatment medication. Another ethical issue arises of whether to give a placebo which mimics the side effects of the medication (ie: deliberately make the participants ill in some way) just to maintain blinding.

The best design of a study of a medication would have a treatment group (or groups if different dosages are being tested), a placebo group, and a control group (receiving no medication or placebo).

1.3.2. Blinding

Concealing who receives the treatment and who receives the placebo is a way to reduce the impact of expectations.

The main types of blinding include:

i) None - eg: "open label" studies. The participant and the researcher know who is receiving the treatment and who is part of the control group. Though there is a risk of expectations influencing the findings, it may be study the only way to study the treatment or therapy. There is no ethical issue with deception here.

ii) "Blind-at-randomisation" - Participants are randomised blindly to a group, and then it is revealed to the participant and/or the researcher which group is treatment and which group is placebo (Huson 2000).

iii) Single blind - The participants do not know their group, but the researchers do. There is a risk of the influence of researcher (or experimenter) expectancy effects.

iv) Double-blind - Neither the researcher nor the participant know who is receiving treatment or placebo. This removes the expectancy effects on both sides, in theory. It is achieved by having researchers who interact with the participants not knowing which group is which.

In other words, different researchers administer the trial to those who collect the data.

v) "Expectancy control design" - The participants are divided into a treatment and a control group, but also divided again based on the information told to the data collectors. Half the treatment group is presented to the researchers collecting the data as receiving the treatment, and half as receiving the placebo, and the same for the control group. This design involves four groups (table 1.2) (Coolican 2004). It involves deceiving the researchers who collect the data.

		Information given to data collectors	
		Treatment	Control/Placebo
Actual Group	Treatment	1 (true)	2 (false)
	Control/Placebo	3 (false)	4 (true)

Table 1.2 - Expectancy control design.

An example of this design comes an unpublished thesis from Burnham (1966 quoted in Rosnow and Rosenthal 1997). Burnham studied the impact of experimenter expectations on rats learning a maze. Half the rats had a part of the brain removed after learning ("lesioned") and half underwent "sham surgery" ("non-lesioned"). Twenty-three "experimenters" were employed to measure a rat's performance after surgery, either being told the truth or a lie about lesion/non-lesioned. It was expected that non-lesioned rats would perform better. Objectively, non-lesioned rats did perform better than lesioned rats. But lesioned rats who the "experimenters" believed were non-lesioned performed better than lesioned rats described as so, while non-lesioned rats labelled as lesioned performed worse than non-lesioned presented as such to the "experimenters". Rosnow and Rosenthal (1997) commented: "What makes this experiment of special interest is that the effects of expectancy... were somewhat larger than those of the actual removal of brain tissue..." (p61).

1.3.3. The Sample

There are a number of issues related to the

individuals who take part in the study (ie: the sample). Ideally, they will be a small representation of the population as a whole. "Population" here refers to the group that the therapy or treatment is aimed at - it may be the general population, or all individuals with major depression, say.

1. Means of recruitment - eg: via specialist clinics (which includes only individuals who seek help or were referred there); advertisements (where and how done).

2. Size.

3. Inclusion and exclusion criteria - Most importantly, there should be equality of symptoms (and severity) between the control and treatment groups.

4. Drop-out

- Before randomisation
- After randomisation, but discovered later that should have been excluded
- Given wrong treatment
- Due to side effects
- Leave during trial for some reason

1.3.4. Outcome Measure

The outcome measure(s) chosen is crucial. It is the yardstick by which success or failure is assessed. McLeod (1997) outlined four issues here:

i) Selecting measures that are sensitive to the different dimensions of change. For example, improvements in depressive symptoms includes changes in mood, thoughts, and behaviours. A single measure may miss change compared to multiple outcomes.

ii) Having self-report measures that are not too long (ie: “not overburdening the client”; McLeod 1997 p495). A very long questionnaire to measure outcomes may be a sign of having too many outcome measures.

iii) Assessing potential change that is consistent with the treatment/therapy used. For example, cognitive-behavioural therapy (CBT) primarily aims to change thinking about a problem.

iv) Having reliable and valid outcome measures for the group studied.

Particularly for assessing counselling, but relevant generally with therapy and treatment, Lambert et al (1992) recommended three types of outcomes to measure - intrapersonal (mood, thoughts, behaviours), interpersonal relations (eg: with significant others), and social role (ie: "contribution to society"; McLeod 1997 p495).

The length of time of follow-up is also relevant here. In other words, when to apply the measure of success or failure - at the end of the treatment or some time later? Good practice is to have multiple measurement time points. A follow-up, say, one year after treatment ends also shows the long-term benefits of the treatment and the risk of relapse.

Surrogate Markers

"Surrogate markers are often used in clinical trials settings where obtaining a final outcome to evaluate the effectiveness of a treatment requires a long wait, is expensive, or both. Informally, a surrogate marker can be thought of as a replacement for the final outcome in that the treatment will impact the surrogate in the same manner as the actual outcome of interest" (Elliott 2023 p76).

Surrogate markers (also called "surrogate end points") are common in health research, as in CD4 cell counts as an indicator of immune response in AIDS treatment, and prostate specific antigen (PSA) levels in treatment for prostate cancer (Elliott 2023).

"The ideal surrogate is one in the causal pathway between the treatment effect and the final outcome, so that whatever impact the treatment has on the outcome will have to affect the surrogate as well" (Elliott 2023 p76).

There is a problem of the "surrogate paradox" (Chen et al 2007) - "the situation where surrogate markers badly mis-estimate the treatment effect, potentially leading to harm if damaging treatments are approved on the basis of a positive treatment effect in a surrogate marker" (Elliott 2023 p76).

1.3.5. Other Issues

1. Baseline – eg: what measures taken; when the study begins.

2. Dosage of medication or frequency of therapy to use.

3. Randomisation.

1.4. THERAPIST

"Clients of the best therapists improve at a rate at least 50 per cent higher and drop out at a rate at least 50 per cent lower than those of average clinicians" (Miller et al 2008 p14). This observation came from a study of nearly 600 psychologists, psychiatrists, and therapists, and over 6000 clients (Wampold and Brown 2006). The successful individuals have been called "supershrinks". Ricks (1974) coined this term in a study of "high disturbed" adolescents.

Success of therapy outcome is not linked to client's age, gender, diagnosis, level of functional impairment, or prior treatment history, nor therapist's age, gender, training, professional discipline, and years of experience (Miller et al 2008).

Miller et al (2008) viewed the therapist's "performance" with a client through the lens of sports performance, for example. In particular the use of feedback, especially negative, to improve subsequent performance. "Supershrinks... are exquisitely attuned to the vicissitudes of client engagement. In what amounts to a quantum difference between themselves and average therapists, they are more likely to ask for and receive negative feedback about the quality of the work and their contribution to the alliance" (Miller et al 2008 p20).

More worryingly, the least effective therapists believe themselves to be on a par with the best (Hiatt and Hargrave 1995).

A drug is not a drug, as Wampold and Brown (2006) found that medication prescribed by the best therapists "achieved gains from the drugs 10 times greater than those seen by the less effective practitioners. Among the latter group, the drugs virtually made no difference. So, in the chemistry of mental health treatment, orientations, techniques, and even medications are inert.

The clinician is the catalyst" (Miller et al 2008 pp15-16).

1.4.1. Therapeutic Alliance

The "therapist-client alliance" (or therapeutic alliance; TA) can be viewed as an indicator of the potential success of therapy. Put simply, it is a feeling that both parties are working together ("collaborative engagement") to solve the client's problems, or as Fluckiger et al (2020) put it, "the subjective experience of a collaborative, trusting and goal-oriented environment" (p706). So a high alliance should predict successful therapy outcomes. But there are many confounders like intake characteristics (eg: initial distress level), and process variables (eg: theoretical approach of therapist).

One meta-analysis of sixty studies (Fluckiger et al 2020) found their TA was an independent predictor of outcome of therapy (eg: symptom reduction). In summary: "The alliance is a robust predictor of outcome at the between-patient level. Patients who report a stronger alliance during treatment are also likely to report better treatment outcome. This association remains significant when controlling for patients' intake characteristics and therapists' adherence and competence" (Fluckiger et al 2020 p706).

Goldsmith et al (2015) claimed to "the first ever demonstration that TA has a causal effect on symptomatic outcome of a psychological treatment, and that poor TA is actively detrimental" (p2365). These researchers concentrated on individuals with first episode psychosis, and compared CBT, supportive counselling, and routine care (control group) over six weeks.

TA was measured by a therapist-rated measure, and the client completing the "California Therapeutic Alliance Scale" (CALPAS) (Marmar et al 1989) at Week 3 [4]. The outcome of the psychosis was measured with the "Positive and Negative Syndrome Scale" (PANSS) (Kay et al 1987) at 18 months. There were 308 participants involved in the study in the UK.

The best TA scores were associated with improvements in symptoms, while the worst TA scores were associated

[4] Examples of items: "Did you feel pressured by your therapist to make changes before you were ready?", and "Did you feel that you were working together with your therapist, that the two of you were joined in a struggle to overcome your problem?" (Gaston and Marmar 1993).

with declines in PANSS score.

1.4.2. Outcome vs Process

Most clinical trials are outcome research (ie: the focus is upon the improvement on the outcome measure, like severity of symptoms). But there is also process research which investigates the experience of therapy rather than the end point.

There is process-outcome research which combines both approaches, and can include isolating the variables that characterise effective therapy, like a warm relationship between client and therapist, a special setting encouraging help, and activities that produce change (Frank 1974).

For example, Bowman and Fine (2000) explored the perceptions of positive therapy experiences for five heterosexual couples undergoing relationship-related therapy. The characteristics included:

- The therapist seems caring (irrelevant of the suggestions of help made).
- Empathy and acceptance felt.
- Homework.
- A safe place to talk.

Dallos and Cullen (1990) suggested characteristics like empathy by the therapist, a good relationship and communication between client and therapist, and an analysis of behaviours that maintain the problem or create cycles of behaviour.

1.5. PATIENTS

The focus upon the therapy or treatment as working or not can ignore a key variable, namely the receiver of the therapy/treatment – ie: the patient, client, service user, or other terms used.

There are a number of factors relevant here.

i) Demographics – eg: sex/gender; ethnicity/race; social class/socio-economic status; sexual orientation.

Each of these will influence the experience of the therapy in some way, particularly in terms of the

interaction with the therapist. Put simply, a male therapist with a female client, say, will be different to a male therapist with a male client.

iii) Individual differences – eg: personality.

iii) Motivation to improve – Individuals will vary here, and this can influence whether the therapy works, particularly where the patient has to do something as part of their treatment (eg: keep a diary of negative thoughts; practice positive thinking).

The attitude towards therapy or treatment is also a potential factor here. Some individuals may have been pressurised in some way to undertake the treatment, and so have a different attitude compared to voluntary help-seekers. Individuals may vary in their beliefs about whether the treatment will work as well as of treatment generally.

It has been found that "intelligent and educated individuals whose problems are not too great and who have a positive attitude towards therapy will benefit most" (Brewer 2001 p76).

iv) External circumstances – During the period of a study or during treatment or therapy, things will happen to the individual outside of the focus that impact the individual or their therapy/treatment. This includes life events (serious and minor), for example.

v) Symptom severity – Good studies should have comparable symptom severity levels in their comparison groups (ie: treatment vs placebo). A treatment group containing individuals with low severity may show improvements over a placebo group full of individuals with high severity, irrelevant of the therapy or treatment.

There will be individual differences in how symptom severity is experienced – ie: some individuals can tolerate more than others – as well as the response to the amount of improvement in symptoms. For example, disappointed patients who feel that their improvement is not as great as expected may give more negative ratings of their perceived improvement in symptoms. This is one of the risks of using a subjective measure scale. The opposite applies with patients where improvements are beyond expectations. Expectations of improvement are linked to attitudes towards treatment and therapy.

1.6. MISCELLANEOUS

Wirtz et al (2022) made this observation about randomised controlled trials: "Although considered the gold standard, they are increasingly critiqued as expensive and logistically complicated, given the large sample sizes required for some trials, long periods of enrolment, and significant staffing and infrastructure for biologic and survey data collection . Barriers to participation in trials have historically been attributed to the time commitment, costs associated with lost productivity and transit for study participation, and competing priorities that hinder enrolment and participation" (p87).

So, the possibility of "digital clinical trials" (DCTs) is very attractive. "The expression digital clinical trials is an umbrella term that largely captures clinical trials that range from partial to full integration of technology in trial implementation, interventions, and/or data collection" (Wirtz et al 2022 p88).

Whatever the advantages of DCTs, there is an inequity based on access to digital technology. This includes ownership of the necessary equipment (eg: smartphone), as well as lack of infrastructure (eg: broadband or Wi-Fi access in rural areas). These inequities will vary with gender, age, income, social class [5], ethnicity, education [6], and geographical divide both within and between societies.

Mobile phone use is rising rapidly throughout the world which is a positive here. Wirtz et al (2022) noted caution, however: "Estimates of ubiquitous phone ownership mask a critical appraisal of consistent mobile phone, smartphone, and internet connection over time that is often much lower than estimates of ownership or use for many populations and persist despite expansion of device ownership. Consistent mobile phone and internet access can be affected by individual inability to pay for service, theft or lending of devices, temporary use of unregistered SIM cards in countries that require registration to an identity document, and by other structural factors" (p88).

Wirtz et al (2022) reviewed the studies of DCTs in HIV/AIDS treatment (table 1.3). For example, a DCT in

[5] "Digital redlining" is a term used to describe the "creation and maintenance of technological policies, practices, pedagogy, and investment decisions that enforce class boundaries and discriminate against specific groups" (Gillard 2017 quoted in Wirtz et al 2022).

[6] There is approximately one-seventh of the world's population that is not literate (Wirtz et al 2022).

South Africa (Venter et al 2018) using short messages via an Android-based app screened 4500 potential participants, of which 3540 passed, but only 350 were available for participation. The reasons for exclusion included (in order): lack of an Android phone, lack of a mobile phone generally or an active SIM card, not having the correct version of Android phone, adequate RAM (memory space) or mobile data allowance (Wirtz et al 2022).

- Potential participants are more likely to be members of "marginalised" groups in society in many cases
- Recruitment - eg: not digitally accessible; type of device required
- Eligibility to participate - eg: no device; failed digital literacy screening
- Retention on study - eg: access barriers like Wi-Fi; lost/stolen device
- General issues for marginalised groups - eg: concerns about providing information about self; mobile population and loss of contact

Table 1.3 - Inequity in DCTs in HIV/AIDS treatment.

1.7. CONCLUDING THOUGHTS

There are a few points of conclusion that can be made about therapies and treatments for mental health problems, which echo the starting point that there is no simple answer. It is admitted that these points are a personal view from the evidence.

i) There is no single therapy or treatment that is best for every condition. It is better to look for particular treatments/therapies for particular conditions. But even then, there will be differences in "success" depending on variables related to the individual and therapist/treatment provider.

ii) In studies that evaluate therapies and treatments, it depends on what is being compared:

- Therapy or treatment vs none.
- Therapy A vs therapy B (table 1.4).

- Therapy A vs treatment A.
- Treatment A vs treatment B.

- Luborsky et al (2002) found seventeen meta-analyses that compared different therapies for the common psychiatric conditions among adults. Altogether, there was little different between the therapies in outcome, thereby suggesting the "Dodo bird verdict".

- Luborsky et al (2002) proposed four possible explanations for the findings:

- i) All therapies have common ingredients beyond their specific "techniques" which benefit patients.

- ii) The researchers evaluating the different therapies support one approach over another and this impacts the findings (in the form of "expectancy effects"), but this evens out over the many studies.

- iii) The methodological problems with comparing different therapies leads to similarities in effect.

- iv) Patient variables, not controlled for, explain the findings.

Table 1.4 – Luborsky et al (2002).

iii) It is possible to establish in some situations that one treatment/therapy is "superior" to another. This avoids the situation called the "Dodo bird verdict" (Luborsky et al 1975). This quotes the Dodo bird in "Alice in Wonderland" that "everybody has won, and all must have prizes" [7].

There is a difference, however, between characteristics common to effective therapy, and that there is equivalence among them (which is the "Dodo bird verdict"). Meta-analyses that combine treatments and therapies in order to produce a single effect size risk producing such a verdict, warned Beutler (2002). This is because patient variables are collapsed in a single group, and the assumption is made that "these various qualities do not differentially determine patients' response to treatment..." (Beutler 2002 p31). Beutler et al (2000) found over three hundred studies where patient variables "exerted a significant mediating effect on

[7] This phrase is attributed to Rosenzweig (1936), who was the first to argue that common factors in the different therapies would mean little difference between them in terms of outcomes (Luborsky et al 2002).

treatment – 'different strokes for different folks'" (Beutler 2002).

Therapies based on the same theoretical principles are also combined in many meta-analyses as well as therapies from different schools of thought. Beutler (2002) commented that "on one hand, it is relatively easy to justify collapsing cognitive and cognitive-behavioural treatments. On the other hand, it is more difficult to rationalise the consolidation of such widely different theories as transpersonal, systemic, client-centred, experiential, existential, and so on, under a label of 'dynamic'... without feeling that something was sacrificed in the translation" (p31).

iv) When evaluating therapy or treatment, the questions asked are important. "Does it work?" or "which is better?" may not be as helpful as "what kinds of patients can change in what kinds of ways through what kinds of therapy?" and "how does therapy effect these changes?" (Ryle 1975).

1.8. APPENDIX 1A – PLACEBO

"Placebo is Latin for 'I shall please' and can be defined as any medically inactive [8] or sham component of treatment that is knowingly used by a therapist for its non-specific psychological and psychophysiological effect" (Curtis 2000 p25) [9].

Placebo can be viewed more as the effect of the context on a patient as well as their expectations, desires and emotions (Price et al 2008) [10].

The role of expectancy in the placebo effect can be seen in a standardised design of experiment on perception

[8] There is a paradox here: "If a placebo is inert, it can't cause an effect, as something that is inert has no inherent properties that allow it to cause an effect" (Price et al 2008 p567).

[9] Kunkel et al (2024) expressed this view: "Placebo effects are an individual's psychophysiological response to contextual information and associated expectations to treatments that are physically and pharmacologically inert. The strength of placebo effects varies considerably among and within individuals depending on contextual factors, prior experiences of treatment benefit, and expectations regarding the treatment" (pp1-2).

[10] Botvinik-Nezer et al (2024) commented: "Throughout history, placebo effects have been variously considered as mysterious healing forces and tricks played upon the gullible by medical practitioners. Scientific research over the past decades has shown that placebo effects are neither of these. Rather, they are now understood to result from active, endogenous brain processes related to expectation, meaning, and predictive regulation of the body. A substantial part of the benefit of many kinds of treatments — including conventional drug therapies, surgery, acupuncture, psychotherapy, and more — is related to these psychological and brain processes. The study of placebo effects is thus the study of the internal brain processes that promote health and healing" (p1). The placebo effect can be seen in neuroimaging of the brain (ie: physiological changes in brain activity) (Botvinik-Nezer et al 2024).

of pain. Participants receive a heat stimulation to the arm, for example, and rate the pain. This is the baseline measure. Then a placebo cream, say, is applied to the arm and the heat stimulation continues. In order to deceive the participants, the intensity of the stimulus is reduced without telling them. Thus the expectancy of the cream's pain relief properties is established. The intensity of the heat stimulus is returned to the baseline level. Participants tend to now rate the stimulus as less painful than their baseline rating (eg: Price et al 1999) [11].

The placebo effect can be very powerful. For example, sufferers of Parkinson's Disease showed improvements in motor skills (eg: hand movements) which are disrupted by the disease, when given a placebo that they believed was an anti-Parkinson's drug (eg: Goetz et al 2000).

In the case of acupuncture, studies have shown that individuals who believe in its effectiveness and/or expected benefits from it had improvements in conditions like migraine or chronic low back pain, after sham acupuncture (Price et al 2008).

Placebo pain relief does not work for everyone, and individual responses can vary from none to large. For example, Levine et al (1978) found that "responders" had an average improvement in pain of three out of ten on a scale used after a dental operation. More recently, Furmark et al (2008) reported gene differences related to the amygdala in responders and non-responders to a placebo for reducing social anxiety about a public speaking task.

1.8.1. Actual Effectiveness of Placebos

When assessing the effectiveness of the treatment, the effect of a placebo (ie: expectations) have to be disentangled (Brooks 2008) [12]. Meta-analyses have shown

[11] Pain relief from an inert treatment (ie: placebo analgesia) has been studied to understand the neurochemistry of it. An increase in dopamine in particular areas of the brain in response to expectations is one possibility. Though, contrary to research expectations, Kunkel et al (2024) found no support for this.

[12] Establishing the actual placebo effect is hampered by a concept known as "regression to the mean". For example, with pain intensity, extreme ratings will move towards the average over time. Generally, individuals tend to report more pain at an initial assessment than at subsequent assessments even without treatment. So improvements could be "regression to the mean" effects or placebo effects in a clinical trial (Price et al 2008).

that the "true" placebo effect is limited.

The actual effect of placebo can be assessed by a comparison with a no-treatment group in a clinical trial of a treatment. Hrobjartsson and Gotzsche (2001) reviewed 114 randomised clinical trials that included placebo groups and no-treatment groups as well as the treatment groups. The authors found no benefit for placebo over no-treatment.

Hrobjartsson and Gotzsche (2004) updated this study with data from 42 more trials to give a total of 156 studies covering forty-six clinical conditions. Overall, placebo groups were no better than no-treatment groups, but on subjective patient measures (eg: self-rating of pain) there were significant benefits. But Hrobjartsson and Gotzsche (2004) were cautious about this positive finding - "It is important to realise that the difference between placebo groups and no-treatment groups does not equal the effect of placebo as such a comparison is unblinded. Thus, even if there were no true effect of placebo, one would expect to measure differences due to reporting bias, attrition bias [13] and other forms of bias related to lack of blinding. Reporting bias is particularly problematic. Most patients are polite and prone to please the investigators by reporting improvement, even when no improvement was felt. It is difficult to separate such reporting bias from true effects of placebo, but we suspect reporting bias occurred" (p97).

Interestingly, differences were found between studies measuring outcomes on binary scales (eg: yes/no) and continuous scales (eg: 1-5). While the no-treatment group may not be entirely untreated as patients may seek alternative treatments, and this group does interact with physicians, even if only in receiving attention when measuring health outcomes (a placebo itself?).

1.8.2. Physiology of Placebos

The downside of placebo is nocebo, where negative expectations can produce harmful effects. This can be seen in clinical trials where the placebo group reports side effects similar to the drug group (because of the expectations of such side effects) (Pilcher 2009).

Both placebos and nocebos produce a physiological effect as seen in PET scans, like a change in dopamine and opioid activity. With nocebos it is a decrease which

[13] Individuals dropping out of the study.

explains increased pain, and the opposite for placebos (Scott et al 2008) [14]. Neural activity decreases in pain processing areas of the brain (eg: thalamus) after placebo pain relief (Price et al 2008).

Price et al (2008) concluded their review of the placebo effect:

> Our understanding and conceptualisation of the placebo effect has changed in recent times, shifting from a focus on the inert content of a physical placebo agent to the overall simulation of a therapeutic intervention. Research has allowed for the identification of not one but many placebo responses, each of which may be driven by different psychological and neurobiological mechanisms depending on the particular context in which the placebo is given. We are still some way from understanding the relationships between the identified psychological variables and their neurobiological underpinnings, although a body of literature is emerging that identifies the roles of certain cognitive and emotional factors and various biochemical and neuroanatomical mechanisms in driving placebo responses. This literature also shows that placebos have actual biological effects on the brain and body and are more than response biases (pp584-585).

1.9. REFERENCES

Beutler, L.E (2002) The dodo bird is extinct Clinical Psychology: Science and Practice 9, 1, 30-34

Beutler, L.E et al (2000) Guidelines for Systematic Treatment of the Depressed Patient New York: Oxford University Press

Botvinik-Nezer, R et al (2024) Placebo treatment affects brain systems related to affective and cognitive processes, but not nociceptive pain Nature Communications 15, article 6017

Bowman, L & Fine, M (2000) Client perceptions of couples therapy: Helpful and unhelpful aspects American Journal of Family Therapy 28, 295-310

Brewer, K (2001) Clinical Psychology Oxford: Heinemann

Brooks, M (2008) The power of belief New Scientist 23/8, 36-39

Chen, H et al (2007) Criteria for surrogate end points Journal

[14] Ten individuals were given a placebo (based on positive expectations - "This medication is thought to have analgesic effects through the activation of natural brain systems that suppress pain"), and five individuals a nocebo (with negative expectations) for pain relief.

of the Royal Statistical Society B 69, 919-932

Choi, B & Pak, A (2007) Hello-goodbye effect. In Salkind, N.J (ed) Encyclopaedia of Measurement and Statistics (volume 2) Thousand Oaks, CA: Sage

Coolican, H (2004) Research Methods and Statistics in Psychology (4th ed) London: Hodder & Stoughton

Cronbach, L (1990) Essentials of Psychological Testing (5th ed) New York: Harper & Row

Curtis, A (2000) The placebo effect Psychology Review September, 24-26

Dallos, R & Cullen, C (1990) Clinical psychology. In Roth, I (ed) Introduction to Psychology Milton Keynes: Open University

Elliott, M.R (2023) Surrogate endpoints in clinical trials Annual Review of Statistics and Its Application 10, 75-96

Ephron, S (2023) Therapy in the spotlight New Scientist 15th April, 38-42

Fluckiger, C et al (2020) Assessing the alliance-outcome association adjusted for patient characteristics and treatment processes: A meta-analytic summary of direct comparisons Journal of Counselling Psychology 67, 6, 706-711

Foa, E.B et al (1997) The validation of a self-report measure of post-traumatic stress disorder: The Post-traumatic Diagnostic Scale Psychological Assessment 9, 445-451

Frank, J.D (1974) Therapeutic components of psychotherapy: A 25-year progress report of research Journal of Nervous and Mental Diseases 159, 5, 325-342

Furmark, T et al (2008) A link between serotonin-related gene polymorphisms, amygdala activity, and placebo-induced relief from social anxiety Journal of Neuroscience 28, 49, 13066-13074

Gaston, L & Marmar, C.R (1993) Manual of California Psychotherapy Alliance Scales (Unpublished: https://www.researchgate.net/publication/321732093_CALPAS_Manual)

Gilliard, C (2017) Pedagogy and the logic of platforms Educause Review 52, 4 (https://er.educause.edu/articles/2017/7/pedagogy-and-the-logic-of-platforms)

Goetz, C.G et al (2000) Objective changes in motor function during placebo treatment in PD Neurology 54, 710-714

Goldsmith, L.P et al (2015) Psychological treatments for early psychosis can be beneficial or harmful, depending on the therapeutic alliance: An instrumental variable analysis Psychological Medicine 45, 2365-2373

Hiatt, D & Hargrave, G.E (1995) The characteristics of highly effective therapists in managed behavioural provider networks

Behavioural Healthcare Tomorrow 4, 19-22

Hrobjartsson, A & Gotsche, P.C (2001) Is the placebo powerless? An analysis of clinical trials comparing placebo treatment with no treatment New England Journal of Medicine 344, 1594-1602

Hrobjartsson, A & Gotzsche, P.C (2004) Is the placebo powerless? Update of a systematic review with 52 new randomised trials comparing placebo with no treatment Journal of Internal Medicine 256, 91-100

Huson, L (2000) Study design. In Cohen, A & Posner, J (eds) A Guide to Clinical Drug Research (2nd ed) Dordrecht: Kluwer Academic Publishing

Kay, S.R et al (1987) The Positive and Negative Syndrome Scale (PANSS) for schizophrenia Schizophrenia Bulletin 13, 507-518

Kolassa, I-T et al (2010) Spontaneous remission from PTSD depends on the number of traumatic event types experienced Psychological Trauma: Theory, Review, Practice and Policy 2, 3, 169-174

Kunkel, A et al (2024) Dopamine has no direct causal role in the formation of treatment expectations and placebo analgesia in humans PLoS Biology 22, 9, e3002772 (Freely available at https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3002772)

Lambert, M.J et al (1992) Choosing outcome assessment devices: An organisational and conceptual scheme Journal of Counselling and Development 70, 527-532

Levine, J.D et al (1978) The mechanisms of placebo analgesia Lancet 2, 654-657

Luborsky, L et al (1975) Comparative studies of psychotherapies Archives of General Psychiatry 32, 995-1008

Luborsky, L et al (2002) The dodo bird verdict is alive and well – mostly Clinical Psychology: Science and Practice 9, 1, 2-12

Marmar, C.R et al (1989) Towards the validation of the California Therapeutic Alliance Rating System Psychological Assessment 1, 46-52

McLeod, J (1997) Research and evaluation in counselling. In Palmer, S & McMahon, G (eds) Handbook of Counselling (2nd ed) London: Routledge

Miller, S.D et al (2008) Supershrinks: What is the secret of their success? Psychotherapy in Australia 14, 4, 14-22

Pilcher, H (2009) Beware witch doctors New Scientist 16/5, 30-33

Price, D.D et al (1999) An analysis of factors that contribute to the magnitude of placebo analgesia in an experimental paradigm Pain 83, 147-156

Price, D.D et al (2008) A comprehensive review of the placebo effect: Recent advances and current thought Annual Review of Psychology 59, 565-590

Ricks, D.F (1974) Supershrink: Methods of a therapist judged successful on the basis of adult outcomes of adolescent patients. In Ricks, D.F & Roff, M (eds) Life History Research in Psychotherapy Minneapolis: University of Minnesota Press

Rosenzweig, S (1936) Some implicit common factors in diverse methods of psychotherapy: "At last the Dodo said, 'Everybody has won and all must have prizes'" American Journal of Orthopsychiatry 6, 412-415

Rosnow, R.L & Rosenthal, R (1997) People Studying People: Artifacts and Ethics in Behavioural Research New York: WH Freeman

Ryle, A (1975) Psychotherapy research: The role of the repertory grid New Behaviour August 28th, 326-328

Scott, D.J et al (2008) Placebo and nocebo are defined by opposite opioid and dopaminergic responses Archives of General Psychiatry 65, 2, 220-231

Venter, W et al (2018) Improving linkage to HIV care through mobile phone apps: Randomised controlled trial JMIR mHealth and uHealth 6, 7, e155

Walters, G.D et al (2002) Assessing change with the Psychological Inventory of Criminal Thinking Styles: A controlled analysis and multi-site cross-validation Criminal Justice and Behaviour 29, 3, 308-331

Wampold, B.E & Brown, J (2005) Estimating variability in outcomes attributable to therapists: A naturalistic study of outcomes in managed care Journal of Consulting and Clinical Psychology 73, 5, 914-923

Wirtz, A.L et al (2022) Addressing health inequalities in digital clinical trials: A review of challenges and solutions from the field of HIV research Epidemiologic Reviews 44, 87-109

2. EXAMPLES OF STUDIES AND RESEARCH

2.1. OVERALL

When there are a number of studies on a particular topic, researchers attempt to summarise the patterns within them. In terms of the effectiveness of psychotherapy, Eysenck (1952) drew together studies available at the time on psychoanalysis (5 studies) and "eclectic" (mixed) therapies (19 studies). He calculated that 44% of individuals were cured, much improved or improved [15] with psychoanalysis and 64% with eclectic therapies. This compared with 72% of the control group (individuals hospitalised or "treated" by GPs with sedatives, tonics, and reassurance).

This negative conclusion for psychoanalysis that individuals were better off without it began "an explosion of research into the effects of psychotherapy" (Gross 1990). Subsequent research has become more sophisticated using meta-analysis (eg: Smith et al 1980). This re-analyses the different studies to produce standardised effect sizes. However, it is still dependent on the quality of the individual studies.

A meta-analysis and review brings together individual randomised controlled trials and provides an overview, while an umbrella review synthesises meta-analyses on the subject. Dragioti et al (2017) performed the latter on 247 meta-analyses of randomised controlled trials of psychotherapies.

Four-fifths of the meta-analyses found a significant improvement from psychotherapy compared to no psychotherapy. But there were many confounders which meant "the effectiveness of psychotherapy is often exaggerated" (Dragioti et al 2017). In the end, sixteen meta-analyses provided "convincing evidence". These tended to be for specific situations (eg: Cognitive Behavioural Therapy (CBT) for depressive symptoms; CBT

[15] Recovery was defined as "(i) return to work, and ability to carry on well in economic adjustments for at least five years; (ii) complaint of no further or very slight difficulties; (iii) making successful social adjustment" (Eysenck 1952).

and counselling for smoking cessation; cognitive remediation therapy for cognitive functioning in schizophrenia).

There was so much variability in the studies, including in type of psychotherapy, disorders, and outcome measures used, in particular. There were also differences sample sizes, and make-up (eg: children; adults). Studies finding no differences were often unpublished, producing "publication bias". It has been calculated that studies with statistically significant findings are five times more likely to be published in psychiatry and psychology than studies with null (non-significant) findings (Fanelli 2010).

2.2. DYNAMIC PSYCHOTHERAPY

Chambless and Hollon (1998) (C&H) proposed criteria for assessing a treatment as efficacious:

i) The treatment is statistically superior to no treatment or a placebo in randomised controlled trials (RCTs).

ii) The treatment is at least equivalent to alternative "bona fide" treatments in RCTs.

iii) The RCTs involved sound methodology (eg: specific diagnostic groups, standardised techniques of therapy).

iv) The superiority of the treatment demonstrated in at least two independent studies.

Connolly Gibbons et al (2008) applied the C&H criteria to "dynamic psychotherapy". This term covered all forms of therapy that are directly or indirectly related to the ideas of Sigmund Freud. So, it includes traditional Freudian psychoanalysis through to modern brief dynamic therapies. In an ideal world, each form of psychotherapy would be evaluated separately.

Connolly Gibbons et al (2008) evaluated dynamic psychotherapy for different mental disorders:

a) Major depressive disorder - The researchers found no recent studies (in last decade) that met the C&H criteria showing psychotherapy to be more effective than no treatment. However, a number of the studies did suggest its effectiveness, particularly as combined with

anti-depressants and for a set period of time (eg: sixteen sessions).

b) Generalised anxiety disorder (GAD) - "There is no reliable evidence in the empirical literature to support or deny the efficacy of dynamic psychotherapy for GAD" (Connolly Gibbons et al 2008 p99).

c) Social anxiety, obsessive-compulsive disorder, agoraphobia, post-traumatic stress disorder, schizophrenia - No relevant recent RCTs.

d) Panic disorder - The C&H criteria were met here with a study by Milrod et al (2007). Panic-focused psychodynamic psychotherapy twice weekly for twelve weeks reduced panic attacks for nearly three-quarters of the group compared to improvements for only 39% of the control group (performing applied relaxation).

e) Borderline personality disorder - Though studies do not meet the C&H criteria, there is evidence that dynamic psychotherapy is "positively efficacious" here. This means giving the idea further consideration and research.

f) Substance abuse - Dynamic psychotherapy was found to fulfil the C&H criteria for opiate dependence with one study (Woody et al 1987).

Many therapies do not meet the C&H criteria, often because RCTs are not traditionally used in the "therapy world" as in the "medical world" of drug treatments. Does this mean that therapies that fail these criteria are not effective for helping individuals with different mental disorders?

It has been argued that some therapies are better assessed in investigations that link process to outcome (ie: process research, or process-outcome research) rather than the simple focus on outcome in RCTs.

2.3. TELEPSYCHIATRY VERSUS FACE-TO-FACE TREATMENT

"Telemedicine" (ie: health care provided online) expanded greatly with the covid-19 pandemic and lockdowns. One type of this is "telepsychiatry", where videoconferencing is used for diagnosis and symptom assessment, and therapy/counselling.

"Telepsychiatry increases access to psychiatric care. It can be a necessity for patients living in underserved areas and those for whom visiting hospitals/clinics is difficult because of physical difficulties and/or psychiatric symptoms. Telepsychiatry can also lead to early intervention and can simplify the co-ordination of treatment involving multiple healthcare providers. On the other hand, the potential disadvantages of telepsychiatry include a negative impact on doctor-patient rapport, drop out from long-term treatment, the cost of infrastructure, the possibility of misdiagnosis/maltreatment, essential proficiency in operating web conferencing systems, and limited availability owing to financial and IT skills/ability issues. Moreover, telepsychiatry might be well suited for some psychiatric disorders but not others, depending on the specific disease characteristics" (Hagi et al 2023 p408).

How does telepsychiatry compare to face-to-face psychiatry? Hagi et al (2023) found thirty-two randomised controlled trials that had compared the two modes of psychiatry for their meta-analysis. Eleven mental disorders were compared (eg: depression; insomnia; mild dementia).

Using the outcome measure of symptom score improvement, telepsychiatry was significantly better with depression, but significantly worse with eating disorders than face-to-face treatment. There was no significant difference for other disorders. Combining all studies, there was no difference between the two modes of treatment.

In the studies, "there was insufficient consideration of the impact of the quality of communication. A system with professional specifications and no communication delays would naturally provide the same therapeutic effect as a face-to-face treatment (although there would be some disadvantages such as not being able to see the whole body and not being able to smell it), but on a small screen, where the communication tends to be choppy, the therapeutic effect may be less good. On the other hand, recent smartphones may be able to achieve clear and low-latency communication even with small screens" (Hagi et al 2023 p413). Also many relevant variables were not measured, including the quality of the rapport between psychiatrist and client/patient, the presence of co-morbidities, and adherence to medication (where prescribed). Only randomised controlled trials with twenty or more participants were included in the review.

The studies varied in methodological issues,

including sample size, sample characteristics, mental disorders, length of follow-up, type of treatment, and outcome measures.

2.4. BRAIN CHANGES AFTER THERAPY

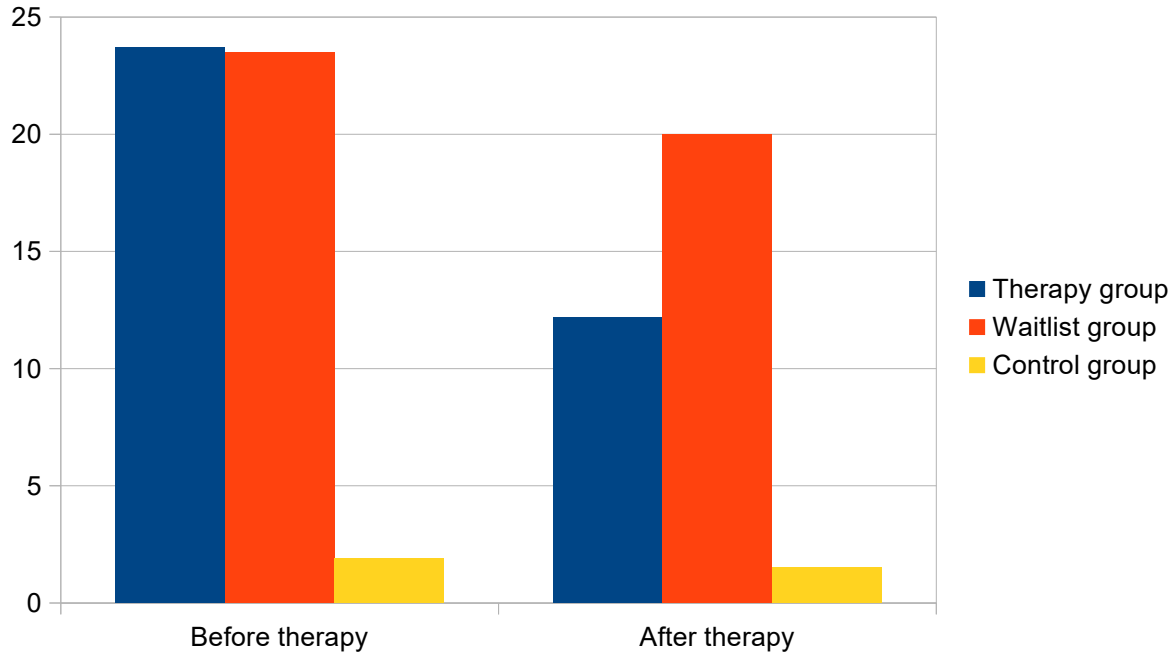
Self-report measures are the usually means of assessing an outcome of therapy or treatment. An alternative is to measure changes in the brain using neuroimaging, as Wittmann et al (2018) did for CBT with agoraphobia sufferers.

Fifty-one patients at clinics in Germany who received CBT, and fifteen who did not (wait-list control group), as well as 51 healthy controls, were recruited for the study. Functional magnetic resonance imaging (fMRI) scans were performed before and after treatment (twelve sessions of CBT). During scanning the participants were presented with 48 agoraphobia-specific pictures (eg: crowds) and 48 neutral pictures.

Fear-related brain structures (eg: amygdala) are known to be more active in individuals with phobias when thinking about or faced with feared stimuli. The patients showed a reduction in brain activation of two fear-related brain areas in particular between before and after treatment (namely, the right ventral striatum during anticipation of the feared stimuli, and the right amygdala during observation of the agoraphobia-specific pictures).

Self-report data (Hamilton Anxiety Rating Scale; HAM-A) collected at the same time showed a reduction in fear rating for agoraphobia-specific pictures from before to after treatment (figure 2.1). This study showed that CBT altered brain activity in a positive way.

Santarnecchi et al (2019) also measured brain changes in their comparison of trauma-focused cognitive behavioural therapy (TF-CBT), and eye movement desensitisation and reprocessing (EMDR) with thirty-one post-traumatic stress disorder (PTSD) sufferers in Italy after an earthquake in 2002. TF-CBT concentrates on changing the thoughts related to the traumatic event, while EMDR uses specific eye movements to produce "a decoupling between external attention and internal reprocessing of traumatic memories, which prevents patients from feeling overwhelmed" (Santarnecchi et al 2019 p2).



(Data from table 2 p354 Wittmann et al 2018)

Figure 2.1 - Mean HAM-A scores before and after treatment.

The study took place ten years after the traumatic event, and half of the participants received TF-CBT (n = 14), and half EMDR (n =17). PTSD was measured using standardised questionnaires, and self-reports of trauma symptoms. The outcome measure in this study related to brain activity as measured by magnetic resonance imaging (MRI) scans.

Both therapies produced significant changes in the self-reports, and in the brain (specifically connectivity between different areas), between the baseline (pre-treatment) and post-treatment. Santarnecchi et al (2019) summed up: "Results point to a similar, beneficial psychological impact of EMDR and TF-CBT for treatment of natural-disaster PTSD patients" (p2).

This study had no control (or wait-list) group, and the participants were "pseudo-randomised" to the two conditions (ie: based on practical availability of therapy, and symptom severity)
(Santarnecchi et al 2019).

2.5. REFERENCES

Chambless, D.L & Hollon, S.D (1998) Defining empirically

supported therapies Journal of Consulting and Clinical Psychology 66, 7-18

Connolly Gibbons, M.B et al (2008) The empirical status of psychodynamic therapies Annual Review of Clinical Psychology 4, 93-108

Dragioti, E et al (2017) Does psychotherapy work? An umbrella review of meta-analyses of randomised controlled trials Acta Psychiatrica Scandinavica 136, 3, 236-246

Eysenck, H.J (1952) The effects of psychotherapy: An evaluation Journal of Consulting Psychology 16, 319-324

Fanelli, D (2010) "Positive" results increase down the hierarchy of the sciences PLoS ONE 5, 4, e10068 (Freely available at https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0010068)

Gross, R.D (1990) Key Studies in Psychology London: Hodder & Stoughton

Hagi, K et al (2023) Telepsychiatry versus face-to-face treatment: Systematic review and meta-analysis of randomised controlled trials British Journal of Psychiatry 223, 407-414

Milrod, B et al (2007) A randomised controlled clinical trial of psychoanalytic psychotherapy for panic disorder American Journal of Psychiatry 164, 265-277

Santarnecchi, E et al (2019) Psychological and brain connectivity changes following trauma-focused CBT and EMDR treatment in single-episode PTSD patients Frontiers in Psychology 10, article 129

Smith, M.L et al (1980) The Benefits of Psychotherapy Baltimore: John Hopkins University Press

Wittmann, A et al (2018) Effects of cognitive behavioural therapy on neural processing of agoraphobia-specific stimuli in panic disorder and agoraphobia Psychotherapy and Psychosomatics 87, 350-365

Woody, G.E et al (1987) Twelve month follow-up of psychotherapy for opiate dependence American Journal of Psychiatry 144, 590-596