

The Equivalence of Psychotherapies

'Why,' said the Dodo, 'the best way to explain it is to do it.' (Lewis Carroll [1896] *Alice's Adventures in Wonderland*)

Introduction

In October 2009, *Newsweek* magazine in the United States published an article by Sharon Begley called 'ignoring the evidence'. In it, psychotherapy researchers are quoted berating psychotherapists for taking no notice of science and instead treating patients with whatever intervention they are familiar. There is a 'widening gulf between researcher and clinician'. One researcher, Timothy Baker, argues that clinicians 'give more weight to their personal experiences than to science'. The tone of the article is that this is clearly a bad thing. The implication is that science has progressed since the early days of therapy, and that we now know more about what therapies work and why. Clinicians should, therefore, keep up to date in their knowledge and change their practice accordingly.

If this is all true, it is pretty damning. The truth, though, is a bit more complicated.

Maximising the Benefits of Psychotherapy: A Practice-Based Evidence Approach,
First Edition. David Green and Gary Latchford.

© 2012 John Wiley & Sons, Ltd. Published 2012 by John Wiley & Sons, Ltd.

Psychotherapists naturally want to be good at their job, and their clients equally naturally want to see someone who has the best chance of helping them. What is the most important factor in success? It's a reasonable assumption that this is the particular form of psychotherapy or counselling being provided. If so, can science tell us which therapies have been shown to be best? Is the *Newsweek* article right? Should therapists be directed to follow the science and train in whatever therapy has the best evidence? Are newer therapies more effective than older ones?

It's very tempting to assume that anything newer must be better. The belief in progress is perhaps strongest in medicine. Although this can sometimes lead to unrealistically optimistic expectations of the power of modern medicine to cure, there have undoubtedly been some striking and high-profile successes such as transplant surgery. In medicine a great deal of effort has gone into finding ways to evaluate and improve treatments, and then to disseminate scientific findings to clinicians in the form of guidelines.

Psychotherapy and counselling are now well established as effective frontline treatments for mental health problems. In many countries they are seen as being a part of the framework of mental health services and often offered as an alternative to drug treatments. We usually think of psychotherapy, then, as being a kind of medical intervention often (but not always) delivered alongside other forms of treatment in health settings. It's not surprising, then, that questions about effectiveness and cost benefits asked of medical treatments should also be asked of psychotherapy. It is also understandable that the same methods used to evaluate medical treatments should be used to evaluate psychotherapies. Such an approach leads directly to the views expressed in the *Newsweek* article. Therapists who do not change their practice in line with science, this argument goes, are like medical doctors prescribing out-of-date medication. At best they are less effective than they should be, at worst they are dangerous.

So what is a therapist to do? It all hinges, of course, on what the evidence actually says, and whether the evidence is reliable. The next chapter will look at whether the research methods we use to evaluate

therapy are fit for purpose, and will explore some of the controversies in the area. This chapter will go straight to the heart of the matter: what is the evidence that some therapies are better than others?

Guidelines and Evidence-Based Practice

If I have diabetes and of the two drugs available one has been shown in scientific research to be better for the condition and with fewer side effects, that's the one I'd want my doctor to be aware of and to prescribe for me. How can we make sure that doctors are kept informed about best practice, and don't rely solely on their own experience (which might be out of date)? This issue has been the focus of the evidence-based practice movement. It has had two aims: to improve the quality of research and to develop methods of combining findings so that dissemination is more efficient. Perhaps the group most associated with this movement is the Cochrane Collaboration, founded in 1993. This is a not-for-profit international group of specialists who conduct systematic reviews – 'critical summaries' – of different areas of healthcare to establish current evidence for treatment and prevention. It is named after Scottish epidemiologist Archie Cochrane who is recognised as giving the initial impetus for the movement in his book *Effectiveness and Efficiency* in 1972.

The most notable impact of the evidence-based practice movement for clinicians has been the emergence of guidelines that review the current research evidence and direct them on which treatments to use. This is massively influential in medicine, providing an up-to-date overview of research that was previously beyond the scope of any individual doctor.

The success of the approach has inevitably led to it being applied to other areas of practice, including psychotherapy. In the United Kingdom, this has most prominently been in the form of guidelines from the National Institute for Health and Clinical Excellence (NICE) on the treatment of various mental illnesses. The depression guideline revised in 2009, for example, recommends cognitive behavioural

therapy (CBT), interpersonal therapy (IPT), behavioural activation and couples therapy. No other therapies are supported, though the guidelines admit that counselling and short-term psychodynamic psychotherapy are ‘limited options’ for those people who refuse one of the recommended treatments, and go on to make it clear that the therapist needs to discuss the lack of evidence for these approaches with the patient before starting therapy. Whether therapists really tell their clients they are about to offer a treatment that NICE thinks won’t work is not known.

Meanwhile in the United States, in 1994, Division 12 (Clinical Psychology) of the American Psychological Association published a list of empirically supported Treatments (ESTs) – psychotherapies for which there was thought to be supporting evidence (Chambless *et al.*, 1998). Significantly, they used the same criteria for evaluating psychotherapies as the US Food and Drug Administration (FDA) uses for drug trials – for example, at least two trials showing that the therapy is better than no treatment or produces equivalent results to any established treatment. There are now around 300 officially sanctioned ESTs, and the Division lists which therapies are recommended for particular conditions such as depression.

The key question, of course, is whether the evidence informing these guidelines – and the *Newsweek* article – is convincing. Is there clear evidence for the superiority of any one psychotherapy over another? Curiously, although this has been the focus of a growing debate since the 1970s, the question was first asked of psychotherapy in 1936, and the answer given then remains as relevant today as it was when it was written.

Saul Rosenzweig, Dodos and Common Factors

Saul Rosenzweig got his doctorate from Harvard in 1932 and remained active in psychological research until shortly before his death 72 years later. His most influential paper – ‘Some implicit common factors in diverse methods of psychotherapy’ was published in 1936. Just to give some context, one of Rosenzweig’s classmates at Harvard was B.F. Skinner, one of the founders of

behaviourism. Rosenzweig's paper was a response to the claims of proponents of the various psychotherapies popular at the time that their particular therapy was more effective than all the others, and that the reason for this was that the theory of change on which their therapy was based was right (and the others wrong).

In contrast, Rosenzweig was the first to propose that the mechanisms of change in psychotherapy might not be specific to the form of therapy, but instead be agents of change common to all therapies, which he termed 'the common factors'. It is tempting to think that one of the reasons this paper is still regularly cited in contemporary research is the quotation with which it begins: 'At last the Dodo said, "Everybody has won, and *all* must have prizes."'

The allusion was to the caucus race in Lewis Carroll's 1865 book *Alice in Wonderland*, in which the Dodo instigates a confused race with no rules in order for the participants to shake themselves dry. Rosenzweig argued that competition between different schools of therapy was similarly misguided, because all were equally effective. This became known as 'the dodo bird hypothesis', a poetic touch that we suspect has captured the imagination of many subsequent researchers.

Rosenzweig's paper was a reflective one – it contained no data to support his assertions. Reading it now, in some ways it also seems very dated – the psychotherapies mentioned include psychoanalysis and Christian Science rather than CBT and IPT, for example. In other ways, however, it has not dated at all. Rosenzweig's opening argument is that all proponents of psychotherapies can produce successful cases to support assertions that they work. Unfortunately, he argues, they tend then to imply that this evidence also demonstrates that their own brand of therapy is better than the others. As someone surveying this from a more detached position, he argues that the logical conclusion to reach is that, if so many different therapies based on conflicting theoretical approaches can produce successful outcomes, then the reasons for success are unlikely to lie within any one theory.

His basic logical point was that the success of any therapy cannot be used as evidence that the therapy has brought about change in the way it claims to – it provides evidence that the therapy works,

but not why it works. Rosenzweig's belief was that therapies that appeared to be very different actually had more in common than the proponents realised – that the effectiveness of therapies was a result of unrecognised factors common to them all.

What did he think these factors might be? Rosenzweig makes a few suggestions. He argues that the relationship with the therapist may allow for some social reconditioning to occur, and also suggests that catharsis may be a common consequence of different therapeutic approaches. He also focuses on the effect of the therapist's personality. Though hard to quantify, he argues that there is a shared understanding of the qualities needed in a good therapist, and that good therapists are distributed across different schools of therapy.

Interestingly, his next possible factor is the coherent structure for understanding that all therapies offer. Rosenzweig describes this in terms of personality – that therapies offer a consistent and persuasive schema for the client to achieve greater personality organisation. In 1936 personality was a key concept in psychotherapy – Rosenzweig's definition of an established psychotherapy was one based on a general theory of personality, and he refers to mental disorder as a conflict of disintegrated personality constituents. Although the terminology of therapy has changed over the years, the notion that therapies may be providing a coherent and believable structure by which clients can begin to understand and solve their problems remains a persuasive idea.

Rosenzweig's next argument comes even now as a refreshing acknowledgement of the complexity of psychological problems and the limits of our understanding. Put bluntly, he argues that psychological events are so complex that many different, equally justified formulations are possible, and that each may contain a certain amount that is accurate. Thus no one therapy or interpretation has a monopoly on truth. In addition, personality is so complex that it is likely that there are very many ways to effect a change in organisation. Again, to put it bluntly, different therapies may target different aspects but have a similar overall impact. The implication is that therapists' formulations do not have to be completely accurate to

have a therapeutic effect, and that different foci in therapy can bring about similar therapeutic change.

Rosenzweig's paper is still immensely provocative. Though not based on research itself, it anticipates much of the psychotherapy research to take place over the subsequent 70 years – from studies comparing different therapies, examining the therapeutic alliance or the accuracy of a formulation, to dismantling studies aiming to identify the active ingredients in therapy. It also anticipates many of the concepts which have become associated with the processes of therapy, most obviously the common or non-specific factors that may underlie effectiveness. He also stressed the importance of the confidence that therapist and client have in the therapy—that it needs to appear credible to the client and evoke allegiance in the therapist. Both of these factors are reflected in the consistency of the therapist in adhering to the treatment approach.

Three other things are worth noting from this paper. First, Rosenzweig recognised that equivalence of therapies was only true when they were appropriately used. He limits his conclusions to accepted therapies, competently applied. Second, he did not rule out the possibility that some forms of treatment may be better suited to particular kinds of cases. He also considers the potential importance of matching patient and therapist in terms of personality. Finally, there is another possible solution to the logical conundrum of many therapies claiming differential effectiveness but all producing successful outcomes: rather than the explanation being that success is due to non-specific factors common to them all, perhaps each type of therapy utilises specific and distinct factors, but these are equally effective. In Rosenzweig's terms, these factors may have an impact in different ways but produce similar overall changes. We'll return to this point later.

After Rosenzweig: Does Therapy Actually Work?

Although he had particular views about differences between therapies, Rosenzweig never doubted that psychotherapy was effective.

This has not always been accepted, however, and 20 years after his paper the majority opinion seemed to reject it. In two papers, Hans Eysenck (1952, 1965), perhaps the most famous psychologist of his generation, argued that 75% of patients get better regardless of whether they receive psychotherapy (though he was more hopeful for treatments based on ‘modern learning theory’). Although his conclusions were founded on reviews of a small number of studies, they were extremely influential at the time, leading many to conclude that traditional psychotherapy was ineffective. By the end of the 1960s, however, the weight of evidence was turning decisively against this view, as exemplified by Bergin’s (1971) review of a much larger number of well-conducted studies. By this time the methods used to study psychotherapy outcome had grown increasingly sophisticated, taking their lead from medical trials. Even so, reviews that listed the outcome of different studies tended to add to the debate about the effectiveness of psychotherapy rather than end it. The decisive blow finally came in the 1970s and was made not by a psychotherapy researcher, but by an educational psychologist called Gene Glass, whose innovation in the way information can be summarised from different studies would change not just psychotherapy, but the way the whole of medicine is practised. Before we come to this, we need to understand the methods for conducting research that had by then become increasingly popular, and that provided the raw data for Gene Glass’s innovation.

A Well-conducted Study: The Randomised Controlled Trial

What is a well-conducted study? How can a reader determine that an investigator’s conclusions are justified, that the research has been conducted without bias and presents a fair and truthful account? As psychotherapies became more popular, and increasingly formed an accepted part of mental health services, it was natural for investigators to draw on the research methods used to evaluate other interventions in health. In fact, medicine had addressed the problems of trustworthiness some time before with the randomised controlled

trial (RCT), and this became the method of choice for psychotherapy investigators looking at outcome.

The first RCT in medicine is usually thought to date from 1946 and examined whether streptomycin was an effective treatment for pulmonary tuberculosis. Since then countless RCTs have been conducted, and have undoubtedly made a huge contribution to the development of medical practice. Put simply, RCTs enable a fair and unbiased comparison of treatments for particular conditions so that policy makers and clinicians have evidence rather than opinion on which to base treatment decisions.

To illustrate, suppose that a new drug is invented to treat the common cold. In an RCT the experimenter attempts to control for all the things that can bias a fair comparison so that a treatment can be properly evaluated. First, the new treatment is compared with something else – usually an existing treatment, but sometimes no treatment. Second, the appropriate group is targeted – in this case, people with a cold. The treatment is designed for the common cold so it would be unhelpful and unfair to include patients with the 'flu. Third, everyone agreeing to take part is randomised – that is, they are randomly allocated to either the new treatment or the old one with everyone having an equal chance of being in either group. They are randomised so that the two groups are likely to be comparable, and no one group is stacked with body builders with a fierce resistance to the common cold, for example. This randomisation is double blind – neither the patient nor the experimenter knows which treatment they are getting. This is to prevent the knowledge influencing either the patient ('I'm on the new wonder drug so I must be feeling better') or the experimenter ('I've spent years researching this great new drug and they must be feeling better'). The trial is then carried out, with each group receiving one or other of the treatments. At the end of the trial, an appropriate assessment is taken – in this case, whether they still have a cold. At this point the experimenter is able to crack the code on who was receiving which treatment, and compares the two groups to see if one has better outcomes than the other. In RCTs, then, the aim is to eliminate potential confounds so that a pure comparison between two treatments can be carried out.

When reading an RCT, then, although the methodological rigour with which they are carried out can vary considerably, you may have a certain degree of confidence in the conclusions. It therefore seems absolutely appropriate to utilise these methods in evaluating psychotherapy. Unfortunately, techniques that are relatively unproblematic when comparing two drug treatments become much more complicated when comparing psychotherapies – but more on this later.

How have these methods been used in psychotherapy research? Generally, to show that a particular psychotherapy is effective (i.e., comparing it with a no-treatment control) or to compare two or more psychotherapies to see if one is differentially more effective. The latter, of course, also provides evidence about the effectiveness of psychotherapies in general

The problem with individual RCTs, however, is that even the best of them remains just a single study, and when the sample size is small the results are often inconclusive – traditional inferential statistics usually require quite large numbers of participants if there is to be a good chance of detecting a significant difference between groups. Worse, when several similar studies are carried out, it is not uncommon for the results to vary, making an objective interpretation difficult.

An obvious solution is to write reviews gathering together the results of many RCTs in order to summarise their results. The first reviews of existing research are now known as narrative reviews. In them, the authors (usually experts in the field) offer an overview and interpretation of the existing literature. Though such reviews can be valuable and interesting, the opinions are just opinions – there is no attempt to present a mathematical summary of the findings of previous studies. This means, of course, that two reviews may potentially feature very different opinions and reach opposite conclusions about what current research is saying.

In essence, this was the situation described earlier regarding the effectiveness of psychotherapy, with Eysenck reviewing research and finding no evidence for effectiveness, and others arguing the opposite.

An important way to improve reviews would be to find some way of representing the data from individual studies. Luborsky, Singer

and Luborsky (1975) reviewed 40 psychotherapy outcome studies using a 'voting system' whereby each study that produced a statistically significant result in favour of a particular therapy was added to a tally. When they looked at the votes for different therapies, they found little evidence for any difference in effectiveness between therapies and resurrected Rosenzweig's use of the dodo bird metaphor in what became a highly influential paper.

Although this method did acknowledge the importance of the data in individual studies, reducing the data to a tick in favour of a treatment clearly meant that a lot of potentially important information was lost. A way of retaining this information and producing a more comprehensive way of summarising the data was the logical next step and appeared soon after. It was devised by Gene Glass, who called his method 'meta-analysis'.

Gene Glass and Meta-analysis

Psychotherapy research was the very first area in which the new method of meta-analysis was tested. As the inventor, Gene Glass, was an educational researcher, this seems quite puzzling. In fact, Glass was someone who felt that he had personally benefited from psychotherapy early in his career and was quite irritated by Eysenck's (1965) criticisms. He also argued that Eysenck's conclusions were based on poor methodology – arbitrary decisions about what studies to include (just 11 studies), and very crude interpretation of available data. In response, Glass chose to apply the ideas he was developing on synthesising data from multiple studies to the issue of the effectiveness of psychotherapy.

A key innovation in Glass's approach was to use effect sizes in his calculations. Studies tend to report whether there were any statistically significant differences between groups but, as said earlier, this is influenced by the number of people in the study and can be misleading. It also leads to a focus on whether the results were significant or not, with any non-significant results tending to be ignored. Within a research article, however, there is usually enough information to

calculate an effect size, an incredibly useful statistic reflecting the magnitude of the effect of the therapy. Effect sizes are based on standardised mean differences between groups – Glass defined this as the mean difference between the treatment and control group divided by the standard deviation (a measure of variability) of the control group. Using an effect size as the unit of outcome for each study gives meta-analysis a number of advantages: effect sizes are not influenced by sample size; effect sizes tell you *how large* the difference was between groups and not just whether it was significant or not (to give you an idea about how to interpret effect sizes, if you compared the heights of 15 and 16 year olds you would get a small effect size, if you compared the difference between the heights of 13 and 18 year olds you would get a large effect size); effect sizes enable you to compare the results of different studies even when they have used different outcome measures. Most important of all, effect sizes from different studies can be combined to produce an overall effect size, thus mathematically representing a summary of them. A common way of representing an effect size is Cohen's (1992) d , which reflects the size of the difference between groups.

Glass's meta-analysis was published in a paper with Mary Smith in the *American Psychologist* in 1977, and expanded in book form three years later (Smith, Glass and Miller, 1980). It presented a meta-analysis of nearly 375 controlled studies of psychotherapy and counselling representing nearly 50,000 participants. The authors found very clear evidence of the effectiveness of psychotherapy – they cite that on average the typical therapy client is better off than 75% of people receiving no treatment. Glass pointed out that this figure is by coincidence exactly the same figure that Eysenck had cited as the number of people who recover without therapy. Eysenck's attempt at a rebuttal, calling the new technique 'mega-silliness', now seems quaint in an era in which meta-analysis is pivotal in evidence-based medicine.

Subsequent researchers have confirmed Glass's work. Grissom (1996), for example, analysed data from 46 meta-analyses and calculated that the probability that any random client was better off receiving psychotherapy was $.7 \pm .2$, which he referred to as 'the magical number'. This figure sounds impressive, and it is. Rosenthal

(1990) compared the effect size obtained by Glass for psychotherapy with the effect size obtained in one of the key studies on azidothymidine (AZT) for AIDS – generally regarded as one of the most significant medical breakthroughs of the last 20 years. He found that the effect size for psychotherapy was significantly greater than that for AZT.

So Psychotherapy Works – Are Therapies Equivalent?

Glass's meta-analysis went further than examining the effectiveness of psychotherapy. He gave his new method a challenge – to determine whether some therapies are more effective than others. This is actually much trickier than determining whether psychotherapy works, which involves simply pooling data from all studies comparing a psychotherapy with a control group. Summarising different comparisons of therapies would be meaningless – for example, one study might find a behavioural therapy is more effective than a psychodynamic; another might find the opposite, so pooling the effect size is nonsense. In fact, there are many different studies comparing therapies, and often huge differences in study quality, duration, severity of problem and a host of other factors. Glass's solution was to group similar psychotherapies together into small categories (based on expert opinion) and compare these. He reports on four in the book, two (behavioural versus non-behavioural) in the paper. He found little evidence for any difference between therapy types when this was done, though the method for categorising different therapies is inevitably open to differing opinion. More convincing is a subset of his analysis, which looks at the studies that have both a behavioural and a non-behavioural arm (around 50) and finds similar effect sizes for both.

Glass's study was followed by a number of similar papers replicating and refining his approach, as well as, initially, arguments from behaviourally oriented researchers that many key behavioural papers had been ignored. A key response to this and other criticisms was the work of husband and wife team David and Diana Shapiro, based

in Sheffield in the United Kingdom (Shapiro and Shapiro, 1982). They refined Glass's method – for example, they included only studies that directly compared two or more treatments and had a control group – and made a particular effort to include the behavioural studies alluded to by critics of Glass's paper. They identified 143 outcome trials, a selection dominated by behavioural interventions (featured in 134 studies) and group therapy approaches (52% of studies). Most studies were of treatment of phobias and anxiety-related disorders, with only ten looking at depression, for example. Their meta-analysis confirmed the effectiveness of therapy compared with a placebo, and found that in comparison differences between therapies were modest, accounting for less variance than the target problem.

One of the most thoughtful uses of Glass's method has been by Bruce Wampold and colleagues (Wampold *et al.*, 1997) who tried a different solution to the problem of using meta-analysis to compare different therapies: they looked at the pattern of differences rather than the size. They only included studies (114) that directly compared two or more established therapies, and randomly assigned a positive or negative to the effect sizes from each comparison. The overall effect size for all the studies will therefore be zero, but they reasoned that, if there really is no difference between therapies, the distribution of scores will be small. This is, indeed, what they found.

A final example is one of the most comprehensive – Luborsky *et al.*'s (2002) attempt to revisit the equivalence question. They conducted a meta-analysis of meta-analyses, examining data from 17 meta-analyses of studies where active treatments were compared with each other. These covered common problems such as depression and anxiety, and common therapies, such as cognitive and psychodynamic. They found that the difference between them was small. Chambless (2002) has argued that the differences between therapies might be larger for particular problems and raises the possibility of matching treatment to problem. The idea that some therapies work much better for particular problems is a seductive notion, though others have argued that this is misguided – that the way the therapy is delivered is much more important (Messer and Wampold,

2002). A comprehensive examination of this issue would require a huge amount of effort – analysing good quality data on every combination of condition and competing therapy – and we don't have this kind of data. Luborsky *et al.* titled their 2002 paper 'The dodo bird verdict is alive and well – mostly', and this seems to us an apt summary. So far, the balance of evidence from these large meta-analyses suggests that the dodo still reigns.

Curiously, these large-scale meta-analyses do not find their way into treatment guidelines. Instead, guidelines rely on meta-analyses of treatments for particular conditions such as depression, and on individual studies.

Meta-analysis for Particular Conditions

There are now many meta-analyses summarising research into effective treatments for particular conditions. Some find no difference between the treatments studied – Rosa-Alcázar *et al.* (2008), for example, looked at 19 studies of the treatment of obsessive compulsive disorder, finding similar effect sizes for exposure with response prevention alone, cognitive restructuring alone and a combination (although all the treatments studied were different aspects of behavioural interventions). Other reviews find an advantage for a particular approach – Wolitzky-Taylor *et al.* (2008), for example, conducted a meta-analysis of 33 RCTs of treatments for phobia and found exposure-based treatments to be more effective than relaxation and cognitive-based approaches, but all were better than placebo or no treatment. Such reviews depend, of course, on the original research available. Wolitzky-Taylor *et al.* found no psychodynamic studies to include in their review, for example.

In other areas, such as depression, the findings of such meta-analyses are particularly interesting. Ekers, Richards and Gilbody (2008) conducted a meta-analysis of 17 RCTs, finding that behaviour therapy was as effective as CBT, and better than supportive counselling and brief psychotherapy. Cuijpers *et al.* (2008), looking at a larger pool of studies (53 RCTs) and comparing seven different

treatments (CBT, non-directive supportive treatment, behavioural activation, psychodynamic, problem-solving, interpersonal psychotherapy and social skills training) found no large differences between them, though the dropout rate was higher in CBT.

Perhaps most provocative of these meta-analyses, however, has been Benish, Imel and Wampold (2008), who conducted a meta-analysis of 15 trials of the treatment of post-traumatic stress disorder (PTSD). For some time it has been widely assumed that exposure is a necessary ingredient of a successful intervention in PTSD. Their review, surprisingly, found no difference between the therapies and no specific, crucial ingredient necessary for a successful outcome, including exposure. A long rebuttal followed (Ehlers *et al.*, 2010), arguing that the meta-analysis was flawed in that the selection criteria was biased and that the results did not show that treatments were effective compared with natural recovery. The implication is that the researchers were also biased – they knew what they wanted to find and made the evidence fit. An even longer reply followed this (Wampold *et al.*, 2010), pointing out in great detail that the meta-analysis was notably transparent and coherent in all criteria applied and avoided the conceptual fudging that, they argued, undermined the existing guidelines on what therapies work for PTSD (the authors of which contributed to the earlier rebuttal). This is a topic that clearly raises temperatures, and in many ways it reflects the ongoing debate in psychotherapy more generally. We urge you to read all three papers and make up your mind about this. Is exposure necessary for the treatment of PTSD? Having read these papers, we would argue that at the very least this remains not proven.

The Big Psychotherapy Trials

We end, however, with attempts to study the effectiveness of different approaches within a single RCT. This has a number of advantages – it means that the conditions for the comparison can be carefully controlled, and the type of patients seen by therapists using the different therapies will be similar in terms of severity. In order to get meaningful data, however, the studies need to be large. Such studies

span several different sites – with teams of therapists working on the same project in different cities, in a series of ‘mini-RCTs’. When all these data are pooled together, the study has the power to detect differences between treatments. Few such studies have been conducted because of expense, but those that have been completed have proved to be very influential.

We’ll begin with the largest of them: Project MATCH (1997), a US study of three treatments for alcohol abuse, had 1726 participants. It involved multiple sites and compared 12-step facilitation (based on Alcoholics Anonymous), CBT coping skills, and motivational enhancement therapy (MET, based on motivational interviewing), all offered as either outpatient or follow-up to inpatient treatment. It also included an attempt to match patient characteristics with the treatment offered. The results showed no differences between treatments – all were effective in terms of abstinence at one year follow-up – and very little evidence for matching.

The UK Alcohol Treatment Trial (UKATT, 2005) was a British version of Project MATCH, conducted at five centres offering either MET or a new intervention called social network behavioural therapy. The researchers found no differences in outcome between the treatments, though interestingly there were differences at individual sites (Davidson, 2008) so if these individual RCTs at the different sites had been published separately the results would have been misleading.

The most well known of these trials, however, is the US National Institute for Mental Health (NIMH) Treatment of Depression Collaborative Research Project (Elkin *et al.*, 1989; Shea *et al.*, 1992). It is generally considered a landmark study – well planned and well executed. It involved 250 patients with unipolar depression randomly allocated to therapists (psychologists and psychiatrists) at different sites administering cognitive therapy, interpersonal therapy, medication (imipramine) with clinical management, or a placebo with clinical management. The therapists were experienced, followed treatment manuals, had expert supervision and were subject to fidelity tests to ensure they were providing what they were supposed to be providing. The results showed no significant differences between treatments. Interestingly, there were again some differences between treatments at some sites.

Other Sources of Evidence?

We've now looked at the main sources of research evidence for examining differences in effectiveness of psychotherapies. We're going to finish with something slightly different. A common criticism of research trials is that they do not reflect the everyday experience of therapists – patients tend to have single problems, therapists are highly trained in one approach, etc. If routine clinical practice involves routine outcome assessment, however, it has the potential to be another source of research evidence. This was the approach taken by a group in West Yorkshire in the United Kingdom. In two studies (Stiles *et al.*, 2006; 2008), outcomes were compared for the six approaches offered in this service (CBT, person-centred or psychodynamic therapy alone or any one of these plus one additional approach: integrative, supportive or art therapy). These studies involved a very large number of clients: 1309 in the first, 5613 in the second. In both papers they reported no difference between the treatments – they were all effective.

The usefulness of this naturalistic approach to gathering evidence on effectiveness has been cited by Jacques Barber in his inaugural address as president of the American Psychological Association in 2009 (Barber, 2009) as a bridge to the better controlled RCTs. Of course it is not an RCT – as Clark, Fairburn and Wessely (2008) have pointed out, we don't actually know what the therapists were doing. The point is, however, that they were doing what the vast majority of therapists do in real life, and their choice of therapy made no difference to outcome. The point is also, of course, that the properly controlled RCTs and meta-analyses discussed earlier have obtained the same result.

Conclusions

What are we to make of all this? Research on the effectiveness and equivalence of the psychotherapies has a long history. Psychotherapy

is proven to be very effective, but is there now any scientific evidence for the superiority of one therapy over another? At the moment, we think that it is difficult to read this literature and argue convincingly that there is. The message to therapists then is to hold fast, and not switch therapies to a newer ‘improved’ version.

Why might research show all therapies to be equivalent? This result is, on the face of it, bizarre – that all therapies, new, old, those based on nineteenth-century ideas of the unconscious, those based on modern research into cognitive psychology, all therapies work just about as well as any other. How? This is an interesting question, and one no one knows the answer to. There are many possibilities, including the four we outline as follows:

1. All therapies really are equivalent and the reason for this is they all share the same active ingredients (i.e., common factors), mostly concerning the relationship between therapist and client.
2. All therapies really are equivalent and the reason for this is they all work in very different ways (i.e., uncommon factors), but each of these ways is equally effective.
3. Therapies are not equivalent at all and the reason we haven’t spotted this yet is that our research methods are not yet good enough.
4. All therapies are roughly equivalent, but there might be some conditions where some types of therapy tend to be more effective or quicker at achieving a result; however, these differences are small and tricky to spot.

Of course the reality might be a messy combination of several of these possibilities.

At the moment the evidence we have leaves us with a default position of assuming no important differences in therapy effectiveness. This leaves us with two questions. The first is why does this large amount of evidence not find its way into evidence-based guidelines in psychotherapy? The second is, if it is not the form of therapy that makes a difference to outcome, what does?

We'll ponder these two questions in the next chapter, when we'll look more closely at the research methodologies underlying psychotherapy research, and see how much research can help us in understanding not just whether a psychotherapy works but how.

References

- Barber, J.P. (2009) Toward a working through of some core conflicts in psychotherapy research. *Psychotherapy Research*, 19, 1–12.
- Begley, S. (2009) Ignoring the evidence. Why do psychologists reject science? *Newsweek* (Oct 2), <http://www.newsweek.com/id/216506> (accessed 11 September, 2011).
- Benish, S., Imel, Z.E. and Wampold, B.E. (2008) The relative efficacy of bona fide psychotherapies of post-traumatic stress disorder: a meta-analysis of direct comparisons. *Clinical Psychology Review*, 28, 746–758.
- Bergin, A.E. (1971) The evaluation of therapeutic outcomes, in *Handbook of Psychotherapy and Behaviour Change* (eds A.E. Bergin and S.L. Garfield), John Wiley & Sons, Inc, New York.
- Chambless, D.L., Baker, M.J., Baucom, D.H., *et al.* (1998) Update on empirically validated therapies, II. *The Clinical Psychologist*, 51 (1), 3–16.
- Chambless, D.L. (2002) Beware the dodo bird: the dangers of overgeneralization. *Clinical Psychology: Science and Practice*, 9, 13–16.
- Clark, D., Fairburn, C. and Wessely, S. (2008) Psychological treatment outcomes in routine NHS services; a commentary on Stiles *et al.* (2007). *Psychological Medicine*, 38, 629–634.
- Cochrane, A.L. (1972) *Effectiveness and Efficiency. Random Reflections on Health Services*, Nuffield Provincial Hospitals Trust, London.
- Cohen, J. (1992) A power primer. *Psychological Bulletin*, 112, 155–159.
- Cuijpers, P., van Straten, A., Andersson, G. and van Oppen, P. (2008) Psychotherapy for depression in adults: a meta-analysis of comparative treatments. *Journal of Consulting and Clinical Psychology*, 76, 909–922.
- Davidson, R. (2008) Accredited, brand named psychotherapies and the standard of evidence. *Clinical Psychology Forum*, 191, 48–52.
- Ehlers, A., Bisson, J., Clark, D.M., *et al.* (2010) Do all psychological treatments really work the same in posttraumatic stress disorder. *Clinical Psychology Review*, 30, 269–276.

- Ekers, D., Richards, D. and Gilbody, S. (2008) A meta analysis of randomized trials of behavioural treatment of depression. *Psychological Medicine*, 38, 611–623.
- Elkin, I., Shea, M.T., Watkins, J.T., *et al.* (1989) National Institute of Mental Health treatment of depression collaborative research program. General effectiveness of treatments. *Archives of General Psychiatry*, 46, 971–982.
- Eysenck, H.J. (1952). The effects of psychotherapy: an evaluation. *Journal of Consulting Psychology*, 16, 319–324.
- Eysenck, H.J. (1965) The effects of psychotherapy. *International Journal of Psychiatry*, 1, 97–178.
- Grissom, R.J. (1996) The magical number .7 +/- .2: meta-meta-analysis of the probability of superior outcome in comparisons involving therapy, placebo, and control. *Journal of Consulting and Clinical Psychology*, 64, 973–982.
- Luborsky, L., Singer, B. and Luborsky, L. (1975) Comparative studies of psychotherapy. *Archives of General Psychiatry*, 32, 995–1008.
- Luborsky, L., Rosenthal, R., Diguier, L., *et al.* (2002). The dodo bird verdict is alive and well – mostly. *Clinical Psychology: Science and Practice*, 9, 2–12.
- Messer, S.B. and Wampold, B.E. (2002), Let's face facts: common factors are more potent than specific therapy ingredients. *Clinical Psychology: Science and Practice*, 9, 21–25.
- National Institute for Health and Clinical Excellence (2009) *The Treatment and Management of Depression in Adults (update)*. Clinical Guideline 90. <http://www.nice.org.uk/CG90niceguideline> (accessed 11 September, 2001).
- Project MATCH (1997). Matching alcoholism treatment to client heterogeneity. *Journal of Studies on Alcohol*, 58, 7–29.
- Rosa-Alcázar, A., Sánchez-Meca, J., Gómez-Conesa, A. and Marín-Martínez, F. (2008) Psychological treatment of obsessive–compulsive disorder: a meta-analysis. *Clinical Psychology Review*, 28, 1310–1325.
- Rosenthal, R. (1990) How are we doing in soft psychology? *American Psychologist*, 45, 775–777.
- Rosenzweig, S. (1936) Some implicit common factors in diverse methods of psychotherapy. *American Journal of Orthopsychiatry*, 6, 422–425.
- Shapiro, D.A. and Shapiro, D. (1982) Meta-analysis of comparative therapy outcome studies: a replication and refinement. *Psychological Bulletin*, 92, 581–604.

- Shea, M.T., Elkin, I., Imber, S.D., *et al.* (1992) Course of depressive symptoms over follow up: the NIMH TDCRP. *Archives of General Psychiatry*, 49, 782–787.
- Smith, M.L. and Glass, G.V. (1977) Meta-analysis of psychotherapy outcome studies. *American Psychologist*, 32, 752–760.
- Smith, M.L., Glass, G.V. and Miller, T.I. (1980) *The Benefits of Psychotherapy*, Johns Hopkins University Press, Baltimore.
- Stiles, W.B., Barkham, M., Mellor-Clark, J. and Connell, J. (2008) Effectiveness of cognitive-behavioural, personcentred, and psychodynamic therapies in UK primary care routine practice: replication in a larger sample. *Psychological Medicine*, 38, 677–688.
- Stiles, W.B., Barkham, M., Twigg, E., *et al.* (2006) Effectiveness of cognitive-behavioural, person-centred and psychodynamic therapies as practised in UK National Health Service settings. *Psychological Medicine*, 36, 555–566.
- UKATT Research Team (2005) Effectiveness of treatment for alcohol problems: findings of the randomised UK alcohol treatment trial (UKATT). *British Medical Journal*, 331, 541–546.
- Wampold, B.E., Mondin, G.W., Moody, M., *et al.* (1997) A meta-analysis of outcome studies comparing bona fide psychotherapies: empirically, ‘all must have prizes.’ *Psychological Bulletin*, 122, 203–215.
- Wampold, B.E., Imel, Z.E., Laska, K.M., *et al.* (2010) Determining what works in the treatment of PTSD. *Clinical Psychology Review*, 30, 923–933.
- Wolitzky-Taylor, K., Horowitz, M., Powers, M. and Telch, M. (2008) Psychological approaches in the treatment of specific phobias: a meta-analysis. *Clinical Psychology Review*, 28, 1021–1037.