Counselling *for* Depression

A Person-centred and Experiential Approach to Practice

Pete Sanders and Andy Hill

bacp
British Association for
Counselling & Psychotherapy

SAGE
Los Angeles | London | New Delhi
Singapore | Washington DC

2

Evidence-Based Practice and Person-Centred and Experiential Therapies

Andy Hill and Robert Elliott

Evidence-based practice (EBP)

What is EBP?

During the last 20 years the evidence-based paradigm has gained pre-eminence and in the UK context has been enshrined in the process of developing clinical guidelines, as set out by NICE. In the world of UK public healthcare a rigorous assessment of the efficacy – and increasingly the cost-effectiveness – of treatments is necessary before recommendations can be made. This is to ensure that healthcare interventions are effective and represent value for money for the taxpayer. Wasting taxpayers' money by providing ineffective treatments can open governments and their departments to political and ethical criticism, which they are anxious to avoid.

EBP is a scientific paradigm which has been adopted as a defence against such criticism. Originating in the world of medicine, as opposed to psychological therapy, it was described by Sackett et al. (1996: 71–2), as a combination of rigorous science and professional judgement:

> the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of Evidence-based Medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research. Good doctors use both individual clinical expertise and the best available external evidence, and neither alone is enough. Without clinical expertise, practice risks becoming tyrannised by evidence, for even excellent external evidence may be inapplicable to or inappropriate for an individual patient. Without best current evidence, practice risks becoming rapidly out of date, to the detriment of patients.

The paradigm aims to integrate clinical judgement with the findings of high-quality research to ensure healthcare interventions are guided by the best contemporary knowledge of effectiveness in order to maximise outcomes for service users. The implementation of the paradigm within the UK healthcare context includes a number of elements:

- A definition of what constitutes high-quality efficacy research (i.e. randomised controlled trials)
- A technology for aggregating and synthesising the findings from multiple high-quality research studies (i.e. systematic reviews and meta-analyses)
- A method for translating research findings into practice and identifying where there are gaps in research (i.e. guideline development groups)
- Infrastructure to ensure research is carried out where gaps exist (i.e. grant funding programs)

Randomised controlled trials (RCTs)

As EBP is currently defined in the UK, the RCT has been accorded with significant status as a highly rigorous research design capable of producing reliable evidence on the effectiveness of treatments. In terms of the interventions they provide, practitioners need to know whether a treatment has been shown to produce beneficial effects and which of the wide range of therapies available appears to work best (Bower and King, 2000). Although not without its limitations, the RCT design is well placed to answer these questions, resulting from its ability to establish cause and effect relationships between interventions and outcomes. There are many reasons other than therapy why a client may improve while in therapy. The simple passage of time may be enough to produce recovery (so-called 'spontaneous remission'); extra-therapy factors such as the client's social circumstances may change, leading to greater levels of support and better interpersonal relationships; involvement in new self-help activities such as regular exercise may also be a factor. These variables need to be controlled in order to establish a causal link between therapy and its outcomes. Within a research study, the degree of confidence that a causal link has been established by ruling out other causes is known as the *internal validity* of the study. This contrasts with *construct validity*, the degree to which the therapy delivered is what it is claimed or theorised to be, and *external validity,* the degree to which the relationships within the study can be generalised beyond the research to practical situations: whether the findings of the study hold true in other settings, with other therapists and clients in the real world.

A number of strategies are employed to protect the internal validity of RCTs. Clients need to be selected on the basis of the type of problem that is being targeted and according to specified levels of severity. There may be additional criteria governing the recruitment of clients including the absence of comorbid problems such as drug or alcohol misuse. It is clear to see how these kinds of controls, which are necessary in trials, would be impractical in routine practice. Where a trial seeks to establish whether or not an intervention is effective a *no-treatment control group* will be introduced, primarily to control for the passage of time and spontaneous remission. It is well known that a proportion of psychological problems will improve over time without any professional

intervention. Control groups prevent spontaneous remission being confused with the effects of an intervention by ensuring that it affects both the control and the active treatment group. If it is found that the levels of recovery in the intervention group are not significantly different from those in the control group, then, all things being equal, it can be concluded that the intervention did not provide benefits over and above what would have happened anyway without therapy and hence that the treatment was not effective. If participants in the active treatment group did worse than those in the no-treatment control group then it could be concluded that the intervention had negative effects and was harmful. Without a control group it would be difficult to discern these effects with any degree of certainty.

A somewhat different issue is the construct validity of the treatments, whether the therapy is delivered as intended and what the active nature of the therapy is. Thus, adherence and competence checks are used to evaluate whether the therapists within the trial are delivering the therapy they are supposed to. An example of such a check is the Person-Centred and Experiential Psychotherapy Scale (PCEPS), which is described in Chapter 10. A further issue that can threaten the construct validity of trials is the so-called placebo effect of being offered therapy as opposed to being put on a waiting list. Unlike with drug trials where participants receive either a drug or placebo and are unaware of which of these they are receiving, in trials of psychological therapies it is not possible to disguise whether or not participants are receiving therapy. This sets up psychological differences between the groups, where those assigned to the waiting list may feel disappointed and those assigned to treatment may feel hopeful of a resolution of their problems. This instillation of hope may be the actual active ingredient in the treatment rather than the theorised change processes, such as the empathic exploration of depressive experiences. A third important construct validity issue is the fact that there are nonspecific factors common to all therapies that are likely to be responsible for effects. The common factors issue derives from the fact that all therapies share common features such as attention, empathy, supportive listening and a coherent theoretical framework. The extent to which these factors produce benefits as opposed to more technical aspects of therapy is an important question. Some trials have addressed this question by introducing sessions of non-directive, supportive listening to the control group, in an attempt to separate out the effects of the more technical aspects of the therapy. The use of placebo and non-specific control groups in therapy research has been widely criticised, especially when applied to relational therapies such as psychodynamic and humanistic approaches, where it can be difficult, if not impossible, to sort out common factors from those unique to the particular approach.

This leads us on to *comparative trials,* which seek to ascertain whether one type of therapy is superior to another by using two or more different active treatments in order to identify the most effective one. This type of 'horse race' study assumes that psychological therapy is generally beneficial, but aims to establish which type of therapy works best for a particular client population. The advantages of this study design are that the positive expectations that come from receiving treatment (as opposed to being assigned to a waiting list control group) are equal in all groups within the trial and that common factors are controlled as they, by definition, exist in all treatment groups. The

main downside of this design is that differences between active treatments are typically fairly small, and so it often takes a large, expensive study with many clients to detect differences.

Another key feature of RCTs is randomisation, used to reduce the likelihood of significant pre-treatment differences between the groups that could produce post-treatment differences in outcomes. As long as the number of participants in a trial is large enough, then the random allocation of participants to the different groups within a trial will make it unlikely that there will be any systematic differences between the groups on any variable (either known or unknown). If this is the case, then between-group differences in post-treatment outcome can be attributed to differences between treatments rather than pre-existing differences between groups.

External validity is a term used to describe the extent to which the findings of a trial can be applied to contexts outside of the study, particularly to routine practice settings. There tends to be an inverse relationship between internal and external validity; high levels of internal validity typically lead to low external validity and vice versa. While both types of validity are desirable for RCTs, the inherent conflict between the two has led to the development of a dichotomy in the design of trials. Those that privilege internal validity, often referred to as *explanatory* trials, emphasise strict control of variables to ensure that the causal relationship between intervention and effects is not compromised. Usually this degree of control is neither feasible nor desirable in routine settings and so such studies are more akin to laboratory-type experiments than evaluations of what happens in routine practice: interventions are clearly specified; the number of therapy sessions is specific; participants are recruited according to rigid criteria. In contrast, *pragmatic* trials seek to strike more of a balance between internal and external validity: therapy is provided more flexibly; recruitment of participants to the trial is more reflective of the clients who would normally present to a service for therapy. The aim here is less to establish a cause-effect relationship between intervention and outcome, and more to evaluate interventions in routine settings to identify benefits to clients using services. In making judgements about the results of trials it is important to take on board the type of trial in question, as different criteria apply in assessing the two different types of trial. For more information on RCT methodology see Torgerson and Torgerson (2008).

Systematic reviews

An individual RCT study, although producing significant results, does not usually represent adequate evidence on which to base major policy decisions that may affect millions of health service users. A body of evidence is required, bringing together multiple studies and weighing their findings. The traditional literature review has in the past provided this function, where a subject expert would select a group of papers and provide a summary of these. The problem with this type of review is that they are open to charges of bias both in the selection of the original papers and the interpretation of their findings. A hallmark of the RCT study design is to reduce bias to a minimum; for the

review of such studies to be so susceptible to bias undermines this intention. Systematic review methodology has been developed to address this problem and support EBP.

Unlike traditional reviews, systematic reviews are explicit in their methodology and transparent in how conclusions are derived. They can take the form of narrative reviews, where research evidence is analysed and summarised in verbal form, or they can employ a statistical technique known as meta-analysis. This latter technique aggregates statistical data from multiple studies to produce a calculation of effectiveness and has the benefit of pooling the results of RCTs with small sample sizes that may, in themselves, lack the statistical power to detect small but important effects. When the results of these small studies are pooled, significant and robust findings may emerge. Systematic reviews set out to answer specific questions and follow a detailed protocol for how the review will be conducted. The question should specify who the participants are, the intervention and the outcome(s) of interest. For example:

1. adult users of primary care
2. receiving counselling
3. for depression

A search strategy should specify the search terms used and which of the electronic databases of published literature are to be searched. If a search of the unpublished literature is to be conducted then this also should be specified. Inclusion and exclusion criteria are necessary to clarify how the individual studies were selected for inclusion in the review and what quality criteria were used to ensure that the review findings are based upon the most reliable studies. Given that some degree of bias is inevitable, strategies should be in place to reduce bias and all stages of the review process should be as transparent as possible. While systematic reviews aim to guide healthcare interventions based on robust and reliable evidence, it is important to bear in mind that the quality of a review's findings depends on the quality of the primary research (i.e. the individual RCTs and other studies) on which the review is based. It is therefore quite common, where there is an absence of reliable primary research, for reviews to be inconclusive as regards recommendations for clinical practice, but specific in identifying areas where new research studies are needed. Hence systematic reviews can be very useful in telling what we don't know.

As long ago as 1972, a British epidemiologist, Archie Cochrane, recognised healthcare professionals' need for rigorous and reliable reviews of scientific literature. The Cochrane Collaboration was established in response to this call. It is an international organisation aiming to help health professionals make well-informed decisions by supplying them with up-to-date systematic reviews of RCT evidence. There are a number of groups within the collaboration specialising in particular areas of healthcare, including mental health. A rigorous and standardised method of meta-analysis is used across all the groups to ensure the research evidence is of the highest quality. Reviews are published in the Cochrane Library, which is updated on a regular basis.[1] For more information on systematic review methodology see Petticrew and Roberts (2006).

[1] www.thecochranelibrary.com/view/0/index.html (retrieved 04/05/2013).

The National Institute for Health and Care Excellence

The role of NICE is to provide national guidance and advice to improve health and social care. Its scope is very wide, covering bio-medical interventions, mental health, public health and social care. Along with the provision of information services for those working in health and social care, it produces evidence-based guidance and develops quality standards and metrics to measure the performance of services.

Originally set up in 1999 as the National Institute for Clinical Excellence, in 2005, after merging with the Health Development Agency, it began developing public health guidance to help prevent ill health and promote healthier lifestyles, changing its name to the National Institute for Health and Clinical Excellence. In April 2013 NICE took on responsibility for developing guidance and quality standards in social care, and its name changed once more to reflect these new responsibilities. NICE is accountable to the Department of Health, but is operationally independent of government. The National Collaborating Centre for Mental Health has been commissioned by NICE to produce guidelines relating to psychological and other interventions for mental health problems.

NICE guidance and recommendations are made by independent committees. Topics for guidance and appraisals are decided by the Department of Health, based on a number of factors, including the burden of disease, the impact on resources and whether there is inappropriate variation in practice across the country. The process of guideline development uses the best available evidence and includes the views of experts, patients and carers, and industry. Guidance is reviewed regularly to ensure it is up to date and a consultation process is in place to allow individuals, patient groups, charities and industry to comment on recommendations. An independent guideline development group is established for each clinical guideline being developed, including health professionals and patient/carer representatives with relevant expertise and experience. Registered stakeholders are invited to nominate people to join the group. There is a specific protocol for guideline development[2] which uses rigorous systematic review methods, looking at the evidence available and considering comments made on draft versions of the guideline issued for consultation before producing the final version. Not only does NICE provide guidance on health and social care practice, but also it identifies gaps in the evidence base and makes recommendations for future research.

National Institute for Health Research (NIHR)

For new research to take place infrastructure is needed to support it. Funded through the Department of Health and established in 2006, the NIHR[3] aims to set research

[2] http://publications.nice.org.uk/the-guidelines-manual-pmg6 (retrieved 12/11/2013).
[3] www.nihr.ac.uk/about/Pages/default.aspx (retrieved 12/11/2013).

priorities and provide funding for social care and public health research in order to improve treatments for the benefit of patients. The Health and Social Care Act 2012 places a statutory duty on the Secretary of State to promote and support research and the NIHR provides a key means through which this duty is discharged. Its role is to develop the research evidence to support decision-making by professionals, policy makers and patients, make this evidence available and to encourage its uptake and use. It is for other organisations, such as NICE, to use the research evidence to provide national guidance on promoting good health and preventing and treating ill health. NIHR's key objective is to improve the quality, relevance and focus of research in the NHS and social care by distributing funds in a transparent way after open competition and peer review. The NIHR funds a range of programmes addressing a broad range of health priorities. Funding is based on the quality and relevance of the research to social services, public health and the NHS.

Criticisms of EBP

Whereas providing therapy to clients provides the foreground for working as a counsellor in health and social care settings, EBP provides the backdrop. A fundamental assumption in such settings is that practice should be informed by research evidence and so open to revision and adaptation in the light of new findings. The nature of evidence tends to be defined by positivist epistemology, supported by RCT and meta-analytical methods and implemented by organisations established to generate research knowledge and practice guidelines. While laudable in its intentions to improve treatments for patients and avoid harmful and wasteful practices, EBP has presented significant challenges, particularly for counsellors working in NHS settings. A number of criticisms have been levelled at EBP, particularly with regard to its reliance on RCTs.

To begin with, EBP had its origins in evidence-based medicine (Sackett et al., 1996), which means that it uses methods originally developed to test biomedical treatments. These methods are now being applied to the evaluation of psychosocial interventions. Thus, drug-trial methods are being used to test the effectiveness of relational and interpersonal interventions as opposed to those that rely on biochemical mechanisms and direct physical intervention (e.g. surgery). Unlike in drug trials, where those in the control condition receive a placebo, in trials of psychological therapies participants cannot be blinded as to whether or not they are receiving an active treatment. This in itself can produce psychological differences in the control and the intervention group, with those being offered therapy feeling more positive and hopeful for recovery, as discussed previously.

The fact that there are many different 'brands' of therapy (like different brands of drugs) gives the impression that these are all distinctive interventions with their own particular techniques and mechanisms of change. However, they are all talking therapies, relying on interpersonal relating, which means that they share important common or

non-specific factors. Examples are: being related to in a warm and collaborative manner; being listened to supportively; being empathised with; being offered hope within a theoretical framework that explains problems and how to get better. These factors are common across therapies and are responsible for a proportion of the positive outcomes experienced by clients. In RCTs it is difficult to disaggregate specific effects (those techniques and methods specific to particular types of therapy) from non-specific effects (those elements common to all therapies) as a therapeutic approach tends to be delivered as a package in a trial. If indeed it is the non-specific factors as opposed to the specific factors which produce the majority of therapeutic change for clients then it follows that all types of therapy will be more or less equally effective, a common finding where different therapies have been tested against each other. In a reference to the story of Alice in Wonderland, this is often referred to as the *Dodo Bird Verdict* (Wampold, 2001). By casting doubt upon exactly what the therapeutic ingredients are that have been tested, the common factors argument tends to undermine RCT evidence supporting the different 'brands' of therapy.

Triallists (that is, people who believe in the centrality of RCTs as a scientific method) have responded to this challenge by creating an *attention/placebo* to be given to the control group, to compensate for common factors. Most often referred to as *supportive therapy* (or sometimes, confusingly, as *non-directive supportive therapy*), this is an invented control treatment in which clients receive various kinds of psychological support, including sympathetic, supportive listening, with varying degrees of empathy, sharing and practical problem-solving; in contrast, those in the intervention group receive a structured and theoretically grounded therapy such as CBT, psychodynamic therapy or person-centred therapy. Supportive therapy conditions often have specific treatment elements such as exploration of traumatic events systematically removed, in a research strategy that has been referred to as 'intent to fail', s o that they can make a contrasting treatment look better. While this represents a useful step towards controlling for common factors, it may still not provide a full placebo effect, if those receiving it are aware that they are not receiving a full, bona fide therapeutic intervention.

A further problem lies in the delivery of the intervention. Whereas in drug trials the health professionals administering the drug are independent of the intervention (generally it does not matter who supplies the drug to the trial participant), in RCTs of psychological therapies the therapist is confounded with the therapy (therapists work in particular ways and have different personal qualities, so it does matter who delivers the therapy). To control for this, for example, where two therapies are being compared, the same therapists may deliver both therapies in the trial. While this helps to eliminate therapist effects it does raise the problem of whether therapists have the ability to deliver different therapeutic approaches with equal amounts of skill and commitment.

Further criticisms levelled at RCTs relate to their lack of external validity and the cost of conducting them. The fact that conditions are often carefully controlled in RCTs (particularly explanatory trials) means the applicability of their findings to routine settings where conditions are generally uncontrolled is questionable.

Therapists in routine practice rarely have the luxury of selecting their clients according to their having single diagnoses and specific levels of problem severity; in routine practice complexity and comorbidity are the norm. Likewise, in routine practice, therapy is delivered flexibly, responding to clients' needs as opposed to according to a manual or therapeutic protocol. This lack of external validity has led to the argument that the findings of RCTs are only applicable to other RCTs and are irrelevant to routine practice. The cost of conducting this type of study is also considerable, making it unfeasible that all models of psychological therapy could be tested in RCTs. The net result of this is we are left with a 'first past the post' scenario with CBT recommended as a frontline therapy, based on its extensive RCT evidence base, and other therapies struggling to find funding for research and occupying very marginal positions in clinical guidelines, or being excluded completely. This state of affairs does not mean that these excluded or marginal therapies do not work (a lack of evidence of effectiveness is not the same as evidence of ineffectiveness). It simply means that they have not yet been tested by RCT methodology; whether, at some point in the future, the funding and resources will become available for this to occur is a moot point. The resulting scenario has resulted in a narrowing of practice within health-care settings and the marginalisation of a range of not yet tested, but likely to be effective, therapeutic approaches, with a consequent reduction in choice of therapies available for service users.

The focus on RCTs has led to two other important kinds of scientific evidence being ignored. First, systematic case studies use rich case records of qualitative and quantitative data to assess client outcome and the causal role of therapy. They can be used by counsellors and psychotherapists to study their own practice and to document the possible effectiveness of new approaches and new client populations. Second, pre-post studies (also known as open clinical trials) are useful for documenting the amount of pre-post client change, which can be useful for evaluating the possible effectiveness of new or emerging treatments and for creating benchmarks against which to assess routine practice. Thus, an exclusive focus on RCT evidence slights both innovative, emerging approaches and also real-world practice.

The RCT design itself suffers from both excessive support, with some viewing it as the only useful type of research on which to base policy decisions, and excessive denigration, with others viewing the method as reductionist and inappropriate for the field of counselling and psychotherapy. A more balanced position would be to recognise the contribution RCTs can make to our knowledge of the field and also their many limitations. RCT evidence can provide a useful counterbalance to clinical judgement, which at times can be flawed, and also an antidote to blind adherence to particular approaches based on faith or ideology. Such evidence also provides some protection for the public in identifying treatments that work and those that may be ineffective or harmful. It should also be recognised that an accumulation of RCT evidence is responsible for psychological therapies now being recommended as frontline treatments for common mental health problems (as opposed to medication alone). Without RCTs this may not have happened.

Research on the effectiveness of PCE therapy with depression

In this section we review two different sets of research on the effectiveness of PCE therapy with depression: the evidence reviewed by the Guideline Development Group (GDG) in constructing the NICE guideline for depression with adults NICE (2009a) and Elliott et al.'s (2013) meta-analysis of 27 studies of humanistic psychotherapy and counselling, the latter having a much broader focus than the former.

Evidence reviewed for the NICE guideline for depression

This section presents a discussion of the evidence for the effectiveness of counselling reviewed by the GDG for the production of the depression guideline. A full summary of the results of the studies is available (NICE, 2009a).[4] Table 2.1 is extracted from the full version of the NICE depression guideline document. (Note that although it had been included in the previous guideline and was used for utility analyses, the Ward et al., 2000 study was excluded from the main analyses in the 2009 update.)

Characteristics of included studies

Several studies reviewed by the depression GDG were rejected as they did not meet the criteria for inclusion. A study using a non-RCT method (Marriott and Kellett, 2009) compared counselling, cognitive analytic therapy and CBT in routine service settings. The study used neither randomisation to treatment group nor a no-treatment control group. Depression pre- and post-treatment was measured using the Beck Depression Inventory (BDI). This study was excluded by the GDG from the analysis on the grounds that its sample size was too small to reach any definitive conclusion on the differential effectiveness of the treatments. Additionally, just 34% of the sample had a diagnosis of depression, making it difficult to draw any conclusions about the interventions' effectiveness as treatments for depression. A further non-RCT study (Stiles et al., 2006) compared CBT, psychodynamic therapy and person-centred therapy in routine NHS settings, using CORE as the outcome measure. As with Marriott and Kellett (2009) randomisation and a control group were not part of the study design. This study was excluded by the GDG because, once again, not all participants in the study met the criteria for depression and other diagnoses were included in the sample, making it difficult to draw conclusions about the effectiveness of the interventions with depression. Ward et al. (2000) was initially excluded on similar grounds; only 62% of the participants

[4] www.nice.org.uk/nicemedia/live/12329/45896/45896.pdf (retrieved 12/11/2013).

Table 2.1 Counselling studies reviewed in the Depression Guideline (NICE 2009a)

Study	Participants	Interventions	Depression measures used
Bedi et al. (2000) (RCT)[1] (UK)	clients recruited via GP n=103 diagnosed by GP for depression using RDC[2]	• antidepressants • six sessions of counselling using a flexible approach according to needs of clients	BDI[3] and RDC taken at 8 weeks and 12 month follow-up
Goldman et al. (2006) (RCT) (Canada)	n=38 all with major depression measured by DSM-IV	• client-centred therapy: 9–20 sessions • EFT[4]: 9–20 sessions	SCL-90[5], BDI
Greenberg et al. (1998) (RCT) (Canada)	n=34 all with major depression measured by DSM-III-R	• client-centred therapy: 15–20 sessions • process-experiential therapy (EFT): 15–20 sessions	SCL-90, BDI
Simpson et al. (2003) (RCT) (UK)	clients recruited from 9 GP practices n=145 all with depression score>14 on BDI	• 6–12 sessions of psychodynamic counselling + usual care • usual care (note usual care in some cases involved use of medication)	BDI at 6 and 12 months
Watson et al. (2003) (RCT) (Canada)	n=93 all with major depression measured by DSM-IV	• CBT: 16 sessions • process-experiential therapy (EFT): 16 sessions	SCL-90, BDI
Ward et al. (2000)[6] (RCT) (UK)	clients referred by GPs, n=464, 62% diagnosed with depression (BDI>14)	• usual GP care • CBT • non-directive counselling based on Rogers' approach • Duration of therapy: 6–12 weekly sessions	BDI at baseline, 4 months and 12 month follow-up

[1]randomised control trial
[2]Research diagnostic criteria. See: Spitzer, R.L., Endicott, J. and Robins, E. (1978) 'Research diagnostic criteria: rationale and reliability.' *Archives of General Psychiatry,* 35: 773–82.
[3]Beck Depression Inventory. See: Beck, A.T., Ward, C.H., Mendelson, M. et al. (1961) 'An inventory for measuring depression.' *Archives of General Psychiatry,* 4: 561–71.
[4]Emotion-focused therapy.
[5]Symptom Checklist 90. For more information see: http://psychcorp.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=PAg514 (retrieved 12/11/2013).
[6]Initially excluded from the review, but a subgroup analysis of the data from the trial was included following this initial decision.

met the diagnosis for depression, and also the study was not completely randomised. However, a subsequent subgroup analysis focusing only on those participants who met the criteria for depression in this trial was submitted and was included in the evidence review. The results of this analysis indicated a significant medium-sized effect on depression scores post treatment but no significant effect at follow-up.

Five studies met the criteria for inclusion and were included in the evidence review (all in Table 2.1 except Ward et al., 2000). These studies form the basis of the evidence for the effectiveness of counselling in the NICE (2009a) guideline. Bedi et al. (2000) compared the effectiveness of counselling versus antidepressants. No significant differences between the two types of treatment were found and at 12-month follow-up clinician-reported depression scores were significantly lower in the antidepressant group when compared with counselling. On this the GDG viewed the study as inconclusive and not supporting a conclusion that counselling and antidepressants were equivalent. They also stated that this study should be treated with some caution as the introduction of a patient preference element to the trial led to considerable differences in baseline severity measures between the two arms.

Two studies (Goldman et al., 2006; Greenberg and Watson, 1998) compared two different types of PCE therapy (client-centred and emotion-focused therapy). In Goldman et al. (2006) the comparison of client-centred counselling and EFT favoured EFT. In Greenberg and Watson (1998) the comparison of client-centred counselling and EFT (referred to as process-experiential counselling)[5] findings indicated that there was no significant difference between treatments in the reduction of self-reported depression scores. The GDG urged caution in the interpretation of these results because of what it considered to be small sample sizes.

Simpson et al. (2003) compared the combination of psychodynamic counselling plus GP care with usual GP care alone and found no important clinical benefit of therapy plus GP care. Watson et al. (2003) compared EFT with CBT. The GDG criticised this study on the basis of its sample size, judging it to be small and concluding that the study produced insufficient evidence to reach any definite conclusion about the relative effectiveness of the two treatments.

Despite a number of studies having major depression as a criterion for the recruitment of participants (see Table 2.1), the GDG concluded that participants in the reviewed studies were predominantly drawn from groups in the mild-to-moderate range of depression (mean baseline BDI scores between 18 and 26) and two trials included people with minor depression (BDI scores starting from 14) (Bedi et al., 2000 and Ward et al., 2000). Because of this the GDG concluded that evidence supported the effectiveness of counselling for mild-to-moderate depression but not for severe depression. The evidence was also seen to be limited by the small size of the samples of participants recruited into the studies, resulting in studies with low *power* to reliably detect differences between groups within trials. The concept of *power* refers to whether the sample size of a trial is large enough to detect differences that might exist between groups. A relatively small sample size can be used where the differences between groups

[5] Note: *process-experiential* and *emotion-focused* are two names for the same therapy.

are expected to be large, such as where a therapy is compared with no-treatment. Where two active treatments are compared such as EFT and CBT (Watson et al., 2003) and differences would be expected to be small, a much larger sample size would have been needed to detect differences. The evidence reviewed was thus judged to be limited partly because of small sample sizes. Another issue, relating to sample selection, is whether participants meet the criteria for depression. NICE guidelines are disorder specific and so it follows that the guidelines for depression should be based on studies of participants who were clearly depressed. Hence studies of participants who did not meet the full diagnostic criteria for depression or who had other prominent psychological problems besides depression were excluded from the evidence review.

The fact that some clients may have strong preferences for particular treatments presents a further problem for the randomisation process in RCTs. If allocated to a treatment they do not want, these participants may become demoralised, hence affecting the outcomes of the treatment they receive. Random allocation to groups is thus predicated on the notion that clients have no strong preference for treatment. Where strong preferences exist *patient preference* trials have been designed to enable those without strong preferences to be randomly allocated and those with strong preferences to be given the intervention they wish to receive. Whereas this is more ethical and helps with recruitment, it can have the effect of setting up the differences between groups that random allocation intends to prevent. This is a criticism levelled at the Bedi et al. (2000) study.

The GDG considered studies that compared two different forms of PCE therapy (client-centred and EFT) (Goldman et al., 2006; Greenberg and Watson, 1998) as problematic because they only evaluated the effects of two quite similar interventions. Had these therapies been compared with a no-treatment control group or comparison with a recommended treatment such as CBT, then the effectiveness of these therapies would have been more clearly established.

The definition of *counselling* presented some difficulty in assessing the evidence. A number of terms were used in the studies reviewed for the depression guideline, including client-centred, psychodynamic, process-experiential, emotion-focused (the latter two being different terms for the same therapy). This underlines some of the complexities around how *counselling* is defined. Within the counselling profession the term is viewed as something of an umbrella term embracing a variety of approaches. However, outside of the profession counselling tends to be viewed as a type of intervention distinct from other therapies such as CBT and psychodynamic therapy. This dual perspective tends to perpetuate misunderstandings about the nature of counselling. From an EBP perspective it also presents fundamental difficulties: reviewing research evidence and developing clinical guidelines require a more precise definition of the intervention in question.

Apart from one study (Simpson et al., 2003), the evidence that supports the inclusion of counselling in the NICE depression guideline consists mainly of RCTs of either person-centred therapy or EFT. This would suggest that the term *counselling* as defined in the NICE depression guideline is based on predominantly PCE types of therapy (namely client-centred, based on Carl Rogers' theories, and EFT). The fact that there is evidence for both person-centred and EFT also suggests that elements from both of these

approaches are effective in the treatment of depression and should form a basis for a more precise definition of *counselling* as specified in the guideline. The fact that the GDG viewed the evidence for counselling as limited underlines the need for more RCTs with larger samples, more strict selection of participants to meet the criteria for depression, and with more severely depressed populations. Furthermore, the focus should be on comparisons with waiting list control groups or with established treatments such as CBT.

The Elliott et al. (2013) meta-analysis

As noted earlier, a meta-analysis is an analysis of analyses, carried out in a number of stages. Initially, a number of studies examining the same research question are collected and their relevant characteristics coded (for example, number of participants, measures used, whether participants were randomly assigned to treatments). The next step is complex and involves putting the measures in all the individual studies onto the same metric, so that they can be combined and compared. This common metric is called an *effect size* (ES). The most frequently used ES measure is called the *standardised mean difference*, which is, in the case of this meta-analysis, the difference between the average pre-therapy score and the average post-therapy score divided by the *pooled standard deviation*. Standard deviation is a measure of the variability associated with an average.[6] The last step in a meta-analysis is analysing all the analyses (that's the 'meta' part of the process), running various corrections, coming up with summary values, and looking for variables that might explain differences in effect sizes (such as randomisation or level of therapists' experience).

In their meta-analysis of outcome studies of humanistic-experiential therapies, Elliott et al. (2013) took a much more inclusive approach to the evidence review process than was used by the NICE GDG. There were several reasons for this. First, when Elliott and colleagues began meta-analysing person-centred-experiential outcome research (Greenberg et al., 1994), there was very little research available and so they wanted to use all the available data, including evidence for emerging versions of PCE therapy and applications to new client populations. Second, they were concerned that selecting studies based on judgements of quality would introduce bias: if you don't like the results of a study, it is very easy to find faults with the statistics and design. Third, following the original philosophy behind meta-analysis (Glass et al., 1981), a wide range of studies using different methods was included and methodological features of studies were coded in order to make it possible to see what difference these made for the results. For example, there is an assumption that non-randomised studies are biased and thus produce different results from randomised studies.

In any event, Elliott et al. (2013) looked at approximately 200 studies of the outcome of PCE therapies (which they referred to as humanistic-experiential psychotherapies, or

[6] For more information on the calculation of standard deviation and effect size see Sanders and Wilkins (2010).

HEPs). Within this data set were five types of client presenting problem: depression, relationship problems, coping with chronic medical problems (e.g. HIV), habitual self-damaging behaviours (substance misuse, eating disorders) and psychosis. Of these, there were more studies of depression than any other client presenting problem. Twenty-seven studies of depression were included in the meta-analysis and form the basis of the discussion here. There were 34 samples of clients within the 27 studies, comprising a total of 1,287 clients. The types of therapy tested were most commonly person-centred therapy (10 samples), supportive therapy (often used as a control condition, as discussed earlier in this chapter) (9 samples), or EFT (8 samples). Other types of experiential therapy, such as gestalt or psychodrama, were also included. For a more detailed analysis of the results see Appendix 2 and Elliott et al. (2013).

The 27 studies fell into two broad categories: those which measured levels of distress pre- and post-therapy without the use of a control or comparison group (n=19) and similar studies which made use of comparison/control groups (n=8). Analyses were based on the calculation of effect size, where 0.2 and is viewed as small, 0.5 as medium and 0.8 as large (Cooper, 2008). The weighted mean pre-post effect size across all 34 samples was large. On the other hand, the effect size across just the 8 comparison/control studies was somewhat weaker, but still a statistically significant weighted effect in the small to medium range. Within this latter group of studies were two outliers (Maynard, 1993; Tyson and Range, 1987), where negative outcomes were found for the interventions compared with no-treatment groups. Both of these studies had small samples and used group interventions which were not bona fide PCE therapies.

Where PCE is compared with other types of therapy (23 studies), most commonly CBT, the outcomes are broadly equivalent: positive and negative comparative results are evenly balanced across the studies. Within the range of PCE therapies there is some preliminary support that process-guiding approaches may have some superiority over approaches that do not use these methods with depressed clients. Four studies made comparisons between more and less process-guiding therapies involving depressed clients (Beutler et al., 1991; Goldman et al., 2006; Greenberg and Watson, 1998; Tyson and Range, 1987). A significant small to medium mean effect size was found across these studies. However there was a degree of heterogeneity in the interventions tested: Greenberg and Watson (1998) and Goldman et al. (2006) compared EFT with client-centred therapy; Beutler et al. (1991) compared focused-expressive group therapy with a supportive group involving bibliotherapy; and Tyson and Range (1987) compared group gestalt therapy with an active expression group.

Two clusters of evidence on depression are worth noting. First, there are the three well-designed comparative treatment RCTs testing EFT for depression (Goldman et al., 2006; Greenberg and Watson, 1998; Watson et al., 2003), comparing EFT with other therapies in the treatment of major depressive disorder, using medium-sized samples and conducted by two different research teams. These studies, also discussed earlier in this chapter, were brought to the attention of the NICE GDG in the review process but were generally dismissed for being relatively small and for comparing related treatments. Goldman et al. (2006) found that EFT had significantly better outcomes (including very low relapse rates) when compared with person-centred therapy. Watson et al. (2003)

found generally equivalent (and on some measures better) results compared with CBT. Second, there were four well-designed RCTs of person-centred therapy for perinatal depression with medium to large sample sizes that either showed superiority to treatment as usual (Holden et al., 1989; Morrell et al., 2009; Wickberg and Hwang, 1996), or no difference in comparison with CBT (Cooper et al., 2003) or short-term psychodynamic therapy (Cooper et al., 2003; Morrell et al., 2009). Both of these clusters of well-controlled studies met Chambless and Hollon's (1998) criteria for *efficacious and specific* treatments: that is, they were well-designed, conducted by at least two different research teams, and were either superior to some other treatment or superior to a recognised efficacious treatment.

Key recent studies include Cooper et al. (2003) and Morrell et al. (2009), both with perinatal depression as mentioned above, and two studies by Mohr and colleagues on depression in a medical population (Mohr et al., 2001; 2005). The other recent substantial study is Stice et al. (2010), in which adolescents with mild to moderate depression were randomised to one of four conditions: supportive group therapy vs. CBT group therapy vs. CBT bibliotherapy vs. controls. Participants seen in supportive therapy showed benefits comparable with those in CBT which were sustained to two year follow-ups and did much better than control group clients.

In summary, the evidence-based paradigm has become part of the fabric of systems of healthcare delivery, both in the UK and elsewhere. The principles of this approach inform clinical guidelines and decisions about the commissioning of treatments. This has presented significant challenges to PCE therapies because of the relative paucity of RCT evidence compared with CBT and the consequential marginalisation of the PCE approach, particularly in healthcare settings. The narrowness of the scope used by NICE for the inclusion of research evidence into clinical guidelines inevitably excludes significant areas of research evidence supporting the effectiveness of PCE therapies, prompting the need for NICE's methods to be reviewed. A more comprehensive review of the research on the effectiveness of PCE therapies with depression, including perinatal depression, suggests PCE therapies have a significant positive effect, with effect sizes varying between small and large depending on the type of studies analysed. In comparisons with other types of therapy PCE approaches have broadly similar outcomes and there is some preliminary support for the superiority of process-guiding approaches, which, in turn, needs to be tested in further studies. Drawing upon this evidence base, and in response to the crisis evidence-based practice has presented to PCE therapies, CfD has been developed in an attempt to delineate an evidence-based form of PCE therapy, which is specifically adapted for working with depression and can help to consolidate the position of PCE therapists in UK healthcare settings.