

ORIGINAL ARTICLE

How should we evaluate research on counselling and the treatment of depression? A case study on how the National Institute for Health and Care Excellence's draft 2018 guideline for depression considered what counts as best evidence

Michael Barkham^{1,*} , Naomi P. Moller^{2,3}  & Joanne Pybis³

¹ Centre for Psychological Services Research, University of Sheffield, Sheffield, UK

² Open University, Milton Keynes, UK

³ British Association for Counselling and Psychotherapy, Lutterworth, UK

*Corresponding author. Email: m.barkham@sheffield.ac.uk

Abstract

Background: Health guidelines are developed to improve patient care by ensuring the most recent and 'best available evidence' is used to guide treatment recommendations. The National Institute for Health and Care Excellence's (NICE's) guideline development methodology acknowledges that evidence needed to answer one question (treatment efficacy) may be different from evidence needed to answer another (cost-effectiveness, treatment acceptability to patients). This review uses counselling in the treatment of depression as a case study, and interrogates the constructs of 'best' evidence and 'best' guideline methodologies. *Method:* The review comprises six sections: (i) implications of diverse definitions of counselling in research; (ii) research findings from meta-analyses and randomised controlled trials (RCTs); (iii) limitations to trials-based evidence; (iv) findings from large routine outcome datasets; (v) the inclusion of qualitative research that emphasises service-user voices; and (vi) conclusions and recommendations. *Results:* Research from meta-analyses and RCTs contained in the draft 2018 NICE Guideline is limited but positive in relation to the effectiveness of counselling in the treatment for depression. The weight of evidence suggests little, if any, advantage to cognitive behaviour therapy (CBT) over counselling once risk of bias and researcher allegiance are taken into account. A growing body of evidence from large NHS data sets also evidences that, for depression, counselling is as effective as CBT and cost-effective when delivered in NHS settings. *Conclusion:* Specifications in NICE's updated guideline procedures allow for data other than RCTs and meta-analyses to be included. Accordingly, there is a need to include large standardised collected data sets from routine practice as well as the voice of patients via high-quality qualitative research.

Keywords: counselling, depression, large data sets, NICE guidelines, qualitative research, randomised controlled trials

doi: 10.1002/capr.12141

The views expressed in this article are our own and do not necessarily reflect the views of our respective organisations.

Introduction

English health guidelines are created and regularly updated with the aim of improving patient care by ensuring that the most recent and 'best available

evidence' is used to guide treatment (National Institute for Health and Care Excellence Guidance, 2017a). As stated on its website: 'National Institute for Health and Care Excellence (NICE) guidelines are evidence-based recommendations for health and care

in England' (NICE Guidelines, 2017b). Although some NICE guidance is also adopted by Wales, Scotland and Northern Ireland, a separate UK-based body equivalent to NICE exists; namely the Scottish Intercollegiate Guidelines Network (2017). Mental health treatment guidelines are also developed by other international organisations, such as the World Health Organization (2017) and professional/scientific bodies such as the American Psychiatric Association (2017), and by European and other countries (Vlayen, Aertgeerts, Hannes, Sermeus & Ramaeker, 2005).

This article focuses on: (i) NICE guidelines because of the organisation's impact in shaping mental health care, not only in the UK but internationally (Hernandez-Villafuerte, Garau & Devlin, 2014); (ii) depression, as NICE is currently updating their depression guideline (NICE, 2017d), and; (iii) counselling as the intervention, as different guidelines have drawn different conclusions (Moriana, Gálvez-Lara & Corpas, 2017). Specially, we focus on the selection and use of evidence. In terms of overall methodology, in their procedural manual NICE state: 'Guidance is based on the best available evidence of what works, and what it costs' (NICE, 2014/2017, p. 14). Although the procedural manual states that randomised controlled trials (RCTs) are often the most appropriate design, it also states: 'However, other study designs (including observational, experimental or qualitative) may also be used to assess effectiveness, or aspects of effectiveness' (NICE, 2014/2017, p. 15). Accordingly, we assess the extent to which NICE has adhered to its own methods manual in drawing up the draft guideline. While NICE's depression guideline is used as the example, arguments in this article are intended to have broad relevance for any organisation developing guidelines across mental health treatments.

The new revision of the *NICE Guideline for Depression in Adults: Recognition and Management* is scheduled to be published in January 2018 and available as a consultation document at the time of writing (NICE, 2017d). The previous 2009 NICE Guideline stated: 'For people with depression who decline an antidepressant, CBT [cognitive behaviour therapy], IPT [interpersonal psychotherapy], behavioural activation and behavioural couples therapy, consider: counselling for people with persistent subthreshold depressive symptoms or mild to moderate depression' (NICE, 2009, p. 23). Counselling was included in the 2009 Guidelines but only for those who declined other recommended treatments; the guidelines were accordingly critiqued on the basis of limiting patient

choice (British Association for Counselling and Psychotherapy, 2009). In addition, practitioners offering counselling to adults with depression were recommended to: 'Discuss with the person the uncertainty of the effectiveness of counselling and psychodynamic psychotherapy in treating depression' (p. 24). This recommendation was criticised as research suggests that both patient hope and a good therapeutic relationship are important in creating good patient outcomes (Barber, Connoll, Crits-Christoph, Gladis & Siqueland, 2000). Accordingly, this recommendation would likely have negatively impacted on early engagement in counselling as well as on outcomes for counselling, if practitioners had implemented this guidance.

The consultation document for the 2018 proposed guideline states: 'Consider counselling if a person with less severe depression would like help for significant psychosocial, relationship or employment problems and has had group CBT, exercise or facilitated self-help, antidepressant medication, individual CBT or BA for a previous episode of depression, but this did not work well for them, or does not want group CBT, exercise or facilitated self-help, antidepressant medication, individual CBT or BA' (NICE, 2017d; Recommendation 64, p. 252). It also recommends that the counselling 'is based on a model developed specifically for depression, consists of up to 16 individual sessions each lasting up to an hour, and takes place over 12–16 weeks, including follow-up' (NICE, 2017d; Recommendation 65, p. 252). Importantly, the 'uncertainty' directive has been removed. Hence, the proposed guideline is arguably an improvement on before, as it moves towards a principle of matching counselling with specific issues (i.e., psychosocial, relationship and employment) together with a crucial note about the specificity of the counselling model to be adopted.

Historically, the NICE Guideline for Depression has been highly influential in shaping healthcare provision for those experiencing depression. As described by Clark (2011), the NICE recommendations for depression from 2004 onwards contributed to the development and roll-out of the Improving Access to Psychological Therapies (IAPT) programme, which in England now provides the bulk of treatment for depression in primary care (Gyani, Pumphrey, Parker, Shafran & Rose, 2012). One example of the impact of the revised 2009 Guideline appears to have been the cutting of counselling jobs in the NHS, with IAPT workforce census data suggesting a 35% decline in the number of qualified counsellors working as high-intensity therapists between 2012 and 2015, in a

period where the total IAPT workforce grew by almost 18% (IAPT Programme, 2013; NHS England & Health Education England, 2016). Workforce shifts that apparently follow revised NICE guidelines (e.g., counselling not being recommended as a first-line treatment for depression) underline the importance of scrutinising guideline recommendations since a core assumption is that using 'best' evidence and guideline methodologies will lead to NICE recommendations that improve patient care. An implicit question in the remainder of this article is whether the positioning of counselling as a second-tier treatment for mild-to-moderate depression (only) through NICE recommendations is likely to lead to improved outcomes for clients with depression.

Defining counselling as a psychological intervention

The NICE depression guidelines (2009, 2017d) have included recommendations for 'counselling', but the definition of 'counselling' is unclear. The British Association for Counselling and Psychotherapy (BACP) adopts a generic definition for both counselling and psychotherapy as umbrella terms for 'a range of talking therapies' (BACP, 2017). Equivalent professional organisations, such as the American Counseling Association (ACA) and the European Association for Counselling (EAC) define counselling in terms of a professional relationship that seeks to aid patients (ACA, 2017; EAC, 2017). What these definitions have in common is that they are nonspecific: counselling is a broad family of interventions that includes subtypes of counselling such as person-centred therapy (PCT) or cognitive behaviour therapy (CBT). However – and problematically – the 2009 NICE Guideline for Depression directly compared 'counselling' with subtypes of counselling.

The 2009 NICE Guideline for Depression did not specify a definition of counselling; however, various definitions for counselling are provided in the empirical literature. For example, King, Marston and Bower (2014) reported on a reanalysis of the Health Technology Assessment-funded trial (Ward et al., 2000), comprising a head-to-head RCT comparing 'nondirective counselling' and cognitive behaviour therapy (CBT), and defined the counselling used in their study as 'a nondirective, inter-personal approach' (p. 1836) derived from the work of Carl Rogers. In this context, the therapy 'counselling' has clear theoretical and empirical roots and is a synonym for a type of talking therapy.

In contrast, a 2012 meta-analytic study by Cuijpers et al. examined the efficacy of 'nondirective supportive therapy' (NDST) – which they stated is 'commonly described in the literature as counselling' (p. 281). They defined NDST as an approach that utilises the shared attributes (or common factors) of all talking therapies 'without (utilizing) specific psychological techniques...' (p. 281), which characterise particular types of therapy. Cuijpers et al. (2012) point out that many RCTs that include counselling do so as a nonspecific control group and suggest researchers appear to treat counselling as not being a bona fide active treatment. In this context 'counselling' is neither a category nor an example of a category, but a shared nonspecific attribute of psychological therapies in general.

The outcome of the 2009 NICE guidance recommendations spurred the development of a model of counselling for the treatment of depression designed to be effective as a high-intensity intervention within IAPT that took the form of a person-centred experiential therapy named Counselling for Depression (CfD; Sanders & Hill, 2014). The aim was to develop a bona fide psychological therapy using an established methodology that involved defining a range of basic, generic, specific and meta-competencies for this model of therapy (Roth, Hill & Pilling, 2009). The CfD (person-centred experiential) model, which is now available to IAPT patients (NHS England, 2017), also meets the recommendations in the 2018 draft guidelines for a model of counselling developed for depression.

The reviewed definitions suggest there are potentially two distinct forms of counselling: a nonspecific counselling that utilises generic and basic competences common to all forms of therapy, and a model-specific form of counselling, such as person-centred experiential counselling, which includes CfD. This distinction between generic counselling and a bona fide/active intervention potentially implies critical differences in the level of training and competencies of a practitioner (comparable to the differences between low and high-intensity treatment in IAPT) and in the specificity of the model of intervention used. The 2018 proposed guideline does not utilise such distinctions, however, the only recommendation in the draft guidelines is that the counselling intervention should be one developed specifically for depression (yet CfD is not named). This suggests that guideline developers need to make a concerted effort to use definitions that specify the theoretical approach and potentially the level of professional training or competencies.

The current evidence for the clinical efficacy and effectiveness of counselling in the treatment of depression

NICE guidelines for depression draw on two main classes of data to arrive at clinical recommendations, namely meta-analyses and RCTs. NICE's methodological procedures state: 'NICE prefers data from head-to-head RCTs to compare the effectiveness of interventions' (NICE, 2014/2017, p. 103). Further, the procedures require the detailing of the methods and results of individual trials. If direct evidence from treatment comparisons is not available, then indirect comparisons can be made using network meta-analysis (see Mills, Thorlund & Ioannidis, 2013). This procedure, which combines direct and indirect treatment comparisons, focuses on classes of interventions (i.e., broader headings of approaches rather than specific therapy brands) to arrive at recommendations when comparing multiple interventions. The interventions are judged against an appropriate comparator, that is, a common standard. The draft 2018 Guideline uses a pill placebo condition as the appropriate comparator. The Guideline also considers the cost-effectiveness of interventions. In this section, we provide an overview of the current status of evidence regarding counselling as derived from meta-analyses and RCTs.

Meta-analyses of counselling in the treatment of depression

In terms of meta-analyses, the aim is to combine data from multiple studies and to statistically synthesise the results to create conclusions that are more robust. There are three meta-analyses of direct relevance.

First, Cape, Whittington, Buszewicz, Wallace and Underwood (2010) carried out a meta-analysis and meta-regression of 34 studies focusing on brief psychological interventions for anxiety and depression, involving 3962 patients. Most interventions were brief cognitive behaviour therapy (CBT; $n = 13$), counselling ($n = 8$) or problem solving therapy (PST; $n = 12$). Results showed effectiveness for all three types of therapy: studies of CBT for depression ($d: -.33$, 95% CI: $-.60$ to $-.06$) and studies of CBT for mixed anxiety and depression ($d: -.26$, 95% CI: $-.44$ to $-.08$); counselling in the treatment of depression alone as well as mixed anxiety and depression ($d: -.32$, 95% CI: $-.52$ to $-.11$); and PST for depression and mixed anxiety and depression ($d: -.21$, 95% CI: $-.37$ to $-.05$). Controlling for diagnosis, meta-regression found no difference between CBT, counselling and PST. The

authors concluded that brief CBT, counselling and PST are all effective treatments in primary care, but that effect sizes are low compared to longer length treatments. Nonetheless, it should be pointed out that for the analysis of the four studies of counselling for the treatment of depression only, the results were not statistically significant. However, four studies are not sufficient to yield reliable results.

Second, Cuijpers et al. (2012) found that studies in which NDST was compared with CBT resulted in a small and nonsignificant difference between NDST and CBT. The authors commented that NDST has been treated as a proxy for counselling, although it specifically excludes active elements that may be present in bona fide counselling interventions. However, they found that the studies with researcher allegiance in favour of the alternative psychotherapy resulted in a considerably larger effect size than studies without researcher allegiance. Moreover, in studies without an indication of researcher allegiance, the difference between NDST and other therapies was virtually zero. The authors argued that such results suggested that NDST is effective and deserved more respect from the research community.

Third, the most recent relevant study by Barth et al. (2013) adopted a network meta-analysis – the same method used by the NICE Guideline Development Group – using 198 trials comparing seven forms of psychotherapeutic interventions, one of which was 'supportive counselling'. The analysis found significant effects for supportive counselling compared against waitlist and that the evidence base for supportive counselling was broad. However, when that analysis focused only on the network of large trials, for four of the interventions, including supportive counselling, significant effects were no longer found. Barth et al. (2013) themselves invoked the results of the Cuijpers et al. (2012) meta-analysis that found no difference between NDST and other treatments. They stated it was 'unjustified' to dismiss supportive counselling as a suboptimal treatment because, although the evidence for this intervention was less strong, the size of the differences between the interventions studied was small. They concluded that different psychotherapeutic interventions for depression have comparable, moderate-to-large effects.

In summary, when studies with a low researcher allegiance against counselling together with evidence from bona fide counselling interventions are considered, the meta-analytic studies comparing counselling with CBT for depression suggest either broad equivalence of patient outcomes or, where differences do exist, that they are small.

RCTs of counselling in the treatment of depression

As a tradition, counselling in the UK is often associated with Humanistic/Experiential therapies, and there are a few RCTs which report evidence for the efficacy for these therapies with depressed patients (Goldman, Greenberg & Angus, 2006), including one that compared process-experiential therapy (now referred to as emotion-focused therapy) with CBT and found comparable outcomes (Watson, Gordon, Stermac, Kalogerakos & Steckley, 2003). However, only one recent report directly compared counselling (defined as nondirective person-centred counselling) to CBT in the treatment of depression. The original study reported comparisons between nondirective counselling and CBT for mixed anxiety and depression and found no significant difference in outcomes for the two therapies (Ward et al., 2000). A subsequent reanalysis of the subsample of patients meeting a diagnosis of depression only, found similar results with both therapies being equally effective and both being superior to usual General Practice care at 4 months but not at 12 months (King et al., 2014).

The findings from this study are important because of the lack of RCT research that might provide direct head-to-head trial evidence for the efficacy of counselling. The 2009 NICE Guideline for Depression development process identified six relevant studies for consideration. One was excluded due to the mixed diagnosis (Ward et al., 2000) although, as stated, a subanalysis focusing on patients reporting depression only was considered (and subsequently published as King et al., 2014). Data from five other trials were also used (Bedi et al., 2000; Goldman et al., 2006; Greenberg & Watson, 1998; Simpson, Corney, Fitzgerald & Beecham, 2000; Watson et al., 2003). However, they were all either low powered in terms of patient numbers, had patient samples drawn from the mild-to-moderate range of depression only with some including subthreshold patients, or compared outcomes for similar (Humanistic/Experiential) therapies. The 2009 guideline recommendation was that counselling should not be considered as a first-line intervention, as it had more limited evidence, and should only be considered for patients experiencing subthreshold, mild or moderate depression who declined the other treatments available. As stated, the guideline also added the qualification about the uncertainty of the evidence for counselling, and suggested patients should be advised on this matter.

In summary, while there is minimal recent RCT evidence comparing counselling as a bona fide

intervention with CBT, the evidence that does exist supports the general efficacy of counselling. However, apart from the Ward/King reports, RCT studies are generally small-scale and lack a standard comparator such as CBT. The lack of new data may explain why the recommendations for counselling in the 2009 published and 2018 draft guidelines are broadly similar. However, unlike the 2009 Guideline, the draft 2018 Guideline is based on network meta-analyses. As some commentators have noted: 'Nonetheless, a network meta-analysis is not a substitute for a well conducted randomized controlled trial' (Kanters et al., 2016, p. 783). More immediately, perhaps, there needs to be a debate as to the appropriateness of using pill placebo as the appropriate comparator in relation to decision-making. To use a nonclinically viable intervention as the appropriate comparator – something a patient experiencing depression would never be offered – does not appear to be the most useful benchmark for informing decision-making regarding differing interventions (see Dias, 2013).

Yet, beyond meta-analyses and RCTs, other potentially valuable sources of evidence exist that are defined by NICE as within the scope of evidence that could be considered but, unfortunately, have not been in the 2018 draft recommendations. In the next section, we argue that there has been an overreliance on the RCT design, before then presenting a case for including relevant non-RCT data.

The limitations of currently considered evidence in guideline development

An overreliance on RCTs

Within the counselling and psychotherapy outcomes literature, there has been a long-standing debate regarding what counts as evidence (Kazdin, 2008). Evidence from RCTs has traditionally been favoured due to specific features that control for systematic biases, leading them to be judged as providing the most stringent form of evidence. In short, randomisation protects against any systematic biases in the assignment of patients to treatments. The component of randomisation is probably the hallmark most often cited as underpinning the superiority of trials data in the field of the psychological therapies. However, the other central element of RCTs – participants being double-blinded – can only be utilised in drug trials where the content of the drug can be hidden to patients and to the professional providing the medication. Hence, while trial designs

in the psychological therapies are not the strongest form that the RCT design allows, it has long been held as the design that yields the most reliable and valid findings (Wessley, 2007).

While the strengths of RCT designs are well accepted, no research method is immune from criticism and one of the abiding criticisms of RCTs concerns their lack of generalisability (Kennedy-Martin, Curtis, Faries, Robinson & Johnston, 2015). While statistical work is taking place to develop procedures in an attempt to address this issue (Stuart, Bradshaw & Leaf, 2015), by design, RCTs involve the careful screening of patients to ensure that all trial participants fully meet diagnostic criteria for the presenting condition under study. Typically, this involves screening out patients presenting with any comorbidities, something that leads to the criticism that RCT participants are atypical of patients in actual practice, since, for example, depression is highly comorbid with anxiety (Kaufman & Charney, 2000). In addition, by their very nature RCTs draw on a specific subgroup of the population of patients, namely those who are willing to be trial participants. A major reason patients decline to be participants in trials is that they do not wish to be research subjects (Barnes et al., 2013). In addition, there has been a long-term concern about the lack or underrepresentation of minorities in research studies (Hussain-Gambles, Atkin & Leese, 2004; Stronks, Wieringa & Hardon, 2013). Hence, while a well-conducted RCT will state that the intention to offer treatment *X* (from an intent-to-treat analysis) or receipt of treatment *X* (from a per-protocol analysis) is better than treatment *Y* in a specific setting, it will not address the question a commissioner asks, namely: will it work for us? (Cartwright & Munro, 2010).

Jadad and Enkin (2007), the authors of the standard guide to designing RCTs, state: '... randomized trials are not divine revelations, they are human constructs, and like all human constructs, are fallible. They are valuable, useful tools that should be used wisely and well' (p. 44). Indeed, Jadad and Enkin list over 50 specific biases that are possible when carrying out a trial and go on to provide a strong warning that unless their weaknesses are acknowledged, there is a 'risk of fundamentalism and intolerance of criticism, or alternative views (that) can discourage innovation' (p. 45).

Despite such criticisms, trials have become the dominant source for informing clinical guidelines. Yet, as the previous Chairman of NICE, Sir Mike Rawlins, stated: 'Awarding such prominence to the results of RCTs, however, is unreasonable' (2008, p.

2159). Rawlins further argued in relation to the hierarchy of evidence used by NICE that privileges trials data, that 'Hierarchies of evidence should be replaced by accepting a diversity of approaches.' (p. 2159). And indeed, the word *hierarchy* does not appear at all in the NICE methods manual (NICE, 2014/2017). Rawlins' argument was not to abandon RCTs in favour of observational studies; rather what he sought was for researchers to improve their methods and for decision makers to avoid adopting entrenched positions about the nature of evidence. However, given the dominance of RCT evidence and the absence of relevant and available observational data in the draft 2018 guidelines, it would appear that Rawlins' call has not been heeded.

Considering statistical power and nonindependence of patients in RCTs

A separate but major issue concerning trials, as identified earlier, is the extent to which they are appropriately powered to detect any hypothesised differences. To have confidence in the findings from RCTs that test the superiority, noninferiority or equivalence of one treatment condition against another, studies must have the required statistical power (sufficient numbers of patients in the trial) to detect such a difference if one exists. The standard criterion that defines sufficient power for a superiority trial requires that a study will have at least an 80% chance of detecting a difference at $p < .05$ if one exists.

Cuijpers (2016) reviewed the statistical power needed both for individual RCTs and for meta-analytic studies focused on adult depression. His analysis should be considered alongside the three classes of between-group effect sizes traditionally postulated by Cohen (1992): small ($d = .2$), medium ($d = .5$), and large ($d = .8$). He identified that a sample size of 90 trial patients (i.e., 45 patients per arm) was required to find a differential effect size of $d = .6$ (i.e., a medium effect size). Having established in an earlier article that an effect size of $d = .24$ could be considered as a 'minimally important difference' from the patient's perspective (Cuijpers, Turner, Koole, van Dijke & Smit, 2014), he calculated that for a trial to determine such a minimally important difference between two active treatments for depression would require 548 patients – that is, 274 patients in each arm of the trial.

Yet in Cuijpers' (2016) analysis, the mean number of patients included in RCT comparisons between CBT and another psychotherapy for depression was 52,

with a range from 13 to 178. The effect size that can be detected with the average trial comprising 52 patients was $d = .79$, an effect size similar to that comparing CBT with untreated control groups (i.e., $d = .71$). For nondirective counselling, the analysis found that the largest study had sufficient power to detect a differential effect size of $d = .34$. The largest comparative trial found in three comprehensive meta-analyses of major types of psychotherapy comprised 221 patients. This is about 40% of the 548 patients needed to detect a clinically relevant effect size of $d = .24$. Taking these statistics together, it is uncertain whether there can be sufficient confidence in the results of RCTs for adult depression conducted to date that compare CBT with another therapy because they likely lack sufficient statistical power (Cuijpers, 2016).

Meta-analyses are, like single RCTs, subject to considerations of power. For meta-analyses of RCTs focused on treatment of depression, Cuijpers (2016) suggests that for CBT (based on a mean of 52 patients per study), 18 trials would be needed to detect a significant effect of $d = .24$ with a power of .8, or 24 trials with a power of .9. According to his analysis, the actual number of trials was 46, which was sufficient to detect a clinically relevant effect. However, he concluded that only 13 of these trials had a low risk of bias. This is important, as 'bias' is an agreed index of factors that reduce confidence in the results of RCTs. For example, a potential source of bias is the degree to which assessors or data analysts have prior knowledge of the specific intervention any individual study participant received. Hence, meta-analyses are also vulnerable to low power once only studies with a low risk of bias are considered.

For nondirective supportive counselling (based on a slightly higher mean of 59 patients per trial), 16 trials would be needed to detect an effect of $d = .24$ with a power of .8 or 21 trials with a power of .9. The 32 trials comparing counselling with other therapies therefore had sufficient power to detect a clinically relevant effect. However, only 14 trials had low risk of bias, yielding the same conclusion that there were not enough trials to detect such an effect.

In addition to issues of bias and low power, the statistical analysis applied to the data assumes that the data – that is, patients – are independent of each other. However, patients are *not* independent of each other as they are nested within therapists. Patient outcomes for one therapist will be correlated with the other patients from the same therapist and differ from the outcomes with other therapists. It is likely that there will be variability between the outcomes of therapists, a

phenomenon known as therapist effects (Barkham, Lutz, Lambert & Saxon, 2017). Failure to take account of therapist effects results in this effect being attributed to the treatment effect and, thereby, inflating it (or deflating it if the therapists are not effective).

In summary, despite numerous comparative trials being conducted, from this data it is unclear whether one therapy for adult depression is more effective than another *to an extent that is clinically relevant*. Trials are underpowered and require much greater statistical power and less bias to determine differential effectiveness. In the light of this position, we now consider arguments for including very large data sets from routine practice.

Incorporating very large routine practice-based data sets in guideline development for depression

As stated earlier, the NICE methods manual states that while RCTs may often be the most appropriate design, 'other study designs (including observational, experimental or qualitative) may also be used to assess effectiveness, or aspects of effectiveness' (NICE, 2014/2017, p. 15). And in terms of the development work in network meta-analysis, the aim is to move towards 'the inclusion of studies of various designs, including observational studies, within one analysis' (Kanters et al., 2016, p. 783). Accordingly, there appears to be little reason, if any, for NICE not to consider high-quality and relevant observational data.

One key development over the past decade or more has been the growth in the availability of very large data sets. For the psychological therapies, this is best exemplified by the implementation of the IAPT programme in England (London School of Economics and Political Science, 2006). The IAPT programme comprises a stepped care approach in which patients are initially referred for low-intensity interventions such as psychoeducational interventions delivered by psychological wellbeing practitioners (PWPs). If not successful, these are 'stepped up' to high-intensity interventions comprising CBT and several non-CBT therapies, including CfD (person-centred experiential therapy), a standardised model of intervention focused on depression with standards of training and supervision. Some patients, based on their presenting issues, are assigned directly to high-intensity interventions. The IAPT programme, which was piloted in 2006 and independently evaluated (Parry et al., 2011), has been rolled out nationally and has focused largely on patients experiencing depression and anxiety but is being expanded to other patient groups.

A key feature of the IAPT programme is the administration of a common set of outcome measures – a minimum data set (MDS) – at each attended session. The MDS comprises the following: the Patient Health Questionnaire-9 (PHQ-9: Kroenke, Spitzer & Williams, 2001), which acts as a proxy measure for depression; the General Anxiety Disorder-7 (GAD-7; Spitzer, Kroenke, Williams & Löwe, 2006); and the Work and Social Adjustment Scale (WSAS; Mundt, Marks, Shear & Greist, 2002). The per-session administration of the PHQ-9, GAD-7 and WSAS in IAPT has yielded potential standardised data sets from routine practice of unprecedented size. In 2015–2016 (the last year for which there is currently data), almost a million people entered IAPT treatment, with over half a million completing a course of treatment (NHS Digital, 2016).

The numbers of IAPT patients for whom systematic data has been collected potentially makes this one of the largest standardised data sets on the psychological therapies in the world. Kazdin (2008), on observing the general waste from data in practice settings not being used stated: ‘we are letting knowledge from practice drip through the holes of a colander’ (p. 155). Indeed, the collection and use of such large-scale routinely collected standardised data are a hallmark of the research paradigm termed practice-based evidence (Barkham & Margison, 2007; Barkham, Stiles, Lambert & Mellor-Clark, 2010). While the privileging of trials data ahead of observational data may have been appropriate when the latter comprised small-scale and unsystematic studies, this is no longer the case. In the same way that narrative reviews have developed a clear and systematic methodological underpinning to yield systematic reviews, the methods of collection and analyses of ‘routine data’ have developed a level of sophistication that can arguably no longer be dismissed (or labelled) as simply observational data.

Consistent with this practice-based paradigm, the proposed 2018 Guideline states: ‘For all interventions for people with depression: use sessional outcome measures; review how well the treatment is working with the person; and monitor and evaluate treatment adherence’ (NICE, 2017d; Recommendation 37, p. 248). In addition, healthcare professionals delivering interventions for people with depression should: ‘receive regular high-quality supervision; and have their competence monitored and evaluated, for example, by using video and audio tapes, and external audit’ (NICE, 2017d; Recommendation 38, p. 248). These recommendations provide the underpinning not only for enhancing the quality of

clinical practice but also of ensuring the collection of high-quality standardized data that would complement trials-based data. However, despite the potential size of the IAPT data set and its quality, the data are not currently considered in NICE guideline developments. Given that the IAPT initiative was shaped by iterations of the NICE Guidelines for depression, the IAPT data itself may contribute to a better linkage between practice in routine settings, the yield from RCTs, and guideline development. It also enables practitioners in routine practice to contribute directly via their standardised data to informing the very guidelines that they will have to implement.

The IAPT data set: effectiveness of counselling in the treatment of depression in the NHS

The potential value of the IAPT data set in contributing to the evidence base on effective treatment for depression in adults is illustrated by examining reports and studies derived from IAPT data. Since 2013–14, IAPT have published annual reports comparing the number of referrals, average number of sessions and recovery rates between the available psychological therapies (NHS Digital, 2014, 2015, 2016). As demonstrated in Table I, whilst a greater proportion of referrals (approximately 60–65%) received CBT as compared with counselling, patient outcomes (i.e., recovery rates) have been virtually equivalent between the two interventions.

Research studies carried out by different academic groups that have accessed different portions of the IAPT data set to undertake more detailed analyses have also reported comparable outcomes between CBT and counselling in relation to the treatment of depression (Gyani, Shafran, Layard & Clark, 2013; Pybis, Saxon, Hill & Barkham, 2017). In more sophisticated studies using multilevel modelling to

Table I: Data extracted from successive NHS digital reports on comparisons between cognitive behaviour therapy (CBT) and counselling/counselling for depression (CfD).

Year	Intervention	Number of referrals for depressive disorder	Average number of sessions	Recovery rate (%)
2013–14	CBT	21,622	5.7	45.1
	Counselling	13,369	5.4	
2014–15	CBT	28,350	5.1	44.1
	Counselling	14,994	4.4	45.2
2015–16	CBT	35,589	5.8	45.9
	Counselling (CfD)	20,011	5.3	47.6

account for patient case mix and the nested nature of data, where differences have been observed these have been small and clinically insignificant (Pybis et al., 2017; Saxon, Firth & Barkham, 2017). These data demonstrate that for patients accessing psychological therapy throughout the NHS, counselling is, to all intents and purposes, as effective as CBT in the treatment of depression for both moderate and severe levels of depression. These studies, as well as the publicly available evidence from NHS Digital, confirm the findings of earlier studies using the Clinical Outcomes in Routine Evaluation measure (CORE-OM; Evans et al., 2002). These studies used routinely collected CORE-OM data from naturalistic settings before the implementation of IAPT and yielded comparable patient outcomes between counselling and CBT (Stiles, Barkham, Mellor-Clark & Connell, 2008; Stiles, Barkham, Twigg, Mellor-Clark & Cooper, 2006).

In summary, the evidence from the IAPT data set is that counselling is as effective as CBT as an intervention for depression. This evidence of effectiveness in NHS practice settings across England accords with the conclusions of Cuijpers (2017), who reviewed over 500 depression RCTs from four decades of research, and concluded that there were no significant differences between the main interventions, once biases and allegiances were considered. The consistency of the trials-based and practice-based findings is important in supporting the value of counselling as an intervention for depression offered in the NHS in England. However, we argue that the key conclusion for guideline development from these findings is that focus of research attention should not be on repeatedly re-evaluating the evidence for different interventions. Instead the focus should move to other factors such as therapist effects or site effects where there appear to be noticeable differences in patients' outcomes (e.g., Saxon & Barkham, 2012). This refocusing away from treatment differences and towards other factors is a position endorsed by the American Psychological Association (2012).

The IAPT data set: efficiency and cost-effectiveness of counselling in the treatment of depression

A 2010 report calculated the annual cost of depression in England to be almost £11 billion in lost earnings, demands on the health service and the cost of prescribing drugs to address the depression (Cost of Depression in England, 2010). In this context, the cost-effectiveness of treatment is important to

consider. Determining cost-effectiveness with acceptable degrees of certainty requires large samples, which the IAPT data set offers in a way that trials do not. Given the NICE procedural manual states that, for example, observational data can be used for 'aspects of effectiveness', the potential contribution of the IAPT data set to considerations of cost-effectiveness is significant.

Improving Access to Psychological Therapies data suggest patients accessing counselling attend fewer sessions on average than those accessing CBT (NHS Digital, 2014, 2015, 2016; Pybis et al., 2017; Saxon, Firth et al., 2017). This suggests counselling may well be cheaper and therefore more cost-efficient than CBT as it achieves comparable patient outcomes. To consider this in more detail, a study exploring the cost-effectiveness of IAPT as a service reported data collected from five Primary Care Trusts and found the cost of a high-intensity session was £177 (Radhakrishnan et al., 2013). Using this estimate alongside figures from the latest IAPT report that counselling is typically seeing patients for 5.9 sessions, whereas CBT is seeing patients for 7.1 sessions (NHS Digital, 2016), this would suggest counselling costs approximately £1044 per patient and CBT approximately £1256 per patient. In 2015–16, 152,452 patients completed a course of CBT at an estimated cost of £191 million. If those same patients had received counselling the cost saving could have been over £30 million.

The potential saving of £30 million is calculated only from the fewer sessions (on average) received by counselling patients in IAPT. However, given that counsellors in IAPT are often paid a grade lower than 'IAPT-qualified' therapists (Perren, 2009), this figure may underestimate the potential saving. Moreover, while counselling training is typically self-funded, IAPT CBT trainings have been government funded, initially centrally and more recently locally. This illustrates the potential financial implications of how research evidence is weighed up and then synthesised into guideline recommendations for the treatment of depression.

In summary, the vast data set derived from the IAPT programme needs to be used to complement data from RCTs. And this is particularly true for questions concerning cost-effectiveness that cannot be adequately addressed by RCTs alone. Within years, there will be patient data on millions of patients within IAPT services. Its inclusion in the scope of NICE guideline reviews would be wholly consistent with the NICE guidelines procedure manual.

Considering the role of service users' voices via qualitative research in guideline development

The previous section has argued for guideline developers to consider very large patient data sets. In this section, we argue for guideline developers to incorporate qualitative evidence that gives voice to service users. Doing so would be in accordance with NHS England's business plan for 2016/2017, which sets out a commitment: 'to make a genuine shift to place patients at the centre, shaping services around their preferences and involving them at all stages' (NHS England, 2016, p. 49). NICE has a similar commitment (NICE Patient and Public Involvement Policy, 2017c). Currently, while qualitative research is included in guideline development, NICE processes do not allow such data to be included in the final summative analyses that shape key recommendations. Yet a number of researchers (Hill, Chui & Baumann, 2013; Midgley, Ansaldo & Target, 2014) argue that qualitative outcome studies are important to consider because they 'offer a significant challenge to assumptions about outcome that derive from mainstream quantitative research on this topic, in relation to two questions: how the outcome is conceptualised, and the overall effectiveness of therapy' (McLeod, 2013, p. 65). Reviewing existing literature, McLeod suggested patients themselves conceptualise outcome much more broadly than in terms of symptom or behavioural change (Binder, Holgersen & Nielsen, 2010). Typically, patients acknowledge ways in which therapy has been helpful but also where it has failed, suggesting that quantitative outcome research may overstate therapeutic effectiveness. Qualitative studies can also help answer questions about patient experience and expectations of NHS services, including whether treatments are credible and acceptable to them, which have an impact on outcomes.

Turning to qualitative research focused on depression, there is a growing literature on understanding the experiences of patient populations such as minority ethnic groups (e.g., Lawrence et al., 2006a), women (e.g., Stoppard & McMullen, 1999), men (e.g., Emslie, Ridge, Ziebland & Hunt, 2006) and older adults (e.g., Lawrence et al., 2006b). Such studies elucidate population-specific depression experiences that can be useful in understanding why certain populations benefit less from treatment. There is also a literature that seeks to describe the experience of aspects of depression such as recovery (e.g., Ridge & Ziebland, 2006) or types of depression such as postnatal depression (e.g., Beck, 2002). However,

currently relatively little research focuses on patients' experiences of depression treatment. There is some research on depressed patients' experiences of computer-mediated depression treatment (e.g., Beattie, Shaw, Kaur & Kessler, 2009; Lillevoll et al., 2013), and mindfulness (e.g., Mason & Hargreaves, 2001; Smith, Graham & Senthinathan, 2007). However, there is less research on the major modalities such as CBT (e.g., Barnes et al., 2013), psychodynamic (e.g., Valkonen, Hänninen & Lindfors, 2011) and process-experiential therapies (e.g., Timulak & Elliott, 2003). The lack matters because such qualitative research focusing on treatment experiences provides a method by which theoretical assumptions about how a therapy 'works' can be evaluated against the patient perspective.

Even more rare are comparative qualitative outcome studies (e.g., Nilsson, Svensson, Sandell & Clinton, 2007). Such studies focusing on depression are valuable because they can foster understanding of whether patients experience outcomes differently in different therapies. One example is Straarup and Poulsen's (2015) study, which compared patients' experiences of CBT and metacognitive therapy and found evidence of different understandings of the causes of depression and what had changed as a result of therapy.

In summary, qualitative research has considerable value in terms of capturing patients' experiences of psychotherapy that can inform practice (see Levitt, Pomerville & Surace, 2016). This suggests the need: (1) to consider qualitative outcome studies in guideline development and recommendations, and (2) encouraging further research focused on guideline-recommended treatments and differential patient experiences.

Towards a broader spectrum of best evidence

Whatever the potential pool of data, guideline organisations need to establish and implement procedures for making recommendations. A recent review considered how different national organisations produce clinical guidelines. Moriana et al. (2017) analysed and compiled lists of evidence-based psychological treatments by disorder using data provided by RCTs, meta-analyses, guidelines and systematic reviews of NICE, Cochrane, Division 12 of the American Psychological Association and the Australian Psychological Society. For depression, they found poor agreement with no single intervention obtaining positive consensus agreement from all four organisations. The authors suggested one possible

cause for the lack of agreement might be subtle biases in committee procedures, while evidence considered by both NICE and Cochrane may be overinfluenced by the key meta-analyses that both organisations commission to support their decision-making. Whilst one organisation might favour its own procedures in this way, the process lacks standardisation across the different bodies and leads to discrepancies in guidance.

The finding that guideline processes have led to different treatment recommendations for the same condition underlines the criticisms of an approach to synthesising evidence that rigidly prioritises RCTs. We argue that a rigorous and relevant knowledge base of the psychological therapies cannot be built on one research paradigm or type of data alone but should incorporate both *evidence-based practice* (i.e., trials) and *practice-based evidence* (i.e., routine practice data; Barkham & Margison, 2007). In this conceptualisation, trials provide evidence from a top-down model (RCT evidence generating national guidelines that are implemented in practice settings) while practice-based evidence builds upwards using data from routine practice settings to guide interventions and inform guideline development. Both paradigms are complementary and, most importantly, the results from one paradigm can be tested out in the other. Further, a synthesis of evidence from both paradigms ensures that the data from trials remain directly connected and relevant to routine practice, creating a continual cycle between practice and research and between practitioners and researchers.

Given the points made here, there is little justification for relying *solely* on trials data and dismissing evidence from large standardised routine datasets delivering NICE recommended and IAPT approved psychological therapies. There are issues and vulnerabilities with both paradigms and the evidence they provide, but it is no longer credible to suggest that the term *best* applies only to trials data. To abide by the advice of Rawlins (2008) as well as Jadad and Enkin (2007), views concerning nontrial data need to become more accommodating. Overall, a collective move to a position of considering the *weight of evidence* from a wider bandwidth or spectrum provides a more rounded and inclusive view of available high-quality data. By applying the concept of teleoanalysis – that is, the synthesis of different categories of evidence to obtain a quantitative summary – it is possible to arrive at more robust and relevant conclusions (Clarke & Barkham, 2009; Wald & Morris, 2003). This, we would suggest, is an

approach that would yield both better and more relevant evidence. Accordingly, IAPT data now needs to be considered alongside evidence from trials to form a more complete and accurate picture of the comparative effectiveness of psychological therapies. Further, high-quality qualitative data require inclusion in arriving at recommendations, particularly as it is a primary source for patients' perspectives and experiences.

Conclusions and recommendations

We have argued for greater precision in defining the profession and practice of counselling, provided an overview of research on counselling for the treatment of depression from meta-analyses and RCTs, raised issues arising from a sole reliance on trials, and put the case for broadening the bandwidth of high-quality evidence using large routine standardised data sets and the consideration of high-quality qualitative studies. Overall, with regard to depression, counselling is effective. Some analyses suggest it is somewhat less effective than other therapies for depression (e.g., CBT), but when research findings are adjusted for researcher allegiance and low risk of bias, such differences are minimal and not clinically relevant (Cuijpers, 2017). Results from (very) large standardised data sets in routine practice show counselling to be as effective as CBT in the treatment of patient-reported depression and with a suggestion that it may be more cost-efficient. However, such data are not considered by NICE even though it is consistent with the scope of data defined in their guideline development procedural manual (NICE, 2014/2017).

One clear observation concerning RCTs in the field of depression is the paucity of high-quality head-to-head trials relating to counselling. In addition, there are calls from advocates of RCTs for trials to be larger and pragmatic (Wessley, 2007). In response to such calls, there is a large pragmatic noninferiority RCT comparing CfD (Person-centred experiential therapy) with CBT as the benchmark treatment that will yield initial results late in 2018 (Saxon, Ashley et al., 2017). Particularly significant is the trial's focus on patients diagnosed as experiencing moderate or severe depression. The results regarding any differential effectiveness of counselling between moderate and severe depression will address a key issue as to whether CfD could be considered as a front-line intervention. Funders should call for other therapeutic approaches to be evaluated using CBT as a benchmark – to determine whether another therapy is, in any

clinically meaningful way, noninferior to CBT. In this way, a robust and relevant knowledge base will be constructed that aims to ensure quality and standards of psychological interventions for the treatment of depression while providing choice to patients. This is important given the mounting empirical evidence that improving patient treatment choice improves therapy outcomes (Lindhiem, Bennett, Trentacosta & McLear, 2014; Williams et al., 2016).

Finally, in this article, we have sought to make an argument about re-evaluating the definition of best evidence for guideline development. Using the evidence base for counselling in the treatment of depression as an example, we have argued that guideline developers should move towards integrating differing forms of high-quality evidence rather than relying on trials alone. But this requires change for all stakeholders: for individual researchers in counselling to be strategic and ensure their work builds cumulatively on the work of others; for researchers in organisations to yield larger and more substantive studies; for service providers to collaborate in collating common data through, for example, building practice research networks; for counselling bodies to devise, fund and implement research strategies that will deliver a robust evidence base for practice; and for guideline developers to accept a diversity of substantive research approaches that, combined, will yield best evidence. In doing so, not only will it be possible to draw more robust conclusions about the cost-effectiveness of depression treatment in the NHS and the clinical efficacy and effectiveness of different interventions, but also potentially the community, service, therapist, and patient variables that significantly impact on patient outcomes.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful comments on an earlier draft.

References

- American Counseling Association (2017). *What is counseling?* Retrieved from <https://www.counseling.org/aca-community/learn-about-counseling/what-is-counseling/overview>
- American Psychiatric Association (2017). *APA practice guidelines*. Retrieved from <http://psychiatryonline.org/guidelines>
- American Psychological Association (2012). *Recognition of psychotherapy effectiveness*. Retrieved from <http://www.apa.org/about/policy/resolution-psychotherapy.aspx>
- Barber, J. P., Connoll, M. B., Crits-Christoph, P., Gladis, L., & Siqueland, L. (2000). Alliance predicts patients' outcome beyond in-treatment change in symptoms. *Journal of Consulting and Clinical Psychology*, **68**, 1027–1032.
- Barkham, M., Lutz, W., Lambert, M. J., & Saxon, D. (2017). Therapist effects, effective therapists, and the law of variability. In L. G. Castonguay, & C. E. Hill (Eds.), *How and why are some therapists better than others?: Understanding therapist effects* (pp. 13–36). Washington, DC: American Psychological Association.
- Barkham, M., & Margison, F. (2007). Practice-based evidence as a complement to evidence-based practice: from dichotomy to chiasmus. In C. Freeman, & M. Power (Eds.), *Handbook of evidence-based psychotherapies: A guide for research and practice* (pp. 443–476). Chichester, United Kingdom: Wiley.
- Barkham, M., Stiles, W. B., Lambert, M. J., & Mellor-Clark, J. (2010). Building a rigorous and relevant knowledge-base for the psychological therapies. In M. Barkham, G. E. Hardy, & J. Mellor-Clark (Eds.), *Developing and delivering practice-based evidence: A guide for the psychological therapies* (pp. 21–61). Chichester, United Kingdom: Wiley.
- Barnes, M., Sherlock, S., Thomas, L., Kessler, D., Kuyken, W., Owen-Smith, A., ... Turner, K. (2013). No pain, no gain: depressed clients' experiences of cognitive behavioural therapy. *British Journal of Clinical Psychology*, **52**, 347–364.
- Barth, J., Munder, T., Gerger, H., Nüesch, E., Trelle, S., Znoj, H., ... Cuijpers, P. (2013). Comparative efficacy of seven psychotherapeutic interventions for patients with depression: a network meta-analysis. *PLoS Medicine*, **10**, e1001454.
- Beattie, A., Shaw, A., Kaur, S., & Kessler, D. (2009). Primary-care patients' expectations and experiences of online cognitive behavioural therapy for depression: a qualitative study. *Health Expectations*, **12**, 45–59.
- Beck, C. T. (2002). Postpartum depression: a metasynthesis. *Qualitative Health Research*, **12**, 453–472.
- Bedi, N., Chilvers, C., Churchill, R., Dewey, M., Duggan, C., Fielding, K., ... Williams, I. (2000). Assessing effectiveness of treatment of depression in primary care. *British Journal of Psychiatry*, **177**, 312–328.
- Binder, P. E., Holgersen, H., & Nielsen, G. H. S. (2010). What is a 'good outcome' in psychotherapy? A qualitative exploration of former patients' point of view. *Psychotherapy Research*, **20**, 285–294.
- British Association for Counselling and Psychotherapy (2009). *BACP issues warning that new depression guidelines may harm patients*, October 9 2009. Retrieved from <http://www.bacp.co.uk/media/index.php?newsId=1610>
- British Association for Counselling and Psychotherapy (2017). *What is counselling?* Retrieved from <http://www.bacp.co.uk/crs/Training/whatiscounselling.php>

- Cape, J., Whittington, C., Buszewicz, M., Wallace, P., & Underwood, L. (2010). Brief psychological therapies for anxiety and depression in primary care: meta-analysis and meta-regression. *BMC Medicine*, **8**, 38.
- Cartwright, N., & Munro, E. (2010). The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*, **16**, 260–266.
- Clark, D. M. (2011). Implementing NICE guidelines for the psychological treatment of depression and anxiety disorders: the IAPT experience. *International Review of Psychiatry*, **23**, 318–327.
- Clarke, J., & Barkham, M. (2009). Tribute to Phil Richardson - Evidence *de rigueur*: the shape of evidence in psychological therapies and the modern practitioner as teleoanalyst. *Clinical Psychology Form*, **202**, 7–11.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, **112**, 155–159.
- Cost of Depression in England (2010). *All party parliamentary Group on wellbeing economics*. Retrieved from <https://wellbeingeconomics.wordpress.com/reports/> Retrieved July 19 2017.
- Cuijpers, P. (2016). Are all psychotherapies equally effective in the treatment of adult depression? The lack of statistical power of comparative outcome studies. *Evidence Based Mental Health*, **19**, 39–42.
- Cuijpers, P. (2017). Four decades of outcome research on psychotherapies for adult depression: an overview of a series of meta-analyses. *Canadian Psychology/Psychologie Canadienne*, **58**, 7–19.
- Cuijpers, P., Driessen, E., Hollon, S. D., van Oppen, P., Barth, J., & Andersson, G. (2012). The efficacy of non-directive supportive therapy for adult depression: a meta-analysis. *Clinical Psychology Review*, **32**, 280–291.
- Cuijpers, P., Turner, E. H., Koole, S. L., van Dijke, A., & Smit F. (2014). What is the threshold for a clinically relevant effect? The case of major depressive disorders. *Depression and Anxiety*, **31**, 374–378.
- Dias S. (2013). *Using network meta-analysis (NMA) for decision making*. Paper presented at the 38th International Society for Clinical Biostatistics, Munich, Germany.
- Emslie, C., Ridge, D., Ziebland, S., & Hunt, K. (2006). Men's accounts of depression: reconstructing or resisting hegemonic masculinity? *Social Science & Medicine*, **62**, 2246–2257.
- European Association for Counselling (2017). *Definition of counselling*. Retrieved from <http://eac.eu.com/standards-ethics/definition-counselling/>. Cited 19 June 2017.
- Evans, C., Connell, J., Barkham, M., Margison, F., Mellor-Clark, J., McGrath, G., & Audin, K. (2002). Towards a standardised brief outcome measure: psychometric properties and utility of the CORE-OM. *British Journal of Psychiatry*, **180**, 51–60.
- Goldman, R. N., Greenberg, L. S., & Angus, L. (2006). The effects of adding emotion-focussed interventions to the client-centred relationship conditions in the treatment of depression. *Psychotherapy Research*, **16**, 537–549.
- Greenberg, L. S., & Watson, J. C. (1998). Experiential therapy of depression: differential effects of client-centred relationship conditions and process experiential interventions. *Psychotherapy Research*, **8**, 210–224.
- Gyani, A., Pumphrey, N., Parker, H., Shafran, R., & Rose, S. (2012). Investigating the use of NICE guidelines and IAPT services in the treatment of depression. *Mental Health in Family Medicine*, **9**, 149–160.
- Gyani, A., Shafran, R., Layard, R., & Clark, D. M. (2013). Enhancing recovery rates: lessons from year one of IAPT. *Behaviour Research and Therapy*, **51**, 597–606.
- Hernandez-Villafuerte, K., Garau, M., & Devlin, N. (2014). Do NICE decisions affect decisions in other countries? *Value Health*, **17**, A418.
- Hill, C. E., Chui, H., & Baumann, E. (2013). Revisiting and reenvisioning the outcome problem in psychotherapy: an argument to include individualized and qualitative measurement. *Psychotherapy*, **50**, 68–76.
- Hussain-Gambles, M., Atkin, K., & Leese, B. (2004). Why ethnic minority groups are underrepresented in clinical trials: a review of the literature. *Health and Social Care in the Community*, **12**, 382–389.
- IAPT Programme (2013). *Census of the IAPT Workforce as at August 2012*. Retrieved from <https://www.uea.ac.uk/documents/246046/11919343/iapt-workforce-education-and-training-2012-census-report.pdf/907e15d0-b36a-432c-8058-b2452d3628de>. Cited 19 June 2017.
- Jadad, A. R., & Enkin, M. W. (2007). *Randomized control trials: Questions, answers and musings*, 2nd edn. Oxford, United Kingdom: Blackwell.
- Kanters, S., Ford, N., Druyts, E., Thorlund, K., Mills, E. J., & Bansback, N. (2016). Use of network meta-analysis in clinical guidelines. *Bulletin of the World Health Organization*, **94**, 782–784.
- Kaufman, J., & Charney, D. (2000). Comorbidity of mood and anxiety disorders. *Depression and Anxiety*, **12**(Suppl. 1), 69–74.
- Kazdin, A. E. (2008). Evidence-based treatment and practice. New opportunities to bridge clinical research and practice, enhance the knowledge base and improve patient care. *American Psychologist*, **63**, 146–159.
- Kennedy-Martin, T., Curtis, S., Faries, D., Robinson, S., & Johnston, J. (2015). A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. *Trials*, **16**, 495.
- King, M., Marston, L., & Bower, P. (2014). Comparison of non-directive counselling and cognitive behaviour therapy for patients presenting in general practice with an ICD-10 depressive episode: a randomized control trial. *Psychological Medicine*, **44**, 1835–1844.

- Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (2001). The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, **16**, 606–613.
- Lawrence, V., Banerjee, S., Bhugra, D., Sangha, K., Turner, S., & Murray, J. (2006a). Coping with depression in later life: a qualitative study of help-seeking in three ethnic groups. *Psychological Medicine*, **36**, 1375–1383.
- Lawrence, V., Murray, J., Banerjee, S., Turner, S., Sangha, K., Byng, R., . . . Macdonald, A. (2006b). Concepts and causation of depression: a cross-cultural study of the beliefs of older adults. *The Gerontologist*, **46**, 23–32.
- Levitt, H. M., Pomerville, A., & Surace, F. I. (2016). A qualitative meta-analysis examining clients' experiences of psychotherapy: a new agenda. *Psychological Bulletin*, **142**, 801–830.
- Lillevoll, K. R., Wilhelmsen, M., Kolstrup, N., Høifødt, R. S., Waterloo, K., Eisemann, M., & Risør, M. B. (2013). Patients' experiences of helpfulness in guided internet-based treatment for depression: qualitative study of integrated therapeutic dimensions. *Journal of Medical Internet Research*, **15**(6), e126.
- Lindhiem, O., Bennett, C. B., Trentacosta, C. J., & McLear, C. (2014). Client preferences affect treatment satisfaction, completion, and clinical outcome: a meta-analysis. *Clinical Psychology Review*, **34**, 506–517.
- London School of Economics and Political Science. Centre for Economic Performance. Mental Health Policy Group (2006). *The depression report: A new deal for depression and anxiety disorders* [online]. London, UK: LSE Research Online.
- Mason, O., & Hargreaves, I. (2001). A qualitative study of mindfulness-based cognitive therapy for depression. *Psychology and Psychotherapy: Theory, Research and Practice*, **74**, 197–212.
- McLeod, J. (2013). Qualitative research: methods and contributions. In M. J. Lambert (Ed.), *Bergin and Garfield's handbook of psychotherapy and behavior change*, 6th edn (pp. 49–84). Hoboken, NJ: Wiley.
- Midgley, N., Ansaldo, F., & Target, M. (2014). The meaningful assessment of therapy outcomes: incorporating a qualitative study into a randomized controlled trial evaluating the treatment of adolescent depression. *Psychotherapy*, **51**, 128–137.
- Mills, E. J., Thorlund, K., & Ioannidis, J. P. A. (2013). Demystifying trial networks and network meta-analysis. *BMJ*, **346**(f2914), 1–6.
- Moriana, J. A., Gálvez-Lara, M., & Corpas, J. (2017). Psychological treatments for mental disorders in adults: a review of the evidence of leading international organizations. *Clinical Psychology Review*, **54**, 29–43.
- Mundt, J. C., Marks, I. M., Shear, M. K., & Greist, J. M. (2002). The work and social adjustment scale: a simple measure of impairment in functioning. *British Journal of Psychiatry*, **180**, 461–464.
- National Institute for Health and Care Excellence (2009). *NICE guidelines for depression in adults: recognition and management*. Retrieved from <https://www.nice.org.uk/guidance/cg90>
- National Institute for Health and Care Excellence (2014/updated 2017). *Developing NICE guidelines: the manual: Process and methods*. Retrieved from <https://www.nice.org.uk/process/pmg20>
- National Institute for Health and Care Excellence (2017a). *NICE guidance*. Retrieved from <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance>
- National Institute for Health and Care Excellence (2017b). *NICE guidelines*. Retrieved from <https://www.nice.org.uk/About/What-we-do/Our-Programmes/NICE-guidance/NICE-guidelines>
- National Institute for Health and Care Excellence (2017c). *Patient and public involvement policy*. Retrieved from <https://www.nice.org.uk/about/nice-communities/public-involvement/patient-and-public-involvement-policy>
- National Institute for Health and Care Excellence (2017d). *Depression in adults: recognition and management: Draft guidance consultation*. Retrieved from <https://www.nice.org.uk/guidance/GID-CGWAVE0725/documents/draft-guideline>
- NHS Digital (2014). *Psychological therapies: Annual report on the use of IAPT services*. England, 2013-2014. Retrieved from <http://content.digital.nhs.uk/catalogue/PUB14899>
- NHS Digital (2015). *Psychological therapies: Annual report on the use of IAPT services*. England, 2014-2015. Retrieved from <http://content.digital.nhs.uk/catalogue/PUB19098>
- NHS Digital (2016). *Psychological therapies: Annual report on the use of IAPT services*. England, 2015-2016. Retrieved from <http://content.digital.nhs.uk/pubs/psycther1516>
- NHS England (2016). *Our 2016/2017 business plan*. Retrieved from <https://www.england.nhs.uk/wp-content/uploads/2016/03/bus-plan-16.pdf>
- NHS England (2017). *IAPT workforce*. Retrieved from <https://www.england.nhs.uk/mental-health/adults/iapt/workforce/>
- NHS England and Health Education England (2016). *2015 Adult IAPT workforce census report*. Retrieved from <https://www.england.nhs.uk/mentalhealth/wp-content/uploads/sites/29/2016/09/adult-iapt-workforce-census-report-15.pdf>
- Nilsson, T., Svensson, M., Sandell, R., & Clinton, D. (2007). Patients' experiences of change in cognitive-behavioral therapy and psychodynamic therapy: a qualitative comparative study. *Psychotherapy Research*, **17**, 553–566.
- Parry, G., Barkham, M., Brazier, J., Dent-Brown, K., Hardy, G., Kendrick, T., . . . Lovell, K. (2011). *An evaluation of a new service model: Improving access to psychological therapies demonstration sites 2006–2009*. Final

- report. NIHR Service Delivery and Organisation programme.
- Perren, S. (2009). Topics in training: thinking of applying for the IAPT high or low-intensity training? Sara Perren has some advice. *Healthcare Counselling and Psychotherapy Journal*, **1**, 29–30.
- Pybis, J., Saxon, D., Hill, A., & Barkham, M. (2017). The comparative effectiveness and efficiency of cognitive behaviour therapy and counselling in the treatment of depression: evidence from the 2nd UK national audit of psychological therapies. *BMC Psychiatry*, **17**, 215.
- Radhakrishnan, M., Hammond, G., Jones, P. B., Watson, A., McMillan-Shields, F., & Lafortune, L. (2013). Cost of Improving Access to Psychological Therapies (IAPT) programme: an analysis of cost of session, treatment and recovery in selected primary care trusts in the east of England region. *Behaviour Research and Therapy*, **51**, 37–45.
- Rawlins, M. (2008). De testimonio: on the evidence for decisions about the use of therapeutic interventions. *The Lancet*, **372**, 2152–2161.
- Ridge, D., & Ziebland, S. (2006). ‘The old me could never have done that’: how people give meaning to recovery following depression. *Qualitative Health Research*, **16**, 1038–1053.
- Roth, A. D., Hill, A., & Pilling, S. (2009). *The competences required to deliver effective humanistic psychological therapies*. London, UK: Department of Health.
- Sanders, P., & Hill, A. (2014). *Counselling for depression: A person-centred and experiential approach to practice*. London, United Kingdom: Sage.
- Saxon, D., Ashley, K., Bishop-Edwards, L., Connell, J., Harrison, P., Ohlsen, S., . . . Barkham, M. (2017). A pragmatic randomized controlled trial assessing the non-inferiority of counselling for depression versus cognitive-behaviour therapy for patients in primary care meeting a diagnosis of moderate or severe depression (PRaCTICED): a study protocol for a randomized controlled trial. *Trials*, **18**, 93.
- Saxon, D., & Barkham, M. (2012). Patterns of therapist variability: therapist effects and the contribution of patient severity and risk. *Journal of Consulting and Clinical Psychology*, **80**, 535–546.
- Saxon, D., Firth, N., & Barkham, M. (2017). The relationship between therapist effects and therapy delivery factors: therapy modality, dosage, and non-completion. *Administration and Policy in Mental Health*, **44**, 705–715.
- Scottish Intercollegiate Guidelines Network (2017). *What we do*. Retrieved from <http://www.sign.ac.uk/what-we-do.html>
- Simpson, S., Corney, R., Fitzgerald, P., & Beecham, J. (2000). A randomised controlled trial to evaluate the effectiveness and cost-effectiveness of counselling patients with chronic depression. *Health Technology Assessment*, **4**(36), 1–83.
- Smith, A., Graham, L., & Senthinathan, S. (2007). Mindfulness-based cognitive therapy for recurring depression in older people: a qualitative study. *Aging and Mental Health*, **11**, 346–357.
- Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine*, **22**, 1092–1097.
- Stiles, W. B., Barkham, M., Mellor-Clark, J., & Connell, J. (2008). Effectiveness of cognitive-behavioural, person-centred, 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., Mellor-Clark, J., & Cooper, M. (2006). Effectiveness of cognitive-behavioural, person-centred, and psychodynamic therapies as practiced in UK National Health Service settings. *Psychological Medicine*, **36**, 555–566.
- Stoppard, J. M., & McMullen, L. M. (1999). Toward an understanding of depression from the standpoint of women: exploring contributions of qualitative research approaches. *Canadian Psychology*, **40**, 75–76.
- Straarup, N. S., & Poulsen, S. (2015). Helpful aspects of metacognitive therapy and cognitive behaviour therapy for depression: a qualitative study. *The Cognitive Behaviour Therapist*, **8**, e22.
- Stronks, K., Wieringa, F., & Hardon, A. (2013). Confronting diversity in the production of clinical evidence goes beyond merely including under-represented groups in clinical trials. *Trials*, **14**, 177.
- Stuart, E. A., Bradshaw, C. P., & Leaf, P. J. (2015). Assessing the generalizability of randomized trial results to target populations. *Prevention Science*, **16**, 475–485.
- Timulak, L., & Elliott, R. (2003). Empowerment events in process-experiential psychotherapy of depression: an exploratory qualitative analysis. *Psychotherapy Research*, **13**, 443–460.
- Valkonen, J., Hänninen, V., & Lindfors, O. (2011). Outcomes of psychotherapy from the perspective of the users. *Psychotherapy Research*, **21**, 227–240.
- Vlayen, J., Aertgeerts, B., Hannes, K., Sermeus, W., & Ramaeker, D. (2005). A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. *International Journal for Quality in Health Care*, **17**, 235–242.
- Wald, N. J., & Morris, J. K. (2003). Teleoanalysis: combining data from different types of study. *BMJ*, **327**, 616–618.
- Ward, E., King, M., Lloyd, M., Bower, P., Sibbald, B., Farrelly, S., . . . Addington-Hall, J. (2000). Randomised controlled trial of non-directive counselling, cognitive-behaviour therapy, and usual general practitioner care

- for patients with depression. I: clinical effectiveness. *BMJ*, **321**, 1383–1388.
- Watson, J. C., Gordon, L. B., Stermac, L., Kalogerakos, F., & Steckley, P. (2003). Comparing the effectiveness of process-experiential with cognitive-behavioral psychotherapy in the treatment of depression. *Journal of Consulting and Clinical Psychology*, **71**, 773–781.
- Wessley, S. (2007). Commentary: a defence of the randomized controlled trial in mental health. *BioSocieties*, **2**, 115–127.
- Williams, R., Farquharson, L., Palmer, L., Bassett, P., Clarke, J., Clark, D. M., & Crawford, M. J. (2016). Patient preference in psychological treatment and associations with self-reported outcome: national cross-sectional survey in England and Wales. *BMC Psychiatry*, **16**, 4.
- World Health Organization (2017). *WHO guidelines approved by the Guidelines Review Committee*. Retrieved from <http://www.who.int/publications/guidelines/en/>

Biographies

Michael Barkham is Professor of Clinical Psychology and Director of the Centre for Psychological Services Research at the University of Sheffield.

Naomi P. Moller is Joint Head of Research for the British Association for Counselling and Psychotherapy and Senior Lecturer in the School of Psychology at the Open University.

Joanne Pybis is Senior Research Fellow for the British Association for Counselling and Psychotherapy.