View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by University of Essex Research Repository

# Epistemological flaws in NICE review methodology and its impact on recommendations for psychodynamic psychotherapies for complex and persistent depression

Susan McPherson[1], Felicitas Rost[2], Joel Town[3], Allan Abbass[4]

[1]Dr Susan McPherson, School of Health and Social Care, University of Essex, Colchester, CO4 3SQ, UK. Email: smcpher@essex.ac.uk

[2]Dr Felicitas Rost, Portman Clinic, 8 Fitzjohn's Avenue, Tavistock and Portman NHS Foundation Trust, London, NW3 5NA, UK. Email: frost@tavi-port.nhs.uk.

[3]Dr Joel Town, Department of Psychiatry, Faculty of Medicine, Dalhousie University, 5909 Veterans Memorial Lane, Halifax, NS, Canada, B3H 2E2. Email: Joel.town@dal.ca

[4]Dr Allan Abbass, Centre for Emotions and Health, Faculty of Medicine, Dalhousie University, 5909 Veterans Memorial Lane, Halifax, NS, Canada, B3H 2E2. Email: Allan.Abbass@Dal.Ca.

Correspondence concerning this article should be addressed to Susan McPherson, School of Health and Social Care, University of Essex, Colchester , CO4 3SQ, UK. Contact: smcpher@essex.ac.uk

Published in Psychoanalytic Psychotherapy online 9 April 2018
https://www.tandfonline.com/eprint/awkDyW7UQ5bUrvHABPnn/full

**Abstract**

The UK draft NICE guideline on depression in adults was sent out for stakeholder consultation between July and September 2017. The final guideline publication date currently remains 'to be confirmed'. This paper sets out key concerns with the methodology employed in the guideline and its impact on recommendations for psychodynamic psychotherapies for complex and persistent depression. The draft largely ignored the subjective experiences and voices of service users, carers and members of the public, using out of date limited evidence of service user and carer experiences. The guideline fails to incorporate what limited qualitative evidence it reviewed into any treatment recommendations. The Guideline Committee created its own method for categorising depression by longevity, severity and complexity. This has resulted in erroneous and unhelpful classifications of research studies under groupings which do not match clinical and service user experiences or US and European approaches, rendering analyses and conclusions unreliable. We also outline instances of incorrect classification of psychodynamic treatments (such as inclusion of non bona-fide treatments or exclusion of relevant bona-fide treatment studies) which enables the omission of a recommendation for psychodynamic psychotherapy for complex and persistent depression. Depression is often a long-term condition or can become so if immediate care is inadequate; yet the draft recommendations are all made on the basis of short-term outcome data (with often less than 8 weeks between baseline and outcome). NICE guidelines for long-term physical conditions would treat this evidence as inadequate. Finally, the draft guideline used the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) system of assessing methodological quality in such a way as to produce a systematic bias in favour of drug trials, selectively omitting trial data with long-term follow up points and those which used non-symptom outcomes. Herein we consider the increasingly evident limitations of the paradigm NICE works within for ensuring patient choice and equity of access to a wide range of therapies.

Epistemological flaws in NICE review methodology and its impact on recommendations for psychodynamic psychotherapies for complex and persistent depression

## Background

The draft NICE guideline (2017) on depression in adults was sent out for stakeholder consultation between July and September 2017 with a planned publication date of January 2018, that has since been moved twice, first to March 2018 and currently "to be confirmed". This pause might be an acknowledgement of the gravity of stakeholder concerns and the relevance of the guideline to public mental health. Much new evidence has emerged since the previous 2009 guideline, which, along with significant social and economic changes impacting on the experience of depression and delivery of treatments, clearly merit a new assessment of evidence. Yet, it is critical that any new guideline takes a properly balanced and nuanced approach to evidence synthesis combined appropriately with service user views, given the potential impact it will have on commissioning in the UK and further afield (several European guidelines as well as Canadian guidelines look towards NICE as the authoritative model). Elements of the first 2004 guideline on depression were employed with huge success as a lobbying device to dramatically change the nature and shape of the psychological therapy workforce and delivery in the form of the Improving Access to Psychological Therapy (IAPT) programme. In spite of many stakeholder concerns at that time about the narrow focus of IAPT on Cognitive Behavioural Therapy (CBT), the combination of NICE's skewed methodological approach and lobbying by professional groups with significant support from influential politicians, led to a significant increase in CBT training and provision, whereas training and delivery of other therapies decreased, significantly reducing patient choice.

This paper sets out key concerns regarding the methodology employed to date in the draft guideline, indicating that the guideline will not be fit for purpose and has the potential to seriously impede patient choice. Specifically, stakeholders and commissioners of mental health services require treatment recommendations that address the clinical needs of the entire spectrum of patients who experience depression. We argue that alongside the high prevalence of mild to moderate cases of depression, services also face the significant clinical challenge of treating cases of chronic, recurrent, refractory depressive symptoms that commonly co-occur with other morbid physical and mental health conditions. If adopted without significant redesign, these issues will have a major impact on care for a significant proportion of the population given that, although not everyone experiencing depression will

have a formal diagnosis, 42% of adults report experiencing depression at some time in their life (Mental Health Foundation, n.d.). The guideline will impact on individuals, families and communities as well as people in other nations that have in the past adopted NICE guidelines. The largest impact will be felt as a result of not adequately meeting the challenges of chronic, recurrent and complex cases.

In this paper, we address a number of key issues with the draft guideline and illustrate how these issues impact significantly on the way recent RCTs for chronic depression (CD), treatment resistant depression (TRD) and complex depression have been classified in the draft guideline, leading to a bias against psychological therapies generally and psychodynamic psychotherapies in particular. The focus on CD and TRD is chosen partly because of the limited availability of effective interventions for this group who are therefore already disadvantaged by current services and likely to be most affected by any implementation of the guideline. Because of the focus on TRD and CD, in the more technical aspects of critique we draw mainly on two recent RCTs of psychodynamic psychotherapy for TRD. These are Fonagy et al (2015), an RCT of 18-month individual long-term psychodynamic psychotherapy (LTPP) for TRD with a two-year follow-up; and Town et al (2017), an RCT of short-term psychodynamic psychotherapy (STPP) for TRD with an average of 16 sessions and one-year follow-up.

## Undermining Subjective Experience

NICE made a decision not to update the section in the guideline on service user experience purportedly because of insufficient stakeholder responses supporting an update during the scoping phase of development. This is regrettable given that the section copied over from the 2009 guideline (NICE, 2009) was inadequate. Experience of depression is intertwined with the social and economic context in which people live. It relates to levels of community cohesion, economic circumstances, social support and loneliness. There is growing evidence of the impact of austerity on depression and many clients with depression have been significantly affected by reductions in their benefits, loss of work or changes to employment conditions resulting from the economic downturn and political choices (see for example Psychologists Against Austerity, n.d.). Experiences of depression will have been affected by this and there is no scientific basis for the assumption that experience of depression is a static biological phenomenon (see, for example Kendler, 2016). There have

also been changes which impact on the extent to which stigma features in client experience. Campaigns such a Time to Change (Henderson et al, 2016), may (or may not) have had an impact on the experience of depression and help seeking. Recent policy changes could also have impacted on experiences of carers; the Care Act coming into law in 2014 and benefits changes mean that carers' experiences are unlikely to be the same as in 2004 or 2009. The guideline approach implies these have remained largely static.

The 2009 Guideline Committee (GC) reviewed some primary qualitative literature but the review was based on one existing review of patient experience of guided self-help for depression (Khan et al, 2007) plus two other articles. This focus is skewed and narrow. There is a great deal of primary research on experiences of depression and experiences of treatments going well beyond guided self-help interventions. The review by the GC (transferred verbatim from the 2009 version) is therefore muddled and incomplete. Recent literature since 2009 (not reviewed by the GC) extends client experience data to under-represented groups and has been undertaken in more recent social and economic contexts (e.g. Smith & Rhodes, 2014; van Grieken et al, 2014; Oliffe et al, 2013; Corcoran et al, 2013; Patterson-Kane & Quirk, 2014; Anderson & Roy, 2013; Brown et al, 2012; Oliffe et al, 2011; Körner et al, 2011; Gask et al, 2011; Bryant-Bedell & Waite, 2010; Oliffe et al, 2012). Recent literature also includes a qualitative metasynthesis of the experiences of people caring for partners or family members with depression (Priestley & McPherson, 2015) which could have informed the current guideline. There have also been many case studies published (for example Cohen, 2016; Roberts & Sedley, 2016).

Leaving aside the neglect of recent evidence, patient experience data that was used in the 2009 guideline consisted of seven written accounts collected by the GC and a reanalysis of 38 individual accounts extracted from an online database 'Healthtalkonline'. The seven service user accounts were written by respondents who were asked to compose a written narrative considering a set of closed questions. It is unclear how diverse the sample was, nor what formal method of analysis if any was used. The guideline printed the accounts in full and provided a very brief summary of their content:

> *Although the 6 questions were aimed at people with any form of depression, all of the personal accounts received were from people who have/have had severe and chronic*

*depression, spanning many years. The themes that are most frequently expressed in the testimonies include trauma or conflict in childhood as a perceived cause of depression; the need for long-term psychotherapy for people with severe and chronic depression; the need to take personal responsibility for and understand the illness to improve outcomes; issues around diversity; paid and unpaid employment as an important part of the recovery process; the negative impact on daily functioning; concerns regarding stigma and discrimination in the workplace; and the relationship between people with depression and professionals. (p68)*

This analysis lacks depth and rigour but more importantly, these findings were not incorporated into methodological approaches in the guideline or treatment recommendations. The 2009 review also re-analysed 38 accounts from a secondary data source (Healthtalkonline). No demographic details were given for the individuals whose accounts were taken from the database so it is unclear which elements of the population were represented in the data selected. The extent to which the Healthtalkonline data or the seven personal accounts collected included under-represented populations such as BAME, men, older adults, non-heterosexual clients is unclear.

Given the purpose of a NICE review is to review existing evidence, a full systematic review of primary studies in this field employing formal methodology for synthesis such as meta-ethnographic synthesis (Noblit & Hare, 1988), metasummary (Sandelowski & Barrosso, 2003) or narrative synthesis (Popay et al, 2006) would have been a more appropriate approach to a guideline review. This would have enhanced understanding of service user experiences, a position held by several bodies including the American Psychiatric Association (2006), the Cochrane Collaboration (Noyes et al, 2011) and the Health Foundation (2017). Review of qualitative findings should also be incorporated into treatment recommendations rather than being left as a stand-alone section. As to how this limitation has already disadvantaged psychodynamic psychotherapy, the guideline notes:

T*here was a strong feeling within the service user and carer topic group that … psychological treatment offered by the NHS in the form of CBT does not go far enough in addressing the trauma experienced in childhood. The study by Ridge and Ziebland (2006) confirms the opinions of the topic group and the testimony from the personal*

*accounts that people with 'deep and complex problems felt the need for longer term therapy'. Those that have had long-term psychodynamic therapy report that it has been helpful in their understanding of themselves and their depression and that until they have worked through and repaired the damage experienced in childhood, depression will be a major factor in the person's life. The service user and carer topic group do acknowledge, however, that as there has been little research into the efficacy of long-term psychodynamic therapy, it cannot be recommended as a course of treatment in this guideline. (p97)*

This comment was made around 10 years ago. Since then studies have been carried out on psychodynamic psychotherapies for long-term depression (e.g. Fonagy et al, 2015; Town et al, 2017, both discussed below). None of the recommendations (p.100) deriving from service user and carer experiences relate to interventions and the guideline makes no policy comment about the key issue raised by service users and carers concerning stigma, which effects their help seeking behaviour even before any treatment choice is considered. Many of the new recommendations throughout the guideline do nod towards patient choice by recommending that if the patient 'does not want' the treatment recommended, an alternative (from a specified list) should be offered. This is, however, a limited interpretation of patient choice and fails to acknowledge inequity of access to a range of therapies as a result of the impact of the previous guideline. Any new guideline should be informed by an impact assessment of the previous guideline and this missing step has led to inequity of access to therapies compounding NICE's limited commitment to patient choice.

**Erroneous classification of depression subtypes and bona-fide treatments**

Depression is an extremely diverse presentation, not least in terms of chronicity, severity and complexity. Treatments are also very diverse and inevitably, RCTs of treatments for depression have a wide range of designs, measures and analytic approaches. The task of a systematic review team is therefore hugely complex, an exercise in comparing apples and pears on a mammoth scale. In order to remain within the restrictions of a diagnostic paradigm with a strict evidence hierarchy, some blunt categorical decisions must be made. However, although the GC acknowledges that there are considerable problems when attempting to classify depression and thus trials into categories, the approach to classification in this draft guideline is concerning. We examine the classifications the GC use for severity, complexity

and chronicity and how these impact on their handling of psychodynamic psychotherapy RCTs. We also illustrate how the incorrect categorisation of psychodynamic psychotherapies leads to selective exclusion of bona-fide treatments and selective inclusion of non bona-fide treatments, all of which leads to unreliable analyses and conclusions.

**An idiosyncratic approach to depression severity**

The draft guidelines used a method of dividing trial populations by dichotomising baseline severity as 'more severe' or 'less severe'. This single reductive proxy estimate of severity depends on the un-evidenced assumption that the equivalence algorithm combining different depression rating scales developed by the GC is reliable and valid. Most of the component measures taken account of have their own range of severity categories, validated in the literature. The decision to ignore these risks misrepresenting the true severity of a study population and thus forming inaccurate judgments on the effectiveness of treatments. For instance, the study populations in the Fonagy et al (2015) and Town et al (2017) RCTs of LTPP and STPP respectively are both categorised as 'Less severe' for baseline severity. Fonagy et al (2015) and Town et al (2017) used the 17-item Hamilton Depression rating scale (HAMD) and according to the HAMD validated categories, the mean baseline scores for both studies fell in the 'severe' range. This indicates a significant discrepancy between the guideline method and the method established in the literature.

The GC approach to severity also fails to take account of individual change and amount of change. It is more difficult for a treatment to move an individual from a very high score on a measure to the target low score which would count as 'full remission'. It is easier to move someone into full remission if they start on or closer to the low score one has to reach to count as being in remission. It would therefore have been important for the guideline to look at rates of partial remission as well as full remission, particularly where participants begin the intervention with severe depression.

It would be yet more useful to look at individual change. There is significant variation in individual baseline severity scores within and between studies. Examining Standardised Mean Differences alone is an inadequate approach to looking at change. A method for determining Reliable and Clinically Significant Change for each participant individually (Jacobsen & Truaux, 1991) offers a better assessment of how changes on different measures

taking individual baseline severity into account, might be interpreted. This method is employed in IAPT reports in which recent figures show an overall 'recovery' rate of 46.3% (HSCIC, 2016). IAPT reports then examine 'reliable improvement' (which considers baseline and end-point severity, rather than only whether the case met 'clinical caseness' at either point). This finds that 62.2% of IAPT clients improve – a higher figure than when just reporting % reaching the 'recovery' point. Using 'reliable improvement' methods in the trials included in the guideline meta-analyses would offer a fuller picture; which is particularly important when trials have studied the treatment of markedly severe and refractory populations for whom currently there are few well-evidenced treatments available.

**Chronicity: mistaken augmentation strategies**

The draft guideline separates RCTs of CD and TRD into two discrete categories, whether or not study populations fell into both categories. RCTs that fit both categories were allocated to one category or the other with no explicit criterion for deciding which category takes precedence. The 2009 guideline noted that the overlap of CD and TRD populations is so large as to render questionable their separation as a means of structuring meta-analyses and therefore did not use the TRD category. The 2009 guideline cited evidence for a more loosely defined heterogenous group of long-term, difficult-to-treat depressive conditions, frequently associated with dysthymia and co-morbid common mental disorders, various personality traits and serious psychosocial disability. This position is also followed by the APA in the fifth edition of the Diagnostic and Statistical Manual (DSM, 2013) and the European Psychiatric Association guidelines for persistent depression (EPA; Jobst, 2016).

The separation of CD and TRD in NICE's draft guideline gives rise to considerable confusion. The definition of treatment resistance given states that a study should include people "who have not tolerated previous treatment (for the current episode), and who have been randomised to the further line interventions at the point at which they had no/adequate/limited response" (p343). This definition requires studies to prospectively monitor responses to treatments prior to allocating participants to a study of TRD. The definition of CD is based on a retrospective judgement; studies must meet "criteria for full Major Depressive Disorder (MDD) for 2 years; persistent subthreshold symptoms (dysthymia); double depression (an acute episode of MDD superimposed on dysthymia)". Many subjects in the trials included in the TRD meta-analysis will also meet the definition of

CD. Ruhe et al (2012) note: "because of their chronic clinical course, approximately 40% of CD patients also fulfil criteria for TRD…… usually defined by the number of non-successful biological treatments".

The definitions given of TRD in the guideline (p. 342-3) are exclusively pharmacological requiring operationalisation of dose and duration monitoring. They are not those used in usual clinical settings where case identification is usually descriptive and involves complex evaluations of psychosocial functioning across several domains. They imply that the inadequate response to the agent was immediately recent or current. Trials in this category are classed as 'dose escalation' or 'augmentation strategies', the latter including various sub-types such as augmenting with another antidepressant; augmenting with a psychological intervention etc. Both the Fonagy et al (2015) study of LTPP and the Town et al (2017) study of STPP were classified as trials of 'augmenting the antidepressant with a psychological intervention versus continuing with the antidepressant only'. Other trials classed as TRD generally established TRD by means of a medical model operationalisation. For example, the Kocsis (2009) criteria read:

> *Inadequate response to 12 weeks of antidepressant medication according to a pharmacotherapy algorithm. Inadequate response defined as failing to meet criteria for remission (≥60% reduction in HAMD score, a HAMD total score<8, and no longer meeting DSM-IV criteria for MDD for 2 consecutive visits during weeks 6 through 12)". (Table 120)*

While the participants in the LTPP study are described by the authors as 'treatment resistant', participants do not meet the criteria for TRD defined by NICE and the intervention was not an 'augmentation strategy'. The criteria were broadly inclusive, stating participants should have had an 'Inadequate response to least two different treatments (mean of 3.7 previously failed treatment attempts)' (Fonagy et al, 2015). This represents a fundamentally different approach to 'resistance'. There was no requirement in the inclusion criteria that the medication received need be recent or current. Although the appendices of the guideline indicate that the study met criteria for CD as well as TRD, it is analysed along with other TRD studies. It is similarly misleading to classify Town et al (2017) as an augmentation strategy, particularly given that two thirds in the STPP group reduced or came off medication

while just over half the control group increased or started medication. Conceiving of these studies as 'augmentations' of medications is clearly a misrepresentation with the consequence that they are then compared directly against drug trials in methodologically inappropriate ways.

**Selective inclusion of non bona-fide treatments**

The recent RCT by Town et al (2017) provides evidence for the use of psychodynamic psychotherapy for treatment resistant depression. Based upon the accurate inclusion of only bona-fide therapies for treatment resistant depression, the evidence for STPP as a first line treatment approach is at least comparable to that for Interpersonal Psychotherapy Therapy (IPT) which is recommended based on evidence from one low quality single study of IPT in which IPT did not even outperform 'treatment as usual' (Souza et al, 2016). On first glance, it is therefore unclear what standard has been used to justify the omission of a NICE recommendation for STPP for treatment resistant depression.

On closer examination of the inclusion criteria of studies included under the category of 'STPP versus treatment as usual or antidepressants', we identified a significant mistake. Alongside the Town et al (2017) trial, data for Brief Supportive Psychotherapy (BSP) from another RCT (Kocsis et al, 2009; Klein et al, 2011) were included as data on STPP. BSP is a recognised treatment approach emphasizing common psychotherapy factors but is distinct from other psychotherapies given its lack of structure and specific interventions. The BSP manual for this trial stated that "psychodynamic interventions were strictly proscribed" (Kocsis et al, 2009). It therefore should not be classified as an STPP in this instance. The inclusion of a non bona-fide STPP raises serious problems with the integrity of the analysis. The GC concluded that there was no effect of STPP as an augmentation strategy, clearly taking a lead from the Kocsis et al (2009) trial rather than Town et al (2017) – the latter a bona fide STPP treatment indicating a clinically important and statistically significant benefit of individual STPP. If NICE apply the same standard for evaluating treatments for resistant depression across psychological therapies, revision to the draft guidelines is required in order to include STPP as a treatment option.

**The narrow gaze of complexity and exclusion of bona-fide treatments**

The GC also introduced another category of 'depression with co-morbidites', as distinct from CD and TRD (p585). Within this category are the sub-categories 'complex depression' and 'psychotic depression'. Again, these are not in line with DSM or EPA and create false divisions between study populations. Comorbidity with personality disorder (PD) is the only criterion for 'complex depression'; an arbitrary cut-off of 51% of participants having PD is provided bearing no resemblance to any recognised classification system nor any professional or service user trends. Many studies otherwise classed as TRD or CD would also fit the 'complex' depression category but have arbitrarily been placed in another category. The EPA guidelines for persistent depression (Jobst et al, 2016) adopt a slightly more nuanced approach to complexity in depression that is closer to the complexity generally encountered in clinical services. Complexity is noted in terms of early versus late onset, type of depression, number of episodes, early trauma, symptom severity, patient preference and comorbid personality disorder, and that the type of treatment offered should be individually tailored accordingly (evidence level 4; Good Practice Point). The GC use of PD as the only criterion ignores these aspects of complexity and in doing so denies the reality of complex early lives and experiences of trauma that many individuals within study populations are likely to have experienced.

On examination of the 70 RCTs reviewed at full text for inclusion in the 'complex depression' meta-analysis, the Fonagy et al (2015) study of LTPP was not considered for inclusion and five RCTs of STPP were considered but excluded (see below). Eighty-five percent of the Fonagy et al (2015) study population had one or more Axis II disorder. Participants had high levels of childhood adversity (89% - unpublished data available on request); and high comorbidity: 47% had musculoskeletal problems, 25% had gastrointestinal problems; 91% had at least one other comorbid Axis 1 disorder; 54% were unemployed; the mean baseline Global Assessment of Functioning (GAF) score was 49.1; 45% had made at least one previous suicide attempt etc. Clinically this is a very complex population, as was the Town et al (2017) sample, yet these trials are classed as TRD only – rather than chronic and/or complex, highlighting the systematic ignoring of individual subjectivities which in many cases are likely to be highly predicated on childhood adversity and trauma.

Five RCTs of STPP were excluded on the grounds that these studies included (i) "no extractable outcomes of interest" (Abbass et al, 2008; Vinnars et al, 2005); (ii) less than 50%

of the trial population had comorbid PD (Thyme, 2007); and (iii) less than 10 participants per treatment arm (Maina et al, 2005; Svartberg et al, 2004). However, despite the protocol for the review of interventions for complex depression specifying the inclusion of systematic reviews and disaggregated RCT data for this population, published disaggregated data was overlooked for 79 participants receiving STPP from these 5 RCTs (Abbass et al, 2011). All participants in this excluded group had confirmed diagnoses of depressive disorder with PD and treatment outcome data on validated depression scales were available. Three RCTs of STPP (Abbass et al, 2008; Thyme et al, 2007; Vinnars et al, 2005) were essentially incorrectly excluded based on rationale i and ii. Furthermore, disaggregated data from two further RCTs of STPP (Maina et al, 2005; Svartberg et al, 2004) which identified a true sample of complex depression, were excluded because of less than 10 participants per treatment arm. In contrast, one of the five RCTs included in this review (Liberman et al, 1981 study of Behaviour Therapy) had failed to administer a formal PD assessment and precise numbers on those with an unconfirmed PD diagnosis were not provided. Hence only a 'best guess' could speculate how many participants in this small trial (N=24) may have had PD. The GC conclude that the evidence base for complex depression is limited in volume having only included five small RCTs. If bona-fide studies of therapies are selectively excluded from data analytic approaches to examine the effects of a specific treatment in this way, estimates of within- and between-group differences can be expected to be unreliable and ensuing quality statements and recommendations questionable (Leichsenring et al, 2017). Likewise, including non bona-fide interventions as noted previously in the inclusion of BSP as an STPP, has an equally problematic impact on guideline recommendations.

We acknowledge that some of these technical issues around categorising trials by chronicity, severity and complexity may appear akin to dancing on a pin head and that perhaps a redrawing of the boundaries would likely open yet new potholes. Yet this in itself illustrates that the scientific methods employed by NICE ultimately rely on a series of multiple complex decisions made by a relatively small group of appointed experts forced to come to a consensus to drive through the guideline development process. There are limited objective facts guiding these decisions and there are so many decisions to be made at each stage of review that there are multiple opportunities for bias and/or error to occur. Cumulatively, the result of these decisions for the proposed guideline update is the continued omission of an appropriate and specific place for the implementation of psychodynamic

therapies within the NHS. The aforementioned evidence points to an important place for psychodynamic psychotherapies in the treatment of complex presentations of depression.

We might argue that the paradigm in operation within NICE requires such an extent of ongoing adjustment creating ever-increasing complexity that it cannot sustain itself, as in Thomas Kuhn's idea of paradigm shifts (Kuhn, 1962). An alternative approach may be required which puts patient choice central instead of peripheral and which considers alternative ways of conceptualizing expressions of psychological distress, for example, formulation based approaches or trauma-informed approaches like the recent Power Threat Meaning Framework (Johnstone & Boyle, 2018) published by the British Psychological Society. Such approaches move towards acknowledging that psychological expressions of distress have meaning, can be signposts to what is wrong in a person's life and can be open to change.

**Parity with other Long-Term Conditions?**

The draft guideline states that "*the aim of intervention is to restore health through the relief of symptoms and restoration of function, and in the longer term, to prevent relapse*" (p.40, l.31-32). It furthermore highlights the high likelihood of relapse or deterioration in patients with depression described under the heading of 'treatment resistance' (section 8.1.2). In spite of this, the draft guideline does not take account of long-term follow-up. Indeed, it excluded long-term follow-up data that might demonstrate whether the aims of interventions have indeed been achieved.

The choice to omit long-term data is particularly difficult to comprehend in the sections dealing with TRD and CD and is likely due to most trials not collecting or reporting follow-up data. This is, however, not an adequate justification. Calls for RCTs of interventions for depression to include longer term follow-up years after the end of treatment, given the episodic nature of depression, have been made repeatedly (e.g. McPherson et al, 2005; Goodyer et al, 2008; Goodyer et al, 2011; Goodyer et al, 2017). According to criteria adopted by NICE as well as the APA and EPA, chronic forms of depression must last for at least two years. Studies included in the NICE review report mean durations of illness as long as 7.8 years (Keller et al, 2000); 7.6 years (Kocsis et al, 2009); and 24.4 years (Fonagy et al, 2015). Given the actual mean duration of illness observed, there is an even stronger case for looking at data from follow-up periods in more chronic forms of depression. Moreover,

Westen et al (2004) note that many patients who respond initially to treatments will relapse and/or present to other services subsequently. Hepgul et al (2016) noted that 38% of IAPT attenders had attended IAPT previously, pointing to a high relapse rate in practice for many front-line therapies. Long-term follow-up data is therefore crucial in the evaluation of the therapeutic effects of treatments for depression whether it is already known to have been long lasting or whether it is a first episode that may become long lasting.

On the contrary, most trials in the review have follow-up points of 8 weeks or less. Analyses conducted to evaluate the effectiveness of interventions in the draft guideline have taken the endpoint as the end of treatment. Moreover, the draft guideline has selectively omitted follow-up data in the very few trials of CD and TRD with follow-up and observation periods sufficiently long to offer data about the longer-term durability of treatment effects. The Fonagy et al (2015) study followed participants for two years after the 18-month treatment period giving a total observation period of 182 weeks. The guideline has extracted the end of treatment data from this study (at which point no significant differences between groups were observed) and omitted the important data yielded by the two-year follow-up data, which showed a substantial effect (full remission Numbers Needed to Treat (NNT)=9.6; partial remission NNT = 3.9).

NICE does not treat any other long-term condition in this way. The diabetes (type 2 adults) guideline, for example, includes examination of outcomes ranging from two years up to 10 and over, reflecting the long-term, chronic nature of this condition. The epilepsy guideline and arthritis guideline examined evidence including one and two-year follow-up data and in some cases longer. In order to treat depression, particularly any persistent form of depression, as a long-term condition on a par with long-term physical conditions, it is important for follow-up data to be taken into account. To meet with principles of parity of esteem, where mental health is valued equally with physical health, all treatments deemed to be effective and recommended for long-term depression ought to have demonstrated an impact at least one year (ideally two or more years) beyond the end of treatment. If the effects of the treatment wear off as soon as (or soon after) the treatment finishes (or the long-term effects are unknown) then the treatment can at best be considered a reasonable sticking plaster. Treatments for physical illnesses that stopped working immediately after the end of treatment (or where there was no known effectiveness beyond the end of treatment) would not typically be recommended in a national guideline. The impact of ignoring these issues is

that LTPP has not been recommended partly on the basis of the absence of end of treatment effects; whereas because of its long-term benefits shown in Fonagy et al (2015), LTPP should be recommended as one of the options for service users with persistent depression to choose from.

**Assessing Methodological Quality of Studies: Strategically Inflexible?**

The GC employed the Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach, a widely used method of assessing RCT quality which allocates points for various criteria such as sample size, blinding, analytic approach, reporting quality etc. Authors of the GRADE system note:

> *GRADE is 'outcome centric' in that a rating is made for each outcome, and quality may differ—indeed, is likely to differ—from one outcome to another within a single study and across a body of evidence. (Dijkers, 2013).*

In general, GRADE is designed to be used flexibly:

> *We don't necessarily report on all possible parameters of a study — for example, whether an RCT was single or double blinded, or the precise method of randomisation used — rather, following a critical appraisal of each study, we highlight the methodological or other issues that we feel may affect the interpretation of the results or the weight that might be placed on them. (http://clinicalevidence.bmj.com/x/set/static/ebm/learn/665072.html)*

Yet we are concerned that the team undertaking this assessment has employed GRADE inflexibly and have assessed all trials in the same way, irrespective of whether each criterion is appropriate for the type of trial or intervention being assessed. Assessing the quality of a psychological treatment trial in the same way as a drug trial is epistemologically unsound. For example, it is widely understood that blinding participants and investigators is not possible to apply to psychological treatment trials because a psychological treatment is not like a pill for which you can create another pill identical in all visible ways but with no

active ingredient. In psychological treatment trials the therapy is unique to each individual and therapist pair and involves a unique combination of words exchanged between the two people. It is impossible to create a placebo version of any psychological treatment which will fool participants and investigators. Nonetheless all psychological treatment trials have lost points within the GRADE system for failing to meet this 'standard'. This systematically advantages drug treatment trials.

Use of the GRADE system ought to reflect the complex endeavour of comparing medical and psychological treatments. More specifically, the review should employ quality criteria which take into account the use of a range of outcome measures, in particular the assessment of functioning in addition to targeted symptoms. Functioning is not only important to service users but is also both directly and indirectly related to health service use, social care costs and employment (McCrone et al, 2017) and could therefore be examined as a proxy for health economic outcomes where these are not available. As Dijkers (2013) has stressed, the quality for each outcome may differ between outcomes within a single study and across a body of evidence. A re-analysis of the 2004 NICE review examining outcomes of measures of functioning showed a different order of comparative efficacy amongst interventions (McPherson, Evans & Richardson, 2009). Calls have long been made for RCTs of treatments for depression to include alternative outcome measures (McPherson et al, 2005; Wampold et al, 2017) and would be in line with those who have stressed that the disease burden is primarily due to comorbidities and not merely to the additive effects of having more than one disorder (Wampold et al, 2017). This is important not least because the omission stands in contrast to the emphasis of the draft guideline in the introduction, which clearly states: "Where possible, the key goal of an intervention should be complete relief of symptoms (remission), which is associated with better functioning… For this reason the GC examined a range of outcomes (where available), including response, remission, change in symptoms and relapse." (p. 40, l.32-37)

Service users often report that the outcomes that are most important to them are their quality of life, ability to deal with problems, being able to take care of oneself, feeling independent, feeling able to communicate with others, feeling able to get about, ability to attend education, work or social events (e.g. Zimmerman et al, 2013). Although NICE has acknowledged the importance of these non-medical symptoms in previous versions of the guideline, it has consistently failed to adjust its primary focus away from symptoms. The

current guideline has ignored available data on non-symptom outcomes to illustrate where evidence for treatment effectiveness in these areas exist. If the GC employed the GRADE system to emphasize the importance of long-term outcomes and non-symptom outcomes, those trials that fail to measure or report non-symptom outcomes or long-term outcomes would be rated as lower quality. Instead, GRADE has been used in such a way that a large quantity of narrowly focused short-term trials are given much greater weight than those few trials where broader long-term outcomes are available.

This specifically impacts the way LTPP has been assessed. Fonagy et al (2015) found clinically significant group differences at two-year follow-up on non-symptom outcomes including functioning (GAF), a quality of life measure and the wellbeing subscale of the Clinical Outcome Routine Evaluation (CORE) measure. This trial was given no credit for measuring and reporting these important dimensions of experience. Moreover, as noted earlier in relation to end-of-treatment versus long-term follow-up outcomes, the Fonagy et al (2015) trial was downgraded for quality because "95% Confidence Interval (CI) crosses both line of no effect and threshold for clinically important benefit (SMD -0.5)". In other words, because there was no apparent benefit at the end of treatment, the trial was firstly misrepresented as showing no benefits at all and further misrepresented by being downgraded on quality for the same reason in spite of significant differences emerging at the two-year follow-up.

The GC's application of GRADE in this way systematically biases against trials where the intervention has a long-term effect rather than a short-term effect. This is worse than just ignoring long-term treatment effects – it is in fact rendering long-term efficacy undesirable. The reason why treatment effects may emerge after the end of treatment in trials of psychodynamic psychotherapy is that the aim of treatment is long lasting relational change rather than immediate but temporary symptom reduction. The way GRADE has been applied has systematically prejudiced against this kind of treatment from being recommended.

## Conclusions

This paper has outlined only a small number of the issues we consider problematic in the draft guideline document, focusing specifically on the impact on psychodynamic psychotherapies for complex and persistent forms of depression. There are many more

serious but generic concerns relating to statistical omissions, inadequate reporting and lack of transparency in methodological decisions. In particular, there are serious and unique risks associated with the network meta-analysis approach taken, over and above that of standard meta-analyses that need addressing (Keefe, 2015; Del Re et al, 2013). It appears that these have not been adequately resolved in the approach adopted in the draft guideline, leading to serious apprehension in accepting the treatment recommendations. In addition, there are concerns about GC choices made in many recommendations to weigh cost-effectiveness above clinical effectiveness and a reliance on statistical significance of effects (assessed in the GRADE system) which tends to create a paradox whereby small effects detected in well-powered studies are used to justify a recommendation, whereas much larger clinical effects can be detected in small samples. 'Underpowered' studies "contain valuable information when combined with others like them in a meta-analysis" (Hunter & Schmidt, 2004, p12). Therapist effects have also been ignored in the reviews in spite of clear and growing evidence for their impact. For example, Baldwin and Imel (2013) found that therapist effects account for approximately 5-8% of patient outcomes.

In spite of several additional issues touched on only briefly here, the concerns detailed and elaborated upon in this paper focus primarily on issues around lack of proper use of service user experience data; problematic categorisations of depression and treatments in respect of severity, complexity and chronicity; inaccurate inclusion of non bona-fide treatments or exclusion of relevant bona-fide treatment studies when collating evidence for psychodynamic therapies; issues around the focus on short-term symptom based outcomes to the neglect of crucial long-term follow-up data; and concerns about the system of assessing methodological quality. This is particularly troubling given that there are so few well-evidenced treatments available for these groups of people and thus the approach in the guideline further disadvantages this under-served group and mitigates against them being able to access interventions which may have the potential to help alleviate some of their suffering. Moreover, it disregards patient choice and the increasing preference for psychological treatments over medication by service users (van Schaik et al, 2004; McHugh et al, 2013). The brief eight-week stakeholder consultation process allowed many of these concerns to be aired but we would argue that it is unreliable and unsafe to rely on a one-off short consultation process to correct such serious mistakes and, arguably, the time limits imposed on the consultation and subsequent publication period are too tight for NICE to be

able to reconsider their approach. NICE and the ‘evidence based medicine’ paradigm within which it sits have been shaped by medical science and are now in need of a re-visioning which places patient choice, parity of esteem and equity of access at the centre.

## References


Abbass, A., Sheldon, A., Gyra, J., Kalpin, A. (2008). Intensive short-term dynamic psychotherapy of patients with DSM IV personality disorders: a randomized controlled trial. *Journal of Nervous and Mental Disease, 196*(3), 211-6.

Abbass, A., Town, J. & Driessen, E. (2011). The efficacy of short-term psychodynamic psychotherapy for depressive disorders with comorbid personality disorder. *Psychiatry, 74*(1), 58-71.

Anderson, C. & Roy, T. (2013). Patient experiences of taking antidepressants for depression: a secondary qualitative analysis. *Research in Social and Administrative Pharmacy, 9*(6),884-902.

American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–5*).

APA (2006). Evidence-Based practice in psychology. *American Psychologist, 61,* 271-285.

Baldwin, S.A., Imel, Z.E. (2013). Therapist effects: findings and methods. In M.J. Lambert (Ed), *Bergin and Garfield's handbook of psychotherapy and behavior change (6th ed.)* (pp.258-297). New York: Wiley.

Brown, A., Scales, U., Beever, W., Rickards, B., Rowley, K., O'Dea, K. (2012). Exploring the expression of depression and distress in aboriginal men in central Australia: a qualitative study. *BMC Psychiatry, 1*(12),97.

Bryant-Bedell, K., Waite, R. (2010). Understanding major depressive disorder among middle-aged African American men. *Journal of Advanced Nursing, 66*(9), 2050-60.

Care Act (2014). Available at: http://www.legislation.gov.uk/ukpga/2014/23/contents/enacted

Cohen, R. (2016). Getting into the Acceptance and Commitment Theory with psychoanalytic therapy: the case of "Daniel". *Pragmatic Case Studies in Psychotherapy, 12*(1:1), 1-30.

Corcoran, J., Brown, E., Davis, M., Pineda, M., Kadolph, J., Bell, H. (2013). Depression in older adults: a meta-synthesis. *Journal of Gerontological Social Work, 56*(6), 509-34.

Del Re, A.C., Spielmans, G.I., Flückiger, C. & Wampold, B.E. (2013). Efficacy of new generation antidepressants: Differences seem illusory. *PLoS ONE, 8*(6), e63509.

Dijkers, M. (2013). Introducing GRADE: a systematic approach to rating evidence in systematic reviews and to guideline development. *KT Update,* 1(5). Retrieved from www.ktdrr.org/products/update/v1n5.

Fonagy, P., Rost, F., Carlyle, J., McPherson, S., Thomas, R., Fearon, P., Goldberg, D., Taylor, D. (2015). Pragmatic randomized controlled trial of long-term psychoanalytic psychotherapy for treatment-resistant depression: the Tavistock Adult Depression Study (TADS). *World Psychiatry, 14* (3), 312-321.

Gask, L., Aseem, S., Waquas, A., Wahid, W. (2011). Isolation, feeling 'stuck' and loss of control: understanding persistence of depression in British Pakistani women. *Journal of Affective Disorders, 128*(1-2), 49-55.

Goodyer, I.M., Dubicka, B., Wilkinson, P., Kelvin, R., Roberts, C., Byford, S… Harrington, R. (2008). A randomised controlled trial of cognitive behaviour therapy in adolescents with major depression treated by selective serotonin reuptake inhibitors. The ADAPT trial. *Health Technology Assessment, 12*(14), iii-iv, ix-60.

Goodyer, I.M., Reynolds, S., Barrett, B., Byford, S., Dubicka, B., Hill, J… Fonagy, P. (2017). Cognitive behavioural therapy and short-term psychoanalytical psychotherapy versus a brief psychosocial intervention in adolescents with unipolar major depressive disorder (IMPACT): a multicentre, pragmatic, observer-blind, randomised controlled superiority trial. *Lancet Psychiatry,* 4(2), 109-119.

Goodyer, I.M., Tsancheva, S., Byford, S., Dubicka, B., Hill, J., Kelvin, R…Fonagy, P. (2011). Improving Mood with Psychoanalytic and Cognitive Therapies (IMPACT): A pragmatic effectiveness superiority trial to investigate whether specialised

psychological treatment reduces the risk for relapse in adolescents with moderate to severe unipolar depression: study protocol for a randomised controlled trial. *Trials, 12*(1), 175.

Health and Social Care Information Centre (2016). *Psychological Therapies: Annual Report on the use of IAPT services, England, 2015-16.*

Health Foundation (2017). *Healthy Lives for People in the UK*. Retrieved from www.health.org.uk/publication/healthy-lives-people-uk

Hepgul, N., King, S., Amarasinghe, M., Breen, G., Grant, N., Grey, N… Cleare, A.J.(2016). Clinical characteristics of patients assessed within an Improving Access to Psychological Therapies (IAPT) service: results from a naturalistic cohort study (Predicting Outcome Following Psychological Therapy; PROMPT). *BMC psychiatry, 16*(1), 52.

Hunter, J. & Schmidt, F. (2004).*Methods of Meta-analysis: correcting error and bias in research findings. Second edition*. Sage.

Jacobson, N.S. & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59,* 12-19.

Jobst, A., Brakemeier, E.L., Buchheim, A., Caspar, F., Cuijpers, P., Ebmeier, K.P… Padberg, F. (2016). European Psychiatric Association Guidance on psychotherapy in chronic depression across Europe. *European Psychiatry, 33*, 18–36.

Johnstone, L. & Boyle, M. with Cromby, J., Dillon, J., Harper, D., Kinderman, P., Longden, E., Pilgrim, D. & Read, J. (2018). *The Power Threat Meaning Framework: Towards the identification of patterns in emotional distress, unusual experiences and troubled or troubling behaviour, as an alternative to functional psychiatric diagnosis.* Leicester: British Psychological Society.

Keefe, R. (2015). Correspondence. Heightened risk of false positives in a network meta-analysis of social anxiety. *British Medical Journal* ,*2*(4),292–293.

Keller, M.B., McCullough, J.P., Klein, D.N., Arnow, B., Dunner, D.L., Gelenberg, A.J… Zajeka, J. (2000). A comparison of nefazodone, the cognitive behavioral-analysis system of psychotherapy, and their combination for the treatment of chronic depression. *New England Journal of Medicine, 342*(20),1462-70.

Kendler, K. (2016). The phenomenology of Major Depression and the representativeness and nature of DSM criteria. *American Journal of Psychiatry, 173* (8), 771-780.

Khan, N., Bower, P. & Rogers, A. (2007). Guided self-help in primary care mental health: meta-synthesis of qualitative studies of patient experience. *British Journal of Psychiatry, 191*, 206-11.

Kocsis, J.H., Gelenberg, A.J., Rothbaum, B.O., Klein, D.N., Trivedi, M.H., Manber, R.… Thase, M.E.(2009). Cognitive behavioral analysis system of psychotherapy and brief supportive psychotherapy for augmentation of antidepressant nonresponse in chronic depression: the REVAMP Trial. *Archives of General Psychiatry, 66*(11), 1178-88.

Körner, H., Newman, C., Limin Mao, Kidd, M.R., Saltman, D., Kippax, S.(2011). 'The black dog just came and sat on my face and built a kennel': Gay men making sense of 'depression'. *Health, 15*(4), 417-36.

Leichsenring, F., Abbass, A., Hilsenroth, M.J., Leweke, F., Luyten, P., Keefe, J.R.… Steinert, C. (2017). Biases in research: risk factors for non-replicability in psychotherapy and pharmacotherapy research. *Psychological Medicine, 47*(6), 1000-1011.

Liberman R & Eckman T (1981). Behavior Therapy vs Insight-Oriented Therapy for Repeated Suicide Attempters. Arch Gen Psychiatry,38(10), 1126-1130.

Maina, G., Forner, F., Bogetto, F. (2005). Randomized controlled trial comparing brief dynamic and supportive therapy with waiting list condition in minor depressive disorders. *Psychotherapy and Psychosomatics*, *74*,43-50.

McPherson, S., Cairns, P., Carlyle, J., Shapiro, D., Richardson, P. & Taylor, D. (2005). The effectiveness of psychological treatments for refractory depression: a systematic review. *Acta Psychiatrica Scandinavica, 111,* 331-340.

McPherson, S., Evans, C. & Richardson, P. (2009). The NICE Depression Guidelines and the recovery model: is there an evidence base for IAPT*? Journal of Mental Health, 18*(5), 405-414.

McCrone, P., Rost, F., Koeser, L., Koutoufa, I., Stephanou, S., Knapp, M., . . . Fonagy, P. (2017). The economic cost of treatment-resistant depression in patients referred to a specialist service. *Journal of Mental Health*. Available from doi:10.1080/09638237.2017.1417562.

McHugh, R.K., Whitton, S.W., Peckham, A.D., Welge, J.A. & Otto, M.W. (2013). Patient preference for psychological vs pharmacologic treatment of psychiatric disorders: a meta-analytic review. *Journal of Clinical Psychiatry*, *74*(6), 595-602.

Mental Health Foundation (n.d.). *Surviving or Thriving: the state of the UK's mental health*.

Noblit, G.W., Hare, R.D. (1988). *Meta-ethnography: synthesizing qualitative studies*. Newbury Park: Sage.

Noyes, J., Popay, J., Pearson, A., Hannes, K. & Booth, A. (2011). Qualitative research and Cochrane Reviews. In J. Higgins & S.Green (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions (version 5.1.0)* (chapter 20). The Cochrane Collaboration. Available from www.handbook.cochrane.org.

NICE (2009). Depression in adults: recognition and management.

NICE (2017). Depression in adults: treatment and management. Draft.

Oliffe, J.L., Han, C.S., Ogrodniczuk, J.S., Phillips, J.C., Roy, P. (2011). Suicide from the perspectives of older men who experience depression: a gender analysis. *American Journal of Mens Health, 5*(5), 444-54.

Oliffe, J.L., Ogrodniczuk, J.S., Bottorff, J.L., Johnson, J.L., Hoyak, K. (2012)."You feel like you can't live anymore": suicide from the perspectives of Canadian men who experience depression. *Social Science and Medicine,74*(4),506-14.

Oliffe, J.L., Rasmussen, B., Bottorff, J.L., Kelly, M.T., Galdas, P.M., Phinney, A., Ogrodniczuk, J.S. (2013). Masculinities, work, and retirement among older men who experience depression. *Qualitative Health Research, 23*(12),1626-37.

Patterson-Kane, L. & Quirk, F. (2014). Within the boundary fence: an investigation into the perceptions of men's experience of depression in rural and remote areas of Australia. *Australian Journal of Primiary Health, 20*(2), 162-6.

Popay, J., Roberts, H., Sowden, A., Petticrew, M., Arai, L., Rodgers, M... Duffy, S. (2006). *Guidance on the conduct of narrative synthesis in systematic reviews*. Available at www.lancaster.ac.uk/shm/research/nssr/research/dissemination/publications.php

Priestly, J. & McPherson, S. (2016). Experiences of Adults Providing Care to a Partner or Relative with Depression: A Meta-Ethnographic Synthesis. *Journal of Affective Disorders, 192*, 41–49.

Psychologists Against Austerity (n.d.). *The Psychological Impact of Austerity: A Briefing Paper*. http://www.psychchange.org/uploads/9/7/9/7/97971280/paa-briefing-paper.pdf

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*(8), 1038-53.

Roberts, S.L. & Sedley, B. (2016). Acceptance and Commitment Therapy with Older Adults: Rationale and case-study of 89-year-old with Depression and Generalized Anxiety Disorder. *Clinical Case Studies, 15*(1) 53–67.

Ruhe, H.G., van Rooijen, G., Spijker, J., Peeters, F.P., Schene, A.H. (2012). Staging methods for treatment resistant depression. A systematic review. *Journal of Affective Disorders, 137*, 35–45.

Sandelowski, M., & Barroso, J. (2003). Creating metasummaries of qualitative findings. *Nursing research*, *52*(4), 226-233.

Smith, J.A., Rhodes, J.E. (2015). Being depleted and being shaken: An interpretative phenomenological analysis of the experiential features of a first episode of depression. *Psychology and Psychotherapy, 88*(2),197-209.

Souza,L., Salum, G., Mosqueiro, B., Caldieraro, M., Guerra, T., Fleck, M.P. (2016). Interpersonal psychotherapy as add-on for treatment-resistant depression: A pragmatic randomized controlled trial, *Journal of Affective Disorders*, *193*, 373-380.

Svartberg, M., Stiles, T.C., Seltzer, M.H (2004). Randomized, controlled trial of the effectiveness of short-term dynamic psychotherapy and cognitive therapy for cluster C personality disorders. *American Journal of Psychiatry, 161*(5),810-807.

Thyme, K. E., Sundin, E. C., Stahlberg, G., Lindstrom, B., Eklof, H., & Wiberg, B. (2007). The outcome of short-term psychodynamic art therapy compared to short-term psychodynamic verbal therapy for depressed women. *Psychoanalytic Psychotherapy, 21*, 250-264.

Henderson, C., Robinson, E., Evans-Lacko, S., Corker, E., Rebollo-Mesa, I., Rose, D., Thornicroft G. (2016) Public knowledge, attitudes, social distance and reported contact regarding people with mental illness 2009–2015. *Acta Psychiatrica Scandinavica, 134* (S446),23–33.

Town, J.M., Abbass, A., Stride, C., Bernier, D. (2017). A randomised controlled trial of Intensive Short-Term Dynamic Psychotherapy for treatment resistant depression: the Halifax Depression Study. *Journal of Affective Disorders, 214*, 15-25.

van Grieken, R.A., Beune, E.J., Kirkenier, A.C., Koeter, M.W., van Zwieten, M.C., Schene, A.H. (2014).Patients′ perspectives on how treatment can impede their recovery from depression. *Journal of Affective Disorders, 167*, 153-9.

van Schaik, D.J., Klijn, A.F., van Hout, H.P., van Marwijk, H.W., Beekman, A.T., de Haan, M. & van Dyck, R. (2004). Patients' preferences in the treatment of depressive disorder in primary care. *General Hospital Psychiatry*, 26(3),184-9.

Vinnars, B., Barber, J., Noren, K., Gallop, R., Weinryb, R.(2005). Manualized supportive-expressive therapy versus nonmanualized community-delivered psychodynamic therapy for patients with personality disorders: bridging efficacy and effectiveness. *American Journal of Psychiatry, 162*(10), 1933-40.

Wampold, B. E., Fluckiger, C., Del Re, A.C., Yulish, N.E., Frost, N.D., Pace, B.T. …Hilsenroth, M.J.(2017). In pursuit of truth: A critical examination of meta-analyses of cognitive behavior therapy. *Psychotherapy Research, 27*(1), 4-32.

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

Zimmermann, T.M., Clouth, J., Elosge, M., Heurich, M., Schneider, E., Wilhelm, S., Wolfrath, A.(2013). Patient preferences for outcomes of depression treatment in Germany: a choice-based conjoint analysis study. *Journal of Affective Disorders, 148*(2-3),210-9.