Treating Treatment Non-Responders: A Meta-Analysis of Randomized Controlled Psychotherapy Trials

Andrew T. Gloster[1], Marcia T. B. Rinner[1], Myria Ioannou[2], Jeanette Villanueva[1], Victoria J. Firsching[1], Giovanna Ferrari[3], Charles Benoy[3], Klaus Bader[3], Maria Karekla[2]

[1]University of Basel, Department of Psychology, Division of Clinical Psychology & Intervention Science

[2]University of Cyprus, Department of Psychology

[3]Psychiatric Hospital of the University of Basel, Psychiatric Hospital, Center for Psychosomatics and Psychotherapy, Switzerland

Correspondence author:
Andrew T. Gloster
University of Basel
Department of Psychology
Division of Clinical Psychology and Intervention Science
Missionsstrasse 62A
CH-4055 Basel, Switzerland
Email: andrew.gloster@unibas.ch

Treating Treatment Non-Responders: A Meta-Analysis of Randomized Controlled Psychotherapy Trials

Size of the Problem

The development of treatments for mental disorders is largely a success story. Increasingly sophisticated empirical studies have led to unquestionable advances in treatments. Today, empirically supported treatments are available for numerous disorders (Ballenger et al., 2001; Division 12 of the American Psychological Association, 2016).

Despite these advances, treatment non-response remains an important, yet often-ignored issue (Sage, 2004; Taylor, Abramowitz, & McKay, 2012). Even the best psychotherapy treatments available today have more residual problems than generally acknowledged. Estimated rates of treatment non-response vary across studies due in part to methodological differences (e.g., definition of non-response, research design, assessment time points, treatment types, disorders etc.). Nevertheless, several studies suggest that the proportion of non-responders ranges between 30% and 40% (Clarke, Kingston, James, Bolderston, & Remington, 2014; M. Fava & Davidson, 1996; Souery, Papakostas, & Trivedi, 2006; Taylor et al., 2012; Westen & Morrison, 2001). These rates appear relatively consistent across mood and anxiety disorders. For example, in depressive disorders it has been estimated that 19%-34% of cases are non-respondent (M. Fava & Davidson, 1996), with even high rates of patients that fail to achieve remission following an adequate antidepressant treatment (M. Fava, 2003). Similar rates are estimated in anxiety disorders, where, despite the availability of highly efficacious treatments, approximately one-third of patients are judged to not respond significantly to psychotherapy. These patterns of non-response have been observed across the anxiety disorders including Posttraumatic Stress Disorder (PTSD; Bradley, Greene, Russ, Dura,

& Westen, 2005), panic disorder and agoraphobia (Gloster et al., 2013, 2011; Levitt & Karekla, 2005); generalized anxiety disorder (Westen & Morrison, 2001) and Obsessive Compulsive Disorder (OCD; Eddy, Dutra, Bradley, & Westen, 2004).

Implications of treatment non-response are profound for those patients who do not respond to state-of-the-art treatments. In addition to continued suffering associated with the disorder itself, non-responding patients have decreased levels of quality of life, and increased mortality and suicide rates (Greden, 2001; Schlaepfer et al., 2012; Taylor et al., 2012). For example, treatment non-responding anxiety patients experience a disproportionate burden of illness and have the highest rate of suicidal attempts than any other disorders (Bystritsky, 2006).

Systematizing knowledge about treatment non-response is complicated by overlapping and sometimes indiscriminately applied terminology (Berlim & Turecki, 2007). Nevertheless, general characteristics can be rendered with respect to how each term interprets the role of the treatment and/or the patient. *Treatment non-response* suggests failure of a primary treatment to achieve its treatment goals (usually symptom reduction). The treatment, its stipulated goals or outcomes, and their measurement vary greatly across studies. Conceptually, the focus of this term can be on both the treatment that failed to help the patient, and/or the patient who failed to respond to the given treatment. *Treatment-resistant*, in contrast, suggests the same phenomenon, but puts the focus solely on the patient instead of the treatment. That is, for various reasons the patient does not respond to one or more viable treatments. Other terms sometimes used to describe variants of these issues include *chronic*, *partial response*, and *residual symptoms*. The term "chronic" (e.g., chronic depression) once again places the focus more on the patient than the treatment, with the suggestion that the symptoms have been present for a substantial amount of time and that the treatment goal (usually symptom

reduction) is more difficult to attain than non-chronic variants of the same disorder. "Partial response" and "residual symptoms" are terms used in studies to indicate that the goals (outcomes) have only been partially obtained. To date, there is no consensus as to which term should be used in which situation.

Likewise, there is no consensus about what exactly constitutes treatment "response" vs. "non-response". Some broad suggestions have been proposed, such as when a treatment fails to initiate change towards a partial remission (Frank et al., 1991) or when "there are residual symptoms or when symptoms do not improve at all after some form of therapeutic intervention" (Pollack et al., 2008, p. 468). In reality, however, each research study and likely each therapist make idiosyncratic decisions. Furthermore, very few conceptualizations of treatment non-response include explicit discussions of well-being or functioning. This is important because well-being and functioning do not mean the absence of symptoms, but rather are at least partially independent (World Health Organization [WHO], 1948, 2014) and therefore needs to be measured and tested.

Current Recommendations for Treatment Non-response

Despite these definitional difficulties, empirical studies have tested the efficacy of introducing a second treatment option for patients not responding to a previous treatment. Based on this work, some treatment guidelines (mostly pharmacological in nature) exist for some disorders, in particular mood disorders and anxiety disorders. These treatment guidelines generally suggest that after ruling out misdiagnosis, either the dosage should be increased or the patient should switch to a different class of medication (Bandelow et al., 2008). Some guidelines (Anderson et al., 2008; Bandelow, Lichte, Rudolf, Wiltink, & Beutel, 2015; Kennedy et al., 2016) also suggest adding psychotherapy, although this seems to be based either on anecdotal evidence or on the

psychotherapy literature that shows psychotherapies (with or without pharmacology) are efficacious treatments for mental disorders (i.e., as a first treatment, not explicitly as a response to non-response). The empirical basis for using psychotherapy as a response option for treatment non-response remains inadequately explored.

Purely psychological recommendations on how to treat treatment non-responders have also begun to be explored (McKay, Abramowitz, & Taylor, 2010; Taylor et al., 2012). Here, too, it has been suggested to rule out misdiagnosis as a first step and then to determine whether the treatment was implemented correctly (Taylor et al., 2012). This last step might be considered analogous to increasing medication dosage. Thereafter, it is suggested that practitioners consult their peers, consider additional theoretical variables to target or switch to another option, or refer out to another colleague (Taylor et al., 2012). Whereas these are reasonable and logical suggestions, the empirical basis remains inadequately investigated. To our knowledge, no meta-analysis to date has examined the efficacy of using psychotherapy to treat treatment non-responders.

Aim, Objectives, and Need for this Review

Based on estimates that approximately one-third of patients do not adequately respond to treatment and the paucity of empirical information regarding the efficacy of using psychotherapy as a means of treating these patients, systematic information about this sizable group is urgently needed. Given the lack of research on psychotherapy alternatives to treatment non-response, we conducted this study with the following aims: First, we aimed to document the proportion of published studies on treatment non-responders that examined psychotherapeutic and pharmacological options. Second, we aimed to determine the efficacy of adding psychotherapy for patients who were non-responsive to a previous treatment. For this, we examined only studies that utilized

randomized controlled trials (RCTs). We examined this for patients diagnosed with mood and anxiety disorders because these are the most prevalent group of disorders. In addition to overall efficacy, we aimed to document whether certain types of psychotherapy (e.g., cognitive behavioral therapy, psychodynamic therapy, mindfulness and acceptance therapies) were more efficacious than others for treatment non-responding patients. As such, our overarching goal was to determine whether psychotherapy is a viable option to treatment non-response, identify any promising signals, while simultaneously identifying further research needs. The information is important for treatment providers and ultimately to the patients who continue suffering even if they received the best available treatment.

METHODS

Protocol and Registration

We followed the PRISMA guidelines for reporting (Moher, Liberati, Tetzlaff, & Altman, 2009). The methods of the review were specified in advance and the meta-analysis was pre-registered with PROSPERO[1] (Nr.: CRD42018094064).

Eligibility Criteria

Types of study designs. Only studies published in English were considered up to January 2019. No publication status was imposed. Furthermore, we only included RCTs studying the treatment alternatives for patients who did not respond to at least one preliminary treatment. In order to determine the relative number of studies that added psychotherapy vs. pharmacology, we first documented studies testing any intervention for treatment non-response (i.e., pharmacology, psychotherapy, etc.). Thereafter, we limited studies only to those in which patients were randomized to a psychotherapy for the purpose of treating non-response.

---

[1] http://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42018094064

Types of participants. Patients diagnosed with any mood or anxiety disorder of any age were considered provided they were randomized to one or more alternative treatments following an unsatisfactory preliminary treatment. Given the lack of consensus on what constitutes treatment non-response, we chose to use the definitions of the respective researchers. To be included in this meta-analysis, however, study participants had to have been previously in treatment for their disorder and still meeting diagnostic criteria after this or those previous treatment(s). One study explicitly indicated that they utilized a transdiagnostic sample of mixed mood, anxiety, and personality disorders. The other studies concentrated on a single disorder, but included patients with co-morbid disorders.

Types of intervention. Studies were included that examined psychotherapy, either in isolation or in combination with medication, provided the psychotherapy element that was added was not part of the preliminary unsuccessful treatment. We understood treatment non-response as being the focal point of a study when patients were treated for a condition for which they were previously treated, but where the previous treatment did not achieve its treatment goals (e.g., reduction of symptomatology, still meeting diagnostic criteria, etc.). Studies that only examined pharmacological responses to treatment non-response were examined for preliminary analysis, but were subsequently excluded because we were interested in the effects of adding psychotherapy. We did not limit the type of psychotherapy.

Types of outcome measure. Studies were included that reported at least one validated instrument to measure symptom reduction (primary outcome) of the specific population tested at both baseline and post. Primary outcome measures were not limited *a priori* due to the heterogeneity of outcomes across disorders. However, to increase the comparability between the studies, when multiple primary outcomes were reported for

the same disorder we chose the primary outcome that was the most frequently used instrument within the studies included in the meta-analysis. Information about further instruments measuring broader outcomes of well-being/ functioning/ quality of life at pre and post (secondary outcome) were also included in the meta-analysis, but studies that had not reported secondary outcomes were not excluded.

Types of comparisons. Studies were included that compared the primary and/or secondary outcomes of a psychotherapy group (i.e., those receiving psychotherapy in response to treatment non-response) to a control group. Studies reported two types of control groups: 1) no treatment at all (e.g., wait list control group, WL); and 2) treatment as usual (TAU). Expectedly, TAU varied across studies and included a) individuals who continued with their current medication treatment, but without starting a new medication); b) self-designed control programs (e.g., health-enhancement program omitting mindfulness exercises); c) another controlled psychotherapy (e.g., stress reduction); and d) one study that used CBT as the TAU.

Exclusion Criteria

We excluded studies that did not assess non-responder patients or where our definition of non-responders was not fully met. Studies that treated chronic cases who had not *previously* been treated and/or studies that did not give information about the previous treatment(s) were excluded. Also excluded were studies treating medical conditions (e.g., cancer). Further, we excluded uncontrolled trials, case studies, and studies where randomization was not clearly implemented. Studies were also excluded if they only utilized pharmacological treatments or reported insufficient data to conduct the analysis and we were unable to obtain those data.

Information Sources

Search. To identify possible studies for the meta-analysis we did a systematic search on the electronic databases Ovid, PsycARTICLES, PsycCRITIQUES and PsycINFO up to January 2019. To extend our search, we included various terms such as resistance, refractory, chronic, recurrent, non-response with depression, dysthymia, antidepressant, anxiety, PTSD, post-traumatic stress disorder, social phobia, agoraphobia, panic disorder, obsessive-compulsive disorder, generalized anxiety disorder, mania, and bipolar (see Appendix A for a detailed description). We also used specific quotation at the end of the search terms to enable all possible combinations of word endings (e.g., non-responder, non-respond, non-response, etc.). Furthermore, we specified our search to randomized controlled trials. Finally, additional hand searched literature was identified by searching the references list of identified studies.

Study selection. The study team extensively discussed the concepts involved in treatment non-response before extracting. A range of different definitions of non-response was used by the studies (see Table 1 for a detailed description of the non-response definition of the included studies). First, we screened the title and abstracts of the papers found within the electronic search and checked for potential match with inclusion criteria. Two team members conducted this process separately. A senior third team member rated those with discrepancies independently. Secondly, the remaining studies were examined for type of treatment (e.g. psychotherapy vs. pharmacology) used to treat non-responders. Two raters of the team conducted this process independently, and a senior third team member rated discrepancies independently. Thirdly, two raters of the team independently examined the full text of each paper to determine whether it would be included in the meta-analysis and a senior third team member rated discrepancies independently. See the flow diagram (Figure 1) for more details.

Data Extraction and Management

Data collection process. In a last step, data were extracted by two independent raters and checked with a senior rater. Data were extracted into Excel.

Data extraction sheet. The following information was extracted from each included trial: Paper registration information (year of publication, title, authors, ethics, and funding), demographics (description of patients' disorders, gender, age, marital status, employment status, and type of study population), primary outcome, secondary outcome, description of treatment (length, type, setting, and format), description of the control group, assessor qualifications, therapist qualifications, description of study procedure, and description of dropouts. The primary and secondary outcomes were extracted for pre-treatment, post-treatment, and follow-up (last reported time point).

Outcome measures. The following primary measures were extracted from each included trial: Depression symptomatology was measured using the Beck Depression Inventory (BDI-II Beck, Steer, & Brown, 1996; BDI-I; Beck, Ward, Mendelson, Mock, & Erbaugh, 1961), Hamilton Depression Severity Rating Scales (HAM-D 17 & 24 items; Hamilton, 1967; Furukawa, Akechi, Ozaki, et al., 2003) and the depression subscale of the Symptom Checklist (SCL-20; Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 1974). Bipolar disorder was measured with the Young Mania Rating Scale (YMRS; Young, Biggs, Ziegler, & Meyer, 1978). Anxiety symptoms were measured using the Clinician-Administered PTSD Scale (CAPS; Weathers, Keane, & Davidson, 2001 and CAPS-1; Blake et al., 1990), Panic Agoraphobia Scale (PAS; Bandelow, 1997), and the Global Severity Index (Derogatis & Spencer, 1982). For the secondary outcomes measuring quality of life/ functioning, we extracted data from following measures: the World Health Organization Quality of Life (WHOQOL; Skevington, Lotfy, & O'Connell, 2004), and WHOQOL-brief, specifically from the psychological

subscale, (Fleck et al., 2000), Global Assessment of Functioning (GAF; Hilsenroth et al., 2000), the Clinical Global Impression Scale (CGI; Guy, 1976a), the Longitudinal Interval Follow-up Evaluation Range of Impaired Functioning Tool (LIFE-RIFT, specifically from the satisfaction subscale (Leon et al., 1999)), the Patient-Rated Global Improvement (PGI; Guy, 1976b), the Short Form Health Survey 12 (SF-12; Ware, Kosinski, & Keller, 1996), and the physical component summary score of the 36-Item Short-Form Health Survey (SF-36; Brazier et al., 1992; Fukuhara, Ware, Kosinski, Wada, & Gandek, 1998).

Risk of bias in individual studies. Working independently, two reviewers assessed the quality and risk of bias of each included study. The quality assessment criteria were adapted from the available quality checklists for RCTs, the Centre for Reviews and Dissemination checklist and the CONSORT checklist (Schulz, Altman, Moher, & CONSORT Group, 2010). Following the extraction of data necessary for quality assessment, two senior team members separately assigned a quality score (0-2) based on adequacy of their consideration of each quality criterion, in order for a summative quantitative and comparable quality score to be available for each study. The inter-rater agreement was very high ($r = .95$). Discrepancies between the two-team members were discussed until an agreement was reached. Variables extracted for the quality assessment can be viewed in Appendix B.

Statistical Analyses

The Comprehensive Meta-Analysis software v3.0 (Borenstein, Hedges, Higgins, & Rothstein, 2014) was used for the estimation of the mean effect size. The number of participants at baseline and the mean scores for each group before and after the treatment phase were inserted in the program. For the studies reporting the number of

participants who completed the treatment, an additional mean effect size was computed using the sample size of completers.

Calculation of overall effect sizes and within-group effect size. The standardized mean difference (SMD) was used as the effect size, which considers that the main outcomes were assessed using different measurement tools. The standardized mean difference is calculated by the following equation:

SMD= Difference in mean outcome between groups / Change of SD of outcome between groups

Within-group effect size referred to the SMD of each group, based on the subgroup variable examined each time.

Calculation of statistics for between-group comparisons. The total effect size based on the standardized mean difference was compared between groups. The comparison was made using the Q statistic, which computes a p-value. The between-group differences in effect size were determined based on the corrected p-value and specifically whether this value was lower than or equal to the overall significance level of $p < .05$. Mixed-effects models were used to calculate subgroup analyses, which aimed to detect the predictive ability of factors in estimating the effect size (Cohen's d).

Subgroup analyses. Subgroup analyses were conducted to test whether some variables moderate the outcomes. Subgroup differences were evaluated first in terms of between-group difference, as tested by the Q statistic. In cases where between-group differences (based on the Q statistic) were not significant, the next level of examination concerned the evaluation of the within-subgroups effect (based on SMD/standardized mean difference). When all the within-subgroup effects where in the same direction (were all significant), the level of examination emphasized the quantitative differences between subgroups effectiveness, which was determined by the non-overlapping

confidence intervals of the SMD in each subgroup. These subgroup analyses were based on the following variables: disorder, treatment setting, type of comparison group, primary outcome measure, and type of original treatment. The variables to be tested in subgroup analyses were decided a priori and included in the published PROSPERO protocol of the present meta-analysis. However, when the total number of studies per subgroup category was small, conducting analyses can lead to spurious results. Indeed, the number of studies in each subgroup category should be ≥ 5 in order to have adequate statistical power (Borenstein & Higgins, 2013; Sedgwick, 2013). In such cases it is suggested to explore heterogeneity statistics in each subgroup analysis and to adjust the interpretation of the findings based on the sources of heterogeneity Bigger, (2003). We decided to conduct subgroup analyses only with the variables that had at least one category with more than five studies, and to interpret the findings only for those categories. Also, we explored heterogeneity and interpreted the findings of the categories with low heterogeneity, consistent with recommendations for Cochrane reviews (Donegan, Williams, Dias, Tudur-Smith, & Welton, 2015).

Meta-regression & sensitivity analyses. Meta-regression analyses were performed to examine the predictive role of continuous variables that could explain the effect size, including percentage of female participants, length of original treatment to which the participants did not respond, and treatment duration. Sensitivity analyses were conducted to test whether decisions made during the meta-analysis impacted the results. These sensitivity analyses were used in two ways. First, in order to determine whether dropouts affected the estimated effect size, we performed sensitivity analyses for participants who were recruited into the study and for participants who completed treatment. Second, sensitivity analyses were applied to technical and grouping decisions made during the meta-analysis (e.g. data format reported by the studies).

Testing homogeneity. The fixed-effects model was computed first in order to examine the level of heterogeneity between the studies. The existence of significant heterogeneity was determined using the Q statistic and the level of heterogeneity (small, moderate, large) was estimated using the I_2 statistic. Based on the Q statistic, significant heterogeneity is determined by examining the p value ($p<.05$). The I_2 statistic reflects the percentage of variation existing between the studies that cannot be explained by chance and is due to heterogeneity (Higgins & Thompson, 2002; Higgins, Thompson, Deeks, & Altman, 2003), thus a small percentage ($<25\%$) is preferable.

Risk of bias across studies. Publication bias analysis was computed to investigate the existence of asymmetries on the funnel plot that would indicate a bias in publishing studies favoring positive results in treatment groups.

RESULTS

Study Selection

During the first step, a total of 924 abstract and titles were examined. Of those, 905 where identified through the electronic data search and 19 through hand-search. 497 studies did not meet our inclusion criteria and were excluded during this first step. The remaining 427 studies were examined for type of treatment (e.g. psychotherapy vs. pharmacology) that was used to treat non-responder patients. During this second step, 26 studies were excluded leaving a total of 401 studies. These 401 studies were included for the preliminary analysis (i.e., ratio of psychotherapy vs. pharmacology studies). For further analyses, only RCTs of psychotherapy were of interest, leaving 79 studies for the full-text review. During this final review a total of 61 studies had to be excluded, leaving a total of 18 studies for the final meta-analysis.

Study Characteristics

A review of the literature showed that 80% of published trials on treatment non-responders used exclusively pharmacological/medical interventions whereas only 20% examined psychotherapy as an option for treating treatment non-response.

The studies retained in the meta-analysis included clinical trials that examined the following types of psychotherapy: Cognitive Therapy and Cognitive Behavioral Therapy (CBT) Cognitive Behavioral Analysis System of Psychotherapy (CBASP), Behavioral Therapy (BT), third-wave therapies (i.e., mindfulness based cognitive therapy (MBCT), Acceptance and Commitment Therapy (ACT), and Dialectical Behavioral Therapy (DBT)), and other therapies (i.e. Short-Term Dynamic Psychotherapy (ISTDP), Psychodynamic therapies long-term psychoanalytic psychotherapy (LTPP), Interpersonal Psychotherapy (IPT) and Care management. Some studies included pharmacological agents, but were included here only if the psychotherapy was a new element following the randomization.

In total, 18 RCTs were included that tested the efficaciousness of psychotherapy in treating treatment non-response. These studies included data from n = 1734 participants at pre-treatment (n = 939 in active treatment groups; n = 795 in control groups).

Outcome on Primary and Secondary Outcome Measures

Every study included in this meta-analysis measured symptomatology. Studies that examined depression utilized primarily the BDI or the Hamilton rating scales and one study used the depression subscale of the SCL to measure depression symptoms. One study measuring bipolar disorder measured it with the YMRS. Measurement of anxiety symptoms varied as a function of each anxiety disorder. Unfortunately, only 8 of 18 studies measured quality of life/ functioning. When it was measured, a wide range of instruments as listed in the assessment section. With respect to follow-up

assessments, only seven studies reported the means and SDs of primary outcomes and only five for secondary outcomes.

Risk of Bias Within Studies

The judged quality of the studies varied widely, with overall quality scores ranging between 0.9 and 1.9 (see Table 2). High variability could also be seen between studies for each individual criterion on the quality checklist (Appendix B). 11 studies clearly stated study aims and hypotheses, however, six studies did not clearly state their study aims, and one study did not report them at all. Most studies adequately explained and described the study interventions. However, for two studies, details about the intervention, its duration, dosage, and protocol of the intervention were poorly reported and inadequately explained. Further, only eight studies gave detailed information about patients' previous treatments. In five studies the protocol's existence and adherence was not reported, and for three studies the adherence of intervention providers was limited or unclear. Most studies (all but three) applied intent-to-treat analysis. Finally, three studies did not report blinding procedures. See Table 2 and Appendix B for more details.

Synthesis of Results

Treatment effect. First, we computed the model for participants who filled out the outcome measures (i.e., based on the sample size at baseline). The analyses included participants who dropped out before the end of treatment. We chose to use "intent to treat" samples because they are generally more conservative than "completer" analyses. Significant heterogeneity was detected among the eighteen studies, as shown by the $Q(17)=90.679$, $p<.001$, which reached a high level with $I_2= 81.253\%$. Therefore, due to the heterogeneity detected and that the random error differed from study to study, the results from the random-effects models were considered more reliable. Overall, a positive

moderate to large impact of psychotherapy treatment groups targeting non-response was found, with *SMD*=0.818; 95% *CI* 0.556, 1.081; *p*<0.001; Figure 2).

The publication bias analysis showed a significant publication bias (Egger's test *t*(16)=3.179, *p*=.006, Kendall's *S statistic*= 89, tau=0.575, *p*<.001) favoring studies with positive effect sizes in treatment groups (Figure 3). After adjusting for potential publication bias using the Duval and Tweedie's trim and fill, the estimate for the standardized mean effect size would remain significant even when assuming that seven studies were missing from the literature (*SMD*=0.452; 95% *CI* 0.157, 0.748).

Subgroup Analyses for Symptom Reduction

Since heterogeneity statistics supported high significant heterogeneity we performed subgroup analyses to explore the sources of heterogeneity (Table 3). The results discussed below concern the between-group differences only for the variables with at least one category with more than five studies, the within-group effects for categories with a larger number of studies (≥ 10 studies), and within-group effects for the categories with low heterogeneity.

Do the results differ by disorder? Subgroup analyses showed no between-group significant differences in effect size based on the overall disorder category (i.e., anxiety or mood disorders; *Q*(2)=2.319, *p*=.31). Within-group examination showed that studies targeting populations with mood disorders (14 studies; *SMD*=0.760; 95% *CI* 0.484, 1.037; *p*<.001) were effective with a moderate to large effect size. Subgroup analyses showed non-significant between-group differences between specific disorders (i.e., MDD, other disorders), with *Q*(1)=3.129, *p*=.077. The category of MDD studies had only moderate levels of heterogeneity and supported moderate treatment effect size for patients with MDD (13 studies; *SMD*=0.535; 95% *CI* 0.378, 0.691; *p*<.001).

Do the results differ based on previous treatment type? The type of previous treatment (medication, psychotherapy, combined therapy) did not result in significant between-group differences in terms of effect size $Q(2)=2.079$, $p=.35$. Heterogeneity remained high in the group of studies which received medication as previous treatment, though the effect size for those trials was high and significant using the random-effects model (13 studies; $SMD=0.800$; 95% CI 0.489, 1.110; $p<.001$).

Do the results differ by type of new treatment? Examination of the subgroup analyses based on the type of new treatment offered to the treatment non-responders showed that the overall between-group differences were not significant ($Q(3)=3.840$, $p=.279$) and all categories of studies with different types of treatment produced significant within-group effects. Heterogeneity was low only for the studies implementing a third-wave treatment and for those implementing a treatment other than combined, CBT or third-wave (i.e., psychodynamic, IPT). The treatment effect was moderate and significant for both categories. Examination of the larger category in this subgroup analysis, which was the combined treatment (7 studies) supported very large effect sizes ($SMD=1.126$; 95% CI 0.581, 1.670; $p<.001$), but the studies were highly heterogeneous.

Do the results differ by comparison group? We performed subgroup analyses on comparison groups in three different forms. Examining the treatment approach followed by the comparison group (WLC, TAU, or HEP-Pharmacological) did not produce significant between-group differences, with $Q(2)=3.504$, $p=.17$. The difference between comparison groups that were rated as passive or active also had non-significant between-group differences, with $Q(1)=0.204$, $p=.65$.

Do the results differ by outcome measure? Between-group differences based on the instrument used to measure the primary outcome were not significant

(Q(3)=4.740, p= .192). An examination of the within-group effects showed that the effect size was moderate and significant when the primary outcome measure was the BDI, which was used by most studies and the group of those studies was largely homogeneous (8 studies; SMD=.485, 95% CI 0.355, 0.615, p<.001).

Results for Symptom Reduction by Sample Characteristics

The meta-regression analyses failed to find significant predictors of the overall effect size on primary outcomes for any of the tested variables: patients' mean age (b=0.005, 95% CI=− 0.006, 0.072, z=0.16, p=0.87), percentage of females (b=− 0.013, 95% CI−0.035, 0.008, z=− 1.22, p= 0.22), percentage of married participants (b=-0.002, 95% CI−0.013, 0.009, z=-0.32, p=0.75), percentage of employed participants (b=−0.001, 95% CI−0.012, 0.010, z=−0.21, p=0.83), length of comparison group (b=-0.005, 95% CI−0.023, 0.013, z=-0.52, p=0.60) and length of new treatment (b=-0.005, 95% CI−0.023, 0.013, z=−0.52, p=0.60). Examination of the studies' quality scores showed that this variable was a significant predictor of the overall effect size (b=− 1.631, 95% CI −2.749, -0.514, z=−2.86, p=.004), such that the higher the quality score the lower the effect size. However, only a very small proportion of 3% of total between-study variance could be explained by the model and significant heterogeneity remained, with Q(16)= 84.09, p< .001, I_2= 80.97%, τ_2= .210. Length of the original treatment was also a significant moderator of the effect size (b=0.078, 95% CI 0.048, 0.108, z=5.12, p<.001), suggesting that the longer the previous non-responsive treatment was, the larger the effect of the new treatment. Importantly, this result was based on only the seven studies that included this information. Within these studies, a proportion of 51% of total between-study variance could be explained by the model, but significant heterogeneity remained even after accounting for this variable, with Q(6)= 28.74, p< .001, I_2= 79.12%, τ_2= .169.

Sensitivity Analysis: Are the Effects on Symptom Reduction Impacted by

Decisions in Earlier Stages of the Review?

Sensitivity analyses were conducted to examine whether the results were robust against decisions made during the coding and analysis. For example, one study (Town, Abbass, Stride, & Bernier, 2017) did not provide means and SDs for each group, which necessitated a different computation than the other studies. When this study was excluded from the main analysis, the random-effects model still produced a significant treatment effect size (*SMD* (17)=.830; 95% *CI* 0.552, 1.108; *p*<.001). Sensitivity analyses were also conducted to examine the impact of including a study with bipolar disorder (i.e., Isasi et al., 2010). The mean effect size when excluding this study was not significantly different, with the model still producing a significant treatment effect size *SMD*=0.634; 95% *CI* 0.449, 0.818; *p*<0.001. This supported the decision to include this study, as it did not significantly alter the results. The inclusion of a study with transdiagnostic population (Clarke et al., 2014) was also examined under the context of sensitivity analyses, supporting again a non-significantly different mean effect size with *SMD*=0.850; 95% *CI* 0.575, 1.124; *p*<0.001. When all those three studies were excluded from the analysis (n= 15), the results were consistently not significantly different, with *SMD*=0.648; 95% *CI* 0.444, 0.853; *p*<0.001. Importantly, heterogeneity among the studies remained significant and high, even after excluding those three studies, with $Q(14)= 39.412$, p< .001, $I_2= 64.478$. Thus, the inclusion of these studies was not enough to explain the high heterogeneity among the studies. Exploration of heterogeneity through the subgroup analyses was discussed above and is presented on Table 3.

Also, computation of the overall model including only the studies that reported the number of participants who completed treatment (n=16) showed that the overall SMD was increased. Following the random-effects model due to the high heterogeneity

between the studies ($Q(15)=92.646$, $p<.001$, $I_2=83.809\%$), the pre-post treatment effect size was large, with $SMD=0.920$ (95% CI 0.607, 1.233, $p<.001$; Figure 4). The difference between the model with the participants who began treatment ("intent to treat") and the "completers" was not significant, with $p>.05$. Publication bias remained significant, and the number of studies needed to be trimmed was maintained at seven. Examination of the differences between subgroups of studies showed no qualitative changes compared to the analyses with the participants who were enrolled in treatment ("intent to treat"). All between and within-group effects from the subgroup analyses and the meta-regression analyses were retained.

Effects of Psychotherapy at Follow-Up

Analyses based on follow-up assessment showed that significant heterogeneity existed between the seven studies that included data on follow-up for symptom reduction ($Q(7)=78.724$, $p<.001$), which was high ($I_2= 92.378\%$.) This suggested the need to adopt the random-effects model. This model resulted in a significant effect favoring treatments targeting non-responders ($SMD=1.189$; 95% CI 0.503, 1.875, $p<.001$; see Figure 5). The publication bias analysis showed non-significant bias, with Egger's test $t(5)=2.175$, $p=.08$, as well as with Kendall's S statistic ($S=13$, $tau=0.571$, $p=.07$). The Duval and Tweedie's trim and fill showed that 3 studies were theoretically missing from the literature. When adjusting for the hypothetically missing studies, however, the main effect was no longer significant, with $SMD=.359$; 95% CI -.349, 1.068, $p>.05$.

Subgroup Analyses for Follow-Up

Given the small number of studies that included follow-up information, none of the categories included more than five studies, so subgroup analyses were not performed for follow-up effectiveness.

Follow-Up Results on Symptom Reduction by Sample Characteristics

Meta-regression analyses showed that the length of follow-up assessment did not moderate the results (b= -0.005, 95% *CI* −0.062, 0.052, z=-0.16, p=0.87). The meta-regressions analyses showed that the findings were consistent with the ones for pre-post change. Quality score continued to be a significant predictor of the effect size (b=−7.978, 95% *CI* −11.629, -4.328, z=−4.28, p<.001), such that that the higher the quality score the lower the effect size. However, a minimal proportion (0.1%) of total between-study variance could be explained by the model and significant heterogeneity remained, with $Q(5)$= 65.95, p< .001, I_2= 92.42%, τ_2= .728. Not enough studies were available to run the meta-regression analysis based on the length of the original treatment (reported by 3 studies only).

Effects of Psychotherapy on Changes in Quality of Life

Only eight studies assessed quality of life. Analyses based on the total number of participants who received treatment (including non-completers) showed that significant heterogeneity existed between the studies ($Q(7)$=16.782, p=.02) that was moderate (I_2= 58.289%). This suggested the need to adopt the random-effects model. This resulted in a significant effect favoring treatments targeting non-responders (*SMD*=0.411; 95% *CI* .177, .645, p<.001; see Figure 6). Importantly, no significant publication bias existed, with Egger's test $t(6)$=0.386, p=.71 and Kendall's *S statistic*=8, *tau*=0.250, p=.39, while the Duval and Tweedie's trim and fill showed that no studies were missing from the literature that could alternate the significant effect found.

Subgroup Analyses for Quality of Life

Similarly to the procedures followed above, we present only the subgroup analyses with at least one category with more than five studies.

Do the results on quality of life differ by disorder? Significant subgroup differences existed with regard to the category of disorder targeted ($Q(2)=7.384$, $p=.03$). The studies targeting mood (MDD) disorders resulted in a significant moderate effect size (6 studies; $SMD=0.360$; 95% CI 0.150, 0.570; p=.001), which was significantly smaller than the effect size for the other disorders.

Do the results on quality of life differ by comparison group? No significant between-group differences were detected $Q(1)=1.445$, $p=.23$, between treatments that were compared to either active or passive comparison groups. Significant between-group differences did emerge, however, when examining the specific type of comparison group, with $Q(1)=10.353$, $p=.001$. The effect size for the studies comparing treatments targeting non-response to TAU was moderate and significant (6 studies; $SMD=0.328$; 95% CI 0.163, 0.494; $p<.001$), which was significantly smaller than the effect size for the WLC comparison groups.

Do the results on quality of life differ by sample characteristics? The meta-regression analyses failed to find significant predictors of the overall effect size on quality of life for any of the tested variables: patients' mean age ($b=-0.043$, 95% $CI=-0.114$, 0.028, $z=-1.19$, $p=0.23$), percentage of females ($b=0.010$, 95% $CI-0.007$, 0.027, $z=1.19$, $p=0.24$), percentage of married participants ($b=-0.014$, 95% $CI-0.029$, 0.000, $z=-1.94$, $p=0.52$), percentage of employed participants ($b=-0.001$, 95% $CI-0.021$, 0.019, $z=-0.12$, $p=0.91$). There were not enough studies reporting the length of original treatment, but the length of current treatment was not a significant predictor of the effect size either ($b=-0.001$, 95% $CI-0.016$, 0.015, $z=-0.06$, $p=0.95$). The quality score given for the methodology of the studies did not moderate the treatment effect on quality of life either ($b=0.121$, 95% $CI-1.749$, 1.99, $z=0.13$, $p=0.90$).

Sensitivity Analysis: Are the Effects on Quality of Life Retained Regardless of Authors' Decisions in Earlier Stages of the Review?

Among the eight studies that included a quality of life measure to examine treatment effectiveness of non-responders, only six included the total number of participants who completed treatment. Conducting an analysis using these studies only as a sensitivity analysis, the random-effects model produced very similar results, with SMD=.447 (95% CI .143, 0.751, p=.004), Q(5)=14.810, p=.011, I_2= 66.238%. Accordingly, no significant publication bias was detected.

Follow-Up Quality of Life

Only five studies that assessed quality of life also included a follow-up assessment. The meta-analysis showed moderate (I_2= 42.603%) but non-significant (Q(4)=6.969, p=.14) levels of heterogeneity. Therefore, the fixed-effects model was followed, which favored treatments targeting non-responders (SMD=0.398; 95% CI 0.245, 0.551, p<.001; see Figure 7). No significant publication bias was detected, with Egger's test t(3)=0.426, p=.70 and Kendall's S $statistic$=0, tau=0.00, p=1.00, while the Duval and Tweedie's trim and fill showed that the significant treatment effect would be maintained even after considering the estimated one study missing from the literature (SMD=0.316; 95% CI 0.175, 0.457, p<.001).

No subgroup, sensitivity or meta-regression analyses were conducted due to the low number of included studies.

DISCUSSION

This meta-analysis was, to our knowledge, the first to examine whether psychotherapy is efficacious in treating treatment non-responders. Specifically, we examined studies that randomized previously non-responding patients diagnosed with any mood or anxiety disorder to a subsequent psychotherapy trial. Overall, we identified

18 RCTs with 1734 participants. Results showed that psychotherapy was efficacious in treating treatment non-responders both in terms of decreasing symptoms and increasing quality of life. The degree of improvement was largely equivocal across disorders and treatment types. The treatment effects were maintained at follow-up assessments. This study further documented that psychotherapy trials that attempt to treat treatment-non-response are rare, especially in comparison to pharmacological trials. Based on these findings, more psychotherapy trials for treatment non-responders are both warranted and needed.

Results showed that psychotherapy that was administered following non-response led to a moderate treatment effect on patients' symptoms from baseline to post-treatment. Based on these analyses there do not seem to be meaningful and stable differences between studies targeting depression and anxiety. This is promising in that mood and anxiety disorders are the most common disorders both at a population level (Wittchen et al., 2011) and as seen in treatment settings (Wang et al., 2005). It is equally promising that the effects were maintained at follow-up assessments.

Results further showed that administering psychotherapy to treatment non-responders also led to overall moderate improvements on measures of functioning and quality of life. The importance of this finding lies in the fact that treatment non-responders are at risk for functional disability (G. A. Fava et al., 2001). Although only half the studies reported measures of functioning and quality of life, we found preliminary evidence that especially third-wave treatments showed positive treatment effects on measures of functioning and quality of life. The improvement of quality of life following treatments using third-wave psychotherapies is consistent with previous studies that have shown that these treatments improve well-being of patients (G. A. Fava, 1999; Trompetter, Lamers, Westerhof, Fledderus, & Bohlmeijer, 2017; Wersebe, Lieb, Meyer, Hofer, & Gloster,

2017) even in treatment non-responders (Gloster et al., 2015). The third-wave treatments may be especially useful in that they aim to alter how a patient relates to their symptoms – as opposed as to the symptoms themselves – while actively promoting value-related behaviors. We agree with previous recommendations to include measures of functioning/quality of life and not exclusively of symptoms (Barton, Thompson, Burgess, & Grant, 2015; Boschen, Drummond, Pillay, & Morton, 2010).

The findings of the present meta-analysis suggested the studies included were highly heterogeneous. Only three subgroups of studies had low levels of heterogeneity, namely studies using BDI as a primary outcome measure and the studies implementing a third-wave or a psychodynamic treatment targeting non-response. Exploration of heterogeneity resulted in no consistent conclusions about its sources. Studies with MDD patients were less heterogeneous than the other disorders, but heterogeneity for those was still significant. None of the authors' decisions to include studies with slightly different characteristics (i.e., different statistical format, transdiagnostic patient population, bipolar disorder population) was enough to explain the high and significant heterogeneity. One should keep in mind the impact of the various definitions of non-response and also the different comparison groups used. We followed a systematic procedure to report and describe the non-response inclusion criteria used by each of the included studies. Journals should have specific requirements about the definitions used by the studies and researchers implementing work with non-response patients should clearly state what they consider as non-response in a homogeneous and consistent way. We provide more specific recommendations on this aspect in the recommendations section below.

What is Necessary to Successfully Treat a Treatment Non-Responder?

This meta-analysis did not find consistent and compelling evidence that any type of psychotherapy was better or worse than any other for reducing symptoms. Further,

subgroup analyses did not reveal evidence that any aspect of the examined treatments led to preferential outcomes. Leaving aside methodological and power considerations for a moment, it is possible that the simple act of switching from one treatment to another is itself a key ingredient. This interpretation is tentatively supported by results from this meta-analysis showing that the longer the length of previous-treatment the better the new treatment – although this finding was based on a subsabmple (i.e., seven of the included studies). At the point a patient reaches the status of non-responder, they have experienced a combination of disappointment, frustration, and distrust of treatments and treatment providers and have likely engaged in self-criticism or, worse, self-stigmatizing evaluations. Patients can easily conclude that their lack of response confirms their suspicion that something really is wrong with them as evidenced by the fact that they do not respond to a treatment that works for others. Offering patients a clearly demarcated "re-start" may help them start anew while reinstalling hope by framing this new treatment as something different from that which did not work for the patient up to this point. In short, it gives the patient another chance to achieve the changes they seek. To the degree possible, and when appropriate, it is advisable to take any blame from the patient and place it on the previous treatment. It remains an empirical question whether all switches are equal or if some sequences are better than others (e.g., pharmacotherapy to psychotherapy; CBT to third-wave; third-wave to psychodynamic, etc.).

Alternately, specific treatment elements are indeed needed to help treatment non-responders, but that the effects of such specific elements were not detectable in this meta-analysis above and beyond variance due to methodology, treatment setting, non-specific factors, therapist characteristics including quality, patient engagement, social support of the patient, and inert and iatrogenic techniques included in the therapy packages. Indeed, the quality of the studies varied, particularly with respect to details they provided about

the therapy. Likewise, it is conceivable, if also difficult to empirically capture, that the type of previous treatment the patient experienced helped prime them for change in the subsequent treatment. Future research is strongly encouraged to collect and explicitly, clearly, and systematically present information on previous treatments. Until this is established in the literature, however, this remains speculative.

Despite the methodological difficulties of detecting group level differences in outcomes, a comprehensive approach to establishing a principled approach to treating treatment non-responders needs to include a systematic understanding of the maintaining factors of the mental disorder/ lack of functioning as well as the linked mechanisms of treatment action involved changing these maintaining factors (Gloster & Karkela, in press). In our opinion, these targets should ideally go beyond descriptive counts of symptoms or comorbidity. If, for example, we find that treatment non-responders are more rigid, process information differently, have difficulties with their self-concept, are overly entangled in their cognitions, have difficulties in interpersonal interactions, or difficulties in their social environment, then the treatments for these patients need to be able to target these issues. Some treatment approaches such as ACT (Clarke et al., 2014; Gloster et al., 2015), Mindfulness-Based Cognitive Therapy (Kenny & Williams, 2007), Cognitive Behavioral Analysis System of Psychotherapy (Brakemeier et al., 2014), and others target some of these issues. It is also likely that various patient constellations will react differently and systemic research is needed to successively learn how to ideographically tailor treatment.

Why so Few Studies of Psychotherapy?

This meta-analysis documented that relatively few empirical studies of treatment-non-response examined psychotherapy approaches that deal with this subgroup of patients. There are likely several reasons for this: First, funding for psychotherapy trials

lags behind pharmacological approaches. Second, it may reflect larger patterns in health care. For example, when faced with mental health issues most people consult their general practitioner (Gater et al., 1991; Rickwood, Deane, & Wilson, 2007; Steel et al., 2006) and an increasing proportion of mental health outpatients are treated with psychotropic medication without any psychotherapy (Olfson & Marcus, 2010). Third, anecdotal evidence points to the perception that psychotherapy is not a viable option for treatment non-responders. The results of this study stand in contrast to this perception and can contribute to the important task of developing a systematic response for the sizable group of patients who do not respond to first-line treatments.

Limitations

The results of this meta-analysis need to be interpreted with several limitations in mind. First, the literature is currently characterized by idiosyncratic definitions of treatment non-response and inconsistent details about previous treatments. This heterogeneity is reflected in the results of this meta-analysis. We attempted to counteract this by using random effect models, but future studies are encouraged to clearly and explicitly report these data in all trials of treatment non-response. Second, only half of the studies reported information about functioning or quality of life. The results from this small sample size need to be interpreted with caution. Future studies should report measures of functioning/ quality of life in addition to reporting on symptomatology. Third, our results are limited to studies published in English. Fourth, our results are limited to studies indexed in Ovid, PsycARTICLES, PsycCRITIQUES and PsycINFO. We augmented our search by examining reference sections of relevant empirical and theoretical papers, but we cannot exclude the possibility that we overlooked a relevant study. Finally, we examined mood and anxiety disorders because they are most prevalent. It remains to be tested whether these results can be generalized beyond these disorders.

Recommendations

The research literature as a whole needs to work towards arriving at a consensus about what constitutes treatment non-response. All studies reviewed in this meta-analysis used symptomatology as a basis for making this decision. We strongly suggest that functioning/ quality of life play a larger role in definitions, as symptomatology and well-being are partially independent (Keyes, 2005). Given that no therapy works for all patients, it is likely that treatment options will need to include options for individualization, similar to personalized medicine. This will require more information about what works, how, for whom following initial treatment non-response. It is also recommended that future studies differentiate between treatments specifically designed to address non-adherence and studies that use a preexisting treatment as an alternative. Future research should also document as many details about previous treatments as possible and include information about when people drop out of treatment as well as follow-up data (Fekadu et al., 2009; Karekla, Konstantinou, Ioannou, Kareklas, & Gloster, 2019).

Conclusion

The strength of this study is that it is the first to examine whether psychotherapy is efficacious in treating treatment non-responders. Results suggest that psychotherapy can be used for a range of treatment non-responders including mood and anxiety disorders. Reflecting the state of the literature, the heterogeneity across studies was high and quality of studies was relatively low. Clinical research – both efficacy and effectiveness trials – would profit if some of the recommendations stated above are implemented. No treatment works for every patient. Addressing treatment failure empirically and systematically is an important task for future research in order to better serve our patients.

Acknowledgments

References

Anderson, I. M., Ferrier, I. N., Baldwin, R. C., Cowen, P. J., Howard, L., Lewis, G., … Tylee, A. (2008). Evidence-based guidelines for treating depressive disorders with antidepressants: A revision of the 2000 British Association for Psychopharmacology guidelines. *Journal of Psychopharmacology*, *22*(4), 343–396. doi: 10.7554/eLife.03397

Ballenger, J. C., Davidson, J. R., Lecrubier, Y., Nutt, D. J., Borkovec, T. D., Rickels, K., … Wittchen, H. U. (2001). Consensus statement on generalized anxiety disorder from the International Consensus Group on Depression and Anxiety. *Journal of Clinical Psychiatry*, *62*(Suppl 11), 53–58.

Bandelow, B. (1997). *Panik- und Agoraphobieskala (PAS)*. Göttingen: Hogrefe.

Bandelow, B., Lichte, T., Rudolf, S., Wiltink, J., & Beutel, M. E. (2015). The German guidelines for the treatment of anxiety disorders. *European Archives of Psychiatry and Clinical Neuroscience*, *265*(5), 363–373. doi: 10.1007/s00406-014-0563-z

Bandelow, B., Zohar, J., Hollander, E., Kasper, S., Möller, H. J., Allgulander, C., … Vega, J. (2008). World Federation of Societies of Biological Psychiatry (WFSBP) guidelines for the pharmacological treatment of anxiety, obsessive-compulsive and post-traumatic stress disorders - First revision. *World Journal of Biological Psychiatry*, *9*(4), 248–312. doi: 10.1080/15622970802465807

Barton, H., Thompson, S., Burgess, S., & Grant, M. (2015). *The Routledge handbook of planning for health and well-being: Shaping a sustainable and healthy future*. New York: Routledge.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck Depression Inventory-II*. San Antonio, TX: Psychological Corporation.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry*, *4*(6), 561–571. doi: 10.1001/archpsyc.1961.01710120031004

Berlim, M. T., & Turecki, G. (2007). Definition, assessment, and staging of treatment-resistant refractory major depression: A review of current concepts and methods. *Canadian Journal of Psychiatry*, *52*(1), 46–54.

Bigger, J. T. (2003). Issues in subgroup analyses and meta-analyses of clinical trials. *Journal of Cardiovascular Electrophysiology*, *14*(9 SUPPL.), S6–S8. doi: 10.1046/j.1540-8167.14.s9.1.x

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Klauminzer, G., Charney, D. S., & Keane, T. M. (1990). A clinician rating scale for assessing current and lifetime PTSD: the CAPS-1. *Behavior Therapist*, *13*, 187–188.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2014). Comprehensive meta-analysis (Version 3.0) [Computer software]. Englewood, NJ: Biostat.

Borenstein, M., & Higgins, J. P. T. (2013). Meta-Analysis and Subgroups. *Prevention Science*, *14*(2), 134–143. doi: 10.1007/s11121-013-0377-7

Boschen, M. J., Drummond, L. M., Pillay, A., & Morton, K. (2010). Predicting outcome of treatment for severe, treatment resistant OCD in inpatient and community settings. *Journal of Behavior Therapy and Experimental Psychiatry*, *41*(2), 90–95. doi: 10.1016/j.jbtep.2009.10.006

Bradley, R., Greene, J., Russ, E., Dura, L., & Westen, D. (2005). Multi-dimensional meta-analysis of psychotherapy for refugees. *American Journal of Psychiatry*, *162*(2), 214–227. doi: 10.1176/appi.ajp.162.2.214

Brakemeier, E. L., Merkl, A., Wilbertz, G., Quante, A., Regen, F., Bührsch, N., …

Bajbouj, M. (2014). Cognitive-behavioral therapy as continuation treatment to sustain response after electroconvulsive therapy in depression: A randomized controlled trial. *Biological Psychiatry*, *76*(3), 194–202. doi: 10.1016/j.biopsych.2013.11.030

Brazier, J. E., Harper, R., Jones, N. M., O'Cathain, A., Thomas, K. J., Usherwood, T., & Westlake, L. (1992). Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ (Clinical Research Ed.)*, *305*(6846), 160–164. doi: 10.1136/bmj.305.6846.160

Bystritsky, A. (2006). Treatment-resistant anxiety disorders. *Molecular Psychiatry*, *11*(9), 805–814. doi: 10.1038/sj.mp.4001852

Clarke, S., Kingston, J., James, K., Bolderston, H., & Remington, B. (2014). Acceptance and Commitment Therapy group for treatment-resistant participants: A randomized controlled trial. *Journal of Contextual Behavioral Science*, *3*(3), 179–188. doi: 10.1016/j.jcbs.2014.04.005

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCL): A measure of primary symptom dimensions. In P. Pichot & R. Oliver-Martin (Eds.), *Psychological measurements in psychopharmacology* (pp. 79–110). Oxford, England: S. Karger. doi: 10.1159/000395070

Derogatis, L. R., & Spencer, P. M. (1982). *Administration and procedures: BSI. Manual I*. Baltimore, MD: Clinical Psychometric Research.

Division 12 of the American Psychological Association. (2016). Psychological Treatments. Retrieved January 30, 2019, from https://www.div12.org/treatments/

Donegan, S., Williams, L., Dias, S., Tudur-Smith, C., & Welton, N. (2015). Exploring Treatment by covariate interactions using subgroup analysis and meta-regression

in Cochrane reviews: A review of recent practice. *PLoS ONE*, *10*(6), 1–17. doi: 10.1371/journal.pone.0128804

Eddy, K. T., Dutra, L., Bradley, R., & Westen, D. (2004). A multidimensional meta-analysis of psychotherapy and pharmacotherapy for obsessive-compulsive disorder. *Clinical Psychology Review*, *24*(8), 1011–1030. doi: 10.1016/j.cpr.2004.08.004

Fava, G. A. (1999). Well-being therapy: Conceptual and technical issues. *Psychotherapy and Psychosomatics*, *68*(4), 171–179. doi: 10.1159/000012329

Fava, G. A., Rafanelli, C., Ottolini, F., Ruini, C., Cazzaro, M., & Grandi, S. (2001). Psychological well-being and residual symptoms in remitted patients with panic disorder and agoraphobia. *Journal of Affective Disorders*, *65*(2), 185–190. doi: 10.1016/S0165-0327(00)00267-6

Fava, M. (2003). Diagnosis and definition of treatment-resistant depression. *Biological Psychiatry*, *53*(8), 649–659. doi: 10.1016/S0006-3223(03)00231-2

Fava, M., & Davidson, K. G. (1996). Definition and epidemiology of treatment-resistant depression. *Psychiatric Clinics of North America*, *19*(2), 179–198. doi: 10.1016/S0193-953X(05)70283-5

Fekadu, A., Wooderson, S. C., Markopoulo, K., Donaldson, C., Papadopoulos, A., & Cleare, A. J. (2009). What happens to patients with treatment-resistant depression? A systematic review of medium to long term outcome studies. *Journal of Affective Disorders*, *116*(1–2), 4–11. doi: 10.1016/j.jad.2008.10.014

Fleck, M. P., Louzada, S., Xavier, M., Chachamovich, E., Vieira, G., Santos, L., & Pinzon, V. (2000). Application of the Portuguese version of the abbreviated instrument of quality life WHOQOL-bref. *Revista de Saúde Pública*, *34*(2), 178–183. doi: 10.1590/S0034-89102000000200012

Frank, E., Prien, R. F., Jarrett, R. B., Keller, M. B., Kupfer, D. J., Lavori, P. W., … Weissman, M. M. (1991). Conceptualization and rationale for consensus definitions of terms in major depressive disorder: Remission, recovery, relapse, and recurrence. *Archives of General Psychiatry*, *48*(9), 851–855. doi: 10.1001/archpsyc.1991.01810330075011

Fukuhara, S., Ware, J. E., Kosinski, M., Wada, S., & Gandek, B. (1998). Psychometric and clinical tests of validity of the Japanese SF-36 Health Survey. *Journal of Clinical Epidemiology*, *51*(11), 1045–1053. doi: 10.1016/S0895-4356(98)00096-1

Furukawa, T., Akechi, T., Ozaki, N., & Al., E. (2003). *GRID-HDRS-17*. Tokyo, Japan: The Japanese Society of Clinical Neuropsychopharmacology.

Gater, R., De Almeida E Sousa, B., Barrientos, G., Caraveo, J., Chandrashekar, C. R., Dhadphale, M., … Sartorius, N. (1991). The pathways to psychiatric care: A cross-cultural study. *Psychological Medicine*, *21*(3), 761–774. doi: 10.1017/S003329170002239X

Gloster, A. T., Hauke, C., Höfler, M., Einsle, F., Fydrich, T., Hamm, A., … Wittchen, H. U. (2013). Long-term stability of cognitive behavioral therapy effects for panic disorder with agoraphobia: A two-year follow-up study. *Behaviour Research and Therapy*, *51*(12), 830–839. doi: 10.1016/j.brat.2013.09.009

Gloster, A. T., & Karkela, M. (n.d.). A multi-level, multi-method approach to testing and refining intervention targets. In S. C. Hayes & S. G. Hofmann (Eds.), *Beyond the DSM*. Oakland, CA: Context Press/ New Harbinger.

Gloster, A. T., Sonntag, R., Hoyer, J., Meyer, A. H., Heinze, S., Ströhle, A., … Wittchen, H. U. (2015). Treating treatment-resistant patients with panic disorder and agoraphobia using psychotherapy: A randomized controlled switching trial.

Psychotherapy and Psychosomatics, *84*(2), 100–109. doi: 10.1159/000370162

Gloster, A. T., Wittchen, H. U., Einsle, F., Lang, T., Helbig-Lang, S., Fydrich, T., …
Arolt, V. (2011). Psychological treatment for panic disorder with agoraphobia: A
randomized controlled trial to examine the role of therapist-guided exposure in
situ in CBT. *Journal of Consulting and Clinical Psychology*, *79*(3), 406–420.
doi: 10.1037/a0023584

Greden, J. F. (2001). The Burden of Recurrent Depression : Causes, Consequences,
and Future Prospects. *Health (San Francisco)*, *62*(suppl22), 5–9.

Guy, W. (1976a). *ECDEU Assessment Manual for Psychopharmacology*. Rockville:
Department of Health, Education, and Welfare.

Guy, W. (1976b). *ECDEU Assessment Manual for Psychopharmacology* (Pub no
ADM). Rockville, Md: National Institute of Mental Health.

Hamilton, M. (1967). Development of a Rating Scale for Primary Depressive Illness.
British Journal of Social and Clinical Psychology, *6*(4), 278–296. doi:
10.1111/j.2044-8260.1967.tb00530.x

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-
analysis. *Statistics in Medicine*, *21*(11), 1539–1558. doi: 10.1002/sim.1186

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring
inconsistency in meta-analyses Testing for heterogeneity. *BMJ*, *327*(6), 557–560.
doi: 10.1136/bmj.327.7414.557

Hilsenroth, M. J., Ackerman, S. J., Blagys, M. D., Baumann, B. D., Baity, M. R.,
Smith, S. R., … Holdwick, D. J. (2000). Reliability and Validity of DSM-IV
Axis V. *Am J Psychiatry*, *157*, 1858–1863.

Karekla, M., Konstantinou, P., Ioannou, M., Kareklas, I., & Gloster, A. T. (2019). The
phenomenon of treatment dropout, reasons and moderators in Acceptance and

Commitment Therapy and other active treatments: A meta-analytic review. *Clinical Psychology in Europe.*, *1*(3), e33058. doi: 10.32872/cpe.v1i3.33058

Kennedy, S. H., Lam, R. W., McIntyre, R. S., Tourjman, S. V., Bhat, V., Blier, P., … Uher, R. (2016). Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 clinical guidelines for the management of adults with major depressive disorder: Section 3. Pharmacological Treatments. *Canadian Journal of Psychiatry*. doi: 10.1177/0706743716659417

Kenny, M. A., & Williams, J. M. G. (2007). Treatment-resistant depressed patients show a good response to Mindfulness-based Cognitive Therapy. *Behaviour Research and Therapy*, *45*(3), 617–625. doi: 10.1016/j.brat.2006.04.008

Keyes, C. L. M. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology*, *73*(3), 539–548. doi: 10.1037/0022-006X.73.3.539

Leon, A. C., Solomon, D. A., Mueller, T. I., Turvey, C. L., Endicott, J., & Keller, M. B. (1999). The Range of Impaired Functioning Tool (LIFE-RIFT): A brief measure of functional impairment. *Psychological Medicine*, *29*(4), 869–878. doi: 10.1017/S0033291799008570

Levitt, J. T., & Karekla, M. (2005). Integrating Acceptance and Mindfulness with Cognitive Behavior Treatment for Panic Disorder. In M. M. Anthony (Ed.), *Acceptance and Midfulness-Based Approaches to Anxiety: Conceptualization and Treatment* (pp. 165–188). New York, NY, US: Kluwer Academic/Plenum Publishers.

McKay, D., Abramowitz, J. S., & Taylor, S. (2010). *Cognitive-behavioral therapy for refractory cases: Turning failure into success.* Washington DC: American Psychological Association. doi: 10.1037/12070-000

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Physical Therapy*, *89*(9), 873–880. doi: 10.1016/j.ijsu.2010.02.007

Olfson, M., & Marcus, S. C. (2010). National trends in outpatient psychotherapy. *American Journal of Psychiatry*, *167*(12), 1456–1463. doi: 10.1176/appi.ajp.2010.10040570

Pollack, M. H., Otto, M. W., Roy-Byrne, P. P., Coplan, J. D., Rothbaum, B. O., Simon, N. M., & Gorman, J. M. (2008). Novel treatment approaches for refractory anxiety disorders. *Depression and Anxiety*, *25*(6), 467–476. doi: 10.1002/da.20329

Rickwood, D. J., Deane, F. P., & Wilson, C. J. (2007). When and how do young people seek professional help for mental health problems? *Medical Journal of Australia*, *187*(7), 35–39.

Sage, W. M. (2004). The forgotten third: Liability insurance and the medical malpractice crisis. *Health Affairs*, *23*(4), 10–21. doi: 10.1377/hlthaff.23.4.10

Schlaepfer, T. E., Ågren, H., Monteleone, P., Gasto, C., Pitchot, W., Rouillon, F., … Kasper, S. (2012). The hidden third: Improving outcome in treatment-resistant depression. *Journal of Psychopharmacology*, *26*(5), 587–602. doi: 10.1177/0269881111431748

Schulz, K. F., Altman, D. G., Moher, D., & CONSORT Group. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMJ*, *340*(mar23 1), c332–c332. doi: 10.1136/bmj.c332

Sedgwick, P. (2013, June 24). Meta-analyses: Heterogeneity and subgroup analysis. *BMJ (Online)*. British Medical Journal Publishing Group. doi: 10.1136/bmj.f4040

Skevington, S. M., Lotfy, M., & O'Connell, K. A. (2004). The World Health

Organization's WHOQOL-BREF quality of life assessment: Psychometric

properties and results of the international field trial A Report from the

WHOQOL Group. *Quality of Life Research*, *13*(2), 299–310.

Souery, D., Papakostas, G. I., & Trivedi, M. H. (2006). Treatment-resistant

depression. *The Journal of Clinical Psychiatry*, *67*(suppl 6), 16–22.

Steel, Z., Mcdonald, R., Silove, D., Bauman, A., Sandford, P., Herron, J., & Harry

Minas, I. (2006). Pathways to the First Contact with Specialist Mental Health

Care. *Australian & New Zealand Journal of Psychiatry*, *40*(4), 347–354. doi:

10.1080/j.1440-1614.2006.01801.x

Taylor, S., Abramowitz, J. S., & McKay, D. (2012). Non-adherence and non-response

in the treatment of anxiety disorders. *Journal of Anxiety Disorders*, *26*(5), 583–

589. doi: 10.1016/j.janxdis.2012.02.010

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. doi: 10.1016/j.jad.2017.02.035

Trompetter, H. R., Lamers, S. M. A., Westerhof, G. J., Fledderus, M., & Bohlmeijer,

E. T. (2017). Both positive mental health and psychopathology should be

monitored in psychotherapy: Confirmation for the dual-factor model in

acceptance and commitment therapy. *Behaviour Research and Therapy*, *91*, 58–

63. doi: 10.1016/j.brat.2017.01.008

Wang, P. S., Lane, M., Olfson, M., Pincus, H. A., Wells, K. B., & Kessler, R. C.

(2005). Twelve-month use of mental health services in the United States: Results

from the National Comorbidity Survey Replication. *Archives of General*

Psychiatry, *62*(6), 629–640. doi: 10.1001/archpsyc.62.6.629

Ware, J. E., Kosinski, M., & Keller, S. D. (1996). A 12-item short-form health survey. Construction of scales and preliminary tests of reliability and validity. *Medical Care*, *34*(3), 220–233.

Weathers, F. W., Keane, T. M., & Davidson, J. R. T. (2001). Clinician-administered PTSD Scale: A review of the first ten years of research. *Depression and Anxiety*, *156*, 132–156.

Wersebe, H., Lieb, R., Meyer, A. H., Hofer, P., & Gloster, A. T. (2017). The link between stress, well-being, and psychological flexibility during an Acceptance and Commitment Therapy self-help intervention. *International Journal of Clinical and Health Psychology*. doi: https://doi.org/10.1016/j.ijchp.2017.09.002

Westen, D., & Morrison, K. (2001). A multidimensional meta analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *69*(6), 875–899. doi: 10.1037//0022-006x.69.6.875

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., … Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, *21*(9), 655–679. doi: 10.1016/j.euroneuro.2011.07.018

World Health Organization [WHO]. (1948). *Constitution of the World Health Organization*. Geneva, Switzerland.

World Health Organization [WHO]. (2014). Mental health: a state of well-being. Retrieved January 18, 2019, from https://www.who.int/features/factfiles/mental_health/en/#

Young, R. C., Biggs, J. T., Ziegler, V. E., & Meyer, D. A. (1978). A rating scale for

    mania: Reliability, validity and sensitivity. *British Journal of Psychiatry*,

    133(11), 429–435. doi: 10.1192/bjp.133.5.429

Table 1.

Definition of non-response in each study

Study	Definition of non-response of the 1st Treatment	Current Clinical Characteristics Involved in Defining Non-Response at Time of Non-Responder Trial
Barnhofer et al. (2009)	▪ Previous antidepressant medication ▪ Previous psychotherapy over at least five sessions ▪ History of at least three previous episodes of Major Depression or Chronic Depression	▪ Current diagnosis of Major Depression or presence of residual symptoms
Clarke et al. (2014)	▪ Three or more previous episodes of depression, or one chronic episode lasting 1 year or more ▪ Completion of at least one psychosocial intervention or at least one previous 8-session episode of psychological therapy	▪ Not specified
Eisendrath et al. (2016)	▪ Previous treatment-as-usual (TAU) pharmacotherapy	▪ Current diagnosis of unipolar major depressive disorder

	▪ No full remission after 1 year	▪ HAM-D 17 score of at least 14
	▪ Two or more trials of antidepressant medication	
Fonagy et al. (2015)	▪ At least two failed treatment attempts: one involving antidepressant medication, and the other involving either antidepressant medication or a psychological intervention	▪ Current diagnosis of Major Depression Disorder with a minimum duration of two years ▪ HAM-D 17 score of at least 14 ▪ BDI-II score of at least 21
Gloster et al. (2015)	▪ One or more previous courses of psychological and / or pharmacological treatment (state-of-the-art practice) ▪ Psychotherapy: At least 20 sessions of empirically supported treatment ▪ Pharmacology: intake of an approved drug at least at the minimum dosage and length as recommended by (inter)national guidelines	▪ Mobility Inventory score of at least 1.5 ▪ Clinical Global Impression score of at least 4

Harley et al. (2008)	▪ Ongoing depressive symptoms despite stable, adequate medication treatment for major depressive disorder	▪ Current diagnosis of MDD ▪ Established treatment relationship with a psychiatrist ▪ Stabilized on an adequate dose of antidepressant medication: no change in dosage for at least 6 weeks, receiving at least the standard effective dose of AD
Hartmann Souza et al. (2016)	▪ Failure to respond to one trial of antidepressant medication in adequate dose and duration ▪ Adequate dose equals equivalent of at least 75 mg of amitriptyline ▪ Adequate treatment duration equals at least four weeks	▪ Current diagnosis of MDD
Hinton et al. (2005)	▪ Prior supportive counseling	▪ Current diagnosis of PTSD

- Adequate trial of a selective serotonin reuptake inhibitor (SSRI) (i.e., at least 1 year on the maximally tolerated dosage)

Isasi et al. (2010)

- History of severe or unfavorable progression of the disease despite adequate pharmacological treatment
- Two or more relapses in the preceding year
- Suicide attempts
- Persistent affective symptoms despite appropriate drug treatment

- Meeting inclusion criteria for type I or II bipolar disorder for at least 2 years
- Currently under pharmacological treatment (mainly mood stabilizer)
- Beck's Depression Index score of at least 7
- Young Mania Rating Scale score of at least 6
- Severe difficulties in social-occupational functioning (Inadaptation Scale score of at least 14)

Kocsis et al. (2009)

- Prior antidepressant medication according to a pharmacotherapy algorithm
- Depressive symptoms for more than 2 years without remission

- Current major depressive episode for at least 4 weeks
- HAM-D 24 score of at least 20

	▪ Met criteria for double depression (i.e., current major depression with antecedent dysthymic disorder), chronic major depression, or recurrent major depression with incomplete recovery between episodes	
Ludman et al. (2007)	▪ At least six months of antidepressant treatment prescribed in specialty care ▪ A visit diagnosis of major depressive disorder at time of initial antidepressant prescription	▪ Persistent symptoms of depression
Mantini et al. (2017)	▪ Antidepressant-refractory after taking one or more antidepressants at an adequate dosage for four or more weeks	▪ Primary diagnosis of major depression without psychotic features ▪ BDI-II score of at least 10
Moore & Blackburn (1997)	▪ Recurrent major depression ▪ Failed response to acute treatment with antidepressant medication ▪ HRSD score of at least 14	▪ Current primary, unipolar major depression

Nakagawa et al. (2017)	▪ Antidepressant-refractory after taking one or more antidepressants at an adequate dosage for four or more weeks ▪ Having received adequate therapeutic levels of antidepressant medication for at least 8 weeks ▪ Maudsley Staging method for treatment-resistant depression score of 3 or more	▪ Primary diagnosis of major depression ▪ GRID-HDRS score of at least 16
Otto et al. (2003)	▪ Failed to respond adequately to treatment with clonazepam (0.5 bid to 1 mg bid) in combination with an adequate dose of an SSRI other than sertraline	▪ Current diagnosis of PTSD despite ongoing pharmacotherapy
Town et al. (2017)	▪ At least one trial of antidepressants at the adequate recommended therapeutic dose	▪ Current depressive episode duration of 6 or more weeks ▪ Inadequate response to treatment (assessed by 17-item HAM-D score of at least16)

Watanabe et al. (2011)	▪ Adequate pharmacologic treatment (maximum doses of 2 types of antidepressants for at least 4 weeks each)	▪ Currently partially remitted, mild or moderate depression ▪ Currently significant insomnia ▪ Score of 2 on at least 1 of 3 sleep items on GRID-HAMD ▪ Insomnia Severity Index (ISI) score of 8 or greater ▪ 17-GRID-HAMD of at least 8
Wiles et al. (2013)	▪ Adequate dose of antidepressant medication (based on the British National Formulary and advice from psychopharmacology experts) for at least 6 weeks	▪ Current diagnosis of depression ▪ BDI-II score of at least 15

Note: Only the inclusion and exclusion criteria that contributed to the current clinical characteristics involved in defining non-response are included here. Further inclusion and exclusion criteria exist for all studies

Table 2

Characteristics of studies included in the meta-analysis evaluating psychotherapy for treatment-resistance

Study	Dx	Age	Sex	Married	Sample Size	Tx	Tx length	Outcome Measure (Symptoms)	Quality of Life Outcome	Quality
Barnhofer et al. (2009)	Mood Disorder (MDD)	41.9	67.9	50	AG: 16 CG: 15	Third wave	8	BDI-II	NA	1.8
Clarke et al. (2014)	Transdia gnostic	43.5	67.2	NRD	AG: 25 CG: 16	Third wave	16	GSI	WHOQOL	1.5
Eisendrath et al. (2016)	Mood Disorder (MDD)	46.2	76.25	32.9	AG: 87 CG: 86	Meds & third wave	8	HAM-D 17	NA	1.6
Fonagy et al. (2015)	Mood Disorder (MDD)	44.3	66.4	17.8	AG: 67 CG: 62	Psychoanalytic	72	BDI-II	GAF	1.8
Gloster et al. (2015)	Anxiety (Agoraph obia)	36.9	69.8	23.3	AG: 41 CG: 10	Third wave	4	PAS	CGI	1.9
Harley et al. (2008)	Mood Disorder (MDD)	41.8	75	25	AG: 10 CG: 9	Third wave	16	BDI-I	LIFE-RIFT (subscale satisfaction)	1.5
Hartmann Souza et al. (2016)	Mood Disorder (MDD)	49.2	85	44.7	AG: 17 CG: 23	IPT	19	BDI-I	WHOQOL-brief (subscale psychological)	1.8
Hinton et al. (2005)	Anxiety (PTSD)	51.8	60	NRD	AG: 20 CG: 20	Meds & CBT	12	CAPS	NA	1.4
Isasi et al. (2010)	Mood Disorder (Bipolar	41.3	47.5	NRD	AG: 20 CG: 20	Meds & CBT	20	YMRS	NA	1.2

Study	Dx	Age	Sex	Married	Sample Size	Tx	Tx length	Outcome Measures (Symptoms)	Quality of Life Outcome	
	Disorder I and II)									
Kocsis et al. (2009)	Mood Disorder (MDD)	44.6	53.7	61.7	AG: 195 CG: 94	CBT or CT	12	HAM-D	LIFE-RIFT	1.6
Ludman et al. (2007)	Mood Disorder (MDD)	50.5	73	50	AG: 26 CG: 26	CBT or CT	52	SCL-20 (depression scale)	PGI (NRD)	1.7
Mantani et al. (2017)	Mood Disorder (MDD)	40.9	53.5	48.2	AG: 81 CG: 83	Meds & Smartphone CBT	9	BDI-II	NA	1.8
Moore & Blackburn (1997)	Mood Disorder (MDD)	38	62	NRD	AG: 5 CG: 4	CBT or CT	52	BDI-I	NA	1
Nakagawa et al. (2017)	Mood Disorder (MDD)	40.6	36.3	52.5	AG: 40 CG: 40	Meds & CBT	16	BDI-II	SF-36 (mental component)	1.9
Otto et al. (2003)	Anxiety (PTSD)	47.2	100	NRD	AG: 5 CG: 5	Meds & CBT	NRD	CAPS	NA	0.9
Town et al. (2017)	Mood Disorder (MDD)	41.6	63.3	38.3	AG: 30 CG: 30	Psychodynamic	20	HAM-D 17	NA	1.9
Watanabe et al. (2011)	Mood Disorder (MDD)	50.5	62.2	62.2	AG: 20 CG: 17	CBT or CT	4	HAMD 17	NA	1.8
Wiles et al. (2013)	Mood Disorder (MDD)	49.6	76	53	AG: 234 CG: 235	Meds & CBT	NRD	BDI-II	SF-12 (mental subscale)	1.7

Note. All values reflect the groups included in the meta-analysis. If a study had additional groups that were not relevant for the meta-analysis they are not reflected here; Dx = Diagnostic category; Age = mean age; Sex = percentage of female participants; Married = percentage of the participants that were married or in a relationship; Sample Size = the number of participants that filled out the outcome measures at pre and were included in the meta-analysis; Tx = Overall category of the treatment for the active treatment group; Tx length = number of weeks; Outcome Measures (Symptoms) = measures included in the meta-analysis measuring symptom reduction; Quality of Life Outcome =

measures included in the meta-analysis measuring quality of life or functioning; Quality = quality rating of each study (range 0 [lowest quality] – 2 [highest quality]); MDD = Major Depression Disorder, PTSD = Post Traumatic Stress Disorder, NRD = information was not reported in the study; AG = active treatment group; CG = Control group; Third wave = treatment category including different third wave treatments, such as Mindfulness Based Cognitive Therapy, Acceptance and Commitment Therapy, and Dialectic Behavioral Therapy; Meds & third wave = treatment category including psychopharmacological treatment given to the patients in addition to a third wave psychological treatment; Psychoanalytic = treatment category including long-term psychoanalytic psychotherapy and intensive short-term dynamic psychotherapy; IPT = Interpersonal psychotherapy; CBT or CT = treatment category including cognitive behavioral treatments and cognitive treatments; BDI-I = Beck Depression Inventory – I; BDI-II = Beck Depression Inventory – II ; HAM-D 17 & 24 items = Hamilton Depression Severity Rating Scale; CAPS = Clinician-Administered Post Traumatic Stress Disorder Scale; PAS = Panic Agoraphobia Scale; SCL-20 = Symptom Checklist depression scale; YMRS = Young Mania Rating Scale, GSI = Global Severity Index, WHOQOL = the World Health Organization Quality of Life; WHOWOL brief = brief World Health Organization Quality of Life subscale psychological; GAF = the Global Assessment of Functioning; CGI = the Clinical Global Impression Scale; LIFE-RIFT = the Longitudinal Interval Follow-up Evaluation Range of Impaired Functioning Tool; LIFE-RIF (subscale satisfaction) = the Longitudinal Interval Follow-up Evaluation Range of Impaired Functioning Tool subscale satisfaction; PGI = the Patient-Rated Global Improvement; SF – 12 = the Short Form Health Survey 12; SF – 36 = The 36 Item Short Form Health Survey, NA = not applicable, because authors did not assess an quality of life or functioning measurement.

Grouping variable	Categories	Number of studies	Point Estimate (Fixed)	Standard error	Lower 95 %CI	Upper 95 %CI	Z value	p-value	Q-value	Q(df) p-value	I^2
Diagnostic category	Anxiety	3	1.315	0.248	0.828	1.801	5.293	.000	8.308	(2) .016	75.927
	Mood	14	0.511	0.052	0.409	0.612	9.843	.000	72.042	(13) .000	81.955
	Transdiagnostic	1	0.368	0.323	-0.265	1.001	1.140	.254	0.000	(0) 1.000	0.000
Diagnosis	MDD	13	0.482	0.052	0.380	0.584	9.258	.000	21.315	(12) .046	43.701
	Other	5	1.306	0.189	0.936	1.676	6.922	.000	51.623	(4) .000	92.251
Type of previous non-responsive treatment	Combined	4	0.769	0.138	0.498	1.040	5.563	.000	16.007	(3) .001	81.258
	Medication	13	0.509	0.055	0.402	0.616	9.326	.000	71.325	(12) .000	83.176
	Psychotherapy	1	0.368	0.323	-0.265	1.001	1.140	.254	0.000	(0) 1.000	0.000
Type of comparison group	TAU	14	0.506	0.054	0.399	0.613	9.293	.000	71.679	(13) .000	81.864
	WLC	3	1.294	0.235	0.834	1.754	5.516	.000	8.157	(2) .017	75.481
	Hep&Pharmacological	1	0.486	0.154	0.184	0.789	3.151	.002	0.000	(0) 1.000	0.0000
Modality of comparison group	Active treatment	14	0.553	0.056	0.442	0.663	9.828	.000	67.165	(13) .000	80.645
	Passive treatment	4	0.491	0.111	0.273	0.708	4.421	.000	23.267	(3) .000	87.106
Type of treatment	CBT	4	0.438	0.109	0.223	0.652	4.006	.000	16.781	(3) .001	82.123
	Third wave	4	0.669	0.187	0.303	1.034	3.584	.000	1.335	(3) .721	0.000
	Combined	7	0.558	0.066	0.428	0.688	8.400	.000	70.170	(6) .000	91.449
	Other	3	0.553	0.131	0.297	0.810	4.228	.000	0.951	(2) .622	0.000
Primary outcome measure	BDI	8	0.485	0.066	0.355	0.615	7.321	.000	5.402	(7) .611	0.000
	CAPS	2	1.798	0.342	1.127	2.469	5.254	.000	4.093	(1) .043	75.569
	HAM-D	4	0.478	0.088	0.305	0.650	5.428	.000	15.901	(3) .001	81.133
	Other measure (SCL, YMRS, GSI)	4	0.842	0.176	0.497	1.188	4.779	.000	47.653	(3) .000	93.705

Table 3

Presentation of subgroup analyses with heterogeneity statistics

Note. MDD = Major Depressive Disorder; TAU = Treatment as Usual; WL = Waitlist; Hep = Health Enhancement Program; CBT = Cognitive Behavior Therapy; BDI = Becks Depression Inventory; CAPS = Clinician-Administered PTSD Scale; HAM-D = Hamilton Depression Scale; SCL = Symptom Checklist; YMRS = Young Mania Rating Scale; GSI = Global Severity Index
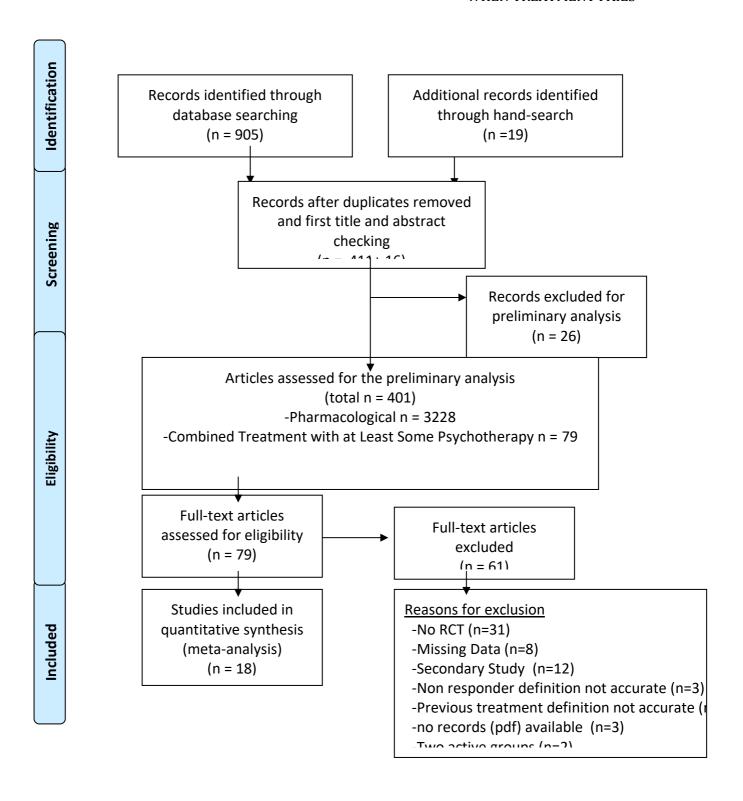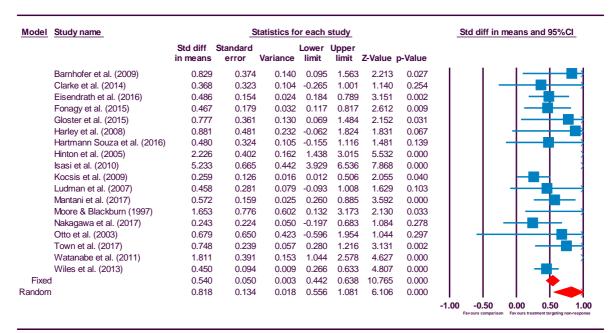
Fig. 1 Study flow chart.

Figure 2. Forest plot of the meta-analysis on primary outcomes (symptoms).

Model	Study name	Statistics for each study							Std diff in means and 95%CI
		Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	
	Barnhofer et al. (2009)	0.829	0.374	0.140	0.095	1.563	2.213	0.027	
	Clarke et al. (2014)	0.368	0.323	0.104	-0.265	1.001	1.140	0.254	
	Eisendrath et al. (2016)	0.486	0.154	0.024	0.184	0.789	3.151	0.002	
	Fonagy et al. (2015)	0.467	0.179	0.032	0.117	0.817	2.612	0.009	
	Gloster et al. (2015)	0.777	0.361	0.130	0.069	1.484	2.152	0.031	
	Harley et al. (2008)	0.881	0.481	0.232	-0.062	1.824	1.831	0.067	
	Hartmann Souza et al. (2016)	0.480	0.324	0.105	-0.155	1.116	1.481	0.139	
	Hinton et al. (2005)	2.226	0.402	0.162	1.438	3.015	5.532	0.000	
	Isasi et al. (2010)	5.233	0.665	0.442	3.929	6.536	7.868	0.000	
	Kocsis et al. (2009)	0.259	0.126	0.016	0.012	0.506	2.055	0.040	
	Ludman et al. (2007)	0.458	0.281	0.079	-0.093	1.008	1.629	0.103	
	Mantani et al. (2017)	0.572	0.159	0.025	0.260	0.885	3.592	0.000	
	Moore & Blackburn (1997)	1.653	0.776	0.602	0.132	3.173	2.130	0.033	
	Nakagawa et al. (2017)	0.243	0.224	0.050	-0.197	0.683	1.084	0.278	
	Otto et al. (2003)	0.679	0.650	0.423	-0.596	1.954	1.044	0.297	
	Town et al. (2017)	0.748	0.239	0.057	0.280	1.216	3.131	0.002	
	Watanabe et al. (2011)	1.811	0.391	0.153	1.044	2.578	4.627	0.000	
	Wiles et al. (2013)	0.450	0.094	0.009	0.266	0.633	4.807	0.000	
Fixed		0.540	0.050	0.003	0.442	0.638	10.765	0.000	
Random		0.818	0.134	0.018	0.556	1.081	6.106	0.000	

-1.00　-0.50　0.00　0.50　1.00

Favours comparison　　Favours treatment targeting non-response

Effectiveness at post-treatment on symptom reduction

Figure 3. Funnel plot showing asymmetry towards studies favoring treatment groups targeting non-responders (red circles represent studies non-existing in the literature).
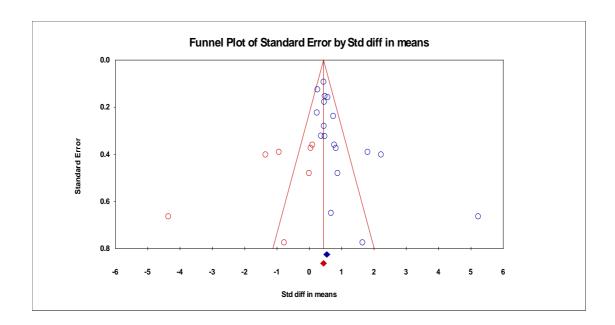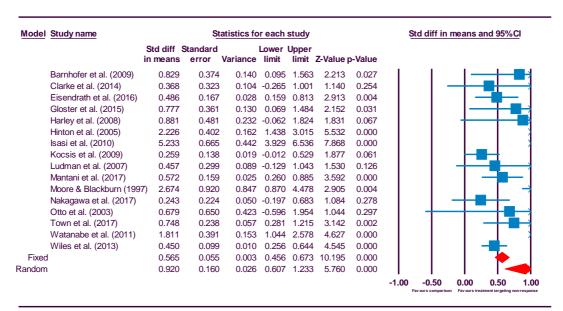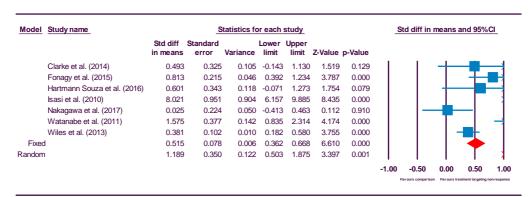
Figure 4. Forest plot of the primary outcomes (symptoms) based on a sensitivity

analysis using treatment completers.

Model	Study name	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Std diff in means and 95% CI
	Barnhofer et al. (2009)	0.829	0.374	0.140	0.095	1.563	2.213	0.027	
	Clarke et al. (2014)	0.368	0.323	0.104	-0.265	1.001	1.140	0.254	
	Eisendrath et al. (2016)	0.486	0.167	0.028	0.159	0.813	2.913	0.004	
	Gloster et al. (2015)	0.777	0.361	0.130	0.069	1.484	2.152	0.031	
	Harley et al. (2008)	0.881	0.481	0.232	-0.062	1.824	1.831	0.067	
	Hinton et al. (2005)	2.226	0.402	0.162	1.438	3.015	5.532	0.000	
	Isasi et al. (2010)	5.233	0.665	0.442	3.929	6.536	7.868	0.000	
	Kocsis et al. (2009)	0.259	0.138	0.019	-0.012	0.529	1.877	0.061	
	Ludman et al. (2007)	0.457	0.299	0.089	-0.129	1.043	1.530	0.126	
	Mantani et al. (2017)	0.572	0.159	0.025	0.260	0.885	3.592	0.000	
	Moore & Blackburn (1997)	2.674	0.920	0.847	0.870	4.478	2.905	0.004	
	Nakagawa et al. (2017)	0.243	0.224	0.050	-0.197	0.683	1.084	0.278	
	Otto et al. (2003)	0.679	0.650	0.423	-0.596	1.954	1.044	0.297	
	Town et al. (2017)	0.748	0.238	0.057	0.281	1.215	3.142	0.002	
	Watanabe et al. (2011)	1.811	0.391	0.153	1.044	2.578	4.627	0.000	
	Wiles et al. (2013)	0.450	0.099	0.010	0.256	0.644	4.545	0.000	
Fixed		0.565	0.055	0.003	0.456	0.673	10.195	0.000	
Random		0.920	0.160	0.026	0.607	1.233	5.760	0.000	

-1.00   -0.50   0.00   0.50   1.00

Favours comparison   Favours treatment targeting non-response

Sensitivity analysis on symptom reduction

Figure 5. Forest plot of effect sizes for primary outcomes at follow-up assessment.

Model	Study name	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Std diff in means and 95%CI
	Clarke et al. (2014)	0.493	0.325	0.105	-0.143	1.130	1.519	0.129	
	Fonagy et al. (2015)	0.813	0.215	0.046	0.392	1.234	3.787	0.000	
	Hartmann Souza et al. (2016)	0.601	0.343	0.118	-0.071	1.273	1.754	0.079	
	Isasi et al. (2010)	8.021	0.951	0.904	6.157	9.885	8.435	0.000	
	Nakagawa et al. (2017)	0.025	0.224	0.050	-0.413	0.463	0.112	0.910	
	Watanabe et al. (2011)	1.575	0.377	0.142	0.835	2.314	4.174	0.000	
	Wiles et al. (2013)	0.381	0.102	0.010	0.182	0.580	3.755	0.000	
Fixed		0.515	0.078	0.006	0.362	0.668	6.610	0.000	
Random		1.189	0.350	0.122	0.503	1.875	3.397	0.001	

-1.00   -0.50   0.00   0.50   1.00

Favours comparison   Favours treatment targeting non-response

Effectiveness at follow-up on symptom reduction

Figure 6. Forest plot of effect sizes for secondary outcomes (quality of life/functioning).

Model	Study name	Statistics for each study							Std diff in means and 95%CI
		Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	
	Clarke et al. (2014)	0.119	0.335	0.112	-0.538	0.776	0.355	0.722	
	Fonagy et al. (2015)	0.532	0.179	0.032	0.181	0.884	2.968	0.003	
	Gloster et al. (2015)	1.367	0.378	0.143	0.626	2.107	3.618	0.000	
	Harley et al. (2008)	1.290	0.505	0.255	0.300	2.280	2.555	0.011	
	Hartmann Souza et al. (2016)	0.024	0.320	0.102	-0.603	0.651	0.075	0.940	
	Kocsis et al. (2009)	0.241	0.140	0.020	-0.034	0.516	1.715	0.086	
	Nakagawa et al. (2017)	0.038	0.224	0.050	-0.400	0.477	0.171	0.864	
	Wiles et al. (2013)	0.473	0.094	0.009	0.289	0.656	5.046	0.000	
Fixed		0.405	0.064	0.004	0.280	0.530	6.341	0.000	
Random		0.411	0.119	0.014	0.177	0.645	3.443	0.001	

-1.00   -0.50   0.00   0.50   1.00

Favours comparison   Favours treatment targeting non-response

Effectiveness at post-treatment on quality of life

Figure 7. Forest plot of effect sizes for secondary outcomes at follow-up assessment (quality of life/functioning).

Model	Study name	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	Std diff in means and 95%CI
	Clarke et al. (2014)	0.292	0.337	0.113	-0.368	0.952	0.867	0.386	
	Fonagy et al. (2015)	0.766	0.183	0.033	0.409	1.124	4.198	0.000	
	Hartmann Souza et al. (2016)	0.688	0.329	0.108	0.043	1.333	2.092	0.036	
	Nakagawa et al. (2017)	0.127	0.224	0.050	-0.311	0.566	0.569	0.569	
	Wiles et al. (2013)	0.322	0.102	0.010	0.122	0.522	3.152	0.002	
Fixed		0.398	0.078	0.006	0.245	0.551	5.103	0.000	
Random		0.426	0.122	0.015	0.187	0.665	3.496	0.000	

Effectiveness at follow-up on quality of life