Accepted Manuscript

The efficacy of Acceptance and Commitment Therapy: An updated systematic review and meta-analysis

Lars-Göran Öst

PII: S0005-7967(14)00121-1

DOI: 10.1016/j.brat.2014.07.018

Reference: BRT 2751

To appear in: *Behaviour Research and Therapy*

Received Date: 7 March 2014

Revised Date: 3 July 2014

Accepted Date: 28 July 2014

Please cite this article as: Öst, L.-G., The efficacy of Acceptance and Commitment Therapy: An updated systematic review and meta-analysis, *Behaviour Research and Therapy* (2014), doi: 10.1016/j.brat.2014.07.018.

The efficacy of Acceptance and Commitment Therapy:

An updated systematic review and meta-analysis

Lars-Göran Öst

Department of Clinical Neuroscience, Division of Psychology, Karolinska Institutet,

SE-171 77 Stockholm, Sweden

Department of Psychology, Stockholm University, Stockholm, Sweden

Abstract

Acceptance and commitment therapy (ACT) has attracted a lot of interest during the last 10-15 years with a strong increase of the number of randomized controlled trials (RCTs). The present review and meta-analysis includes 60 RCTs (4,234 participants) on psychiatric disorders, somatic disorders, and stress at work. The mean effect size across all comparisons was small (0.42). Compared to the Öst (2008) meta-analysis there was no significant improvement in methodological quality and deterioration in effect size (from 0.68). When ACT was compared to various forms of cognitive or behavioral treatments a small and non-significant effect size of 0.16 was obtained. An evidence-base evaluation showed that ACT is not yet well-established for any disorder. It is probably efficacious for chronic pain and tinnitus, possibly efficacious for depression, psychotic symptoms, OCD, mixed anxiety, drug abuse, and stress at work, and experimental for the remaining disorders.

Keywords: ACT, systematic review, meta-analysis, methodological quality, evidence-base

E-mail address: ost@psychology.su.se

Introduction

Acceptance and Commitment Therapy (ACT) has attracted a lot of interest during the last 15 years, since the publication of *Acceptance and commitment therapy* by in 1999 the founders of this treatment, Steven Hayes, Kirk Strosahl, and Kelly Wilson. A search in the database PsycINFO with acceptance and commitment therapy as search word yielded 78 hits 2000-2004, 309 hits 2005-2009, and 500 hits 2010-2014. This also means an almost exponential increment in the number of randomized clinical trials (RCTs). This body of research has been reviewed a number of times, e.g. Hayes (2004), Ruiz (2010), Smout Hayes, Atkins, Klausen, and Duguid (2012), and Swain, Hancock, Hainsworth, and Bowman (2013), which focused specifically on anxiety.

There have been a number of meta-analyses on ACT published during the last decade. Hayes, Luoma, Bond, Masuda, and Lillis (2006) included 18 RCTs and found a mean effect size (ES) of 0.66, Öst (2008) reported a mean ES of 0.68 across 13 RCTs, and Powers, Zum Vörde Sive Vörding, and Emmelkamp (2009) a mean of 0.30 with 18 RCTs. In addition to these general meta-analyses Ruiz (2012) published a meta-analysis which focused on 16 studies comparing ACT and CBT, finding a mean ES of 0.37, that was significant and in favor of ACT.

Why a new meta-analysis? The strong increase in RCTs during the last three years; 9 in 2011, 13 in 2012, and 10 in 2013, means that a large number of RCTs on ACT have never been included in a meta-analysis. This alone warrants an updated meta-analysis which will be able to investigate if the ES of 0.68 in the Öst (2008) paper including 13 RCTs, and the ES of 0.62 in a keynote (Öst, 2009) including 21 RCTs has changed in any direction. It will also enable an updated rating of methodological stringency and a test of whether studies published since the 2008 paper have improved in this respect, and if so in which factors of psychotherapy research methodology.

It is also of interest to update the evaluation of the evidence-base of ACT in light of the many new RCTs that have been published. In my 2008 article and the 2009 keynote I concluded that ACT was not yet a well-established treatment (highest level of empirical support) for any disorder. However, the homepage of the Association of Contextual Behavioral Science refers to websites of various organizations which have information on the evidence base of psychological treatments. Firstly, the Society of Clinical Psychology, Division 12 of the American Psychological Association, states on its website that ACT has strong research support (equals well-established) for chronic and persistent pain in general, and modest research support (equals probably efficacious) for depression, psychotic symptoms, obsessive-compulsive disorder, and mixed anxiety. Secondly, SAMHSA's National Registry of Evidence-Based Programs and Practices listed ACT as an evidence-based treatment in March 2011. However, that decision was based on only three studies (Bach & Hayes, 2002; Bond & Bunce, 2000; Twohig et al., 2010), which is remarkable when 28 RCTs had been published by the end of 2010. There is no information regarding how these three studies were selected.

The aims of the present article were to:

- Update the systematic review and meta-analysis of Öst (2008)
- Compare the early studies (included in Öst, 2008, n = 13) with the later studies (n = 47) regarding methodological stringency and effect size.
- Replicate the Ruiz (2012) comparison of ACT vs CBT in a larger sample of studies.
- Evaluate the evidence-base status of ACT for the different disorders it has been tried for.

Method

Literature search

PsycINFO and PubMed were searched from 1985 to November 2013 with the following search words: Acceptance or ACT, and Randomized controlled trial or RCT or random*. I also used the list of RCTs published on the website of the Association of Contextual Behavioral Science by May 2013.

All abstracts were read and when there was an indication of a group of patients receiving the particular treatment being compared with another group in a randomized clinical trial (RCT) the full-text article was retrieved. Studies using single case designs were excluded since there is no consensus yet regarding the calculation of effect sizes. The reference lists in the retrieved articles were then checked against the database search and any other articles that might fulfil the inclusion criteria were retrieved.

Inclusion criteria

In order to be included in the review and meta-analysis a study had to:

- be published, or in press, in an English language journal
- randomly allocate participants to either treatment and control, or to two or more active treatments
- have participants with either a psychiatric disorder, a somatic disorder, or stress reactions in work situations

Excluded from the review and meta-analysis were:

- Studies with normal people not applying for treatment
- RCTs with only 1-2 components of ACT
- Reanalysis of a subsample from a previously published RCT

Figure 1 shows a flowchart of the inclusions of studies in the current meta-analysis.

Classification of the RCTs

Based on the participants in the studies RCTs were classified as containing a psychiatric disorder (anxiety disorders, depression, mixed anxiety-depression, psychotic symptoms, drug abuse, nicotine dependence, trichotillomania, and borderline personality disorder), a somatic disorder (pain of various types, headache, epilepsy, tinnitus, overweight/obesity, cancer, diabetes and multiple sclerosis), or stress in work situations.

Methodological quality

In order to assess the quality of the research methodology in RCTs various scales have been developed, e.g. the Jadad criteria (Jadad et al., 1996). They are, however, usually constricted to rather few items rated as present or absent. This means that the range of scores is small (e.g. 2-4 in Cavanagh, Strauss, Forder, & Jones, 2014) with ensuing difficulties of showing a relationship between methodological quality and effect size. Based on previous work by Tolin (1999) I developed a scale containing 22 items (Öst, 2008) with a theoretical range of 0-44. When used in my 2008 meta-analysis the total score for the ACT studies ranged from 10 to 27. Thus, there should not be a problem of "restriction-of-range" with this scale.

The psychotherapy outcome study methodology rating scale

The scale consists of the following items: 1. Clarity of sample description, 2. Severity/chronicity of the disorder, 3. Representativeness of the sample, 4. Reliability of the diagnosis in question, 5. Specificity of outcome measures, 6. Reliability and validity of outcome measures, 7. Use of blind evaluators, 8. Assessor training, 9. Assignment to treatment, 10. Design, 11. Power analysis, 12. Assessment points, 13. Manualized, replicable, specific treatment programs, 14. Number of therapists, 15. Therapist training/experience, 16. Checks for treatment adherence, 17. Checks for therapist competence, 18. Control of concomitant treatments, 19. Handling of attrition, 20. Statistical analyses and presentation of

results, 21. Clinical significance, 22. Equality of therapy hours (for non-WLC designs only). Each item is rated as 0 = poor, 1 = fair, and 2 = good, and each step has a verbal description of one or more sentences.

Psychometric data

The internal consistency of the scale was good with a Cronbach's α of 0.81. In order to assess the inter rater reliability of the scale advanced graduate students in clinical psychology received 6 hours of training in the use of the scale by the author, with various outcome studies as training examples. Then the students rated a random selection of 20% of the studies and the ratings were compared with those of the author. The intra-class correlation for the total score was .90, and the *kappa* coefficients on the individual items varied between .50-1.00, with a mean of .73, indicating a good inter-rater reliability.

Meta-analysis

In the current meta-analysis the *primary outcome measure* for each study was used to calculate effect size. If a study did not indicate which was the primary measure I decided which it was based on the disorder focused in the study. In one specific study (Buhrman et al., 2013) the authors designed as primary a measure that other pain studies used as process measure. In this case I decided to use the same measure as most other pain studies described as primary measure. I had to use the data included in each study, which in some studies (mainly older) were completer data and in some studies (mainly more recent ones) were intent-to-treat (ITT) data. When a study presented both sets of data ITT data were used.

The effect size (ES) was calculated as: $(M_{ACT} - M_{comparison})/SD_{pooled,}$ separately for post- and follow-up assessment. Before pooling the effect sizes I screened for statistical outliers, defined as being outside M ± 2SD. Eight (4.1%) of the ESs were outliers. Instead of

deleting those ESs from the analysis *Winsorising* (Lipsey & Wilson, 2001) was used by reducing outliers to the exact value of M + 2SD. The software *Comprehensive Meta-Analysis, version 2.2* (CMA; Biostat Inc., 2010) was used for all analyses and to correct for small samples Hedges's *g* was calculated. Cohen's rule-of-thumb for classification of ES was used; an ES of 0.20-0.49 is considered small, 0.50-079 as moderate, and ≥ 0.80 as large.

Each study contributed with an average of 3.2 ESs (post- and follow-up combined). For studies with more than one measure per time point the ESs were *combined* into a mean ES for that study, in order to include only one ES per study in the pooled mean ES. A random effects model was used since it cannot be assumed that the ESs come from the same population as studies of psychiatric disorders, somatic disorders, and stress at work are included in the meta-analysis.

Heterogeneity among ES's was assessed with the *Q*-statistic and the *I*-square statistic. The possibility of publication bias was analyzed with the trim-and-fill method of Duval and Tweedie (2000) as well as Egger's regression intercept (Egger, Davey Smith, Schneider, & Minder, 1997). Moderator analyses of continuous variables were carried out with meta-regression and for categorical variables with sub-group analysis using the mixed effect model.

Criteria for evidence-based treatments

The criteria developed by the APA Division 12 Task Force (Chambless et al. 1996; 1998) and later modified by Silverman and Hinshaw (2008) were used. The degree of empirical support for various treatments was classified into four categories.

Well-established treatments

1. At least two good group-design studies, conducted by independent research teams, demonstrating efficacy by showing the treatment to be: a) statistically significantly superior to pill or psychological placebo, or to another treatment, or b) equivalent to an already

established treatment in experiments with statistical power being sufficient to detect moderate differences.

2. Treatment manuals were used for the treatment.

3. Conducted with a population, treated for specified problems, for whom inclusion criteria have been delineated in a reliable, valid manner.

4. Reliable and valid outcome assessment measures tapping the problems targeted for change were used.

5. Appropriate data analyses were applied.

Probably efficacious treatments

1. At least two good experiments showing that the treatment is superior (statistically significantly so) to a wait-list control group, or

2. One or more good experiments meeting the well-established treatment criteria with the one exception of having been conducted by at least two by independent investigatory teams.

Possibly efficacious treatments

At least one good study showing the treatment to be efficacious in the absence of conflicting evidence.

Experimental treatments

Treatment not yet tested in trials meeting task force criteria for methodology (modified from Silverman & Hinshaw, 2008).

The Society of Clinical Psychology, which is Division 12 of American Psychological Association, has a website on Research-supported Psychological Treatments. In their evaluation people responsible for this website use the same criteria as the original APA Task Force but with different names. Well-established is called strong research support, probably efficacious is called modest research support, possibly efficacious has no counterpart, and experimental is called no research support. These terms will be used when comparing my

evaluation of ACT's evidence-base status with the evaluation of this website for the disorders that coincide with this review and that of the website.

Results

Description of the ACT studies

The 60 studies originated from USA (n = 29), Sweden (n = 14), Great Britain (n = 7), Australia (n = 5), Finland (n = 2), and one each from Iran, New Zealand, and Spain. A total of 4,234 participants started treatment or control conditions and the attrition rate varied between 0 and 70% with a mean of 21%. The proportion of women varied between 0 (Lappalainen et al., 2013) and 100% (e.g. Zettle & Hayes, 1989) with a mean of 68%. Mean age of the participants across studies was 39.9 years (SD = 9.7; range of study means 14.8-70.8).

Various background data for the ACT RCTs are described in Table 1. Comparisons of studies on psychiatric disorders, somatic disorders and stress in work situations are displayed in Table 2. On 11 of the 13 variables there were no significant differences between the three categories of studies. However, mean age of the samples in psychiatric disorders was significantly lower than that of somatic disorder samples, whereas stress samples did not differ significantly from either of these. The mean number of therapy sessions was significantly higher for studies with psychiatric patients than in studies with stress participants, whereas that of somatic studies did not differ from either.

Methodological data

Table 3 displays the mean scores on the Psychotherapy outcome study methodology rating form for all RCTs and divided on the three categories. An initial test of homogeneity of variances was significant for half of the 22 items in the scale. This is understandable since the

mean for the stress studies was 0.00 for items 4, 7, 8, 11, 16, 17, 18, and 2.00 for items 20, and 22; on these items the variance is zero. Consequently, for comparisons between the three categories the post-hoc test by Games-Howell, which does not assume equal variance, was used. On five of the items studies of psychiatric disorders yielded a significantly higher mean score than studies of somatic studies, which did not differ from stress studies. These were: 7. Use of blind evaluators, 8. Assessor training, 10. Design, 16. Checks for treatment adherence, and 17. Checks for therapist competence. On another two items; 4. Reliability of the diagnosis in question, and 11. Power analysis, studies of psychiatric disorders had higher mean scores than stress studies, while not differing significantly from that of somatic disorder studies. The means for psychiatric and somatic disorders did not differ significantly on six items, whereas both differed from those of stress studies. These were: 1. Clarity of sample description, 2. Severity/chronicity of the disorder, 3. Representativeness of the sample, 5. Specificity of outcome measures, 18. Control of concomitant treatments, and 21. Clinical significance.

Somatic disorder studies never had a higher mean than psychiatric disorders, but stress studies had significantly higher mean than psychiatric disorder studies on two items; 20. Statistical analyses and presentation of results, and 22. Equality of therapy hours. On both of these the difference between psychiatric and somatic disorder studies was nonsignificant. Finally, there were no significant difference between the means for the various disorders on seven items; 6. Reliability and validity of outcome measures, 9. Assignment to treatment ,12. Assessment points, 13. Manualized, replicable, specific treatment programs, 14. Number of therapists, 15. Therapist training/experience, and 19. Handling of attrition. On the total mean score of the scale studies of psychiatric disorders (M = 20.29) did not differ significantly from that of somatic disorders (M = 18.14), whereas both were significantly higher than that of stress studies (M = 12.86).

In order to assess whether recent studies had a more stringent methodology than earlier studies a comparison was made between the 13 studies included in the Öst (2008) review and the 47 studies published since then. Table 4 displays this comparison and there are significantly higher means for recent studies on three of the 22 items. However, on item 11. Power analysis, and item 17. Checks for therapist competence, the early studies had a mean of 0.00, which makes *t*-test unsuitable. Instead the 0-2 scale was dichotomized into 0 vs. 1+2 and tested with Fisher's exact test. This yielded a two-tailed *p*-value of 0.10 for item 11 and 0.18 for item 17. Thus, only item 14. Number of therapists, showed a significantly increased mean in recent studies. The total score increased with only 0.9 points (17.9 to 18.8) a non-significant change.

Designs

The designs used in the RCTs are described in Table 5. There were 66 comparisons in the 60 RCTs and the most common comparison was with some form of CBT (n = 21), followed by WLC (n = 17), and TAU (n = 15).

Specific methodological issues

Treatment-as-usual control groups

A quarter of the RCTs used treatment-as-usual (TAU) as the control condition. This is understandable from two aspects; it is ethically defensible since all patients obtain treatment and the new treatment can be compared with what is the standard care at the clinic in question. However, TAU has a number of drawbacks that rarely are highlighted. Firstly, the treatment is *not constant* but changes across time as the therapists learn new methods. This means that the study does not compare A vs. B, but A_1 vs. A_2, which leads to a lower power and more difficulties showing a significant difference. Secondly, the sessions are usually *not*

recorded and thus adherence and competence cannot be assessed. Thirdly, the patients often get *markedly less therapy hours* than those in the primary treatment. The last factor is illustrated in Table 6. There were 15 RCTs comparing TAU with ACT plus TAU or ACT only. In 10 of the 15 studies the TAU-condition got less therapy hours than the ACT+TAU-condition. This corresponds to a percentage difference varying between 18 and 100, with a mean of 75%. Of the remaining five studies three balanced the treatment time, and two (Petersen & Zettle, 2009; Wicksell, Melin, Lekander & Olsson, 2009) even had somewhat more therapy time for the TAU-condition.

Combining ACT with other treatment(s) or components

In this body of RCTs it is not uncommon that ACT is combined with some empirically supported treatment or components of such a treatment into a package. However, in order to conclude how much, if anything, ACT contributes to the outcome it is necessary to use a *dismantling design*. Twenty of the 60 studies used components from, or complete other treatments, in addition to ACT, but none of the studies used a dismantling design. Examples of this type of studies are Woods et al. (2006) combining ACT with habit reversal training in the treatment of trichotillomania, Gratz and Gunderson (2006) using a combination of ACT, DBT, BT, and emotion-focused therapy for borderline personality disorder, and Lundgren, Dahl, Melin and Kies (2006) and Lundgren, Dahl, Yardi and Melin (2008) combining ACT with behavioral seizure control techniques for patients with epilepsy.

Drawing conclusion of equivalent effects from superiority designs

A non-significant difference on the primary measure does not allow the conclusion that the two compared treatments are equally good. This requires a noninferiority or an equivalence design (e.g. Walker & Nowacki, 2011). However, equivalence can be tested in a superiority

design that yielded a non-significant effect, provided a large enough cell size (at least 30 according to the APA, Division12 Task Force criteria, 1995). The RCTs in this review contained 47 comparisons of ACT with another treatment, and 29 (62%) found no significant difference between them. Reading the abstracts one finds that seven of these described ACT and the compared treatment as yielding equivalent outcomes, however, none did an equivalence test, even if three of the studies (Arch et al., 2012; Forman et al., 2007; Flaxman & Bond, 2010a) had cell sizes of 30 or more.

Lack of statistical power

Psychotherapy outcome studies are usually very expensive and it is questionable to start such a study if it is clearly underpowered, i.e. if the chance of detecting a significant difference is markedly lower than the recommended 80%. The sample power table for *t*-test in Kazdin (2003, p. 444) indicates that if a researcher expects to obtain a large effect size ($d = 0.80$) 26 participants per condition is necessary for 80% power. However, if the expected effect size is moderate ($d = 0.50$) it takes 64, and if it is small ($d = 0.20$) the needed number is a staggering 400 per condition. Using the recommended 80% power and an α of 0.05 at randomization 90% of the ACT-studies would only detect a large effect size, 10% would only detect a moderate effect size, and none a small effect size. When taking attrition into consideration and looking at completers the corresponding figures were 98%, 2%, and 0%, respectively.

Diagnosing the participants

In order for ACT-studies to be compared to other therapies regarding the evidence-base it is important that participants are diagnosed, preferably by employing trained interviewers using established interview schedules (or similar instruments) and assessing inter-rater reliability. Looking at the first issue we find that 23 out of 31 (74%) studies of psychiatric disorders,

13/22 (59%) studies of somatic disorders, and 0/7 stress studies diagnosed the participants, yielding an overall $\chi^2(2) = 13.1$, $p = 0.001$. Pair-wise comparisons with Fisher's exact test showed that psychiatric and somatic studies did not differ from each other, whereas both differed from stress studies ($p = 0.0005$ and 0.008, respectively). However, state-of-the-art in this respect is illustrated in item 4 of the Psychotherapy outcome study methodology rating form on which a score of 2 is defined as "The diagnosis was assessed with structured interview by a trained interviewer *and* adequate inter-rater reliability was demonstrated (e.g. *kappa* coefficient)." In psychiatric disorder studies only six of the 23 studies (26%) that diagnosed the participants got a score of 2 compared to none of the 13 somatic studies, a non-significant difference ($p = 0.068$).

Number of therapists in a study

If only one therapist is used in a RCT there is a complete confounding between therapist and therapy method, and, consequently, it is not possible to ascribe a certain outcome to the therapy applied. In studies (e.g. Zettle, 2003; Zettle & Hayes, 1986; Zettle & Rains, 1989) where only one therapist is doing both the compared therapies the therapist factor is controlled to some extent. However, unless adherence and competence ratings are provided in the article it is impossible to conclude that this single therapist carried out both treatments with equal adherence and competence. Table 7 shows the distribution of number of therapists across type of disorder in the RCTs. As can be seen the mode number is one; fully 33% of the studies had only one therapist, and another 18% had only two therapists. This means that confounding is quite prevalent in the ACT RCTs.

Adherence and competence ratings

Adherence refers to the extent to which *specified procedures* are used by the therapist during the treatment, whereas competence concerns the degree of *skill and judgment* the therapist displays when carrying out the treatment (Barber, Sharpless, Klosterman, & McCarthy, 2007). These constructs are usually highly correlated (e.g. Barber et al., 2003) but cannot replace each other. A therapist can be highly adherent to the procedures in the manual, but not being particularly competent in the therapy situation. The opposite, i.e. a highly competent therapist who is not adherent, is more difficult to envisage. In such a case the therapist is probably doing some other therapy than he/she was supposed to do.

Adherence was assessed by only 13 (23%) of the studies; 11 of the 31 (35%) psychiatric disorder studies and 2 of the 22 (9%) somatic disorder studies, but none of the stress studies. A chi-square test yielded a significant $\chi^2(2) = 7.47$, $p = 0.02$. However, 33% of the cells had expected frequencies of less than 5 and the chi-square is not reliable. Pairwise comparisons with Fisher's exact test showed that psychiatric disorder studies differed significantly from stress studies ($p = 0.0497$), whereas somatic studies did not ($p = 1.0$). Competence was evaluated in only 8 (13%) of the studies; 6 (19%) of the psychiatric, 2 (9%) of the somatic disorder studies, and none of the stress studies. A chi-square test yielded a non-significant $\chi^2(2) = 2.39$, $p = 0.30$.

Lack of credibility ratings

When two treatments are compared to each other in a RCT the patients' perceived credibility of the respective treatments are important to assess since differences in this respect may be a threat to internal validity of the study. Fully 47 of the 60 studies (78%) were comparisons between two, or more, active treatments. However, only 4 (8.9%) of these studies included credibility ratings.

Statistical analyses of dichotomous data

When an article that employs a dichotomous measure, e.g. proportion of patients reaching clinically significant change or being diagnosis-free after treatment, and describes the number of participants achieving this outcome it is possible for a reviewer to recalculate the statistical test. A total of 29 studies reported statistical tests of dichotomous measures. In seven of these ACT did not differ significantly from the comparison condition, and my recalculation gave the same result. In 22 of the studies ACT was found to be significantly better than the comparison condition, and in 12 of these (55%) I got the same result. In 5 studies the choice of statistical test was questionable, and in another 5 an incorrect test was used. This can be illustrated by a couple of examples.

England et al. (2012) compared two forms of exposure rationales for subjects with social anxiety disorder; one acceptance based and one habituation based. At post-treatment the number of diagnosis-free subjects was 21/21 in the acceptance condition compared to 20/24 in the habituation condition, and the authors used Pearson's Chi-square, obtaining a value of 3.84 ($p = 0.05$). However, two of the cells in the 2X2 table had an expected value of less than 5, which means that Chi-square is unreliable. In this situation Fisher's exact test is strongly recommended and using this test the *p*-value is 0.112, which is non-significant.

Lanza and Menéndez (2013) worked with incarcerated female drug-addicts comparing ACT and WLC. At post-treatment 5/18 in the ACT-condition were drug-free compared to 1/13 in the WLC. The authors reported a Pearson Chi-square of 20.48 ($p = 0.000$), ignoring the fact that two of the cells had expected values less than 5. When applying Fisher's exact test a non-significant *p*-value of 0.359 was obtained. At 6 month follow-up the numbers of drug-free subjects were 7/16 in ACT and 2/11 in WLC, with a Chi-square of 6.09 ($p = 0.014$) according to the authors. In this case Fisher's exact test also yielded a non-significant *p*-value of 0.231.

Meta-analysis

Table 8 shows the results of the meta-analysis at post-treatment and follow-up assessment for all comparisons and divided on the different types of comparison conditions. At post-treatment the overall Hedges's *g* was small (0.42) but significantly different from zero. Both indices of heterogeneity were also significant. The effect sizes for comparisons with waiting-list (0.63), placebo (0.59), and TAU-conditions (0.55) were moderate and also significantly heterogeneous. The ESs for WLC- and TAU-comparisons, but not the placebo conditions were significantly different from zero. The ES for comparisons with various active treatments (0.22) just reached the limit for a small effect size. It differed significantly from zero and had significant heterogeneity. Finally, the ES for the comparison between ACT and different forms of CBT- or BT-treatments (0.16) did not reach the lower limit for a small effect size and was not significantly different from zero.

At follow-up assessment, on average 4.8 months after the end of treatment, the overall ES (0.30), as well as those for various comparison groups had been reduced somewhat. Now neither the comparisons with active treatments in general (0.17), nor that with CBT/BT-treatments in particular (0.06), reached the limit for a small ES. The effect sizes for all studies, WLC-, and TAU-studies, were significantly different from zero, whereas the other were not. The heterogeneity was significant for the overall and active treatment ES, but not for the other comparisons.

Publication bias

The analyses of possible publication bias used both the trim-and-fill method and Egger's regression intercept. The results are shown in Figure 2 and Table 9 and it is evident that publication bias is a problem for the ACT RCTs. Regarding the overall ES the trim-and-fill

method suggested that 13 studies should be trimmed which would reduce the mean ES from 0.42 to 0.28. Concerning WLC- and TAU-comparisons seven studies should be trimmed, in each case leading to marked reductions of the ES, and for CBT/BT-comparisons one study should be trimmed. For the overall ES, WLC- and TAU-comparisons Egger's regression intercept also yielded significant *t*-values. For placebo-, active treatment-, and CBT/BT-comparisons publication bias did not seem to be a concern.

Moderator analyses

The following continuous variables were analyzed with the meta-regression module in the CMA program using fixed effect analysis: number of participants starting therapy, number of participants completing therapy, percent attrition in the ACT-condition, proportion of females, mean age of the participants, number of therapists, number of therapy hours, number of additional therapy components, and methodological quality of the study. Three of these yielded a significant point estimate of the slope. Studies with higher proportion of women ($z = 2.08$, $p = 0.038$), were associated with higher ES whereas studies with fewer number of therapists ($z = -2.99$, $p = 0.003$) and lower methodological quality scores ($z = -2.16$, $p = 0.031$) were associated with higher ES.

For categorical variables sub-group analyses were employed in the CMA program and the results are displayed in Table 10. Three of the variables yielded significant $Q_{between}$ values. Regarding type of comparison condition passive conditions (WLC) resulted in higher ES than active treatments. Concerning country studies from the European Union yielded higher ES than studies from USA and other countries (primarily Australia). Finally, type of disorder was also significant with studies on psychiatric disorders having lower ES than studies on somatic disorders and stress in the work place. However, type of outcome measure, way of

recruiting participants, treatment format, and phase of publication did not affect ES significantly.

The evidence-base status of ACT

Psychiatric disorders

The RCTs on psychiatric disorders are summarized in relation to the criteria for well-established empirically supported treatments in Table 11. As in the meta-analysis I have only used the primary outcome measure when evaluating how ACT fared in relation to the various comparison conditions.

Depression. No study compared ACT to a placebo condition. Two compared ACT with an established treatment (cognitive therapy) but did not achieve a significantly better effect. Three studies compared ACT with TAU and two of these got significant effects. However, both of these (Folke et al., 2012; Hayes et al., 2011) used more therapy hours for ACT than for TAU (see Table 6). None of the studies fulfilled criterion 3 (inclusion criteria reliably delineated). Since there are various other methodological problems with these studies my evaluation is that ACT is *possibly efficacious* for depression. In contrast, the website of Division 12 found that ACT had modest research support.

Psychotic symptoms. Shawyer et al. (2012) compared ACT with a placebo treatment (Befriending) without achieving a significantly better effect. The other three studies in this category compared ACT with various TAU-conditions, but only one (Bach & Hayes, 2002) found a significant difference in favor of ACT. However, there is a question mark for the psychometric characteristics of the outcome measures in this study. Furthermore, none of these studies fulfilled criterion 3. There are additional methodological issues with these studies and my evaluation is that ACT is *possibly efficacious* for psychotic symptoms. Division 12 considers ACT to have modest research support.

Anxiety disorders. This is the largest sub-category within psychiatric disorders, which is no wonder since there are eight different anxiety disorders in DSM-IV (APA, 1994). These need to be evaluated separately. There are three studies on generalized anxiety disorder; one comparing ACT with CBT (Wetherell et al., 2011b), one comparing with WLC (Roemer et al., 2008), and one with applied relaxation (Hayes-Skelton et al., 2013). The latter two fulfill criteria 2-5 and I evaluate ACT as *probably efficacious* for GAD. However, the treatment used by the Roemer et al. group is a combination of CBT, dialectical behavior therapy, mindfulness base cognitive therapy and ACT, and in the absence of a dismantling design it is impossible to know what ACT's contribution to the outcome effect is.

There are two studies on social anxiety disorder (SAD). England et al. (2012) did not find ACT to be better than exposure in-vivo and failed criterion 4 and 5, and Kocovski et al. (2013) reported no significant difference between ACT and CBT, but failed criterion 3. My evaluation is that ACT is *possibly efficacious* for SAD.

There are two studies on test anxiety. Zettle (2003) found that ACT did not differ significantly from systematic desensitization but failed criterion 3. Brown et al. (2011) reported no significant difference from CBT and also failed criterion 3. Both of these studies have various methodological problems, e.g. participants not being diagnosed, only one therapist, and no adherence and competence ratings. My evaluation is that ACT is *possibly efficacious*.

There is one study on obsessive-compulsive disorder (OCD). Twohig et al. (2010) found ACT to be significantly better than progressive relaxation training, which, however, never has been an established treatment for OCD. The study fulfilled criteria 2-5 and my evaluation is *possibly efficacious*, which is in disagreement with Division 12 saying that ACT has modest research support for OCD.

Finally, there is one study (Arch et al., 2012) on mixed anxiety (panic disorder, GAD, SAD, OCD and specific phobias). The study found no significant difference between ACT and CBT and fulfilled criteria 2-5. However, since each diagnostic subcategory was too small to allow statistical analysis it is premature to draw strong conclusions about ACT's evidence status based on this study. My evaluation is *possibly efficacious* which disagrees with Division 12 saying modest research support. Mixed anxiety is not a diagnosis and this study cannot be used as evidence for ACT being efficacious across the five anxiety disorders included in the study.

Drug abuse. There are five studies on drug abuse/dependence. Hayes et al. (2004) worked with opiate addicts and found that ACT plus methadone maintenance (MM) was significantly better than MM alone but did not differ from Intensive 12-step facilitation. Smout et al. (2010) treated subjects with methamphetamine abuse/dependence and found that ACT did not differ from CBT. This study had an astonishing attrition of 70% and did not fulfill criterion 4. Louma et al. (2011) worked with various types of drug abusers and found ACT to be significantly better than TAU. Stotts et al. (2012) focused on methadone detoxification in opiate addicts and found no significant difference between ACT and drug counseling. Finally, Lanza and Menéndez (2013) worked with incarcerated women with a mix of drug abuse and found no significant difference between ACT and WLC. This study did not fulfill any of the five criteria. A common feature of the drug abuse studies is that none fulfill criterion 3. My evaluation is that ACT is *possibly efficacious*.

Nicotine dependence. There are three studies on smoking cessation for people with nicotine dependence. Gifford et al. (2004) did not find that ACT was significantly better than nicotine replacement treatment at post-treatment (but at a naturalistic follow-up). Gifford et al. (2011) reported that the combination of ACT, functional analytic psychotherapy, and bupropion was significantly better than bupropion alone. Bricker et al. (2013) tested a web-

based ACT against the so called Smokefree.gov intervention and found no significant difference. The latter study did not fulfill criterion 4 and was questionable regarding criterion 5. None of the studies in this category fulfilled criterion 3. My evaluation is that ACT is *experimental* when it comes to nicotine dependence.

Borderline personality disorder (BPD). There are two studies of ACT for BPD and both found that ACT was significantly better than TAU. Gratz and Gunderson (2006) combined ACT with DBT, BT, and emotion focused therapy, which means that the contribution of ACT is impossible to ascertain. Morton et al. (2012) included participants who fulfilled only 4/9 criteria for BPD when DSM-IV stipulates 5. Both studies gave the TAU-treated subjects markedly less therapy hours (see Table 6) and did not fulfill criterion 3. My evaluation is that ACT is *experimental* regarding BPD.

Various disorders. Woods et al. (2006) worked with trichotillomania and found that the combination of ACT and habit reversal training was significantly better than WLC. Forman et al. (2007) worked with a mix of anxiety and depression in subjects applying for treatment at a student counseling center and found no significant difference between ACT and CBT. This is one of the few studies that did not use a treatment manual. Finally, Lappalainen et al. (2007) included participants with various emotional disorders applying for treatment provided by psychology students at a university training program. They found that ACT was significantly better than CBT. None of the studies in this category fulfilled criterion 3. My evaluation is that ACT is *experimental* in this category.

Somatic disorders

The RCTs on somatic disorders are summarized in relation to the criteria for well-established empirically supported treatments in Table 12.

Pain. This is the largest category with 10 studies focusing on pain in general. Two studies compared ACT with another treatment and not finding a significant difference. Thorsell et al. (2011) used applied relaxation and Wetherell et al. (2011a) used CBT as the comparison condition. The Thorsell study has a question mark regarding the outcome measure since they only used one item to assess pain intensity, whereas the other studies in this section had scales with multiple items.

Four studies compared ACT with different forms of TAU. Two found ACT to be significantly better than TAU (Dahl et al., 2004; Motamedi et al., 2012), whereas two found no significant differences (Wicksell et al., 2009; McCracken et al., 2013). The remaining four studies compared ACT with WLC; three found ACT to be significantly better (Wicksell et al., 2008; Wicksell et al., 2013; Buhrman et al., 2013) and one (Johnston et al. 2010) found no difference. None of the studies in this category fulfilled criterion 3.

It should be noted that ACT was combined with exposure in all three studies from the Wicksell group. Their studies also included different subgroups of pain patients: people with whiplash-associated disorders recruited from the Swedish Association of Survivors of Traffic Accidents and Polio in the 2008 study, children and adolescent patients at a pain treatment service of a children's hospital in the 2009 study, and left-handed women with fibromyalgia in the 2013 study. My EST-evaluation is *probably efficacious* in contrast to the Division 12 web-site that said strong research support (for chronic and persistent pain in general).

Epilepsy. There are two studies on ACT for epilepsy, both coming from the same research group. In both of these ACT was combined with behavioral seizure control techniques which in earlier research (Dahl, Brorson & Melin, 1992) has been found effective for epilepsy. Lundgren et al. (2006) found the combination to be significantly better than a placebo condition, supportive therapy, which has been used in some GAD research but never tested in epilepsy. In the absence of credibility ratings it is impossible to know if the patients

experienced this treatment as credible as the combination patients experienced their treatment. In the Lundgren et al. (2008) study the combination was not significantly better than yoga (unless a questionable statistical analysis is used as the authors do). This study was done in India and it is unclear if yoga should be considered as TAU or an established treatment for epilepsy in that society. Both of these studies fulfilled criterion 3 and my evaluation is *experimental*.

Tinnitus. There are two studies on tinnitus and both come from the research group of Gerhard Andersson at Linköping University, Sweden. Westin et al. (2011) found that ACT was significantly better than WLC as well as tinnitus retraining treatment (TRT) an established treatment in audiology. Hesser et al. (2012) reported that ACT was significantly better than a placebo condition (an online discussion forum) but not better than Internet-based CBT. In the light of strong methodological features in these studies my evaluation is *probably efficacious.*

Overweight/obesity. This category contains a wide variety of studies. Weineland et al. (2012) worked with patients who had undergone bariatric surgery and found ACT to be significantly better than a briefly described TAU. Forman et al. (2013a) focused on craving for sweets in obese women and found no significant difference between ACT and CBT. Forman et al. (2013b) combined both ACT and standard behavior therapy (BT) with 10 different weight loss components and found no significant differences overall. When the treatment was delivered by experts ACT was better than BT. Lillis et al. (2009) found ACT to be significantly better than WLC, whereas Tapper et al. (2010) found no difference between ACT (used alongside the participants' own weight loss plans) and WLC. Three of the studies in this category (Lillis et al., 2009; Weineland et al., 2012; Forman et al., 2013) did not fulfill criterion 3. My evaluation is *possibly efficacious.*

Various disorders. Gregg et al. (2007) worked with diabetic patients and found that the combination of ACT and education was significantly better than education alone. Nordin and Rorsman (2012) studied patients with multiple sclerosis and found that a shortened version of applied relaxation (5 sessions) was significantly better than ACT. Finally, Rost et al. (2012) compared ACT with TAU in late-stage ovarian cancer patients, finding that ACT was significantly better. Both the Nordin and Rost studies failed to fulfill the second criterion, using treatment manuals. My evaluation is that ACT is *experimental* for these disorders.

Stress at work

The RCTs on stress at work are summarized in relation to the criteria for well-established empirically supported treatments in Table 12. One study compared ACT with a placebo condition. Bond and Bunce (2000) found that ACT was significantly better than the Innovation Promotion Program and WLC. Two studies compared ACT with established treatments. Flaxman and Bond (2010a) found no significant difference between ACT and Stress inoculation training, but ACT was better than WLC. Bethay et al. (2012) compared ACT and Applied Behavior Analysis (ABA) with ABA alone and found no significant difference. The remaining studies in this category compared ACT with WLC; Flaxman and Bond (2010b), Brinkborg et al. (2011), and Lloyd et al. (2013) all found ACT to be significantly better than WLC. Finally, Lappalainen et al. (2013) combined ACT with 5 treatment components of different kinds but did not find that the intervention was significantly better than WLC. None of the studies in this category fulfilled criterion 3. My evaluation of ACT for stress at work is *possibly efficacious*.

Summary

To summarize this review of the evidence-base I have to conclude that ACT is not yet a well-established treatment for any disorder. ACT is *probably efficacious* for chronic pain and tinnitus, whereas it is *possibly efficacious* for depression, psychotic symptoms, OCD, mixed anxiety, drug abuse, and stress at work. Finally, ACT is *experimental* for nicotine dependence, borderline personality disorder, trichotillomania, epilepsy, overweight/obesity, diabetes, multiple sclerosis, and ovarian cancer.

Discussion

The aims of this systematic review and meta-analysis were to: (1) Update the systematic review and meta-analysis of Öst (2008), (2) Compare the early studies with the later studies regarding methodological stringency and effect size, (3) Replicate the Ruiz (2012) comparison of ACT vs CBT in a larger sample of studies, and (4) Evaluate the evidence-base status of ACT for the different disorders it has been tried for.

Meta-analysis data

Effect sizes

Concerning the first aim it is possible to conclude that ACT has been tested for a fairly large number of psychiatric disorders (n = 7), somatic disorders (n = 6), and stress at work. A comparison of the ESs obtained 2008 and now gave the following picture. The overall ES decreased from 0.68 to 0.42, the WLC-comparison ES from 0.96 to 0.63, the TAU-comparisons from 0.79 to 0.55, and the active treatment comparisons from 0.53 to 0.22. The lower ES overall and for active treatment comparisons is probably explained by the fact that 18 of the 21 (86%) studies comparing ACT with CBT/BT are included in the later studies and the mean ES for these comparisons is the lowest of all subgroups. The overall ES had

significant heterogeneity which was also the case in the meta-analyses of Öst (2008) and Ruiz (2012), but not that of Powers et al. (2009). This finding was followed by moderator analysis.

Moderators

The meta-regression analyses showed that three variables significantly moderated the effect size. Studies with higher proportion of women were associated with higher ES, whereas studies with lower number of therapists and lower methodological quality scores were associated with higher ES. Why ACT-studies with higher proportion of women should yield higher ESs is difficult to explain and this was not the case in the Ruiz (2012) meta-analysis. The finding that studies with lower number of therapists was associated with higher ES indicates that a study which only has one or two therapists may get an inflated ES, especially if that therapist is one of the originators of ACT (e.g. Zettle, 2003; Zettle & Rains, 1989).

The finding that low methodological quality was associated with high ES is interesting since the majority of meta-analyses that have used this as a moderator have failed to find a relationship. The most parsimonious explanation for this is restriction-of-range, which can be illustrated with some recent meta-analyses. Hofmann, Wu, and Boettcher (2014) used the EPHPP rating system with a range from 1 to 3, Cuijpers et al. (2014) used the "Risk of bias" assessment tool with just 4 dichotomous items with a range from 0 to 4, and Cavanagh et al. (2014) used the Jadad criteria with 5 dichotomous items and an actual range of 2-4. In none of these studies did methodological quality turn out to be a significant moderator of ES. In the present meta-analysis the *Psychotherapy outcome study methodology rating scale* (Öst, 2008*)* with a theoretical range of 0-44, and an actual range of 10-34 was used. It is probable that a number of meta-analyses which have failed to find a significant relationship between methodological quality and ES have incorrectly concluded that no relationship existed when in reality they used a measure of methodological quality that was not sensitive enough.

The sub-group analyses indicated that three variables significantly moderated outcome. The first is trivial; comparisons with a passive condition (WLC) yielded higher ES than those with an active comparison. The second and third are related. Studies on somatic disorders and stress at work gave higher ESs than studies of psychiatric disorders, and studies carried out in Europe yielded higher ES than studies from the USA. It turns out that 14 of the 22 (62%) somatic disorder studies are done in Europe and only 5 (23%) in the USA, whereas only 4 of the 31 (13%) studies on psychiatric disorders emanate from Europe compared to 23 (74%) from USA. Thus, it is not the case that American therapists are worse than European but it seems to be more difficult to obtain a high ES in psychiatric disorders than in somatic disorder.

Publication bias

The analysis of publication bias indicated that this is a real problem for the current meta-analysis. The ES for all ACT studies and the subgroups of WLC- and TAU-controlled studies were all significantly inflated due to publication bias. This means that the obtained ESs have to be interpreted with caution. Similar results have been reported by Cuijpers et al. (2010) in a meta-analysis of psychological treatments for depression. They analyzed 117 trials with 175 comparisons obtaining a mean ES of 0.67, which was reduced to 0.42 after adjustment for publication bias, a reduction of 37%. In the present meta-analysis the mean ES of 0.42 was reduced to 0.28, a reduction of 33%.

Methodological quality

Regarding the second aim I found that the overall methodology score had only increased with 0.9 points from the mean obtained in the 2008 meta-analysis, which was not significant. When testing each individual item only one of the 22 showed a significant improvement with

later studies; Number of therapists in the study had increased. This can be compared with the same kind of analysis done by Smout et al. (2012) comparing the scores for the 13 early studies with 17 later studies (retrieved by January 2012). They found that the items Checks for treatment adherence and competence increased significantly and there was a trend for Reliability of diagnosis and Number of therapists. I also found non-significant trends for the same items (see Table 4).

In my 2008 review the overall mean for 32 CBT-studies, matched for publication year (but not for disorder) with the 3rd wave studies, was 27.8 (SD = 4.2) compared to 18.6 (SD = 5.3) for the 60 ACT-studies in the current review. This difference is statistically significant ($t(90) = 8.49$, $p<0.0001$). The CBT-studies were published 1986-2007 but it can be assumed that if studies from 2008 onwards were included the mean would at least be similar to that of the older studies, and probably increased somewhat as did the ACT-studies' mean, and such a hypothetical comparison would still yield a significant difference. In order for the CBT-mean not to be significantly higher than the ACT-mean it has to be 20.7 (assuming the same SD as for the earlier studies). To reach the same number of studies as we have for ACT 28 CBT-studies have to be added, with a mean score of 12.6. It is hard to envisage such a low score for recent CBT-studies since the range for the 32 CBT studies in the 2008 article was 19-36.

Specific methodological issues

In the Result section I described nine methodological problems that are present in this body of ACT RCTs. Four of these concern the *design* of the study. First is the use of TAU as comparison condition. In two thirds of these studies the TAU-treated patients received markedly less therapy hours that those in the ACT condition and in 11 out of 15 studies the TAU was carried out by other therapists with almost no information provided about their training background and therapy experience. Furthermore, the TAU-sessions were not

recorded, and consequently there were no ratings of therapists' adherence to a treatment manual (if any was used) and competence in doing the therapy. These factors pose threats to the internal validity of the RCT and, thus, I strongly recommend not using TAU in future research on ACT.

Combining ACT with other established treatments or components of such is of course allowed but then it is not possible to ascribe the obtained effect to ACT since no study has used a dismantling design. Unfortunately, some authors are unclear in this respect whereas others label the intervention in a specific way, e.g. Acceptance Based Behavior Therapy (Roemer et al., 2008; Hayes-Skelton et al., 2013).

The remaining design problems are related; lack of power analysis (and in fact a low power in many studies) and drawing conclusions about equivalence from superiority designs yielding a non-significant difference. So many years have passed since Cohen (1962; 1988) enlightened the field about the necessity of doing a power analysis before starting a RCT that you would expect this to be as self-evident as randomization. However, this is not the case and on the Psychotherapy outcome study methodology rating form (item 11) only 6 (10%) of the studies got a rating of 2 "A data informed power analysis was made and the sample size was decided accordingly". Another 4 (7%) got a rating of 1 "A power analysis based on an estimated effect size was used", whereas in an overwhelming majority of studies (83%) no information about power analysis being made prior to the initiation of the study was given. The lack of power analysis led to the situation that 90% of the RCTs had a power of 80% only to detect a large ES and 10% a moderate ES. Since the mean ES (0.42) across all studies was in the small range it is no wonder that 51% of the ACT vs. some comparison group yielded a non-significant difference. A closer look at Tables 11 and 12 also shows that this proportion increased with the strictness of the comparison condition; 24% in WLC-, 25% in placebo, 33% in TAU and 79% in active treatment comparisons.

When a RCT obtains a non-significant difference between two active treatments it is understandable that the authors want to conclude that the treatments are equally effective. However, a design with just two active treatments and no control condition, which is the case for 86% of these studies, does not allow a differentiation between equally effective or equally ineffective. In order to test if the two active treatments are statistically equivalent, and not only non-significantly different due to a small sample size, an equivalence analysis (Rogers, Howard, & Vessey, 1993) is necessary. None of the studies did one, which might be explained with the small samples in most studies. However, the few studies with a cell size of at least 30 did not do it either.

Then there are a number of *procedural* problems. First, a large proportion of studies (40%) did not diagnose the patients and only 10% got a rating of 2 on this item: "The diagnosis was assessed with structured interview by a trained interviewer *and* adequate inter-rater reliability was demonstrated (e.g. *kappa* coefficient)". Just saying in the method section of a RCT that patients with a certain diagnosis, e.g. depression according to the criteria of DSM-IV, were included is not enough. Without indication of sufficient reliability of the diagnostic procedure the reader does not know if the patients really have the disorder in question. This makes comparisons to other studies with reliably diagnosed patients very difficult. It is hard to understand why the issue of diagnosing is handled in this way since there does not seem to be an ideological resistance towards diagnosing in the ACT community.

Another problem is the frequent use of just one (33%) or two (18%) therapists in the RCTs. The confounding of therapists and treatment method makes it impossible to draw unequivocal conclusions from the study. It is hard to understand why studies do not train and hire more therapists and divide the total number of patients among them instead of having just one or two. A related problem is the low frequency of studies assessing adherence (23%) and

competence (13%) of the therapists carrying out the treatment. This is especially important when ACT is compared with an active treatment since we need to know if both treatments have got a "fair chance" in the study; did the therapists adhere to the respective manuals equally well and were they enough and equally competent doing so? Adherence and competence ratings are important in any RCT but when two active treatments are compared it is also necessary to include patient credibility ratings. Only 9% of these studies used credibility ratings and the lack of this makes the reader wonder if the patients in the compared conditions believed in the treatment they were going to get equally much. This is important since the placebo component in therapy is substantial (e.g Hofmann & Smits, 2008). These procedural problems are also threats to the internal validity of the studies.

Finally, there is an issue of statistical tests in some RCTs. My reanalysis of tests used for dichotomous data in 2X2 tables indicated that in 45% of the cases when the authors of studies found that ACT was significantly better than the compared condition the choice of test was questionable or incorrect. This makes me wonder about the statistical tests of continuous variables (e.g. ANCOVA, HLM, Mixed-models) which cannot be reanalyzed without access to the data set. Hopefully these are correct since it takes much more work to carry out these tests than doing a simple Chi-square.

Is ACT better than CBT?

When it comes to the third aim, the replication of the Ruiz (2012) meta-analysis with a larger sample of studies, 21 studies were included. There was an overlap with 12 of the 16 studies in Ruiz's meta-analysis. The following studies failed to fulfill the inclusion criteria for the current meta-analysis; Bond and Bunce (2000) did not use a CBT-method as comparison (innovation promotion program), Block (2002) was an unpublished doctoral dissertation, Páez et al. (2007) was published in Spanish, and Hernández-Lopez et al. (2009) used a quasi-

experimental design. Judging from the forest plot in Figure 1 of the Ruiz (2012) article these four studies had effect sizes between 0.5 and 1.1 approximately, and three of them are in the top half. The deletion of these studies together with 10 new studies not included in Ruiz's meta-analysis led to a mean ES that was not significant, neither at post-treatment (0.16) nor at follow-up (0.06). Thus, the conclusion that can be drawn is that ACT does not lead to significantly higher effect sizes than CBT/BT in randomized studies with direct comparison of these forms of therapy.

The state of ACT's evidence-base

The detailed scrutiny of the ACT RCTs in relation to the criteria for empirically supported treatments led to the following conclusions. (1) ACT is not yet a well-established treatment for any disorder. (2) ACT is *probably efficacious* for chronic pain and tinnitus, whereas it is *possibly efficacious* for depression, psychotic symptoms, OCD, mixed anxiety, drug abuse, and stress at work. (3) Finally, ACT is *experimental* for nicotine dependence, borderline personality disorder, trichotillomania, epilepsy, overweight/obesity, diabetes, multiple sclerosis, and ovarian cancer.

For five of the disorders ACT is included in the evaluation published on the website of Society of Clinical Psychology, Division 12 of APA. Consistently, the authors of that website evaluated ACT's evidence base to be at one step higher than what I arrived at. They said that ACT had strong research support for chronic pain in general, whereas my evaluation was probably efficacious. They said modest research support for depression, psychotic symptoms, OCD, and mixed anxiety, whereas I said possibly efficacious.

Why do we arrive at different conclusions? We both apply the APA Task Force criteria but the difference is probably due to the interpretation of the term "good group-design studies" in criterion 1. To illustrate this issue we can take a closer look at depression for

which there are five RCTs of ACT. The studies by Zettle and Hayes (1986) and Zettle and Rains (1989) suffer from various methodological problems, e.g. patients were not diagnosed, there were no credibility, adherence or competence ratings, and there was only one therapist for all conditions, no power analysis was done, and potential concomitant treatments were not controlled. The Petersen and Zettle (2009) study on depression in alcohol use disorder patients had e.g. unclear diagnostic procedures, only one therapist for ACT and other therapists for TAU, no check for competence, no power analysis, no control of concomitant treatments, and no handling of attrition. The Folke et al. (2012) study did not diagnose the patients but relied on the diagnosis by the referring physician, there was no blind evaluator of outcome, antidepressants and other concomitant treatments were not controlled, no credibility, adherence or competence ratings, and the TAU-condition got 16 hours less therapy. Finally, the Hayes et al. (2011) study on depressed adolescents used DAWBA as diagnostic method and the validity of this web-based method has not been investigated, there was no blind evaluator of outcome, no power analysis, no credibility, adherence or competence ratings, no control of concomitant treatments, and the TAU-condition got 5 hours less therapy. When these methodological problems are taken into consideration I find it impossible to conclude that these studies are "good group-design studies" as stipulated in the criteria for well-established, probably efficacious, and possibly efficacious treatments. Obviously, the editors of the Division 12 website are of another opinion, which they are entitled to. However, it would be interesting to find out how many, and how serious, methodological flaws a study need to have in order for it not to be considered as a good group-design study.

It can be argued that an evaluation of the evidence-base should be done by a committee of people, as was done by the original APA Task Force. However, as a BT- and CBT-researcher of more than 40 years I should be allowed to provide my well-founded opinion on the question. Also, the Division 12 website has only one or two section authors

for each of the various disorders and there is no information about any committee discussion before decisions are made concerning the empirical support of the treatments.

Recommendations for future research

In my 2008 article I listed 15 recommendations for future research and it is interesting to take a look at these to see what, if anything has changed. I am fully aware of the fact that six years might be too short a time period to observe any changes. With this caveat we can compare the 13 early studies with the 47 later studies on the 15 recommendations.

1) Don't use WLC as the control condition, since criterion 1 requires a placebo or another treatment. WLC *increased* from 7.7% to 25.5% of the studies.

2) Don't use TAU as the control condition, since the methodological problems described above are so extensive. TAU decreased somewhat from 30.8% to 23.4%.

3) Use an active treatment as comparison, preferably one that has been established as effective for the disorder in question. Active treatment *decreased* somewhat from 61.5% to 51.1%.

4) Do a proper power analysis before the start of the study and adjust the cell size for the attrition that may occur. Mean score increased non-significantly from 0.00 to 0.34.

5) Use a representative sample of patients, diagnose them using suitable instruments in the hands of trained interviewers, and test the diagnostic reliability. Small changes were seen: representative sample 1.08-1.15, reliability of diagnosis 0.15-0.38.

6) Let an independent researcher or agency use an unobjectionable randomization procedure, and conceal the outcome of it from all persons involved in the study. This is difficult to analyse since randomization is described very briefly in most studies.

7) Use reliable and valid outcome measures; both the ones that are specific to the disorder and general ones. The mean score increased non-significantly 1.54-1.81.

8) Use blind assessors and evaluate their blindness regarding treatment condition of the patients they assess. No change was seen in this respect 0.31-0.28.

9) Train the assessors properly and measure inter-rater reliability on the data collected throughout the study (not just during training). The mean *decreased* somewhat 0.31-0.19.

10) Use three or more properly trained therapists and randomize patients to therapist to enable an analysis of possible therapist effect on the outcome. There was a significant increase in this respect, 0.23-0.72.

11) Include at least a 1 year follow-up in the study and assess any non-protocol treatments that the patients may have obtained during the follow-up period. A small decrease was seen 0.92-0.81.

12) Audio- or videotape all therapy sessions. Randomly select 20% of these and let independent experts rate adherence to treatment manual and therapist competence. There were non-significant increases: adherence 0.08-0.28; competence 0.00-0.19.

13) Insert procedures to control for concomitant treatments that patients in the study may obtain simultaneously as the protocol treatment. A small increase was seen 0.23-0.34.

14) Describe the attrition, do a drop-out analysis and include all randomized subjects in an intent-to-treat analysis. A small increase was observed 0.85-0.96.

15) Assess clinical significance of the improvement on the primary measure. A small *decrease* was obtained 0.69-0.47.

In light of these small changes I can only repeat the recommendations from the 2008 article. I believe that following these will increase the probability that ACT in the future will be evaluated as an evidence-based treatment.

References

*Studies included in the meta-analysis

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders, 4th ed.* Washington, DC: Author.

*Arch, J. J., Eifert, G. H., Davies, C., Vilardaga, J. C. P., Rose, R. D., & Craske, M. G. (2012). Randomized clinical trial of cognitive behavioral therapy (CBT) versus acceptance and commitment therapy (ACT) for mixed anxiety disorders. *Journal of Consulting and Clinical Psychology, 80*, 750-765. doi: http://dx.doi.org/10.1037/a0028310

*Bach, P., & Hayes, S. C. (2002). The use of acceptance and commitment therapy to prevent the rehospitalization of psychotic patients: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 70*, 1129-1139. doi: http://dx.doi.org/10.1037/0022-006X.70.5.1129

Barber, J. P., Liese, B. S., & Abrams, M. J. (2003). Development of the Cognitive Therapy Adherence and Competence scale. *Psychotherapy Research, 13*, 205-221. doi: http://dx.doi.org/10.1093/ptr/kpg019

Barber, J. P., Sharpless, B. A., Klostermann, S., & McCarthy, K. S. (2007). Assessing intervention competence and its relation to therapy outcome: A selected review derived from the outcome literature. *Professional Psychology: Research and Practice, 38*, 493-500. doi: http://dx.doi.org/10.1037/0735-7028.38.5.493

*Bethay, J. S., Wilson, K. G., Schnetzer, L. W., Nassar, S. L., & Bordieri, M. J. (2013). A controlled pilot evaluation of acceptance and commitment training for intellectual disability staff. *Mindfulness, 4*, 113-121. doi: http://dx.doi.org/10.1007/s12671-012-0103-8

Block, J. A. (2002). *Acceptance or change of private experiences: A comparative analysis in college students with public speaking anxiety*. Unpublished doctoral dissertations, University of Albany, State University of New York.

*Bond, F. W., & Bunce, D. (2000). Mediators of change in emotion-focused and problem-focused worksite stress management interventions. *Journal of Occupational Health Psychology, 5*, 156-163. doi: http://dx.doi.org/10.1037/1076-8998.5.1.156

*Bricker, J., Wyszynski, C., Comstock, B., & Heffner, J. L. (2013). Pilot randomized controlled trial of web-based acceptance and commitment therapy for smoking cessation. *Nicotine & Tobacco Research, 15*, 1756-1764. doi: 10.1016/ j.amepre.2009.12.006.

*Brinkborg, H., Michanek, J., Hesser, H., & Berglund, G. (2011). Acceptance and commitment therapy for the treatment of stress among social workers: A randomized controlled trial. *Behaviour Research and Therapy, 49*, 389-398. doi: http://dx.doi.org/10.1016/j.brat.2011.03.009

*Brown, L. A., Forman, E. M., Herbert, J. D., Hoffman, K. L., Yuen, E. K., & Goetter, E. M. (2011). A randomized controlled trial of acceptance-based behavior therapy and cognitive therapy for test anxiety: A pilot study. *Behavior Modification, 35*, 31-53. doi: http://dx.doi.org/10.1177/0145445510390930

*Buhrman, M., Skoglund, A., Husell, J., Bergström, K., Gordh, T., Hursti, T., . . . Andersson, G. (2013). Guided internet-delivered acceptance and commitment therapy for chronic pain patients: A randomized controlled trial. *Behaviour Research and Therapy, 51*, 307-315. doi: http://dx.doi.org/10.1016/j.brat.2013.02.010

Cavanagh, K., Strauss, C., Forder, L., & Jones, F. (2014). Can mindfulness and acceptance be learnt by self-help?: A systematic review and meta-analysis of mindfulness and acceptance-based self-help interventions. *Clinical Psychology Review, 34*, 118-129. doi: 10.2307/1411318 http://dx.doi.org/10.1016/j.cpr.2014.01.001

Chambless, D. L., Baker, M. J., Baucom, D. H., Beutler, L. E., Calhoun, K. S., et al. (1998). Update on empirically validated therapies, II. *The Clinical Psychologist, 51*, 3–16.

Chambless, D. L., Sanderson, W. C., Shoham, V., Johnson, S. B., Pope, K. S., Crits-Christoph, P., et al. (1996). An update on empirically validated therapies. *The Clinical Psychologist, 49*, 5–18.

Cohen, J. (1962). The statistical power of abnormal-social psychological research: A review. *The Journal of Abnormal and Social Psychology, 65*, 145-153. doi: http://dx.doi.org/10.1037/h0045186

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences, 2nd ed*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Comprehensive Meta-analysis (2010). Statistical software, Englewood, NJ: Biostat Inc.

Cuijpers, P., Sijbrandij, M., Koole, S., Huibers, M., Berking, M., & Andersson, G. (2014). Psychological treatment of generalized anxiety disorder: A meta-analysis. *Clinical Psychology Review, 34*, 130-140. doi: DOI: 10.1016/j.brat.2010.05.014].

Cuijpers, P., Smit, F., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). Efficacy of cognitive-behavioural therapy and other psychological treatments for adult depression: Meta-analytic study of publication bias. *The British Journal of Psychiatry, 196*, 173-178. doi: http://dx.doi.org/10.1192/bjp.bp.109.066001

Dahl, J., Brorson, L. O., & Melin, L. (1992). Effects of a broad-spectrum behavioral medicine treatment program on children with refractory epileptic seizures: an 8-year follow-up. *Epilepsia, 33*, 98-102.

*Dahl, J., Wilson, K. G., & Nilsson, A. (2004). Acceptance and commitment therapy and the treatment of persons at risk for long-term disability resulting from stress and pain symptoms: A preliminary randomized trial. *Behavior Therapy, 35*, 785-801. doi: http://dx.doi.org/10.1016/S0005-7894(04)80020-0

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455-463.

Egger M, Davey Smith G, Schneider M, Minder C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629-634.

*England, E. L., Herbert, J. D., Forman, E. M., Rabin, S. J., Juarascio, A., & Goldstein, S. P. (2012). Acceptance-based exposure therapy for public speaking anxiety. *Journal of Contextual Behavioral Science, 1*, 66-72. http://dx.doi.org/10.1016/j.jcbs.2012.07.001

*Flaxman, P. E., & Bond, F. W. (2010). A randomised worksite comparison of acceptance and commitment therapy and stress inoculation training. *Behaviour Research and Therapy, 48*, 816-820. doi: http://dx.doi.org/10.1016/j.brat.2010.05.004

*Flaxman, P. E., & Bond, F. W. (2010). Worksite stress management training: Moderated effects and clinical significance. *Journal of Occupational Health Psychology, 15*, 347-358. doi: http://dx.doi.org/10.1037/a0020522

*Folke, F., Parling, T., & Melin, L. (2012). Acceptance and commitment therapy for depression: A preliminary randomized clinical trial for unemployed on long-term sick leave. *Cognitive and Behavioral Practice, 19*, 583-594. doi: http://dx.doi.org/10.1016/j.cbpra.2012.01.002

*Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D., & Geller, P. A. (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modification, 31*, 772-799. doi: http://dx.doi.org/10.1177/0145445507302202

*Forman, E. M., Butryn, M. L., Juarascio, A. S., Bradley, L. E., Lowe, M. R., Herbert, J. D., & Shaw, J. A. (2013). The Mind Your Health project: A randomized controlled trial of an innovative behavioral treatment for obesity. *Obesity, 21*, 1119-1126. doi: http://dx.doi.org/10.1002/oby.20169

*Forman, E. M., Hoffman, K. L., Juarascio, A. S., Butryn, M. L., & Herbert, J. D. (2013). Comparison of acceptance-based and standard cognitive-based coping strategies for craving sweets in overweight and obese women. *Eating Behaviors, 14*, 64-68. doi: http://dx.doi.org/10.1016/j.eatbeh.2012.10.016

*Gaudiano, B. A., & Herbert, J. D. (2006). Acute treatment of inpatients with psychotic symptoms using acceptance and commitment therapy: Pilot results. *Behaviour Research and Therapy, 44*, 415-437. doi: http://dx.doi.org/10.1016/j.brat.2005.02.007

*Gifford, E. V., Kohlenberg, B. S., Hayes, S. C., Pierson, H. M., Piasecki, M. P., Antonuccio, D. O., & Palm, K. M. (2011). Does acceptance and relationship focused behavior therapy contribute to bupropion outcomes? A randomized controlled trial of functional analytic psychotherapy and acceptance and commitment therapy for smoking cessation. *Behavior Therapy, 42*, 700-715. doi:http://dx.doi.org/10.1016/j.beth.2011.03.002

*Gifford, E. V., Kohlenberg, B. S., Hayes, S. C., Antonuccio, D. O., Piasecki, M. M., Rasmussen-Hall, M. L., & Palm, K. M. (2004). Acceptance-based treatment for smoking cessation. *Behavior Therapy, 35*, 689-705. doi: http://dx.doi.org/10.1016/S0005-7894(04)80015-7

*Gratz, K. L., & Gunderson, J. G. (2006). Preliminary data on acceptance-based emotion regulation group intervention for deliberate self-harm among women with Borderline personality disorder. *Behavior Therapy, 37*, 25-35. doi: http://dx.doi.org/10.1016/j.beth.2005.03.002

*Gregg, J. A., Callaghan, G. M., Hayes, S. C., & Glenn-Lawson, J. L. (2007). Improving diabetes self-management through acceptance, mindfulness, and values: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 75*, 336-343. doi: http://dx.doi.org/10.1037/0022-006X.75.2.336

*Hayes, L., Boyd, C. P., & Sewell, J. (2011). Acceptance and commitment therapy for the treatment of adolescent depression: A pilot study in a psychiatric outpatient setting. *Mindfulness, 2*, 86-94. doi: http://dx.doi.org/10.1007/s12671-011-0046-5

*Hayes-Skelton, S. A., Roemer, L., & Orsillo, S. M. (2013). A randomized clinical trial comparing an acceptance-based behavior therapy to applied relaxation for generalized anxiety disorder. *Journal of Consulting and Clinical Psychology, 81*, 761-773. doi: http://dx.doi.org/10.1037/a0032871

Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavior therapy. *Behavior Therapy, 35*, 639–665.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: model, processes and outcomes. *Behaviour Research and Therapy, 44*, 1-25. doi: 10.1016/j.brat.2005.06.006

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*: Guilford Press, New York, NY.

*Hayes, S. C., Wilson, K. G., Gifford, E. V., Bissett, R., Piasecki, M., Batten, S. V., . . . Gregg, J. (2004). A preliminary trial of twelve-step facilitation and acceptance and commitment therapy with polysubstance-abusing methadone-maintained opiate addicts. *Behavior Therapy, 35*, 667-688. doi: http://dx.doi.org/10.1016/S0005-7894(04)80014-5

Hernández-López, M., Luciano, M. C., Bricker, J. B., Roales-Nieto, J. G., & Montesinos, F. (2009). Acceptance and commitment therapy for smoking cessation: A preliminary

study of its effectiveness in comparison with cognitive behavioral therapy. *Psychology of Addictive Behaviors, 23*, 723-730. doi: http://dx.doi.org/10.1037/a0017632

*Hesser, H., Gustafsson, T., Lundén, C., Henrikson, O., Fattahi, K., Johnsson, E., . . . Andersson, G. (2012). A randomized controlled trial of internet-delivered cognitive behavior therapy and acceptance and commitment therapy in the treatment of tinnitus. *Journal of Consulting and Clinical Psychology, 80*, 649-661. doi: http://dx.doi.org/10.1037/a0027021

Hofmann, S. G., & Smits, J. A. J. (2008). Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychiatry, 69*, 621-632. doi: http://dx.doi.org/10.4088/JCP.v69n0415

Hofmann, S. G., Wu, J. Q., & Boettcher, H. (2014). Effect of cognitive-behavioral therapy for anxiety disorders on quality of life: A meta-analysis. *Journal of Consulting and Clinical Psychology*, Jan 20. [Epub ahead of print]

Jadad, A.R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J., Gavaghan, D. J., et al. (1996). Assessing the quality reports of randomized controlled trials: Is blinding necessary? *Controlled Clinical Trials, 17*, 1-12.

*Johnston, M., Foster, M., Shennan, J., Starkey, N. J., & Johnson, A. (2010). The effectiveness of an acceptance and commitment therapy self-help intervention for chronic pain. *Clinical Journal of Pain, 26*, 393-402. doi: http://dx.doi.org/10.1097/AJP.0b013e3181cf59ce

Kazdin, A. E. (2003). *Research design in clinical psychology, 4th ed*. Boston: Allyn & Bacon.

*Kocovski, N. L., Fleming, J. E., Hawley, L. L., Huta, V., & Antony, M. M. (2013). Mindfulness and acceptance-based group therapy versus traditional cognitive behavioral group therapy for social anxiety disorder: A randomized controlled trial. *Behaviour Research and Therapy, 51*, 889-898. doi: 10.2307/1411318 http://dx.doi.org/10.1016/j.brat.2013.10.007

*Lanza, P. V., & Menéndez, A. G. (2013). Acceptance and Commitment Therapy for drug abuse in incarcerated women. *Psicothema, 25*, 307-312.

*Lappalainen, R., Lehtonen, T., Skarp, E., Taubert, E., Ojanen, M., & Hayes, S. C. (2007). The impact of CBT and ACT models using psychology trainee therapists: A preliminary controlled effectiveness trial. *Behavior Modification, 31*, 488-511. doi: http://dx.doi.org/10.1177/0145445506298436

*Lappalainen, P., Kaipainen, K., Lappalainen, R., Hoffren, H., Myllymaki, T., Kinnunen, M. L., . . . Korhonen, I. (2013). Feasibility of a personal health technology-based psychological intervention for men with stress and mood problems: randomized controlled pilot trial. *JMIR Research Protocols, 2*, e1. doi: 10.2196/resprot.2389

*Lillis, J., Hayes, S. C., Bunting, K., & Masuda, A. (2009). Teaching acceptance and mindfulness to improve the lives of the obese: A preliminary test of a theoretical model. *Annals of Behavioral Medicine, 37*, 58-69. doi: 10.1038/sj .ijo.0802460 http://dx.doi.org/10.1007/s12160-009-9083-x

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*: Thousand Oaks, CA: Sage Publications, Inc.

*Lloyd, J., Bond, F. W., & Flaxman, P. E. (2013). The value of psychological flexibility: Examining psychological mechanisms underpinning a cognitive behavioural therapy intervention for burnout. *Work & Stress, 27*, 181-199. doi: 0142-2774/82/010063 http://dx.doi.org/10.1080/02678373.2013.782157

*Lundgren, T., Dahl, J., Melin, L., & Kies, B. (2006). Evaluation of acceptance and commitment therapy for drug refractory epilepsy: A randomized controlled trial in South Africa--A pilot study. *Epilepsia, 47*, 2173-2179. doi: http://dx.doi.org/10.1111/j.1528-1167.2006.00892.x

*Lundgren, T., Dahl, J., Yardi, N., & Melin, L. (2008). Acceptance and commitment therapy and yoga for drug-refractory epilepsy: A randomized controlled trial. *Epilepsy & Behavior, 13*, 102-108. doi: http://dx.doi.org/10.1016/j.yebeh.2008.02.009

*Luoma, J. B., Kohlenberg, B. S., Hayes, S. C., & Fletcher, L. (2012). Slow and steady wins the race: A randomized clinical trial of acceptance and commitment therapy targeting shame in substance use disorders. *Journal of Consulting and Clinical Psychology, 80*, 43-53. doi: http://dx.doi.org/10.1037/a0026070

*McCracken, L. M., Sato, A., & Taylor, G. J. (2013). A trial of a brief group-based form of acceptance and commitment therapy (ACT) for chronic pain in general practice: Pilot outcome and process results. *The Journal of Pain, 14*, 1398-1406. doi: http://dx.doi.org/10.1016/j.jpain.2013.06.011

*Mo'tamedi, H., Rezaiemaram, P., & Tavallaie, A. (2012). The effectiveness of a group-based acceptance and commitment additive therapy on rehabilitation of female outpatients with chronic headache: Preliminary findings reducing 3 dimensions of headache impact. *Headache: The Journal of Head and Face Pain, 52*, 1106-1119. doi: http://dx.doi.org/10.1111/j.1526-4610.2012.02192.x

*Morton, J., Snowdon, S., Gopold, M., & Guymer, E. (2012). Acceptance and commitment therapy group treatment for symptoms of borderline personality disorder: A public sector pilot study. *Cognitive and Behavioral Practice, 19*, 527-544. doi: http://dx.doi.org/10.1016/j.cbpra.2012.03.005

*Nordin, L., & Rorsman, I. (2012). Cognitive behavioural therapy in multiple sclerosis: a randomized controlled pilot study of acceptance and commitment therapy. *Journal of Rehabilitation Medicine, 44*, 87-90. doi: 10.2340/16501977-0898

Öst, L.-G. (2008). Efficacy of the third wave of behavioral therapies: A systematic review and meta-analysis. *Behaviour Research and Therapy, 46*, 296-321. doi: http://dx.doi.org/10.1016/j.brat.2007.12.005

Öst, L.-G. (2009). *The efficacy and process of ACT: A systematic review and meta-analysis.* Keynote at the Third World Congress of ACT, Enschede, Holland.

Páez, M. B., Luciano, C., & Gutiérrez, O. (2007). Tratamiento psicológico para el afrontamiento del cáncer de mama. Estudio comparativo entre estrategias de aceptación y de control cognitivo. *Psicooncología, 4*, 75-95.

*Petersen, C. L., & Zettle, R. D. (2009). Treating inpatients with comorbid depression and alcohol use disorders: A comparison of acceptance and commitment therapy versus treatment as usual. *The Psychological Record, 59*, 521-536.

Powers, M. B., Zum Vörde Sive Vörding, M. B., & Emmelkamp, P. M. G. (2009). Acceptance and commitment therapy: A meta-analytic review. *Psychotherapy & Psychosomatics, 78*, 73-80. doi: http://dx.doi.org/10.1159/000190790

*Roemer, L., Orsillo, S. M., & Salters-Pedneault, K. (2008). Efficacy of an acceptance-based behavior therapy for generalized anxiety disorder: Evaluation in a randomized controlled trial. *Journal of Consulting and Clinical Psychology, 76*, 1083-1089. doi: http://dx.doi.org/10.1037/a0012720

Rogers, J. L., Howard, K. I., & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin, 113*, 553-565. doi: http://dx.doi.org/10.1037/0033-2909.113.3.553

*Rost, A. D., Wilson, K., Buchanan, E., Hildebrandt, M. J., & Mutch, D. (2012). Improving psychological adjustment among late-stage ovarian cancer patients: Examining the role of avoidance in treatment. *Cognitive and Behavioral Practice, 19*, 508-517. doi: http://dx.doi.org/10.1016/j.cbpra.2012.01.003

Ruiz, F. J. (2010). A review of acceptance and commitment therapy (ACT) empirical evidence: Correlational, experimental psychopathology, component and outcome

studies. *International Journal of Psychology & Psychological Therapy, 10*, 125-162. doi: 10.1016/ j.ejpain.2009.03.004.2010-02145-00710.1016/j.ejpain.2009.03.004

*Ruiz, F. J. (2012). Acceptance and commitment therapy versus traditional cognitive behavioral therapy: A systematic review and meta-analysis of current empirical evidence. *International Journal of Psychology & Psychological Therapy, 12,* 333-357.

*Shawyer, F., Farhall, J., Mackinnon, A., Trauer, T., Sims, E., Ratcliff, K., . . . Copolov, D. (2012). A randomised controlled trial of acceptance-based cognitive behavioural therapy for command hallucinations in psychotic disorders. *Behaviour Research and Therapy, 50*, 110-121. doi: http://dx.doi.org/10.1016/j.brat.2011.11.007ä'

Silverman, W. K., & Hinshaw, S. P. (2008). The second special issue on evidence-based psychosocial treatments for children and adolescents: A 10-year update. *Journal of Clinical Child and Adolescent Psychology, 37*, 1-7. doi: http://dx.doi.org/10.1080/15374410701817725

Smout, M. F., Hayes, L., Atkins, P. W. B., Klausen, J., & Duguid, J. E. (2012). The empirically supported status of acceptance and commitment therapy: An update. *Clinical Psychologist, 16*, 97-109. doi: 10.1002/14651858. CD007407.pub2.

*Smout, M. F., Longo, M., Harrison, S., Minniti, R., Wickes, W., & White, J. M. (2010). Psychosocial treatment for methamphetamine use disorders: A preliminary randomized controlled trial of cognitive behavior therapy and acceptance and commitment therapy. *Substance Abuse, 31*, 98-107. doi: http://dx.doi.org/10.1080/08897071003641578

*Stotts, A. L., Green, C., Masuda, A., Grabowski, J., Wilson, K., Northrup, T. F., . . . Schmitz, J. M. (2012). A Stage I pilot study of acceptance and commitment therapy for methadone detoxification. *Drug and Alcohol Dependence, 125*, 215-222. doi: http://dx.doi.org/10.1016/j.drugalcdep.2012.02.015

Swain, J., Hancock, K., Hainsworth, C., & Bowman, J. (2013). Acceptance and commitment therapy in the treatment of anxiety: A systematic review. *Clinical Psychology Review, 33*, 965-978. doi: http://dx.doi.org/10.1016/j.cpr.2013.07.002

*Tapper, K., Shaw, C., Ilsley, J., Hill, A. J., Bond, F. W., & Moore, L. (2009). Exploratory randomised controlled trial of a mindfulness-based weight loss intervention for women. *Appetite, 52*, 396-404. doi: http://dx.doi.org/10.1016/j.appet.2008.11.012

*Thorsell, J., Finnes, A., Dahl, J., Lundgren, T., Gybrant, M., Gordh, T., & Buhrman, M. (2011). A comparative study of 2 manual-based self-help interventions, acceptance and commitment therapy and applied relaxation, for persons with chronic pain. *Clinical Journal of Pain, 27*, 716-723. doi: http://dx.doi.org/10.1097/AJP.0b013e318219a933

Tolin, D.F. (1999). *A revised meta-analysis of psychosocial treatments for PTSD*. Poster presented at AABT, Toronto, November 11–14, 1999.

*Twohig, M. P., Hayes, S. C., Plumb, J. C., Pruitt, L. D., Collins, A. B., Hazlett-Stevens, H., & Woidneck, M. R. (2010). A randomized clinical trial of acceptance and commitment therapy versus progressive relaxation training for obsessive-compulsive disorder. *Journal of Consulting and Clinical Psychology, 78*, 705-716. doi: http://dx.doi.org/10.1037/a0020508

Walker, E., & Nowacki A. S. (2011). Understanding equivalence and noninferiority testing. *Journal of General Internal Medicine, 26*, 192-196. doi: 10.1007/s11606-010-1513-8.

*Weineland, S., Arvidsson, D., Kakoulidis, T. P., & Dahl, J. (2012). Acceptance and commitment therapy for bariatric surgery patients, a pilot RCT. *Obesity Research & Clinical Practice, 6*, e21-e30. doi: 10.1016/j.orcp.2011.04.004

*Westin, V. Z., Schulin, M., Hesser, H., Karlsson, M., Noe, R. Z., Olofsson, U., . . . Andersson, G. (2011). Acceptance and commitment therapy versus tinnitus retraining

therapy in the treatment of tinnitus: A randomised controlled trial. *Behaviour Research and Therapy, 49*, 737-747. doi: http://dx.doi.org/10.1016/j.brat.2011.08.001

*Wetherell, J. L., Afari, N., Ayers, C. R., Stoddard, J. A., Ruberg, J., Sorrell, J. T., . . . Patterson, T. L. (2011). Acceptance and commitment therapy for generalized anxiety disorder in older adults: a preliminary report. *Behavior Therapy, 42*, 127-134. doi: 10.1016/j.beth.2010.07.002

*Wetherell, J. L., Afari, N., Rutledge, T., Sorrell, J. T., Stoddard, J. A., Petkus, A. J., . . . Atkinson, J. H. (2011). A randomized, controlled trial of acceptance and commitment therapy and cognitive-behavioral therapy for chronic pain. *Pain, 152*, 2098-2107. doi: http://dx.doi.org/10.1016/j.pain.2011.05.016

*White, R., Gumley, A., McTaggart, J., Rattrie, L., McConville, D., Cleare, S., & Mitchell, G. (2011). A feasibility study of Acceptance and Commitment Therapy for emotional dysfunction following psychosis. *Behaviour Research and Therapy, 49*, 901-907. doi: http://dx.doi.org/10.1016/j.brat.2011.09.003

*Wicksell, R. K., Ahlqvist, J., Bring, A., Melin, L., & Olsson, G. L. (2008). Can exposure and acceptance strategies improve functioning and life satisfaction in people with chronic pain and whiplash-associated disorders (WAD)? A randomized controlled trial. *Cognitive Behavior Therapy, 37*, 169-182. doi: 10.1080/16506070802078970

*Wicksell, R. K., Kemani, M., Jensen, K., Kosek, E., Kadetoff, D., Sorjonen, K., . . . Olsson, G. L. (2013). Acceptance and commitment therapy for fibromyalgia: A randomized controlled trial. *European Journal of Pain, 17*, 599-611. doi: http://dx.doi.org/10.1002/j.1532-2149.2012.00224.x

*Wicksell, R. K., Melin, L., Lekander, M., & Olsson, G. L. (2009). Evaluating the effectiveness of exposure and acceptance strategies to improve functioning and quality of life in longstanding pediatric pain--a randomized controlled trial. *Pain, 141*, 248-257. doi: 10.1016/j.pain.2008.11.006

*Woods, D. W., Wetterneck, C. T., & Flessner, C. A. (2006). A controlled evaluation of acceptance and commitment therapy plus habit reversal for trichotillomania. *Behaviour Research and Therapy, 44*, 639-656. doi: http://dx.doi.org/10.1016/j.brat.2005.05.006

*Zettle, R. D. (2003). Acceptance and commitment therapy (ACT) vs. systematic desensitization in treatment of mathematics anxiety. *The Psychological Record, 53*, 197-215.

*Zettle, R. D., & Hayes, S. C. (1986). Dysfunctional control by client verbal behavior: The context of reason giving. *The Analysis of Verbal Behavior, 4*, 30–38.

*Zettle, R. D., & Rains, J. C. (1989). Group cognitive and contextual therapies in treatment of depression. *Journal of Clinical Psychology, 45*, 436-445. doi: http://dx.doi.org/10.1002/1097-4679(198905)45:3<436::AID-JCLP2270450314>3.0.CO;2-L

Table 1
Background data for the included ACT-studies in the meta-analysis.

Disorder	Study	Comparison	N total	N/cell	Attr. Total %	Attr. ACT %	N compl.	Compl. /cell	Percent women	Mean age	# of therapists	# of weeks	# of Sess.	# of hours	F-up months
Depression	Zettle (1986)	CBT	18	9	NI	NI	18	9	100	NI	1	12	12	12.0	3
Depression	Zettle (1989)	CBT	37	12.3	16.2	15.4	31	10.3	100	41.3	1	12	12	10.8	2
Depression	Hayes (2011)	TAU	38	19	21.1	13.6	30	15	71	14.9	3	NI	NI	20.8	3
Depression	Folke (2012)	TAU	35	17.5	28.6	22.2	25	12.5	88	43.2	2	6	6	11.0	18
Depression	Petersen (2009)	TAU	28	14	14	20	24	12	50	37.8	1	4	5	3.1	0
Psychotic symptoms	Bach (2002)	TAU	80	40	12.5	12.5	70	35	36	39.4	1	2	4	4.0	12
Psychotic symptoms	Gaudiano (2006)	TAU	40	20	5.0	5.3	38	19	36	40.0	1	3	4	3.0	4
Psychotic symptoms	White (2011)	TAU	27	13.5	11.1	0	24	12	22	34.1	1	12	10	10.0	0
Psychotic symptoms	Shawyer (2012)	Other	47	15.7	9.3	4.8	39	19.5	44	39.8	5	15	15	12.5	6
Math anxiety	Zettle (2003)	CBT	33	16.5	27.3	14.3	24	12	81	30.5	1	6	6	6.0	0
Test anxiety	Brown (2011)	CBT	16	8	31.3	12.5	11	5.5	69	20.2	1	1	1	2.0	0
GAD	Roemer (2008)	WLC	31	15.5	19.4	13.3	25	12.5	71	33.6	6	17	16	18.0	9
GAD	Wetherell (2011)	CBT	21	10.5	23.8	36.4	16	8	48	70.8	6	12	12	12.0	6
GAD	Hayes-Skelton (2013)	CBT	81	40.5	22.2	25.0	25.0	31.5	65	32.9	11	16	16	18.0	6
OCD	Twohig (2010)	CBT	79	39.5	17.7	14.6	65	32.5	61	37.0	6	8	8	8.0	3
Mixed group	Arch (2012)	CBT	128	64	33.6	35.1	85	42.5	52	38.0	39	12	12	12.0	12
Social anxiety	England (2012)	CBT	45	22.5	22.2	23.8	35	17.5	80	31.9	3	6	6	12.0	1.5
Social anxiety	Kocovski (2013)	CBT/WLC	137	45.7	28.2	30.2	100	33.3	56	34.7	2	12	12	24.0	3
Drug abuse	Hayes (2004)	Other/Drug	124	41.3	37.1	42.9	78	26	51	42.2	4	16	48	24.4	6
Drug abuse	Smout (2010)	CBT	104	52	70.2	72.5	31	15.5	40	30.9	3	12	12	12.0	3
Drug abuse	Luoma (2011)	TAU	133	66.5	24.1	29.4	101	50.5	46	33.6	2	4	3	6.0	4
Drug abuse	Stotts (2012)	Other	56	28	46.4	40.0	30	15	38	39.9	2	24	24	20.0	0
Drug abuse	Lanza (2013)	WLC	31	10.5	0	0	31	10.5	100	32.0	NI	16	16	24.0	6
Nicotine dependence	Gifford (2004)	Med	76	38	35.5	36.4	49	24.5	59	43.0	4	7	14	16.3	12
Nicotine dependence	Gifford (2011)	Med	303	151.5	44.9	40.8	167	83.5	59	46.0	4	10	10	30.0	12
Nicotine dependence	Bricker (2013)	Other	222	111	46.4	45.9	119	59.5	38	45.1	0	12	0	NA	0
Trichotillomania	Woods (2006)	WLC	28	14	10.7	14.3	26	13	89	35.0	1	12	10	12.0	3

Borderline PD	Gratz (2006)	TAU	24	12	8.3	0	22	11	100	33.2	1	14	14	21.0	0
Borderline PD	Morton (2012)	TAU	41	20.5	31.7	33.3	28	14	93	34.8	3	12	12	24.0	3
Mixed group	Forman (2007)	CBT	101	50.5	37.6	33.9	57	31.5	80	27.9	23	NI	18	18.1	18
Mixed group	Lappalainen (2007)	CBT	28	14	0	0	28	14.0	89	41.8	14	10	9	9.1	6
Pain	Dahl (2004)	TAU	19	9.5	0	0	19	9.5	81	40.0	2	4	4	4.0	6
Pain	Wicksell (2008)	WLC	22	11	9.1	0	20	10	76	51.6	3	8	10	10.0	7
Pain	Wicksell (2009)	TAU	32	16	9.4	6.3	29	14.5	78	14.8	2	10	10	10.3	6
Pain	Wicksell (2013)	WLC	40	20	10.0	13.0	36	18	100	45.1	3	12	12	18.0	3
Pain	Johnston (2010)	WLC	24	12	41.6	50.0	14	7	63	43.0	1	6	6	3.0	0
Pain	Thorsell (2011)	CBT	115	57.5	52.2	54.1	55	27.5	64	46.0	8	9	9	6.5	12
Pain	Wetherell (2011)	CBT	114	57	25.4	24.6	85	42.5	51	54.9	3	8	8	12.0	6
Pain	Buhrman (2013)	WLC	76	38	19.7	23.7	61	30.5	59	40.1	3	7	2	0.5	6
Pain	McCracken (2013)	TAU	73	36.5	26.0	27.0	54	27.0	69	58.0	2	2	4	16.0	3
Headache	Motamedi (2012)	TAU	30	15	13.3	26.6	26	13	100	36.0	1	8	8	12.0	0
Epilepsy	Lundgren (2006)	Other	27	13.5	0	0	27	13.5	52	40.7	2	4	4	9.0	12
Epilepsy	Lundgren (2008)	Other	18	9	0	0	18	9	33	23.6	2	5	4	12.0	12
Tinnitus	Westin (2011)	Other/WLC	64	21.3	6.3	4.8	60	20	47	50.9	8	10	10	10.0	6
Tinnitus	Hesser (2012)	CBT	99	33	10.1	8.6	89	29.7	43	48.5	7	8	8	1.2	12
Cancer	Rost (2012)	TAU	47	23.5	34.0	40.0	32	16	100	56.0	1	16	12	12.0	0
Overweight/Obesity	Lillis (2009)	WLC	87	43.5	3.4	7.0	84	42	90	50.8	2	1	1	6.0	3
Overweight/Obesity	Tapper (2009)	WLC	62	31	13	25.8	51	26.5	100	41.0	1	3	3	6.0	3
Overweight/Obesity	Weineland (2012)	TAU	39	19.5	15.4	21.1	33	16.5	90	43.1	NI	8	8	5.5	0
Overweight/Obesity	Forman (2013a)	CBT	48	24	0	0	48	24	100	32.5	NI	1	1	2.0	0
Overweight/Obesity	Forman (2013b)	CBT	128	64	14.1	9.5	110	55	100	45.7	8	40	30	37.5	6
Diabetes	Gregg (2007)	Other	81	40.5	18.5	16.3	66	33	47	50.9	1	1	1	7.0	0
Multiple sclerosis	Nordin (2012)	CBT	21	10.5	4.8	9.1	20	10	80	45.8	2	15	5	5.0	3
Stress	Bond (2000)	IPP/WLC	90	30	27.8	20.0	65	21.7	50	36.4	NI	14	3	9.8	3
Stress	Flaxman (2010a)	SIT/WLC	311	155.5	59.2	64.4	127	63.5	72	41.0	1	14	3	9.0	0
Stress	Flaxman (2010b)	WLC	107	35.7	38.3	48.6	66	22	NI	39.0	1	2	2	6.0	0
Stress	Brinkborg (2011)	WLC	106	53	11.3	10.0	94	47	89	44.0	4	8	4	12.0	0
Stress	Bethay (2013)	CBT	38	19	10.5	10.0	34	17	76	38.0	1	3	3	9.0	3
Stress	Lloyd (2013)	WLC	100	50	26.5	29.5	64	32	83	47.0	1	10	3	9.0	6

Stress	Lappalainen (2013)	WLC	24	12	4.2	8.3	23	11.5	0	43.3	1	12	3	8.0	6

Note: CBT = Cognitive Behavior Therapy, TAU = Treatment-as-usual, WLC = Waitlist control, Med = medication, IPP = Innovation Promotion Program, SIT = Stress Inoculation Treatment, NI = no information.

Table 2
Means (SDs) and *F*-values for different background and therapy variables divided on type of disorder.

Variable	*All studies*	*Psychiatric disorders*	*Somatic disorders*	*Stress at work*	*F-value*
1. Number of participants starting therapy	70.6 (60.6)	70.7 (64.0)	57.6 (34.6)	110.9 (94.4)	2.14
2. Number of participants per condition	33.2 (29.8)	33.2 (31.4)	27.5 (16.8)	50.7 (48.6)	1.64
3. Attrition total (percent of those starting)	21.0 (15.9)	24.5 (15.5)	14.8 (13.9)	25.4 (19.1)	2.86
4. Attrition ACT-condition	21.1 (17.3)	22.9 (16.9)	16.7 (16.0)	27.3 (21.8)	1.35
5. Number of completers	49.8 (32.9)	47.6 (36.1)	47.1 (27.0)	67.6 (35.0)	1.17
6. Cell size (completers/number of conditions)	23.6 (17.8)	22.8 (17.2)	22.5 (12.7)	30.7 (18.5)	0.79
7. Proportion of women	67.9 (23.7)	64.9 (22.8)	73.8 (22.1)	61.7 (33.0)	1.14
8. Mean age of the sample	39.9 (9.7)	36.9 (9.3)[a]	43.6 (10.2)[b]	41.2 (3.7)[ab]	3.44*
9. Number of therapists	4.0 (6.19	5.1 (8.0)	3.1 (2.5)	1.5 (1.2)	1.18
10. Number of therapy weeks	9.6 (6.4)	10.5 (5.2)	8.5 (8.2)	9.0 (4.9)	0.66
11. Number of therapy sessions	9.0 (7.8)	11.6 (8.7)[a]	7.3 (6.2)[ab]	3.0 (0.6)[b]	4.80*
12. Number of therapy hours	11.6 (7.4)	13.9 (7.49	9.3 (7.8)	9.0 (1.8)	3.06
13. Follow-up (months since post-assessment)	4.8 (4.6)	5.4 (5.1)	4.8 (4.2)	2.6 (2.7)	1.07

[a, b] Means with different superscript differs significantly (p <.05 or lower). * p <.05

Table 3
Means (SDs) and *F*-values for the different variables on the Psychotherapy research methodology scale.

Variable	*All studies*	*Psychiatric*	*Somatic*	*Stress*	*F-value*
1. Clarity of sample description	1.22 (0.69)	1.45 (0.57)[a]	1.18 (0.66)[a]	0.29 (0.49)[b]	10.92***
2. Severity/chronicity of the disorder	1.10 (0.84)	1.13 (0.85)[a]	1.36 (0.73)[a]	0.14 (0.38)[b]	6.79**
3. Representativeness of the sample	1.13 (0.65)	1.10 (0.54)[a]	1.45 (0.60)[a]	0.29 (0.49)[b]	11.89***
4. Reliability of the diagnosis in question	0.33 (0.66)	0.48 (0.81)[a]	0.23 (0.43)[ab]	0.00[b]	2.09
5. Specificity of outcome measures	1.80 (0.44)	1.90 (0.30)[a]	1.91 (0.29)[a]	1.00 (0.58)[b]	22.14***
6. Reliability and validity of outcome measures	1.75 (0.51)	1.74 (0.58)	1.73 (0.46)	1.86 (0.38)	0.18
7. Use of blind evaluators	0.28 (0.49)	0.48 (0.57)[a]	0.09 (0.29)[b]	0.00[b]	6.47**
8. Assessor training	0.22 (0.52)	0.39 (0.67)[a]	0.05 (0.21)[b]	0.00[b]	3.73
9. Assignment to treatment	0.98 (0.39)	0.94 (0.44)	1.05 (0.38)	1.00 (0.00)	0.51
10. Design	1.05 (0.85)	1.35 (0.80)[a]	0.77 (0.81)[b]	0.57 (0.79)[b]	4.80
11. Power analysis	0.27 (0.63)	0.42 (0.81)[a]	0.14 (0.35)[ab]	0.00[b]	2.05
12. Assessment points	0.83 (0.64)	0.90 (0.60)	0.82 (0.73)	0.57 (0.54)	0.77
13. Manualized, replicable, treatment programs	1.38 (0.72)	1.55 (0.68)	1.09 (0.68)	1.57 (0.79)	3.11
14. Number of therapists	0.62 (0.56)	0.61 (0.50)	0.73 (0.55)	0.29 (0.76)	1.72
15. Therapist training/experience	0.68 (0.68)	0.58 (0.67)	0.82 (0.73)	0.71 (0.49)	0.80
16. Checks for treatment adherence	0.23 (0.43)	0.39 (0.50)[a]	0.09 (0.29)[b]	0.00[b]	4.85
17. Checks for therapist competence	0.15 (0.36)	0.23 (0.43)[a]	0.09 (0.29)[b]	0.00[b]	1.62
18. Control of concomitant treatments	0.32 (0.50)	0.29 (0.53)[a]	0.45 (0.51)[a]	0.00[b]	2.35
19. Handling of attrition	0.93 (0.66)	0.77 (0.67)	1.09 (0.61)	1.14 (0.69)	1.94
20. Statistical analyses and presentation of results	1.77 (0.53)	1.65 (0.61)[a]	1.86 (0.47)[ab]	2.00[b]	1.90
21. Clinical significance	0.52 (0.77)	0.77 (0.81)[a]	0.14 (0.77)[b]	0.57 (0.98)[ab]	5.04*
22. Equality of therapy hours	1.39 (0.93)	1.33 (0.96)[a]	1.38 (0.96)[a]	2.00[b]	0.69
Total score	18.63 (5.28)	20.29 (5.49)[a]	18.14 (4.24)[a]	12.86 (2.61)[b]	6.99**

[a, b] Means with different superscript differs significantly (p <.05 or lower). * p <.01, ** p <.001, *** p<.0001

Table 4
Means (SDs) and *t*-values for the different variables on the Psychotherapy research methodology scale divided into early and late studies.

Variable	*Studies in the 2008 review*	*Studies since the 2008 review*	*t-value*
1. Clarity of sample description	1.23 (0.73)	1.21 (0.69)	0.08
2. Severity/chronicity of the disorder	1.31 (0.86)	1.04 (0.83)	1.01
3. Representativeness of the sample	1.08 (0.76)	1.15 (0.63)	0.35
4. Reliability of the diagnosis in question	0.15 (0.38)	0.38 (0.71)	1.56
5. Specificity of outcome measures	1.77 (0.60)	1.81 (0.40)	0.28
6. Reliability and validity of outcome measures	1.54 (0.66)	1.81 (0.45)	1.38
7. Use of blind evaluators	0.31 (0.48)	0.28 (0.50)	0.20
8. Assessor training	0.31 (0.63)	0.19 (0.50)	0.71
9. Assignment to treatment	0.85 (0.38)	1.02 (0.39)	1.45
10. Design	1.23 (0.73)	1.00 (0.89)	0.86
11. Power analysis	0.00	0.34 (0.70)	3.33*
12. Assessment points	0.92 (0.64)	0.81 (0.65)	0.57
13. Manualized, replicable, specific treatment programs	1.54 (0.66)	1.34 (0.73)	0.88
14. Number of therapists	0.23 (0.44)	0.72 (0.54)	3.02*
15. Therapist training/experience	0.69 (0.75)	0.68 (0.66)	0.05
16. Checks for treatment adherence	0.08 (0.28)	0.28 (0.45)	1.97
17. Checks for therapist competence	0.00	0.19 (0.40)	3.30*
18. Control of concomitant treatments	0.23 (0.60)	0.34 (0.48)	0.69
19. Handling of attrition	0.85 (0.80)	0.96 (0.62)	0.53
20. Statistical analyses and presentation of results	1.69 (0.63)	1.79 (0.51)	0.57
21. Clinical significance	0.69 (0.75)	0.47 (0.78)	0.93
22. Equality of therapy hours (for non-WLC designs only)	1.45 (0.93)	1.37 (0.95)	0.26
Total score	17.92 (4.99)	18.83 (5.40)	0.55

* p <.001

Table 5

Designs (conditions) in the ACT RCTs.

Comparison	Psychiatric	Somatic	Stress	Total
ACT vs. WLC	4	7	6	17
ACT vs. TAU	9	6	0	15
ACT vs. CBT	13	6	2	21
ACT vs. Other	4	5	1	10
ACT vs. Drug	3	0	0	3
Total	33	24	9	66

Table 6

Therapy hours in TAU-controlled ACT studies.

Study	**Disorder**	**Therapists**	**Therapy hrs ACT+TAU**	**Therapy hrs TAU***	**ACT - TAU difference**
Hayes (2011)	Depression	Same	20.8†	15.6	-25%
Folke (2012)	Depression	Different	16.5	0	-100%
Petersen (2009)	Depression	Different	3.1†	4.3	39%
Bach (2002)	Psychotic	Different	3.3	0	-100%
Gaudiano (2006)	Psychotic	Different	3.0	3.0	0
White (2011)	Psychotic	Different	10.0	0	-100%
Luoma (2011)	Drug abuse	Different	6.0	6.0	0
Gratz (2006)	BPD	Different	50.4	41.3	-18%
Morton (2012)	BPD	Different	24.0	6.0	-75%
Dahl (2004)	Pain	Different	4.0	0	-100%
Wicksell (2009)	Pain	Different	10.3†	10.6	3%
McCracken (2013)	Pain	Different	16.0	0	-100%
Motamedi (2012)	Pain	Same	12.0	8.0	-33%
Rost (2012)	Cancer	Same	12.0†	12.0	0
Weineland (2012)	Obesity	Same	5.5	0	-100%

* 0 in this column means that the number of hours for TAU has not been described.
† Only ACT in this study.

Table 7

Number of therapists in the ACT RCTs.

	Type of disorder			
#	Psychiatric	Somatic	Stress	Sum
1	10	5	5	20 (33%)
2	4	7	0	11 (18%)
3-5	8	3	1	12 (20%)
≥6	7	4	0	11 (18%)
NI	2	3	1	6 (10%)
Sum	31	22	7	60

Table 8

Effect sizes (Hedges' g) for all ACT RCTs and divided on comparison conditions for post-treatment and follow-up assessments.

Comparison	**k**	***g*-value**	**95% CI**	***z*-value**	**Q-value**	I^2
Post-treatment						
All studies	64	0.42	0.31-0.53	7.47[d]	147.9[d]	57
WLC	16	0.63	0.44-0.83	6.35[d]	28.8[d]	48
Placebo	4	0.59	-0.02-1.20	1.90	11.7[b]	74
TAU	14	0.55	0.28-0.83	3.92[d]	32.0[b]	59
Active Tx	30	0.22	0.08-0.36	3.14[b]	57.1[c]	49
CBT/BT	22	0.16	-0.01-0.33	1.82	40.8[b]	48
Follow-up						
All studies	41	0.30	0.19-0.41	5.54[d]	61.2[a]	35
WLC	7	0.39	0.23-0.56	4.72[d]	4.1	0
Placebo	3	0.53	-0.22-1.28	1.39	7.6	74
TAU	7	0.48	0.27-0.69	4.50[d]	4.5	0
Active Tx	23	0.17	0.03-0.32	2.37[a]	34.0[a]	35
CBT/BT	17	0.06	-0.07-0.18	0.84	17.1	6

k = number of comparisons, [a] $p<0.05$, [b] $p<0.01$, [c] $p<0.001$, [d] $p<0.0001$

Table 9

Publication bias data for the ACT RCTs.

Comparison	**Observed ES**	**Trim-and-fill ES**	**# of trimmed studies**	**Egger's regression intercept**	***t*-value**
All studies	0.48	0.26	13	1.54	2.93[a]
WLC	0.63	0.37	7	2.17	3.25[a]
Placebo	0.59	0.59	0	-5.94	0.72
TAU	0.55	0.17	7	4.45	5.16[b]
Active Tx	0.22	0.22	0	0.26	0.33
CBT/BT	0.13	0.11	1	1.55	1.70

[a] $p<0.01$, [b] $p<0.0001$

Table 10

Subgroup analyses of the overall effect size of ACT RCTs at post-treatment.

Variable	N	*g*	95% CI	Q_b-value	*p*-value
Type of comparison				15.69	0.0001
Active treatment	46	0.26	0.18-0.34		
Passive control	17	0.56	0.44-0.68		
Country				12.94	0.002
United States	29	0.24	0.14-0.33		
European Union	28	0.49	0.39-0.59		
Other	6	0.36	0.06-0.66		
Type of disorder				6.26	0.044
Psychiatric	32	0.27	0.17-0.36		
Somatic	22	0.43	0.31-0.56		
Stress	9	0.45	0.29-0.60		
Outcome measure				5.05	0.080
Self-report	15	0.46	0.33-0.58		
Behavioral	7	0.39	0.23-0.55		
Combined	41	0.28	0.19-0.38		
Recruitment				1.59	0.451
Clinical/Referrals	32	0.33	0.23-0.44		
Advertisements	22	0.40	0.30-0.51		
Mixed	9	0.29	0.14-0.44		
Treatment format				3.57	0.168
Individual	26	0.33	0.21-0.45		
Group	30	0.33	0.24-0.42		
Self-help	7	0.52	0.33-0.71		
Phase of publication				0.43	0.511
Early	13	0.41	0.22-0.60		
Late	50	0.34	0.27-0.42		

Table 11
Summary of the ACT-studies for psychiatric disorders in relation to the EST-criteria.

Study	Comparison condition	WLC	Placebo	TAU	Established treatment	Equivalence analysis	Treatment manuals	Inclusion criteria reliably delineated	Reliable and valid outcome measures	Appropriate data analysis
Depression										
Zettle (1986)	CT				=	0	—	—	+	+
Zettle (1989)	CT/PCT				=	0	+	—	+	—
Peterson (2009)	TAU			=		0	+	—	+	+
Hayes (2011)	TAU			>			+	—	+	—
Folke (2012)	TAU			>			+	—	+	+
Psychotic symptoms								—		
Bach (2002)	TAU			>			+	—	?	+
Gaudiano (2006)	TAU			=		0	+	—	?	+
White (2011)	TAU			=		0	+	—	+	+
Shawyer (2012)	Befriending		=			0	+	—	+	+
Anxiety disorders										
Zettle (2003)	SD				=	0	+	—	+	+
Roemer (2008)	WLC	>					+	+	+	+
Twohig (2010)	PR				(>)		+	+	+	+
Brown (2011)	CBT				=	0	+	—	+	+
Wetherell (2011b)	CBT				=	0	—	—	+	—
Arch (2012)	CBT				=	0	+	+	+	+
England (2012)	Exposure				=	0	+	+	—	—
Kocovski (2013)	CBT	>			=	0	+	—	+	+
Hayes-S. (2013)	AR				=	0	+	+	+	+
Drug abuse										
Hayes (2004)	ITSF/MM				=	0	+	—	+	+
Smout (2010)	CBT				=	0	+	—	—	+
Luoma (2011)	TAU			>			+	—	+	+
Stotts (2012)	Drug couns.				=	0	+	—	+	+
Lanza (2013)	WLC	=				0	—	—	—	—
Nicotine dependence										
Gifford (2004)	NRT				=	0	+	—	+	+
Gifford (2011)	Bupropion				>		+	—	+	+
Bricker (2013)	Smokefree				=	0	+	—	—	?
Borderline PD										
Gratz (2006)	TAU			>			+	—	+	+
Morton (2012)	TAU			>			+	—	+	+
Various disorders										
Woods (2006)	WLC	>					+	—	+	+
Forman (2007)	CBT				=	0	—	—	+	+
Lappalainen (2007)	CBT				>		+	—	+	+

Note. AR = applied relaxation, CT = cognitive therapy, ITSF = intensive twelve step facilitation program, NRT = nicotine replacement treatment, PR = progressive relaxation, SD = systematic desensitization, TAU = treatment as usual, WLC = waiting list control.

> = significantly better than the comparison condition, **=** no significant difference between conditions, **0** = equivalence analysis not performed, **+** = criterion fulfilled, **?** = questionable, **—** = criterion not fulfilled.

Table 12
Summary of the ACT-studies for somatic disorders and stress in relation to the EST-criteria.

Study	Comparison condition	WLC	Placebo	TAU	Established treatment	Equivalence analysis	Treatment manuals	Inclusion criteria reliably delineated	Reliable and valid outcome measures	Appropriate data analysis
Pain										
Dahl (2004)	TAU			>			+	—	—	+
Wicksell (2008)	WLC	>					+	—	+	+
Wicksell (2009)	TAU			=		0	+	—	+	+
Johnston (2010)	WLC	=					+	—	+	+
Thorsell (2011)	AR				(=)	0	+	—	?	+
Wetherell (2011a)	CBT				=	0	+	—	+	+
Motamedi (2012)	TAU			>			+	—	+	+
Wicksell (2013)	WLC	>					+	—	+	+
Buhrman (2013)	WLC	>					+	—	+	+
McCracken (2013)	TAU			=		0	+	—	+	+
Epilepsy										
Lundgren (2006)	ST		>				+	+	+	+
Lundgren (2008)	Yoga				(=)	0	+	+	+	?
Tinnitus										
Westin (2011)	TRT/WLC	>			>		+	—	+	+
Hesser (2012)	ICBT/Disc.		>		(=)	0	+	—	+	+
Overweight/Obesity										
Lillis (2009)	WLC	>					+	—	+	+
Tapper (2009)	WLC	=					+	+	+	+
Weineland (2012)	TAU			>			+	—	+	+
Forman (2013a)	CBT				=	0	+	+	+	+
Forman (2013b)	BT				=	0	+	—	+	+
Various disorders										
Gregg (2007)	Education				>		+	+	+	+
Nordin (2012)	AR				(<)		—	+	+	+
Rost (2012)	TAU			>			—	+	+	+
Stress at work										
Bond (2000)	IPP/WLC	>	>				+	—	+	+
Flaxman (2010a)	SIT/WLC	>			=	0	+	—	+	+
Flaxman (2010b)	WLC	>					+	—	+	+
Brinkborg (2011)	WLC	>					+	—	+	+
Bethay (2013)	ABA				(=)	0	+	—	+	+
Lloyd (2013)	WLC	>					+	—	+	+
Lappalainen (2013)	WLC	=				0	—	—	+	+

Note. ABA = applied behavior analysis, AR = applied relaxation, BT = behavior therapy, CBT = cognitive behavior therapy, ICBT = Internet.based CBT, IPP = innovation promotion program, SD = systematic desensitization, SIT = stress inoculation therapy, ST = supportive therapy, TAU = treatment as usual, TRT = tinnitus retraining treatment, WLC = waiting list control,
> = significantly better than the comparison condition, **=** no significant difference between conditions, **0** = equivalence analysis not performed, **+** = criterion fulfilled, **?** = questionable, **—** = criterion not fulfilled.

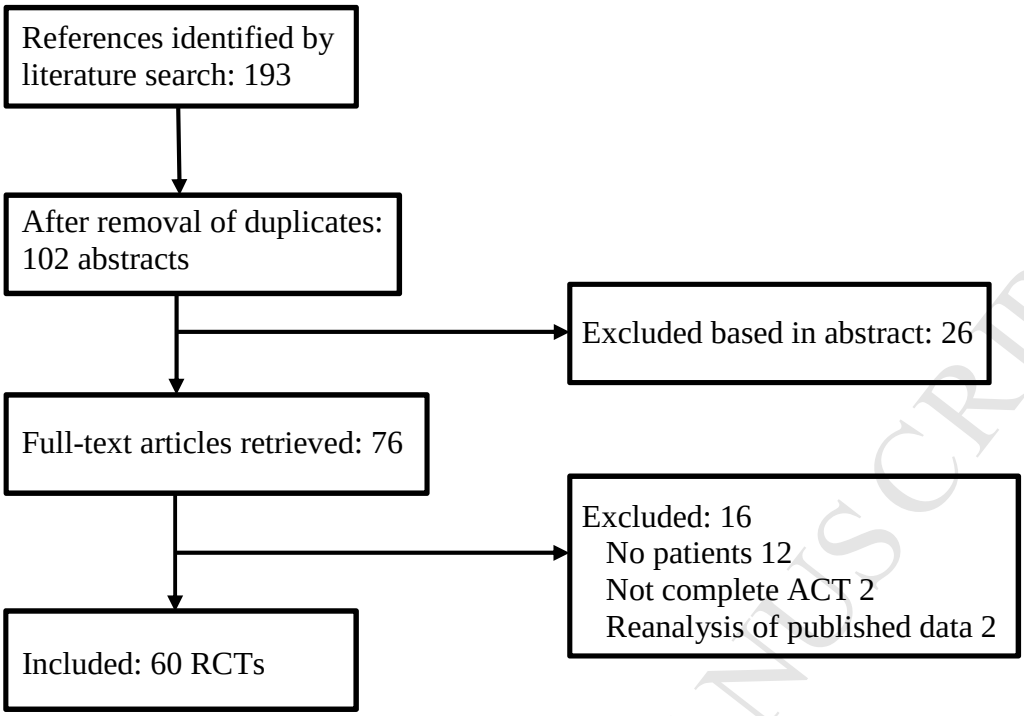

Figure 1. Flowchart of the inclusion of studies.

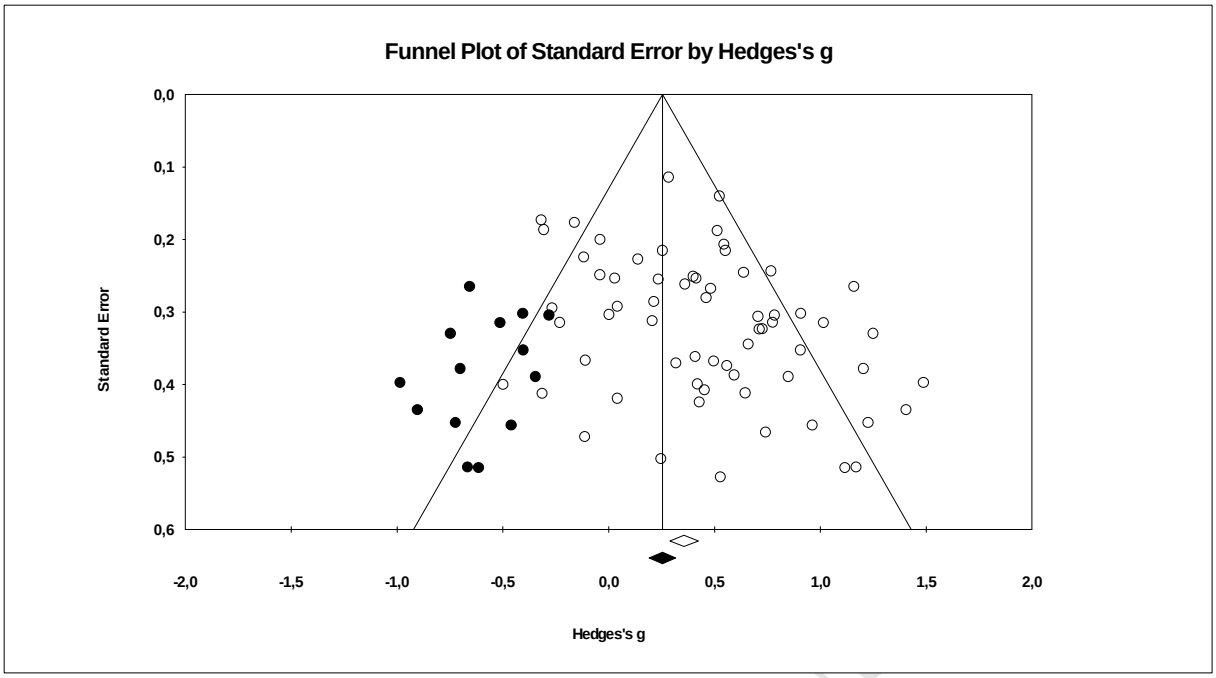

Figure 2. Funnel plot of standard error by Hedges's g. Unfilled circles are observed studies, filled circles are trimmed studies.

Highlights

- ACT RCTs had a number of important methodological problems.
- The overall effect size was small.
- The ES for ACT-CBT comparisons was not significant.
- ACT did not fulfill criteria for well-established treatment for any disorder.