



Rebuttal of Atkins et al. (2017) critique of the Öst (2014) meta-analysis of ACT



Lars-Göran Öst*

Department of Psychology, Stockholm University, Sweden

ARTICLE INFO

Article history:

Received 25 July 2017

Accepted 14 August 2017

Keywords:

Acceptance and commitment therapy

Systematic review

Meta-Analysis

Research methodology

Evidence base

ABSTRACT

Atkins et al. strongly criticize my (Öst, 2014) systematic review and meta-analysis of ACT. The bulk of their re-examination of my article is divided into four parts: a) Selection of studies, b) Ratings of methodological quality, c) Meta-analysis, and d) Judgments of quality of evidence. It is evident from my paper that I have refuted their claims regarding each of these parts. Regarding a) Selection of studies I showed that only four studies had a cell size of less than 10 and their inclusion did not change the mean effect size or increased variability. Concerning b) Ratings of methodological quality I have showed that my ratings were reliable and had accuracy. As for c) Meta-analysis, I have demonstrated that I got very similar results to those of A-Tjak et al. (2015) that Atkins et al. describes as a much better meta-analysis. Regarding d) Judgments of quality of evidence, Atkins et al. brought up 23 studies for which they argued that I have done an incorrect evaluation but for every single study I have disproved their arguments and maintain my 2014 evaluation of the evidence base of ACT. Thus, there is no reason to follow Atkins et al. suggestion that my review “should now be set aside in making decisions regarding the treatment efficacy of ACT.”

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

Atkins et al. (2017) make serious accusations about my systematic review and meta-analysis of ACT (Öst, 2014). In this paper I will refute these claims and show that they are incorrect. The methodology of my review, the actual meta-analysis, and the evaluation of the evidence base of ACT were all done in ways that are in concordance with standards in the field. The only deviation from these is that I was the single author, and the only person evaluating the evidence base. In the discussion of my article I acknowledge this and said: “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.” (Öst, 2014, p. 118). As will be evident from the main part of this paper I have shown that the claims Atkins et al. make are without merit concerning the ACT-studies they mean that I have evaluated incorrectly.

In their introduction Atkins et al. claim that there are 50 errors in Table 1 of my review. This table describes background data of the

included studies and it would require such a lot of journal space to respond to each of these that I have decided not to. Instead I focus on the more serious accusations which Atkins et al. put forth in the main part of their article, particularly questions regarding the evidence base of ACT (Tables 11 and 12 in Öst, 2014).

2. Öst's 2014 review

Atkins et al. claim “Despite written and face-to-face requests, Öst has not provided us with the actual study by study effect size data used in his meta-analyses.” This is not true. I have had no face-to-face request but an e-mail from Paul Atkins in May 2015 asking for the effect size data and the quality rating data. These were directly provided to him via e-mail. Then I did not hear anything on this issue until July 2016 when I got an e-mail from Michelle Craske, editor-in-chief of *Behaviour Research and Therapy* asking for the same information, and I mailed it to her the same day. It does not make any sense that when Atkins asked for effect size and quality rating data I only mailed him the latter. Why didn't he come back with a new request for the effect size data if he only got half of what he asked for?

* Department of Psychology, Stockholm University, S-106 91, Stockholm, Sweden.
E-mail address: ost@psychology.su.se.

3. Part a): selection of studies

One critique of Atkins et al. is “Öst (2014) included studies with fewer than 10 participants per cell in the design.” and that “the decision to examine tiny studies necessarily increases variability, and reduces methodological quality.” The reason why I included all RCTs on ACT with patients was not to be criticized that some important studies were excluded. However, *only four* of the 60 studies (6.7%) had a cell size of less than 10 participants (Öst, 2014, Table 1). In the statistical analysis I used, as recommended, Hedges’ *g*, which corrects for small sample sizes. Furthermore, the software used (Comprehensive Meta-Analysis, v. 2) computes the mean ES by weighing each ES by the inverse of its variance. The four small studies had the following relative weights in the random effects model: Brown et al. (2011) 0.85%, Dahl, Wilson, and Nilsson (2004), 0.97%, Lundgren, Dahl, Yardi, and Melin (2008), 0.96%, and Zettle and Hayes (1986) 0.85%, thus, contributing with only 3.63% to the mean effect size. If these studies were deleted from the analysis, as Atkins et al. suggest, the *g* for all studies is 0.41 (95% CI 0.29–0.52), which should be compared to the *g* = 0.42 (95% CI 0.31–0.53) when they are included (Öst, 2014, Table 8, p. 113). Thus, the confidence intervals are almost exactly the same with these small cell size studies included, and Atkins et al. statement that “tiny studies necessarily increases variability” is incorrect. What about reduction of methodological quality? I calculated the Pearson correlation between the cell size of the RCT and the quality score and it was not significant ($r = 0.17$, $p = 0.20$). Thus, both parts of Atkins et al. critique that I included small studies in the meta-analysis are incorrect.

Atkins et al. go on to say “Öst specifically excluded studies that explored components within ACT” This statement is misleading. I excluded “RCTs with only 1–2 components of ACT” (p. 106) to avoid the possible critique that I did not analyze full ACT treatments, and as can be seen in the flow-chart (Öst, 2014, p. 106) *only two* studies were excluded for this reason.

Then Atkins et al. say that I “failed to examine process or mediational evidence in the randomized trials that were included.” They make it sound as if including such data in a meta-analysis of a form of psychotherapy is compulsory, which is not the case. **It would require a completely separate meta-analysis since such a large number of ACT mediational studies have been published.**

4. Part b): ratings of methodological quality

Atkins et al. strongly criticize the scale I developed (Öst, 2008) for rating of methodological quality saying “Presently, the use of Öst’s list presents an analytic challenge because: a) some of the operational definitions of given features are controversial, b) evidence-based therapy is moving toward a more process-based and transdiagnostic approach which is given short shrift in the list, c) key methodological issues are left off the list, d) over the last eight years this list appears to have been applied to ACT studies and little else, and e) it is unclear whether the scale is reliable and valid.” However, they do not substantiate any of these points. I am aware of two similar scales; the *Quality Rating System* (Moncrieff, Churchill, Drummond, & McGuire, 2001) and the *RCT of Psychotherapy Quality Rating Scale* (Kocsis et al., 2010). The first contains 23 items and the second 25 items, and both use the same 0–2 rating as my scale. If anything, these scales have briefer descriptions of the scale steps than my scale. The three scales are quite similar and do not focus on process as Atkins et al. would like. Furthermore, my scale has been used in at least 11 meta-analyses that I am aware of; four on ACT, two on OCD, one each on Third wave of BT, Depression in children, Brief treatments for anxiety disorders in children, Mindfulness for psychiatric disorders, and Return-to-

work interventions for people on sick leave (the last two under editorial review). I comment on the scale’s reliability and validity in a later section.

Then they describe in length the result that was obtained when the scale was used in the meta-analysis by A-Tjak et al. (2015) on ACT and compare the scores for 36 studies included in both meta-analyses, finding a significant difference; my ratings were overall 10% lower than those of the A-Tjak et al. meta-analysis. In that meta-analysis two raters rated all the included studies and they obtained excellent inter rater reliability (ICC = 0.99). However, this does not say anything about *accuracy* of their ratings. The proper use of any rating scale that is based on expert judgement (clinical or research) requires that the rater has been trained by an expert, e.g. the developer of the scale, and fulfils accuracy criteria before starting to use the instrument in clinical work or research. Examples of this are diagnostic schedules (e.g. Structured Clinical Interview for DSM-5; First, Williams, Karg, & Spitzer, 2015), severity rating scales (e.g. Hamilton Anxiety Rating Scale; Hamilton, 1959), and therapist adherence/competence scales (e.g. Cognitive Therapy Adherence and Competence Scale; Barber, Liese, & Abrams, 2003). In such training multiple examples are displayed to illustrate what the different scale steps of the items mean and the raters learn by discussion with the experts. If the raters have not been trained to accuracy they can be using the scale reliably but in an idiosyncratic way, e.g. systematically inflating their ratings of certain items. This is probably the case with the raters in the A-Tjak et al. (2015) meta-analysis who have not obtained any accuracy training. To exemplify this situation let us look at Table 1 of Atkins et al. The two items with the largest difference between my and the A-Tjak et al. ratings are: 16. Checks for treatment adherence, and 17. Checks for therapist competence. The item descriptions are the same for both items:

- 0 *Poor*. No checks were made to assure that the intervention was consistent with protocol.
- 1 *Fair*. Some checks were made (e.g. assessed a proportion of therapy tapes).
- 2 *Good*. Frequent checks were made (e.g. weekly supervision of each session using a detailed rating form).

What is not spelled out in the description but taught in training regarding these items is that it is not enough to write in the method section that therapy sessions were video or audio taped for assessment of adherence and competence. If the authors don’t present any data on adherence and/or competence they have either not assessed it, or the result was so disappointing that they chose not to publish it. This can be exemplified by the following studies included in my meta-analysis. White et al. (2011) say that “All therapy sessions were recorded and competence and fidelity assessed by an expert in ACT (GM).” (p. 903). Mo’tamedi et al. (2012) say “All sessions were (audio) recorded, then the first author evaluated the contents using a detailed checklist associated to each session.” (p. 1111). Lundgren et al. (2008) say “The sessions were videotaped and audiotaped to ensure treatment integrity.” (p. 105). It is difficult to understand how the mere recording of therapy sessions can *ensure* treatment integrity. Brinkborg, Michanek, Hesser, and Berglund (2011) say “Adherence to the manual was controlled using a checklist after each session.” (p. 392). It is equally improbable that the therapists’ use of a checklist after each session can *control* adherence. Common for these studies, and other RCTs getting a zero rating on these items, is the lack of reporting actual data on adherence and/or competence.

Atkins et al. also submit that “Öst (2014) reported that methodological quality had not improved while A-Tjak et al. (2015, p. 34) reported that it had improved from the 2008 analysis.” However, A-Tjak et al. only say in their discussion “the methodological quality of

the ACT studies seems to have improved over the years,” (p. 34) without providing any data whatsoever for this statement. Should the reader really have more faith in such a statement than in my meta-analysis which provided a full table (Öst, 2014, p. 111) comparing early and recent studies on the total score and each item of the methodological quality scale?

Then Atkins et al. go on to criticize the psychometric evaluation of the quality rating scale in my meta-analysis asking for information that practically never are provided in meta-analyses applying such scales. The A-Tjak et al. (2015) meta-analysis described their procedure in only four lines, whereas I used 10 lines. I apologize for not mentioning that differences between me and the independent, blind, graduate student raters were solved in a discussion to reach consensus. Atkins et al. say that the method I used was “vague and unusual” but without having reviewed a lot of meta-analyses in this respect I argue that the method used by A-Tjak et al., with two raters rating all studies, is unusual. Then they describe a lot of new information about the rating procedure in the A-Tjak et al. study which is not to be found in the published article, and it is impossible for a reader to know if this is an adequate description of what was done. Thus, I maintain that the methodological quality ratings in my meta-analysis are more valid than those in A-Tjak et al. since they are done by the developer of the scale and inter-rater reliability against blind independent raters trained to accuracy is excellent.

5. Part c): effect size data

Atkins et al. also question my analysis of effect size data but fail to compare my results with those of A-Tjak et al. (2015) which are based on 39 studies whereas I had 60. I and A-Tjak got the following mean effect sizes (Hedges' *g*) for the respective comparisons: ACT vs. waitlist: 0.63 and 0.82; ACT vs. placebo: 0.59 and 0.51; ACT vs. Treatment as usual: 0.55 and 0.64; ACT vs. established treatments: 0.22 and 0.32. These differences are small and most likely due to the fact that I included many more studies than A-Tjak did.

6. Part d): judgments of quality of evidence

Atkins et al. claim “In judging the degree of empirical support, Öst (2014) made two sets of quality ratings in his article. The first set was based upon his idiosyncratic set of 22 criteria and has already been discussed in the section on ratings of methodological quality.” This is incorrect. The methodological quality ratings were *not* used to judge empirical support. Under the heading *Criteria for evidence-based treatments* (p. 107) I clearly described that I applied the APA Division 12 criteria for this evaluation.

One criterion that Atkins et al. and I disagree on is “3. Conducted with a population, treated for specified problems, for whom inclusion criteria have been delineated in a reliable, valid manner.” (Öst, 2014, p. 107). In order to know if the participants in a RCT have the disorder in focus for treatment it is not enough to say that they were diagnosed using DSM-IV, even if a structured interview schedule was used. I argue that the inter-rater reliability of the diagnosticians must be evaluated by having blind, independent, and trained clinicians rate a random sample (preferably 20%) of all diagnostic interviews (both included and excluded participants), and the kappa coefficients must be at an acceptable level. A couple of examples will illustrate this. The Gaudio and Herbert (2006) study says “Participants met DSM-IV American Psychiatric Association (APA, 1994) criteria for psychotic disorder or affective disorder with psychotic symptoms.” (p. 419). There is no information on which the diagnostician is, his/her training, use of an interview schedule, or inter-rater reliability. In Folke, Parling, and Melin (2012) the first inclusion criterion is “a diagnosis of unipolar

depressive disorder as defined by the Diagnostic and Statistical Manual of Mental Disorders, 4th edition” and “All diagnoses were established by each participant's M.D. (in most cases a psychiatrist).” Here there is a lack of an interview schedule and assessment of inter-rater reliability. Thus, I argue that in order to judge that a RCT has delineated inclusion criteria in a “reliable, valid manner” inter-rater reliability of the diagnostic procedure has to be provided.

7. Psychiatric disorders

7.1. Depression

Atkins et al. say “For Petersen and Zettle (2009), Öst erroneously compared depression outcomes at discharge between ACT and treatment as usual (TAU) arms and concluded there was no difference. The main outcome variable was *time-to-discharge* as the authors clearly stated in their paper.” However, this is not correct. Petersen and Zettle (2009) do not state anywhere in the introduction or method section of their paper what they have selected as the main or primary outcome variable. Under the heading *Measures*, they describe “a) depression outcome measures at both pretreatment and posttreatment, b) a process measure of experiential avoidance at both measurement occasions, and c) treatment dosage measures.” (p. 525). Since the participants were diagnosed with depression and alcohol use disorder and the authors did not specify the primary outcome I think it is reasonable to consider the depression measures to be primary. On both HRS and BDI-II there were significant time effects but no group or interaction effects, and thus my conclusion is correct.

Atkins et al. say “Öst incorrectly reported that L. Hayes, Boyd, and Sewell (2011) used the web-based Development and Well-Being Assessment (DAWBA) as the diagnostic method, when their paper reported that they used the clinician rated DAWBA from clinical interviews for recruitment of participants into the study.” This is not correct since I did not say that the diagnosis was based only on the patients' online answers. The Hayes et al. study probably used clinician interviews after the patients had answered DAWBA online. Atkins et al. also refer to reliability data that the originator of DAWBA obtained in their initial study (Goodman, Ford, Richards, Gatward, & Meltzer, 2000). However, there are no such data in the Hayes et al. study, and there is no way to know how reliably the diagnoses in this study were ascertained. They also refer to a study by Aebi et al. (2012) showing a good agreement between DAWBA and clinician only diagnosis. However, these diagnoses were not obtained after using a validated interview schedule, e.g. Kiddie-SADS, and thus do not speak to the validity of DAWBA as a diagnostic instrument.

Then Atkins et al. say that “Öst incorrectly classified White et al. (2011) as a study investigating the efficacy of ACT on active psychotic symptoms.” The reason for classifying the White et al. (2011) study under psychotic symptoms is that the authors say that “Participants all met ICD-10 (WHO, 1992) criteria for a psychotic disorder ...” (p. 902), and Table 1 shows that 23 out of 27 patients (85%) had some kind of psychotic disorder. Atkins et al. also claim that I used the measure of psychotic symptoms, the Positive and Negative Syndrome Scale (PANSS), as the primary outcome measure in my meta-analysis, and deleted the measure of depression. This is not correct since I used both the PANSS and the HADS for this study. Finally, they say that “In fact, for depression (the targeted outcome), ACT outperformed TAU for this study.” However, this statement is incorrect since White et al. (2011) did not specify their primary outcome measure and there was only a trend ($p = 0.051$) that ACT did better than TAU on the depression measure and no difference on the anxiety measure. One would assume that more

than this result is required for the conclusion that “ACT outperformed TAU.”

Thus, I maintain that my original conclusion that ACT was *possibly efficacious* for depression at the time of my review early 2014 is correct.

7.2. Psychotic symptoms

Atkins et al. say that “For the study conducted by Gaudiano and Herbert (2006), Öst (2014) incorrectly reported no significant differences from TAU. Gaudiano and Herbert (2006, pp. 427–428) reported a significant difference between groups for distress related to hallucinations,” and that this “was a key outcome for the study”. At the end of the introduction of this study the authors say that “It was hypothesized that the ACT group would show greater improvement on symptom measures at post-treatment” (p. 418). In the method section the first measure described is the Brief Psychiatric Rating Scale (BPRS) (p. 419) and Table 2 in results has BPRS as the first measure. Nowhere have the authors described that distress related to hallucinations was the primary or key outcome measure. Table 2 contains 12 measures and only two yielded a significant difference. Also, there was no significant difference regarding rehospitalization. The authors claim that I “omitted results from psychoses outcomes in the calculation of effect sizes in his meta-analysis.” This is not true since both BPRS and Distress about hallucinations were used to calculate ES. Finally, Atkins et al. say that “Gaudiano and Herbert found significant differences on the SDS and CGI measures.” However, on page 428 of this article we can read “no significant group differences were found on the CGI-S” and further down “a marginally significant difference was found on the CGI improvement scale”.

Atkins et al. claim that “Shawyer et al. (2012, p. 112) clearly indicated the inclusion criteria for their study” but they did not use a validated interview schedule, e.g. SCID, and failed to assess interrater reliability of the diagnostic procedure.

Thus, I maintain that my original evaluation in early 2014 that ACT is *possibly efficacious* for psychotic symptoms is the correct one.

7.3. Anxiety disorders

Atkins et al. contend that “Öst inappropriately included a study by Wetherell, Afari, Ayers, et al. (2011) that investigated whether ACT could be applied to Generalized Anxiety Disorder in older adults. This study was not an RCT. CBT was not mentioned anywhere in the title nor in the abstract; and the paper did not conduct any comparisons of the effectiveness of ACT and CBT.” This makes me wonder if the authors have read the Wetherell et al. article since in the method section under Procedure it says “the 21 participants were randomly assigned to receive either 12 weekly hour-long individual sessions of ACT ($n = 11$) or CBT ($n = 10$).” (p. 129). Thus, it is a RCT with CBT as the comparison condition, and it is correctly included in the review and meta-analysis.

Atkins et al. also claim that I misrepresented the Arch et al. (2012) study by reporting that there was no significant difference between ACT and CBT, which was the case for the primary outcome measure at post-treatment assessment. They say that “at 12-month follow-up, ACT did show significantly lower clinical severity ratings than CBT among completers”. However, Atkins et al. fail to mention the lack of difference in the intent-to-treat analysis, which is used by all modern meta-analyses if available. Also, the follow-up period was uncontrolled, which is the case in most psychotherapy RCTs, and between post-assessment and 6 month follow-up significantly more of the ACT-patients (39%) than the CBT-patients (19%) had used psychotherapy outside of the study. The authors also question my conclusion that “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,” (p. 113). However, since Division 12 of APA and other international organizations classifying the evidence base for psychotherapies use specific diagnostic categories (DSM or ICD) and mixed anxiety is not one of them, it is hard to disregard this problem. My option was to exclude the Arch et al. study but I did not want to do that since it had the highest quality score (34) of all ACT studies in my meta-analysis.

Regarding the England et al. (2012) study Atkins et al. claim that the authors “used the structured clinical interview for DSM-IV (SCID) conducted by blinded raters with established interrater reliability to determine diagnostic status as their primary outcome.” However, reading the article I find that “For training and reliability purposes, a portion of the diagnostic assessments were conducted jointly by two diagnosticians (one of whom was an advanced graduate student); the assessments of the two diagnosticians were then compared as a reliability check. All diagnostic evaluations were reviewed with one or both of the licensed clinical supervisors for the study (JDH and EMF) prior to enrolling participants. Diagnostic interrater reliability was greater than 80% for all assessments, and was 100% for participants enrolled in the study.” (p. 68). This does not qualify as independent and blind assessment since the two supervisors were fully aware of the fact that the study recruited participants with nongeneralized social anxiety disorder (SAD). Furthermore, the three self-report measures had very low internal consistency (Cronbach's α): PCRS 0.41, SSPS 0.34, and STAI 0.36. There were no significant differences on any of the continuous measures (clinician ratings, self-report, or behavioral). The only significant difference the authors found was on the proportion of patients no longer meeting the SAD-criteria. However, this was due to using chi-square in a questionable way, since the expected value of half the cells in their 2×2 table was below 5 (1.9 and 2.1, respectively). In such a situation Fisher's exact probability test is strongly recommended, and it yielded a nonsignificant $p = 0.11$.

Atkins et al. then say “Öst agreed that the (Kocovski, Fleming, Hawley, Huta, & Antony, 2013) study was of high quality.” What I said about the Kocovski et al. study was: “reported no significant difference between ACT and CBT, but failed criterion 3.” (Öst, 2014, p. 113). Thus, my evaluation in early 2014 that ACT is *possibly efficacious* for SAD has to be maintained.

7.4. Drug abuse

Atkins et al. submit that “Öst reported that the outcome measures in Smout et al. (2010) were not reliable and valid, but the self-report instruments used were accompanied by hair analysis, which is an objective, valid, and reliable outcome measure for methamphetamine use”. The primary outcome measure in this study was the self-reported average amount of methamphetamine used in the past month (p. 101). This was “assessed through a semistructured interview, conducted by the second author. This interview was developed for this study.” However, the authors have not provided any psychometric data regarding reliability and validity of this interview measure. When it comes to my comment on attrition in this study I did not single out ACT but said “This study had an astonishing attrition of 70%” (p. 113).

Regarding the study by Stotts et al. (2012) Atkins et al. say that “Öst's method of calculating the effect size was incorrect since the method he applied $[(M_{ACT} - M_{comparison})/SD_{pooled}]$ is appropriate for a continuous outcome whereas this trial used a dichotomous outcome.” I apologize for not specifying in the method section that in the few studies that used a dichotomous measure as the primary outcome that statistic (log Odds ratio) was converted to Cohen's d , as described by the Borenstein, Hedges, Higgins, and Rothstein (2009)

handbook, before using it in the meta-analysis.

Thus, my early 2014 conclusion that ACT is *possibly efficacious* for drug abuse is maintained.

7.5. Nicotine dependence

Atkins et al. argue that “Öst reported the overall outcome of the Bricker, Wyszynski, Comstock, and Heffner (2013) study as non-significant, whereas the study found a significant difference between ACT and Smokefree.gov.” In this study 111 participants were randomized to ACT and 111 to Smokefree, but smoking cessation at posttreatment was analyzed only for the 57 and 58, respectively, who provided data, thus assuming that these subjects were representative for the entire sample. With this questionable assumption they obtained an OR with a $p = 0.05$. When I recalculated these data using Chi-square (3.24, $p = 0.072$) and Fisher's exact probability test (2-tailed, $p = 0.084$; 1-tailed, $p = 0.061$) the difference was not significant. Taking the conservative stance, as is recommended when it comes to addiction, and assuming that the participants who did not provide data were still smoking yielded a Chi-square (2.82, $p = 0.093$) and Fisher's exact probability test (2-tailed, $p = 0.149$; 1-tailed, $p = 0.074$). The authors also question my evaluation “that the Bricker et al. (2013) study did not meet criterion 4 (use of a reliable and valid outcome measure). Self-reported smoking is a standard method for assessing web-based interventions.” I find it difficult to accept that the assessment of the primary outcome measure should vary depending on the intervention; self-report for web-based and biochemical confirmation for face-to-face treatment. What way of assessing should then be used in a study comparing web-based and face-to-face interventions?

Thus my evaluation from early 2014 that ACT is *experimental* when it comes to nicotine dependence has not been changed.

7.6. Borderline personality disorder

Atkins et al. say that “Öst (2014, p. 113) argued that ACT should be rated as “experimental” for BPD because “Both studies gave the TAU-treated subjects markedly less therapy hours.” Regarding the Gratz and Gunderson (2006) study note that I did not say significantly less. However, the difference in total hours of therapy is 9 hours (18%), which may be of importance. I also want to emphasize what I wrote in my article “Gratz and Gunderson (2006) combined ACT with DBT, BT, and emotion focused therapy, which means that the contribution of ACT is impossible to ascertain.” (Öst, 2014, p. 113).

The Morton, Snowden, Gopold, and Guymer (2012) study gave the ACT + TAU participants 12 group session of 2 hours for a total of 24 hours. Regarding the TAU condition it is difficult to find an exact figure but the authors say “Contact with a clinician at least once every 2 weeks was required.” (p. 534). Counting a contact as 1 hour gives a total of 6 hours for TAU, 75% less than the treatment time for ACT + TAU. Morton et al. in the discussion also say “the ACT groups were delivered by experienced staff who specialized in group treatment of people with severe personality disorders. It is possible that it was not the ACT content of the groups that resulted in the improvements, but rather access to specialized treatment, or access to systematized treatment.” (p. 542). The TAU consisted of “public mental health services, and typically consisted of low-key supportive contacts, medication management, with in-patients admissions and crisis contacts if required.” (p. 534). Thus, the TAU-participants got much less therapy, provided by non-experts in personality disorders, and these differences between conditions are a serious threat to the internal validity of the study.

My evaluation from early 2014 that ACT is *experimental*

regarding BPD is maintained.

8. Somatic disorders and stress

8.1. Pain

Atkins et al. claim that “Öst (2014) criticized Dahl et al. (2004) on both inclusion criteria and outcome measures, arguing that they had not used a “structured diagnostic interview.” This is not correct since I never argued that the Dahl et al. study did not use a structured diagnostic interview merely that it did not meet criterion 3. Then they say “In terms of outcome measures, sick leave was the primary dependent measure and it is difficult to understand how number of sick leave days is not a valid measure of sick leave.” The problem with this measure in the Dahl et al. (2004) study is that it is self-report and not objective sick leave data collected by the Swedish Social Insurance Agency, which is the standard way of research on sick leave in Sweden. Thus, there is no way of knowing how reliable and valid the self-report data are.

Regarding the study by Wicksell, Ahlqvist, Bring, Melin, and Olsson (2008) Atkins et al. say that it “was also criticized by Öst for inclusion criteria. Their paper clearly specified the inclusion criteria.” The paper has a single sentence on inclusion criteria: “People older than 20 years who reported being diagnosed with WAD and with pain duration of more than 3 months were considered eligible for inclusion.” (p. 171). This means that WAD can be diagnosed; otherwise the participants would not be able to report having got that diagnosis. Why then did the authors not diagnose the patients before inclusion but relied on their self-report of being diagnosed with whiplash-associated disorder? Since the study was carried out at a Pain treatment service of the Karolinska University Hospital in Stockholm there should not be a shortage of pain specialists to diagnose the patients pre-treatment.

Regarding the Wicksell, Melin, Lekander, and Olsson (2009) study of children experiencing idiopathic pain Atkins et al. once again say that the inclusion criteria were well specified. Here too the inclusion criteria were described in a single sentence: “Patients between 10 and 18 year referred to the PTS with pain duration more than 3 months, were considered eligible for inclusion in the study.” (p. 250). Their first exclusion criterion was: “pain was explained by an identified pathological process (e.g. arthritis, cancer, inflammatory bowel disease)” which requires a diagnostic procedure. Thus, idiopathic pain can be diagnosed as a disorder after excluding other pathological processes, and the reliability of the diagnostic procedure can be assessed. In the Wicksell et al. (2013) study patients with fibromyalgia who fulfilled all of the American College of Rheumatology classification criteria for fibromyalgia were included. Thus, there are specific criteria applied by a clinician and the reliability of this procedure should be possible to ascertain.

Atkins et al. note that “For Thorsell et al. (2011), Öst included a question mark regarding the outcome measure.” The reason for this is that Pain intensity was assessed by a single item rated on a 0–10 scale with no psychometric data provided. Also, Level of function was assessed by five items selected from a longer self-report scale, once again without psychometric data.

Finally, Atkins et al. report that I evaluated the Wetherell, Afari, Rutledge (2011) study to not have reliable inclusion criteria. They say that “The study was purposefully designed to include non-malignant chronic pain from many medical sources (e.g. arthritis, fibromyalgia, etc.)” However, these disorders can be diagnosed using specific criteria as illustrated by the Wicksell et al. (2013) study on fibromyalgia discussed in a previous paragraph.

Thus, I maintain my early 2014 evaluation of ACT as *probably efficacious* for pain.

8.2. Tinnitus

Atkins et al. argue that “For Westin et al. (2011) Öst criticized the inclusion criteria, but the diagnosis of tinnitus was established using a standardized diagnostic interview under the supervision of an ear-nose throat physician.” When a standardized interview is used to ascertain if the patients fulfil the diagnosis it is clearly possible to record this interview and assess the reliability by having another trained and blinded interviewer go through a randomized proportion of the interviews. Furthermore, there is no information on who is doing the interviews and what qualification that person has for carrying out this task.

8.3. Overweight/Obesity

Atkins et al. say that “Öst argued that Lillis, Hayes, Bunting, and Masuda (2009) did not make use of reliable, valid inclusion criteria.” The reason for my position is that in the first paragraph of the method section the authors say “Participants who had completed at least 6 months of any structured weight loss program in the last 2 years were recruited.” They assessed BMI but did not specify a certain score as inclusion criterion and I don't think that the criteria used can qualify as reliable, valid inclusion criteria as Atkins et al. do. My early 2014 evaluation of *possibly efficacious* is thus maintained.

8.4. Stress

Atkins et al. claim that “For Bond and Bunce (2000), Öst excluded the published findings that at post-treatment and follow-up, stress levels (as measured by the General Health Questionnaire) were better for ACT compared with the behavior therapy intervention.” This is not correct for two reasons. In Table 12 I indicated that ACT was significantly better than both the Innovation Promotion Program (IPP) and the waitlist condition. Secondly, IPP is not behavior therapy and Bond and Bunce (2000) do not say that it is in their description of this intervention (p. 159).

Finally, Atkins et al. say that “Öst (2014, p. 110) states “0/7 stress studies diagnosed the participants”, but given that there is no DSM diagnosis available for “stress at work”, this criticism is clearly not justified.” I am fully aware of the fact that there is not a DSM diagnosis for stress at work but it should clearly be possible to use a semi-structured interview, perhaps in combination with a cut-off score on some psychometrically sound self-report scale. They continue saying that “the stress studies listed inclusion and exclusion criteria appropriate for the populations of interest.” Is this statement really correct? If we look at the Bond and Bunce (2000) study discussed in the previous paragraph we find the following under the heading Participants. “Ninety people in a large media organization volunteered to participate ‘in a stress management program that is occurring during working hours.’ Participants were recruited by means of two notices sent through an internal electronic mail system, as well as a briefing paper read at team meetings.” There is not a single sentence on inclusion and exclusion criteria in the Method section of this paper. So I disagree with Atkins et al. that the stress studies have used appropriate inclusion and exclusion criteria and my early 2014 evaluation of ACT for stress at work is *possibly efficacious* is maintained.

In summary, Atkins et al. brought up 23 studies for which they argued that I have done an incorrect evaluation. For every single study I have disproved their arguments, and thus I maintain that my 2014 evaluations of the evidence base of ACT were correct.

9. Independent evaluation of ACT's evidence base

Recently Moriana, Gálvez-Lara, and Corpas (2017) published a review of how four leading organizations evaluated the evidence-base for psychological treatments of psychiatric disorders in adults. These organizations were the National Institute for Health and Care Excellence (NICE), Division 12 (Clinical Psychology) of the American Psychological Association (APA), Cochrane, and the Australian Psychological Society (APS). Somatic disorders, e.g. chronic pain were not included in this review. Below I summarize the findings of Moriana et al. (2017) for the psychiatric disorders included in my meta-analysis.

Mixed anxiety. Since this is not a diagnosis the other organizations have not evaluated the evidence base. However, we can get some information by looking at panic disorder and social anxiety disorder, which are included in mixed anxiety. The ACT evidence base for panic disorder was judged as weak by Cochrane, and that for social anxiety disorder as weak by APS. These evaluations are clearly more in line with my (possibly efficacious) than that of Division 12 (modest research support).

OCD. NICE evaluated that ACT had insufficient evidence and APS that the evidence base was weak. This is also much closer to my evaluation of possibly efficacious than Division 12's modest research support.

Depression. Cochrane's evaluation of ACT's evidence base was weak, whereas APS said modest. Thus, my evaluation (possibly efficacious) is supported by that of Cochrane, whereas that of Division 12 is supported by APS.

Schizophrenia. The evidence base for ACT has not been evaluated by NICE, Cochrane, or APS, perhaps due to insufficient number of studies on this disorder.

Thus, I find more support for my evaluations of the evidence base of ACT for these four psychiatric disorders than for the judgement by Division 12, and I see no reason to change my early 2014 evaluations.

10. Recent evidence

I don't understand how recent evidence, which I interpret as RCTs published after the time period (1986–2013) covered in my meta-analysis, may have any bearing on the conclusions I draw. Atkins et al. say that “there are at least 171 RCTs of ACT https://contextualscience.org/state_of_the_act_evidence”, without any qualification, thus giving the impression that all of these would be included in the type of meta-analysis that I performed (Öst, 2014). However, this is not at all the case since one inclusion criterion was that the study must have “participants with either a psychiatric disorder, a somatic disorder, or stress reactions in work situations”, and one exclusion criterion was “Studies with normal people not applying for treatment”. (p. 106). To illustrate this problem the following RCTs are some of those included in this list of 171 RCTs that would be excluded based on the titles of the articles: on stigma (e.g. Clarke, Taylor, Bolderston, Lancaster, & Remington, 2015), on prevention programs (e.g. Levin, Hayes, Pistorello, & Seeley, 2016), on training of clinicians/staff (e.g. Luoma, & Vilardaga, 2013), on mental health promotion in non-patients (e.g. Fledderus, Bohlmeijer, Smit, & Westerhof, 2010), on depressive symptoms (e.g. Kohtala, Lappalainen, Savonen, Timo, & Tolvanen, 2015), on treatment-resistant participants (Clarke, Kinston, James, Bolderston, & Remington, 2014), on increasing self-compassion (Yadavaia, Hayes, & Vilardaga, 2014), on procrastination (Wang et al., 2015), on parenting (Brown, Whittingham, Boyd, McKinlay, & Sofronoff, 2015), on caregivers (Losada et al., 2015), on physical inactivity (e.g. Ivanova, Yaakoba-Zohar, Jensen, Cassoff, & Knauper, 2016), and on psychological wellbeing of university students

(Räsänen, Lappalainen, Muotka, & Lappalainen, 2016). Naturally, these are all problems well worth studying but RCTs on these cannot be used to evaluate the evidence base of ACT for psychiatric or somatic disorders, or stress at work. I have not had the time to obtain and read the 83 RCTs published 2014–2016 but a tentative evaluation based only on the titles indicates that 58% of them might have fulfilled the inclusion criteria for my meta-analysis.

Then Atkins et al. say that “nearly 50 mediational analyses are currently available on ACT interventions” and that “theoretically coherent ACT processes commonly mediate ACT outcomes”. There are at least two problems with the ACT mediational studies. First, the design of these studies (e.g. Wicksell et al., 2013) is such that the mediator (often a version of Acceptance and Action Questionnaire; AAQ) and the primary outcome variable are assessed pre, post, and at follow-up. Then the change in the mediator pre–post is included in the statistical analysis to test if it mediates the change in the outcome variable pre–follow-up. However, since the authors of these studies have already shown that the outcome variable has changed significantly pre–post there is no way temporal precedence can be shown. Kazdin (2007) clearly specifies that “A timeline must be established to infer a causal relation or mediator of change. Causes and mediators must temporally precede the effects and outcomes.” Thus, a study needs to show that the potential mediator changed significantly before the outcome variable did so otherwise it is not possible to consider it to be a mediator. Second, the validity of the AAQ as a measure of psychological flexibility has been questioned. Wolgast (2014) investigated the validity of AAQ-II and reported that “The findings of the performed exploratory factor analysis indeed showed that the items of the AAQ-II loaded on the same factor as items designed to measure general distress and did not load on the same factor as the items that were designed to measure acceptance/nonacceptance as an explicit attitude or response to aversive psychological states.” (p. 837). He then concluded “the discriminant validity of the AAQ-II is highly questionable.” (p. 837). Thus, if the purported mediator measures distress there is no wonder that it correlates with outcome measures of distress in mediational analyses carried out in a methodologically questionable way.

11. Conclusion

Atkins et al. conclude their critique of my systematic review and meta-analysis by saying that “Its most fundamental empirical errors are the use of an idiosyncratic and unvalidated rating scheme that appears not to have been reliably applied. The review contains numerous factual and interpretive errors in the reporting of trials included in the review.” This is not correct. As I have shown above my rating scale contains most of the important variables to judge the quality of psychotherapy RCTs and it is very similar to two other scales in this field; the Quality Rating System (Moncrieff et al., 2001) and the RCT of Psychotherapy Quality Rating Scale (Kocsis et al., 2010). It is correct that validity data on my scale has not yet been published (in an English language publication). However, in a Swedish book (Öst, 2016) I showed initial data on concurrent validity by a significant correlation with the Cochrane risk-of-bias ratings (Öst, Havnen, Hansen, & Kvale, 2015), and discriminant validity by a significant correlation with the impact factor of the journals publishing the RCTs in the Öst (2014) meta-analysis. Of course, this is only a beginning but it is more than the other two scales can show as of 2016.

The scale was used with excellent inter-rater reliability assessed against raters who had been trained to accuracy by the developer of the scale. Two later meta-analyses on OCD in adults (Öst et al., 2015; ICC = 0.92) and in children (Öst, Riise, Wergeland, Hansen, & Kvale, 2016; ICC = 0.92) got very similar results as in the Öst

(2014) study (ICC = 0.90). Furthermore, for every single study that Atkins et al. brought up under the heading *Part d): Judgement of quality of evidence* and claimed that I have committed factual and interpretive errors I have shown that they are not correct.

Atkins et al. recommend that “future reviews and meta-analyses utilize rating methods that are broadly accepted by the mainstream scientific community” but they don’t describe any such method, probably because none exists. In addition to the three methodology quality scales described above there is the Cochrane Reviews risk-of-bias rating, which has items rated 0, 0.5, and 1. However, meta-analyses using this rarely find it to be a significant moderator of effect size, probably due to restriction of range in this measure.

Atkins et al. divide their re-examination of my article into four parts: a) Selection of studies, b) Ratings of methodological quality, c) Meta-analysis, d) Judgments of quality of evidence. It is evident from my responses that I have refuted their claims regarding each of these parts and there is no reason to follow their suggestion that “both the content of Öst’s (2014) review and the process used to create it should now be set aside in making decisions regarding the treatment efficacy of ACT.”

My conclusion is that the response from Atkins et al. is a plea made by 10 researchers, having strong allegiance with ACT, who all but one have various conflicts of interest; royalties from books, fees for workshops, training courses and therapy. In comparison, I have no conflict of interest regarding ACT since I have never written anything leading to royalty or done any clinical work or teaching on ACT. I leave it up to the readers to decide who they should believe in this situation.

References

- A-Tjak, J. G. L., Davis, M. L., Morina, N., Powers, M. B., Smits, J. A. J., & Emmelkamp, P. M. G. (2015). A meta-analysis of the efficacy of Acceptance and Commitment Therapy for clinically relevant mental and physical health problems. *Psychotherapy and Psychosomatics*, 84, 30–36. <http://dx.doi.org/10.1159/000365764>.
- Aebi, M., Kuhn, C., Metzke, C. W., Stringaris, A., Goodman, R., & Steinhausen, H. C. (2012). The use of the development and well-being assessment (DAWBA) in clinical practice: A randomized trial. *European Child and Adolescent Psychiatry*, 21, 559–567. <http://dx.doi.org/10.1007/s00787-012-0293-6>.
- 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. <http://dx.doi.org/10.1037/a0028310>.
- Atkins, P. W. B., Ciarrochi, J., Gaudiano, B. A., Bricker, J. B., Donald, J., Rovner, G., et al. (2017). Departing from the essential features of a high quality systematic review of psychotherapy: A response to Öst (2014) and recommendations for improvement. *Behaviour Research and Therapy*.
- Barber, J. P., Liese, B. S., & Abrams, M. J. (2003). Development of the cognitive therapy adherence and competence scale. *Psychotherapy Research*, 13, 205–221. <http://dx.doi.org.ezp.sub.su.se/10.1093/ptr/kpg019>.
- 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. <http://dx.doi.org/10.1037/1076-8998.5.1.156>.
- Borenstein, M., Hedges, L. V., Higgins, J. P. Y., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.
- 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. <http://dx.doi.org/10.1016/j.jamepre.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. <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. <http://dx.doi.org/10.1177/0145445510390930>.
- Brown, F. L., Whittingham, K., Boyd, R. N., McKinlay, L., & Sofronoff, K. (2015). Does Stepping Stones Triple P plus Acceptance and Commitment Therapy improve parent, couple, and family adjustment following paediatric acquired brain injury? A randomised controlled trial. *Behaviour Research and Therapy*, 73, 58–66. <http://dx.doi.org.ezp.sub.su.se/10.1016/j.brat.2015.07.001>.

- Clarke, S., Kinston, J., James, K., Bolderston, H., & Remington, B. (2014). Acceptance and commitment therapy group for treatment-resistant participants: A randomized controlled trial. *Journal of Contextual Behavioral Science*, 3, 179–188. <http://dx.doi.org/ezp.sub.su.se/10.1016/j.jcbs.2014.04.005>.
- Clarke, S., Taylor, G., Bolderston, H., Lancaster, J., & Remington, B. (2015). Ameliorating patient stigma amongst staff working with personality disorder: Randomized controlled trial of self-management versus skills training. *Behavioural and Cognitive Psychotherapy*, 43, 692–704. <http://dx.doi.org/ezp.sub.su.se/10.1017/S1352465814000320>.
- 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. [http://dx.doi.org/10.1016/S0005-7894\(04\)80020-0](http://dx.doi.org/10.1016/S0005-7894(04)80020-0).
- 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>.
- First, M. B., Williams, J. B. W., Karg, R. S., & Spitzer, R. L. (2015). *Structured clinical interview for DSM-5 disorders, clinician version (SCID-5-CV)*. Arlington, VA: American Psychiatric Association.
- Fledderus, M., Bohlmeijer, E. T., Smit, F., & Westerhof, G. J. (2010). Mental health promotion as a new goal in public mental health care: A randomized controlled trial of an intervention enhancing psychological flexibility. *American Journal of Public Health*, 100, 2372–2378. <http://dx.doi.org/ezp.sub.su.se/10.2105/AJPH.2010.196196>.
- 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. <http://dx.doi.org/10.1016/j.cbpra.2012.01.002>.
- 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. <http://dx.doi.org/10.1016/j.brat.2005.02.007>.
- Goodman, R., Ford, T., Richards, H., Gatward, R., & Meltzer, H. (2000). The development and well-being assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. *Journal of Child Psychology and Psychiatry*, 41, 645–655. <http://dx.doi.org/10.1111/j.1469-7610.2000.tb02345.x>.
- 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. <http://dx.doi.org/10.1016/j.beth.2005.03.002>.
- Hamilton, M. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology*, 32, 50–55.
- 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. <http://dx.doi.org/10.1007/s12671-011-0046-5>.
- Ivanova, E., Yaakoba-Zohar, N., Jensen, D., Cassoff, J., & Knauper, B. (2016). Acceptance and Commitment Therapy and implementation intentions increase exercise enjoyment and long-term exercise behavior among low-active women. *Current Psychology*, 35, 108–114. <http://dx.doi.org/ezp.sub.su.se/10.1007/s12144-015-9349-3>.
- Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology*, 3, 1–27. <http://dx.doi.org/ezp.sub.su.se/10.1146/annurev.clinpsy.3.022806.091432>.
- 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. <http://dx.doi.org/10.2307/1411318>. <http://dx.doi.org/10.1016/j.brat.2013.10.007>.
- Kocsis, J. H., Gerber, A. J., Milrod, B., Roose, S. P., Barber, J., Thase, M. E., et al. (2010). A new scale for assessing the quality of randomized clinical trials of psychotherapy. *Comprehensive Psychiatry*, 51, 319–324. <http://dx.doi.org/ezp.sub.su.se/10.1016/j.comppsy.2009.07.001>.
- Kohtala, A., Lappalainen, R., Savonen, L., Timo, E., & Tolvanen, A. (2015). A four-session acceptance and commitment therapy based intervention for depressive symptoms delivered by masters degree level psychology students: A preliminary study. *Behavioural and Cognitive Psychotherapy*, 43, 360–373. <https://doi.org/ezp.sub.su.se/10.1017/S1352465813000969>.
- Levin, M. E., Hayes, S. C., Pistorello, J., & Seeley, J. R. (2016). Web-based self-help for preventing mental health problems in universities: Comparing acceptance and commitment training to mental health education. *Journal of Clinical Psychology*, 72, 207–225. <http://dx.doi.org/ezp.sub.su.se/10.1002/jclp.22254>.
- 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. <http://dx.doi.org/10.1038/sj.ijo.0802460>. <http://dx.doi.org/10.1007/s12160-009-9083-x>.
- Losada, A., Marquez-Gonzalez, M., Romero-Moreno, R., Mausbach, B. T., Lopez, J., Fernández-Fernández, V., et al. (2015). Cognitive-Behavioral Therapy (CBT) versus Acceptance and Commitment Therapy (ACT) for dementia family caregivers with significant depressive symptoms: Results of a randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 83, 760–772. <http://dx.doi.org/ezp.sub.su.se/10.1037/ccp0000028>.
- 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, 102e108. <http://dx.doi.org/10.1016/j.yebeh.2008.02.009>.
- Luoma, J. B., & Vilardaga, J. P. (2013). Improving therapist psychological flexibility while training acceptance and commitment therapy: A pilot study. *Cognitive Behaviour Therapy*, 42, 1–8. <http://dx.doi.org/ezp.sub.su.se/10.1080/16506073.2012.701662>.
- Mo'tamedi, H., Rezaeiaram, 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. <http://dx.doi.org/10.1111/j.1526-4610.2012.02192.x>.
- Moncrieff, J., Churchill, R., Drummond, D. C., & McGuire, H. (2001). Development of a quality assessment instrument for trials of treatments for depression and neurosis. *International Journal of Methods in Psychiatric Research*, 10, 126–133. <http://dx.doi.org/10.1002/mpr.108>.
- Moriana, J. A., Gálvez-Lara, M., & Corpas, J. (2017). Psychological treatments for mental disorders in adults: A review of the evidence of leading international organizations. *Clinical Psychology Review*, 54, 29–43. <http://dx.doi.org/ezp.sub.su.se/10.1016/j.cpr.2017.03.008>.
- 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. <http://dx.doi.org/10.1016/j.cbpra.2012.03.005>.
- 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. <http://dx.doi.org/ezp.sub.su.se/10.1016/j.brat.2007.12.005>.
- Öst, L.-G. (2014). The efficacy of Acceptance and Commitment Therapy: An updated systematic review and meta-analysis. *Behaviour Research and Therapy*, 61, 105–121. <http://dx.doi.org/10.1016/j.brat.2014.07.018>.
- Öst, L.-G. (2016). *Att utvärdera psykoterapi och kritiskt granska psykoterapistudier*. Stockholm: Natur & Kultur.
- Öst, L.-G., Havnen, A., Hansen, B., & Kvale, G. (2015). Cognitive behavioral treatments of obsessive-compulsive disorder: A systematic review and meta-analysis of studies published 1993–2014. *Clinical Psychology Review*, 40, 156–169. <http://dx.doi.org/ezp.sub.su.se/10.1016/j.cpr.2015.06.003>.
- Öst, L.-G., Riise, E., Wergeland, G. J., Hansen, B., & Kvale, G. (2016). Cognitive behavioral and pharmacological treatments of OCD in children: A systematic review and meta-analysis. *Journal of Anxiety Disorders*, 43, 58–69. <http://dx.doi.org/ezp.sub.su.se/10.1016/j.janxdis.2016.08.003>.
- 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.
- Räsänen, P., Lappalainen, P., Muotka, J., & Lappalainen, R. (2016). An online guided ACT intervention for enhancing the psychological wellbeing of university students: A randomized controlled clinical trial. *Behaviour Research and Therapy*, 78, 30–42. <http://dx.doi.org/ezp.sub.su.se/10.1016/j.brat.2016.01.001>.
- Shawyer, F., Farhall, J., Mackinnon, A., Trauer, T., Sims, E., Ratcliff, K., et al. (2012). A randomised controlled trial of acceptance-based cognitive behavioural therapy for command hallucinations in psychotic disorders. *Behaviour Research and Therapy*, 50, 110–121. <http://dx.doi.org/10.1016/j.brat.2011.11.007>.
- 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. <http://dx.doi.org/10.1080/08897071003641578>.
- Stotts, A. L., Green, C., Masuda, A., Grabowski, J., Wilson, K., Northrup, T. F., et al. (2012). A stage I pilot study of acceptance and commitment therapy for methadone detoxification. *Drug and Alcohol Dependence*, 125, 215–222. <http://dx.doi.org/10.1016/j.drugalcdep.2012.02.015>.
- Thorsell, J., Finnes, A., Dahl, J., Lundgren, T., Gybrant, M., Gordh, T., et al. (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. <http://dx.doi.org/10.1097/AJP.0b013e318219a933>.
- Wang, S., Zhou, Y., Yu, S., Ran, L., Liu, X., & Chen, Y. (2015). Acceptance and commitment therapy and cognitive-behavioral therapy as treatments for academic procrastination: A randomized controlled group session. *Research on Social Work Practice*, 27, 48–58. <http://dx.doi.org/10.1177/1049731515577890>.
- Westin, V. Z., Schulin, M., Hesser, H., Karlsson, M., Noe, R. Z., Olofsson, U., et al. (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. <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., et al. (2011). Acceptance and commitment therapy for generalized anxiety disorder in older adults: A preliminary report. *Behavior Therapy*, 42, 127–134. <http://dx.doi.org/10.1016/j.beth.2010.07.002>.
- Wetherell, J. L., Afari, N., Rutledge, T., Sorrell, J. T., Stoddard, J. A., Petkus, A. J., et al. (2011). A randomized, controlled trial of acceptance and commitment therapy and cognitive-behavioral therapy for chronic pain. *Pain*, 152, 2098–2107. <http://dx.doi.org/10.1016/j.pain.2011.05.016>.
- White, R., Gumley, A., McTaggart, J., Rattrie, L., McConville, D., Cleare, S., et al. (2011). A feasibility study of Acceptance and Commitment Therapy for emotional dysfunction following psychosis. *Behaviour Research and Therapy*, 49, 901–907. <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. <http://dx.doi.org/10.1080/16506070802078970>.
- Wicksell, R. K., Kemani, M., Jensen, K., Kosek, E., Kadetoff, D., Sorjonen, K., et al. (2013). Acceptance and commitment therapy for fibromyalgia: A randomized controlled trial. *European Journal of Pain*, 17, 599–611. <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. <http://dx.doi.org/10.1016/j.pain.2008.11.006>.
- Wolgast, M. (2014). What does the Acceptance and Action Questionnaire (AAQ-II) really measure? *Behavior Therapy*, 45, 831–839. <http://dx.doi.org.ezp.sub.su.se/10.1016/j.beth.2014.07.002>.
- Yadavaia, J. E., Hayes, S. C., & Vilardaga, R. (2014). Using acceptance and commitment therapy to increase self-compassion: A randomized controlled trial. *Journal of Contextual Behavioral Science*, 3, 248–257. <http://dx.doi.org.ezp.sub.su.se/10.1016/j.jcbs.2014.09.002>.
- 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.