

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	Third wave Cognitive Therapy versus Mentalization-based Treatment for Major Depressive Disorder. A Randomised Clinical Trial
<b>AUTHORS</b>	Jakobsen, Janus; Gluud, Christian; Kongerslev, Mickey; Larsen, Kirsten; Sørensen, Per; Winkel, Per; Lange, Theis; Søgaard, Ulf; Simonsen, Erik

### VERSION 1 - REVIEW

<b>REVIEWER</b>	Ulrich Schweiger Lübeck University Medical School Department of Psychiatry and Psychotherapy Germany
<b>REVIEW RETURNED</b>	19-Mar-2014

<b>GENERAL COMMENTS</b>	<p>The reported pilot trial compares patients treated with two different psychotherapeutic strategies: mentalization based treatment and a third wave cognitive treatment.</p> <p>I am not aware that a similar comparison has ever been made in a study. This makes the study quite interesting!</p> <p>The study is well designed and all methodological aspects are adequate</p> <p>I have the following concerns:</p> <p>1) Recruiting patients for a randomized trial is always difficult and always "harder than expected". This does not constitute a conclusion of the trial. I suggest to eliminate all laments about the difficulty of recruiting from the abstract and the paper and to just describe the facts.</p> <p>The low number of participants will certainly limit the impact of the paper. For example our German Wissenschaftlicher Beirat Psychotherapy suggests not to count studys with less than 30 participants per arm when evaluating the efficacy of a psychotherapeutic method. A low number of participants makes a psychotherapy study more vulnerable to biases that cannot be eliminated by patient randomization and blinding of the interviewers. This should be shortly discussed</p> <p>2) Since I am not able to understand Danish I cannot assess the quality of the two manuals underlying the study. I suggest to BMJ to ask a Danish speaking reviewer.</p> <p>I have the following suggestion:</p> <p>There is a disproportion between the pilot character of the study and the length of the manuscript. A more compact version of the manuscript would be adequate. In particular I consider table 2 redundant. Confounding with pharmacotherapy cannot be controlled given the low number of participants anyway. Equally the figures on page 39 and 40 are not necessary to understand the data.</p>
-------------------------	---

<b>REVIEWER</b>	DR. BERNET KATO IMPERIAL COLLEGE LONDON UNITED KINGDOM
<b>REVIEW RETURNED</b>	31-Mar-2014

<b>GENERAL COMMENTS</b>	<p>This paper reports the results of a randomized clinical pilot trial to compare the effects of two interventions in patients with major depression: Third wave cognitive therapy and mentalization-based treatment. Trials like this one regarding the benefits of interventions can inform decision making in clinical practice and public health. The results are therefore of considerable interest in clinical practice. The paper is well written and includes all the information required when reporting a randomized trial. The paper does provide valuable information that would be useful when planning bigger randomized control trials involving the two interventions.</p> <p><b>MINOR COMMENTS</b></p> <p><b>Abstract</b> First sentence in the second paragraph: The figures do not match up. The sentence should be changed to read: "We planned to randomize 84 consecutive adult participants diagnosed with major depressive disorder but we only managed to get 44 participants who were then randomised to third wave cognitive therapy (n=22) versus mentalization- based treatment (n=22)</p> <p><b>Methods</b> Under methods section the authors ought to mention that they planned the recruitment to take two years rather bringing this up for the first time under limitations.</p> <p><b>Results</b> The authors mention that two missing values in the group randomised to mentalization-based treatment were imputed. More information on how the imputation was done should be provided.</p> <p><b>Discussion</b> Page 41, there is a sentence that reads: "However, when only 44 out of the planned 84 participants (52%) of the projected sample size are obtained in a trial, it is necessary to evaluate the calculated p-values more conservatively". In general, an interpretation of the results does not depend on the p-value alone, the effects sizes are definitely important too. Therefore I suggest that the sentence is changed to "However, when only 44 out of the planned 84 participants (52%) of the projected sample size are obtained in a trial, it is necessary to interpret the results cautiously".</p>
-------------------------	--

<b>REVIEWER</b>	Andrew Mackinnon Head, Biostats Unit Orygen Youth Health Research Centre University of Melbourne Australia
<b>REVIEW RETURNED</b>	06-May-2014

<b>GENERAL COMMENTS</b>	The results of all trial should be made public regardless of their outcomes or deviations from their proposed plan. The authors have followed the CONSORT guide in reporting the trial and I have checked the manuscript against the published protocol. Apart deviations noted in the manuscript itself (principally low recruitment)
-------------------------	--

	<p>and as noted below, the trial has been implemented and analysed as proposed. This forms the basis of an acceptable publication. There are a number of minor matters to be rectified and a number of areas that would benefit from greater explanation. These are raised below.</p> <p>The introduction opens with what amounts to a cherry picking of statistics regarding the prevalence and incidence of depression and suicide. Within their original contexts, the statistics are entirely defensible but they are not immediately applicable to the specific study and its sample. The opening remarks of the paper would be better confined to intervention itself and the specific setting of the trial.</p> <p>There is no convincing rationale for the design of the trial. This becomes apparent later in the paper (p. 22) when the absence of a pure placebo arm is discussed as a limitation. Both third wave cognitive therapy (CBT3) and mentalization therapy were conceived as being potentially efficacious treatments for depression. In these circumstances one may have expected a trial contrasting CBT3 and mentalization to placebo, waitlist, treatment as usual, older forms of CBT in order to assess the superiority of the newly developed treatments compared to no effective treatment or current practice, according to ethical requirement and the goals of the trial. While mentalization is described in the trial protocol (Jakobsen et al., 2012) as the control condition, the trial is effectively a head-to-head comparison of two putatively efficacious interventions when neither has a reliable evidence base for its performance. As such, one might expect the trial to have been designed within an equivalence/non-inferiority framework. As the authors ultimately point out, this comparison gives rise to substantial uncertainty in interpretation. This matter was entirely foreseeable at the protocol development stage and, in my view, casts doubt on why they proceeded with the trial as proposed.</p> <p>The authors state, quite reasonably, that a pure placebo was not ethically acceptable. However, all patients appear to have received or had access to quite a high level of treatment as usual (TAU). This included psycho-social support and mediation as needed. Under these circumstances, a trial including TAU alone and in combination with either CBT3 or mentalization would seem to be defensible and preferable.</p> <p>I do not believe that relegating this to the limitations sections at the end of paper is adequate. It is important that the authors provide a rationale for the decision to compare the two interventions and to do so within a conventional 'superiority' framework. This is reinforced when the authors later refer to the 'feasibility of the design' as a strength of the study. If it is, indeed, a strength, more explanation is required.</p> <p>It is not really clear how this trial is a pilot study as properly described (see Thabane et al., 2010; Leon et al., 2010). I would characterise it as a small scale, single site trial powered for a (slightly?) optimistic effect size. A pilot study should provide insight into trial processes and feasibility. Beyond simply re-establishing Lasanga's Law (van der Wouden et al. (2007), it offers little insight into the reasons of the recruitment problem – was it the setting or the nature of the interventions – and what might be done address this.</p>
--	---

	<p>There is a slight mismatch between the protocol's precise statement of a primary aim and endpoint (HDRS (after 18 weeks of treatment)) and the more general objective stated in the manuscript. Presumably testing the primary outcome was the primary objective of the trial. This should be added to the paper.</p> <p>The authors have formally published the protocol of the trial (Jakobsen et al., 2012). This should be cited in preference to be link currently included in the paper.</p> <p>It is defensible, but unusual to compare outcomes at endpoint without reference to baseline status in either repeated measures ANOVA or ANCOVA. Generally this is more powerful. Personally, I would prefer the reporting of just one approach to analysis, however I note that the authors are closely following their published protocol.</p> <p>In absolute terms, lost to follow-up is minimal (2 out of 44). The authors proceed with the use of imputation, presumably as outlined in their protocol. The imputation procedure described is not multiple imputation and would be better described as a sensitivity analysis. I would suggest not using multiple imputation and carefully following the advice of White and colleagues (2011) regarding ITT analysis.</p> <p>There is a slight inconsistency regarding the reporting of one of the patients with missing data. In the CONSORT diagram, they are described as having dropped out because of low compliance. Was this data not available or withheld due to not compliance? If the latter, this inconsistent with ITT and should not have occurred. Analysis of all available data regardless of compliance is required.</p> <p>I felt that the concept of treating results from an under recruiting trial as an interim analysis and using sequential methods was an innovative idea. However, as only one post-baseline occasion of measurement was available, the impact of this exercise was rather limited.</p> <p>As mentioned above, there is little information or reflection about process and feasibility reported despite the description of the trial as a pilot. Recruitment was harder than expected, but why?</p> <p>Also as indicated above, I do not believe the strengths of the trial regarding feasibility are justifiable. The claim to external validity is also uncertain. The sample seem to me to be unrepresentative of typical patients with depression, in being cases of invalidity from employment and much more predominantly female than the epidemiology of treated depression would suggest. The treatment environment also seems much better than that available in many societies. This can hardly be criticized but may limit any form of generalizability.</p> <p>While I agree with the component of the conclusion that recommends further trials of CBT3, comparison with mentalization does not appear warranted. This conclusion should either be justified or refined.</p> <p>References  Jakobsen, J. C., Gluud, C., Kongerslev, M., Larsen, K. A., Sørensen, P., Winkel, P., . . . Simonsen, E. (2012). Third wave cognitive therapy versus mentalization-based therapy for major depressive</p>
--	--

	<p>disorder. A protocol for a randomised clinical trial. <i>BMC Psychiatry</i>, 12, 232. doi: 10.1186/1471-244X-12-232</p> <p>Leon, A. C., Davis, L. L., &amp; Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. <i>Journal of Psychiatric Research</i>, 45(5), 626–629.</p> <p>Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., . . . Goldsmith, C. H. (2010). A tutorial on pilot studies: the what, why and how. <i>BMC Medical Research Methodology</i>, 10(1).</p> <p>van der Wouden, J. C., Annette H. Blankenstein, Huibers, M. J. H., van der Windt, D. A. W. M., Stalman, W. A. B., &amp; Verhagen, A. P. (2007). Survey among 78 studies showed that Lasagna's law holds in Dutch primary care research. <i>Journal of Clinical Epidemiology</i>, 60(8), 819-824.</p> <p>White, I. R., Horton, N. J., Carpenter, J., &amp; Pocock, S. J. (2011). Strategy for intention to treat analysis in randomised trials with missing outcome data. <i>BMJ</i>, 342, d40. doi: 10.1136/bmj.d40</p>
--	--

### VERSION 1 – AUTHOR RESPONSE

Reviewer 1

Reviewer Name Ulrich Schweiger

Institution and Country Lübeck University Medical School

Department of Psychiatry and Psychotherapy

Germany

Please state any competing interests or state 'None declared': none declared

Reviewer 1: The reported pilot trial compares patients treated with two different psychotherapeutic strategies: mentalization based treatment and a third wave cognitive treatment.

I am not aware that a similar comparison has ever been made in a study. This makes the study quite interesting!

The study is well designed and all methodological aspects are adequate

Our response: We thank the reviewer for these very positive comments.

Reviewer 1: I have the following concerns:

1) Recruiting patients for a randomized trial is always difficult and always "harder than expected". This does not constitute a conclusion of the trial. I suggest to eliminate all laments about the difficulty of recruiting from the abstract and the paper and to just describe the facts.

Our response: We agree with the reviewer. We have kept a description of the facts about the difficulties in recruiting participants, but have deleted the remaining information in the abstract, the conclusion, and in the manuscript:

See 'Abstract' in the revised manuscript page 2-3:

See page 4 in the revised manuscript:

See page 15 in the revised manuscript:

See page 18 in the revised manuscript:

See page 26 in the revised manuscript:

Reviewer 1: The low number of participants will certainly limit the impact of the paper. For example our German Wissenschaftlicher Beirat Psychotherapy suggests not to count studies with less than 30 participants per arm when evaluating the efficacy of a psychotherapeutic method. A low number of participants makes a psychotherapy study more vulnerable to biases that cannot be eliminated by patient randomization and blinding of the interviewers. This should be shortly discussed

Our response: We agree with the reviewer. We have now clarified this in our revised manuscript:

Page 20 in the revised manuscript:

“Our trial has a number of limitations. This small-scale trial was in essence failed because we only included 44 out of the planned 84 participants. The trial inclusion lasted for about two years as planned but we had problems with recruiting participants. Basically, not enough eligible depressed patients were referred to the clinic within the planned trial period. The great advantage of the randomised clinical trial in general is that all known and unknown participant characteristics will be similar at baseline in compared intervention groups.<sup>45</sup> However, even though our baseline characteristics indicate similarity between the two groups on assessed baseline characteristics, it is unlikely that all baseline characteristics will be similar when only 44 participants are randomised. The low number of randomised participants in this small-scale trial increases the risks of wrong results due to type I errors, and type II errors,<sup>54, 55</sup> and our adequate trial methodology cannot necessarily compensate for these increased risks.”

Reviewer 1: 2) Since I am not able to understand Danish I cannot assess the quality of the two manuals underlying the study. I suggest to BMJ to ask a Danish speaking reviewer. I have the following suggestion:

There is a disproportion between the pilot character of the study and the length of the manuscript. A more compact version of the manuscript would be adequate. In particular I consider table 2 redundant. Confounding with pharmacotherapy cannot be controlled given the low number of participants anyway. Equally the figures on page 39 and 40 are not necessary to understand the data.

Our response: We agree with the reviewer that the Table 2 might be moved from the primary manuscript. We have therefore changed Table 2 to be Supplementary Material 1 so the information is still accessible online but the table is deleted from the primary manuscript. However, we do believe that Figure 2 and 3 should remain in the manuscript. These figures are very important because they show why the few significant results should be interpreted with great caution. We have now clarified this in the revised manuscript and included a reference to a publication, which describes why the thresholds for significance should be adjusted if the sample size has not been reached:

Page 18 in the revised manuscript:

“First of all, the trial was conducted with an overall high level of methodological quality and we assessed the validity of the trial results according to the procedure proposed by Jakobsen et al.<sup>40</sup> including adjusting the thresholds for significance according to the number of randomised participants and the planned sample size.<sup>40</sup>”

We have made the manuscript shorter according to the reviewer's comments and have therefore

deleted redundant information about the difficulties in recruiting participants to the trial:

See 'Abstract' in the revised manuscript page 2-3.

See page 5 in the revised manuscript (deleted the first paragraph of the introduction).

See page 15 (deleted the paragraph about the psychopharmacological treatment).

See page 18 in the revised manuscript.

See page 26 in the revised manuscript.

Reviewer 2

Reviewer Name DR. BERNET KATO

Institution and Country

IMPERIAL COLLEGE LONDON

UNITED KINGDOM

Please state any competing interests or state 'None declared': NONE DECLARED

Third wave Cognitive Therapy versus Mentalization-based Treatment for Major Depressive Disorder. A Randomized Clinical Pilot Trial

GENERAL COMMENTS

Reviewer 2: This paper reports the results of a randomized clinical pilot trial to compare the effects of two interventions in patients with major depression: Third wave cognitive therapy and mentalization-based treatment. Trials like this one regarding the benefits of interventions can inform decision making in clinical practice and public health. The results are therefore of considerable interest in clinical practice.

Our response: We thank the reviewer for these very positive comments.

Reviewer 2: The paper is well written and includes all the information required when reporting a randomized trial. The paper does provide valuable information that would be useful when planning bigger randomized control trials involving the two interventions.

Our response: We are very thankful for these positive comments from the reviewer.

MINOR COMMENTS

Reviewer 2:

Abstract

First sentence in the second paragraph: The figures do not match up. The sentence should be changed to read: "We planned to randomize 84 consecutive adult participants diagnosed with major depressive disorder but we only managed to get 44 participants who were then randomised to third wave cognitive therapy (n=22) versus mentalization- based treatment (n=22)

Our response: We have carefully revised the abstract so it is now in accordance with the CONSORT checklist for abstracts (<http://www.consort-statement.org/checklists/view/32-consort/67-abstract>). We

have reported the planned sample size (84 participants) in the description of the design of the trial (included in the paragraph 'Design, participants, and setting') and the number of randomised participants is reported under 'Results'.

Revised abstract page 2-3:

“Objective: To compare the benefits and harms of third wave cognitive therapy versus mentalization-based therapy in a small sample of depressed participants.

Design, participants, and setting: We planned to randomise 84 consecutive adult participants diagnosed with major depressive disorder to third wave cognitive therapy (n=22) versus mentalization-based treatment (n=22) in a superiority randomised clinical trial. The outcome assessors and the statistician were blinded to treatment allocation. The trial was conducted at an outpatient psychiatric clinic for non-psychotic patients in Roskilde, Denmark.

Outcomes: The primary outcome was the Hamilton Rating Scale for Depression (HDRS) at end of treatment (18 weeks). Secondary outcomes were: remission (HDRS < 8), Beck's Depression Inventory, Symptom Checklist 90 Revised, and The World Health Organisation-Five Well-being Index 1999.

Results: The trial inclusion lasted for about two years as planned but only 44 out of the planned 84 participants were randomised. Two mentalization-based participants were lost to follow-up. The unadjusted analysis showed that third wave participants compared with mentalization-based participants did not differ significantly regarding the 18 weeks HDRS score (12.9 versus 17.0; mean difference -4.14; 95% CI -8.30 to 0.03; P = 0.051). In the analysis adjusted for baseline HDRS score, the difference was favouring third wave cognitive therapy (P = 0.039). At 18 weeks, five of the third wave participants (22.7%) were in remission versus none of the mentalization-based participants (P = 0.049). We recorded no suicide attempts or suicides during the intervention period in any of the 44 participants. No significant differences were found between the two intervention groups on the remaining secondary outcomes.

Conclusions: Third wave cognitive therapy may be more effective than mentalization-based therapy for depressive symptoms measured on the HDRS. However, more randomised clinical trials are needed to assess the effects of third wave cognitive therapy and mentalization-based treatment for depression.”

Reviewer 2:

Methods

Under methods section the authors ought to mention that they planned the recruitment to take two years rather bringing this up for the first time under limitations.

Our response: We have described that we estimated that we would need an inclusion period of two years under 'Methods':

Page 14:

“With a 'minimal relevant mean difference' (MIREDF) between the two interventions of 5 HDRS points, an alpha of 0.05 (type I error), a power of 0.90 (type II error of 10%), and a standard deviation (SD) of 7 points, the sample size calculation showed that a total of 84 participants would be necessary. We estimated that we would need an inclusion period of about two years to recruit 84 participants.”

Reviewer 2:

#### Results

The authors mention that two missing values in the group randomised to mentalization-based treatment were imputed. More information on how the imputation was done should be provided.

Our response: We thank the reviewer for this important comment. We had planned (in the protocol) only to impute missing values (using multiple imputation) if more than 5% were missing. Only 2 out of the 44 participants (4.5%) were missing so we have deleted the results including the imputed values and clarified why we did not impute missing values.

Page 16 in the revised manuscript:

“We did not impute missing values because only 2 out of 44 (4.5%) participants had missing values.”

Reviewer 2:

#### Discussion

Page 41, there is a sentence that reads: “However, when only 44 out of the planned 84 participants (52%) of the projected sample size are obtained in a trial, it is necessary to evaluate the calculated p values more conservatively”. In general, an interpretation of the results does not depend on the p value alone, the effects sizes are definitely important too. Therefore I suggest that the sentence is changed to “However, when only 44 out of the planned 84 participants (52%) of the projected sample size are obtained in a trial, it is necessary to interpret the results cautiously”.

Our response: We thank the reviewer for this important comment and we agree totally. We have now revised our manuscript according to the reviewer’s comments:

Page 18 in the revised manuscript:

“However, when only 44 out of the planned 84 participants (52%) of the projected sample size is obtained in a trial, it is necessary to interpret the results cautiously.”

Reviewer 3

Reviewer Name Andrew Mackinnon

Institution and Country Head, Biostats Unit

Orygen Youth Health Research Centre

University of Melbourne

Australia

Please state any competing interests or state ‘None declared’: None declared

Reviewer 3: The results of all trial should be made public regardless of their outcomes or deviations from their proposed plan. The authors have followed the CONSORT guide in reporting the trial and I have checked the manuscript against the published protocol. Apart deviations noted in the manuscript itself (principally low recruitment) and as noted below, the trial has been implemented and analysed as proposed. This forms the basis of an acceptable publication.

Our response: We thank the reviewer for this very positive comment.

Reviewer 3: There are a number of minor matters to be rectified and a number of areas that would benefit from greater explanation. These are raised below.

The introduction opens with what amounts to a cherry picking of statistics regarding the prevalence and incidence of depression and suicide. Within their original contexts, the statistics are entirely defensible but they are not immediately applicable to the specific study and its sample. The opening remarks of the paper would be better confined to intervention itself and the specific setting of the trial.

Our response: We thank the reviewer for this important comment. We have now deleted the first paragraph.

See page 5 in the revised manuscript.

Reviewer 3: There is no convincing rationale for the design of the trial. This becomes apparent later in the paper (p. 22) when the absence of a pure placebo arm is discussed as a limitation. Both third wave cognitive therapy (CBT3) and mentalization therapy were conceived as being potentially efficacious treatments for depression. In these circumstances one may have expected a trial contrasting CBT3 and mentalization to placebo, waitlist, treatment as usual, older forms of CBT in order to assess the superiority of the newly developed treatments compared to no effective treatment or current practice, according to ethical requirement and the goals of the trial. While mentalization is described in the trial protocol (Jakobsen et al., 2012) as the control condition, the trial is effectively a head-to-head comparison of two putatively efficacious interventions when neither has a reliable evidence base for its performance.

Our response: We agree totally with the reviewer. We have now clarified that this is a head-to-head trial.

Page 18 in the revised manuscript:

“Compared with the baseline scores, both intervention groups improved during the trial period on all continuous outcomes. However, we did not include a control group receiving no intervention in this head-to-head trial so it is unclear whether it was trial intervention effects or ‘regression towards the mean’ effects that caused these changes.<sup>45</sup>”

We have also clarified how this limits the interpretation of the trial results.

Page 22 in the revised manuscript:

“All these considerations and practical circumstances led to the choice of the psychotherapeutic interventions and the design of this head-to-head trial comparing third wave cognitive therapy and co-interventions versus mentalization-based therapy and co-interventions. The co-interventions were delivered similarly to both treatment groups and the possible effects of co-interventions will therefore even out between the compared intervention groups unless there are significant interactions. Nevertheless, it is a clear limitation that our interventions are not and have not been compared versus no intervention or a more simple and basic form of psychotherapy plus co-interventions.<sup>45</sup> If a trial comparing the effects of two active interventions shows no difference in effect it is not clear whether the two interventions are equally effective or equally ineffective – and if an experimental intervention seem superior compared with a control intervention then the effect size of the experimental intervention will be unclear because any beneficial or harmful effects of the control intervention might influence the trial results.<sup>45</sup> All interventions should be assessed versus no intervention before being introduced into clinical practice.<sup>45</sup>”

Reviewer 3: As such, one might expect the trial to have been designed within an equivalence/non-inferiority framework. As the authors ultimately point out, this comparison gives rise to substantial

uncertainty in interpretation. This matter was entirely foreseeable at the protocol development stage and, in my view, casts doubt on why they proceeded with the trial as proposed. The authors state, quite reasonably, that a pure placebo was not ethically acceptable. However, all patients appear to have received or had access to quite a high level of treatment as usual (TAU). This included psychosocial support and mediation as needed. Under these circumstances, a trial including TAU alone and in combination with either CBT3 or mentalization would seem to be defensible and preferable.

I do not believe that relegating this to the limitations sections at the end of paper is adequate. It is important that the authors provide a rationale for the decision to compare the two interventions and to do so within a conventional 'superiority' framework. This is reinforced when the authors later refer to the 'feasibility of the design' as a strength of the study. If it is, indeed, a strength, more explanation is required.

Our response: We agree with the reviewer that our choice of comparisons in our trial gives rise to substantial uncertainty. This is very important and we thank the reviewer for this important comment. For ethical reasons we considered it necessary to offer some kind of specialised treatment for all as the patients even though the evidence behind the specialised treatment is lacking. This consideration may be questioned because of the lack of evidence but this was the reason behind the choice of trial interventions. We have now clarified this in our revised manuscript and described the rationale behind the trial:

See Page 22 in the revised manuscript

“All these considerations and practical circumstances led to the choice of the psychotherapeutic interventions and the design of this head-to-head trial comparing third wave cognitive therapy and co-interventions versus mentalization-based therapy and co-interventions. The co-interventions were delivered similarly to both treatment groups and the possible effects of co-interventions will therefore even out between the compared intervention groups unless there are significant interactions. Nevertheless, it is a clear limitation that our interventions are not and have not been compared versus no intervention or a more simple and basic form of psychotherapy plus co-interventions.<sup>45</sup> If a trial comparing the effects of two active interventions shows no difference in effect it is not clear whether the two interventions are equally effective or equally ineffective – and if an experimental intervention seems superior compared with a control intervention then the effect size of the experimental intervention will be unclear because any beneficial or harmful effects of the control intervention might influence the trial results.<sup>45</sup> All interventions should be assessed versus no intervention before being introduced into clinical practice.<sup>45</sup>”

We do not agree that the design of the trial should be equivalence/non-inferiority. Changing the design in such a way would not compensate for the interpretative limitations due to the lack of trials assessing the compared interventions versus no intervention or placebo. Furthermore, we do not believe that equivalence/non-inferiority designs serve the patients' best interests so such design should generally be avoided<sup>1</sup>.

Reviewer 3: It is not really clear how this trial is a pilot study as properly described (see Thabane et al., 2010; Leon et al., 2010). I would characterise it as a small scale, single site trial powered for a (slightly?) optimistic effect size. A pilot study should provide insight in trial processes and feasibility. Beyond simply re-establishing Lasanga's Law (van der Wouden et al. (2007), it offers little insight into the reasons of the recruitment problem – was it the setting or the nature of the interventions – and what might be done address this.

Our response: We believe that our trial does provide some insight into the reasons of the recruitment

problem. Before the randomisation began, we did not systematically assess how many participants it was possible to recruit and this should have been done. On average, we recruited approximately one participant every third week and we expected to be able to recruit approximately one participant every week. So basically not enough eligible participants were referred to the clinic during the inclusion period and we had to terminate the trial due to economical and practical constraints – this was the primary reason why we did not randomise more participants and this is now clarified in the manuscript (see revised manuscript page 24, 25 or below). However, after careful consideration we have decided to follow the reviewer's suggestion that the trial is better characterised as a small-scale randomised clinical trial and we are thankful for this important comment. We have now revised the title and the manuscript:

Title page 1 in the revised manuscript:

'Third wave Cognitive Therapy versus Mentalization-based Treatment for Major Depressive Disorder. A Randomised Clinical Trial'

See page 4 in the revised manuscript.

See page 20 in the revised manuscript.

Page 24-25 in the revised manuscript:

Furthermore, we agree that the anticipated intervention effect used in the sample size calculation was optimistic. To further demonstrate this we have now included a calculation of Bayes factor<sup>2</sup> which confirms that lower anticipated intervention effect should be used to calculate sample sizes in trials assessing the effects of third wave cognitive therapy and mentalization-based therapy.

Page 25 in the revised manuscript:

"We used an anticipated intervention effect of five HDRS points to estimate the necessary sample size and this anticipated intervention effect was optimistic. Calculating Bayes factor based on the anticipated intervention effect, the observed intervention effect, and the standard error of the observed intervention effect shows a Bayes factor of 0.14, which is above the recommended threshold for significance of 0.1.<sup>40</sup> This underlines that our results should be regarded as insignificant and that an anticipated intervention effect lower than five HDRS points ought to be used in sample size calculations in future trials assessing the effects of third wave cognitive therapy and mentalization-based therapy."

Reviewer 3: There is a slight mismatch between the protocol's precise statement of a primary aim and endpoint (HDRS (after 18 weeks of treatment)) and the more general objective stated in the manuscript. Presumably testing the primary outcome was the primary objective of the trial. This should be added to the paper.

Our response: We agree with the reviewer. We have now clearly specified our primary objective

Page 7 in the revised manuscript:

"Objective

Our objective was to compare the effect of third wave cognitive therapy versus mentalization-based therapy in a small sample of participants with major depressive disorder. "

Reviewer 3: The authors have formally published the protocol of the trial (Jakobsen et al., 2012). This

should be cited in preference to be link currently included in the paper.

Our response: We agree with the reviewer. We have now specified this in our revised manuscript.

Page 7 in the revised manuscript:

“Methods

In the following, we briefly describe the methodology of this trial. For details please consult our registered (clinicaltrials.gov: NCT01070134) and published protocol.<sup>27</sup> “

Reviewer 3: It is defensible, but unusual to compare outcomes at endpoint without reference to baseline status in either repeated measures ANOVA or ANCOVA. Generally this is more powerful. Personally, I would prefer the reporting of just one approach to analysis, however I note that the authors are closely following their published protocol.

Our response: We thank the reviewer for this comment and we agree that the results from adjusted analyses in many circumstances show the most precise and unbiased results – especially if stratification variables are used in the randomisation.<sup>3, 4</sup> However, we chose in our protocol to present both adjusted and unadjusted analyses and to discuss discrepancies.<sup>3, 4</sup> We are happy that the reviewer appreciates that we have followed our planned methodology.

Reviewer 3: In absolute terms, lost to follow-up is minimal (2 out of 44). The authors proceed with the use of imputation, presumably as outlined in their protocol. The imputation procedure described is not multiple imputation and would be better described as a sensitivity analysis. I would suggest not using multiple imputation and carefully following the advice of White and colleagues (2011) regarding ITT analysis.

Our response: We thank the reviewer for this important comment. We had planned (in the protocol) only to impute missing values (using multiple imputation) if more than 5% were missing. Only 2 out of the 44 participants (4.5%) were missing so we have deleted the results including the imputed values and clarified why we did not impute missing values.

Page 16 in the revised manuscript:

“We did not impute missing values because only 2 out of 44 (4.5%) participants had missing values.”

Reviewer 3: There is a slight inconsistency regarding the reporting of one of the patients with missing data. In the CONSORT diagram, they are described as having dropped out because of low compliance. Was this data not available or withheld due to not compliance? If the latter, this inconsistent with ITT and should not have occurred. Analysis of all available data regardless of compliance is required.

Our response: We thank the reviewer for this important comment. The excluded participant was not assessed at any time point after randomisation. We have now clarified this in the manuscript and we have revised the CONSORT flowchart.

Page 15 in the revised manuscript:

“Treatment compliance

None of the 22 participants randomised to third wave cognitive therapy were lost to follow-up or excluded due to the fact that they participated in less than 70% of the sessions. One participant out of the 22 randomised to mentalization-based treatment was lost to follow-up and one was excluded, as she did not attend the required 70% of the sessions (Figure 1). The excluded participant was not

assessed on any of the outcomes at end of treatment. “

See the revised ‘CONSORT flowchart’.

Reviewer 3: I felt that the concept of treating results from an under recruiting trial as an interim analysis and using sequential methods was an innovative idea. However, as only one post-baseline occasion of measurement was available, the impact of this exercise was rather limited.

Our response: We thank the reviewer for this positive response. There are a number of reasons why using sequential analysis should be the preferred approach if a trial is stopped before the sample size is reached. We have now included a reference to a former publication, which describes in detail why this approach is valid.<sup>2</sup>

From this publication: “A trial that is stopped prematurely with an effect that is significant (e.g.,  $P < 5\%$ ) may reach this significance level because the estimated difference in effect between the compared trial interventions is larger than anticipated or because the estimated variance is lower than anticipated – or both (see Section 2 about sample size estimation).<sup>5-7</sup> Deviations of intervention effects far from the anticipated values should a priori be regarded as unlikely and this is one reason for using a lower statistical threshold to stop a trial before the planned sample size has been reached.<sup>6</sup> If, e.g., a sample size calculation has shown that a total of 500 patients are needed in a trial and the trial is stopped after only 250 participants are included, it might be necessary to use 1% instead of 5% as statistical threshold for significance in order to avoid undue declarations of statistical significance due to early random high intervention effects or low variance.<sup>8</sup> As mentioned, trials with too small sample sizes often show intervention effect sizes far from the effect sizes shown in larger trials and systematic reviews with meta-analyses.<sup>7, 9</sup> As pointed out by Lindley, the apparent paradox of small trials seemingly contributing with evidence of large intervention effects while large trials tend to rule out smaller intervention effects and thereby also larger intervention effects, is bound to confuse the average clinical researcher and reader.<sup>10</sup> If trialists are allowed to assess statistical significance continuously during a trial (i.e., to conduct interim analyses) and stop at different time points without adjusting the level of statistical significance, this will inevitably increase the risk of falsely negating the null hypothesis.<sup>11</sup> This is due to sparse data and due to repetitive testing on accumulating data both leading to increased risks of random errors. Therefore, the threshold of statistical significance should be related to the fraction of the pre-planned number of participants randomised and the number of tests conducted (see also Section 4) —<sup>12-14</sup> and a number of different methods have been developed for this purpose.<sup>15-18</sup> One example is the O’Brien-Fleming boundaries (and the corresponding adjusted thresholds of the confidence intervals and the P-values),<sup>15, 16</sup> which show the adjusted thresholds for significance if a sample size has not been reached.<sup>15, 19</sup>

Any outcome should only be assessed using the thresholds used in the sample size calculation if there are sufficient data, i.e., that a sample size based on proper acceptable risks of type I and type II errors has been reached. It is, therefore, necessary to perform power calculations for all secondary outcomes (based on an anticipated intervention effect, a variance, and a risk of type I error) before randomisation begins.”

Reviewer 3: As mentioned above, there is little information or reflection about process and feasibility reported despite the description of the trial as a pilot. Recruitment was harder than expected, but why?

Also as indicated above, I do not believe the strengths of the trial regarding feasibility are justifiable.

Our response: See our fifth response to Reviewer 3.

Reviewer 3: The claim to external validity is also uncertain. The sample seem to me to be

unrepresentative of typical patients with depression, in being cases of invalidity from employment and much more predominantly female than the epidemiology of treated depression would suggest.

Our response: We agree with the reviewer. We have now deleted this 'strength' in the Discussion.

See page 19 in the revised manuscript.

Reviewer 3: The treatment environment also seems much better than that available in many societies. This can hardly be criticized but may limit any form of generalizability.

Our response: We agree with the reviewer. We have now described this clear limitation in our revised manuscript:

Page 22 in the revised manuscript:

"Furthermore, the combination of specialised psychotherapy and co-interventions constitute a relatively comprehensive treatment, which might not always be accessible to psychiatric patients in clinical practice – this might limit the generalizability of our results."

Reviewer 3: While I agree with the component of the conclusion that recommends further trials of CBT3, comparison with mentalization does not appear warranted. This conclusion should either be justified or refined.

Our response: We thank the reviewer for this important comment. We agree that future trials randomising other types of participants might assess the effects of third wave cognitive therapy and mentalization-based therapy versus treatment as usual or other control interventions. We have now revised our recommendations:

Page 3 in the revised manuscript:

"Conclusions: Our results suggest that third wave cognitive therapy may be more effective than mentalization-based therapy for depressive symptoms measured on the HDRS. However, more randomised clinical trials are needed to assess third wave cognitive therapy and mentalization-based treatment for depression. Such trials should be multicentre trials to secure adequate enrolment."

Page 18 in the revised manuscript:

"More randomised clinical trials are needed to assess the effects of third wave cognitive therapy and mentalization-based treatment for major depressive disorder."

Page 26 in the revised manuscript:

"Conclusions

Our preliminary results show that third wave cognitive therapy compared with mentalization-based treatment may be a more effective intervention for depressive symptoms measured on the HDRS. The effects of the two interventions did not seem to differ significantly regarding BDI II, SCL 90-R, and WHO 5. More randomised clinical trials are needed to assess the effects of third wave cognitive therapy and mentalization-based treatment."

We hope that we have modified the manuscript to your satisfaction. If you continue to see issues we have overlooked or we still need to engage, please let us know.

Thanks again for your close attention to this manuscript.

On behalf of the authors

Janus Christian Jakobsen

## References

1. Garattini S, Bertele V. Non-inferiority trials are unethical because they disregard patients' interests. *Lancet*. 2007;370(9602):1875-1877.
2. Jakobsen JC, Gluud C, Winkel P, Lange T, Wetterslev J. The thresholds for statistical and clinical significance - a five-step procedure for evaluation of intervention effects in randomised clinical trials. *BMC Med Res Methodol*. 2014;14(34).
3. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. *BMJ*. 2012;345:e5840.
4. Kahan BC, Morris TP. Improper analysis of trials randomised using stratified blocks or minimisation. *Stat Med*. 2011;31:328-340.
5. Jennison C, Turnbull BW. Efficient group sequential designs when there are several effect sizes under consideration. *Stat Med*. 2005;25:917-932.
6. Levin GP, Emerson SC, Emerson SS. Adaptive clinical trial designs with pre-specified rules for modifying the sample size: understanding efficient types of adaptation. *Stat Med*. 2012;32(8):1259-1275.
7. Pereira TV, Horwitz RI, Ioannidis JP. Empirical evaluation of very large treatment effects of medical interventions. *JAMA*. 2012;308:1676-1684.
8. DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. *Stat Med*. 1994;13(13-14):1341-1356.
9. Bassler D, Montori VM, Briel M, Glasziou P, Walter SD, Ramsay T, Guyatt G. Reflections on meta-analyses involving trials stopped early for benefit: is there a problem and if so, what is it? *Stat Methods Med Res*. 2013;22(2):159-168.
10. Lindley DV. A statistical paradox. *Biometrika*. 1957;44(1/2):187-192.
11. Guyatt GH, Briel M, Glasziou P, Bassler D, Montori VM. Problems of stopping trials early. *BMJ*. 2012;344:e3863.
12. Wald A. Sequential tests of statistical hypotheses. *Ann Math Stat*. 1945;16:117-186.
13. Zhang J, Quan H, Ng J, Stepanavage ME. Some statistical methods for multiple endpoints in clinical trials. *Control Clin Trials*. 1997;18:204-221.
14. Imberger G, Vejlbj AD, Hansen SB, Møller AM, Wetterslev J. Statistical multiplicity in systematic reviews of anaesthesia interventions: a quantification and comparison between Cochrane and non-Cochrane reviews. *PLoS One*. 2011;6:e28422.
15. Pocock SJ. When to stop a clinical trial. *BMJ*. 1992;305(6847):235-240.
16. Jennison C, Turnbull BW. Repeated confidence intervals for group sequential clinical trials. *Control Clin Trials*. 1984;5(1):33-45.
17. Todd S, Whitehead J, Facey KM. Point and interval estimation following a sequential clinical trial. *Biometrika*. 1996;83(2):453-461.
18. Jennison C, Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials* (Chapman & Hall/CRC Interdisciplinary Statistics): Chapman and Hall/CRC; 1999.
19. Thorlund K, Engstrøm J, Wetterslev J, Brok J, Imberger G, Gluud C. User manual for trial

sequential analysis (TSA). Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen, Denmark. 2011. p. 1-115. Available from <http://www.ctu.dk/tsa>.

### VERSION 2 – REVIEW

<b>REVIEWER</b>	Ulrich Schweiger Universität zu Lübeck Germany
<b>REVIEW RETURNED</b>	26-Jun-2014

<b>GENERAL COMMENTS</b>	<p>page 5, third sentence is not understandable. Overall the new introduction is better compared to the older one but still somewhat off target with respect to the study. In particular concerns about the effect size of classical cognitive therapy is not a topic advanced by the study. "Furthermore, our trial demonstrated the feasibility of conducting the trial with low risks of bias" This sentence is not really founded because important sources of bias are not eliminated. a) the trial stayed below an appropriate number b) the underlying manuals are not accessible to the international psychotherapeutic community. This creates problems with possible differences of the quality of the underlying manual, therapist training and allegiance. These limitations are still not sufficiently discussed. The paper is considerably improved but still needs major revision.</p>
-------------------------	---

<b>REVIEWER</b>	Bernet Kato Imperial College London United Kingdom
<b>REVIEW RETURNED</b>	16-Jun-2014

<b>GENERAL COMMENTS</b>	The authors have addressed all my concerns. I appreciate the author's close attention to addressing reviewer concerns
-------------------------	---

<b>REVIEWER</b>	Andrew Mackinnon Head, Biostatistics Unit Orygen Youth Health Research Centre University of Melbourne Australia
<b>REVIEW RETURNED</b>	07-Jul-2014

<b>GENERAL COMMENTS</b>	<p>The authors have very diligently addressed the issues I raised, as well as those of the other reviewers. I make only one suggestion. Rather than referring to 'regression to the mean' as a possible explanation for the changes observed (p. 18), I believe it would be more accurate and comprehensive to refer to the nature history of the disorder. For example: However, we did not include a control group receiving no intervention in this head-to-head trial so it is unclear whether it was trial intervention effects or the nature progress of the disorder in this sample is responsible for these changes. (Regression to the mean embraces only measurement effects.)</p> <p>I agree with the authors that it would not be ethically acceptable to undertake their trial with an arm in which patients were offered no or</p>
-------------------------	--

	<p>placebo-only treatment. It would also not answer any clinically relevant question. If they proceed to further trials, I suggest they consider current treatment or a standardized form of conventional CBT as the comparator. In this context, undertaking a comparison of CBT3 and mentalization within an equivalence/non-inferiority framework may be appropriate.</p>
--	--

## VERSION 2 – AUTHOR RESPONSE

Reviewer: 2

The authors have addressed all my concerns. I appreciate the author's close attention to addressing reviewer concerns

Our response: We thank the reviewer for this very positive comment.

Reviewer: 1

Reviewer Name Ulrich Schweiger

Reviewer 1: page 5, third sentence is not understandable.

Our response: We agree with the reviewer. We have now revised the sentence.

Revised manuscript, page 5: "Prior to this trial we carried out a systematic review of randomised clinical trials examining the effects of traditional cognitive therapy versus no intervention for major depressive disorder.<sup>1</sup> We found that cognitive therapy compared with no intervention seems to have a small statistically significant beneficial effect on depressive symptoms."

Reviewer 1: Overall the new introduction is better compared to the older one but still somewhat off target with respect to the study. In particular concerns about the effect size of classical cognitive therapy is not a topic advanced by the study.

Our response: We agree with the reviewer. We have now deleted the information about the effect size of classical cognitive therapy (see 'Introduction' page 5).

Reviewer 1: "Furthermore, our trial demonstrated the feasibility of conducting the trial with low risks of bias" This sentence is not really founded because important sources of bias are not eliminated. a) the trial stayed below an appropriate number

Our response: We think the peer reviewer mixes bias (systematic error) and play of chance (random error). A number of domains may cause systematic error that is overestimation of benefits and underestimation of harms. The problem with not reaching an appropriate number gives, however, rise to random errors. We have dealt extensively with this shortcoming in our manuscript, including in the trial sequential analysis plus amendments below. We have now specified which specific bias risk domains we are referring to in the sentences (the bias risk domains recommended to be assessed in The Cochrane Handbook for Systematic Reviews of Interventions <http://handbook.cochrane.org>):

Revised manuscript, page 4: "It was possible to conduct the trial with a low risk of bias (adequate allocation sequence generation, adequate allocation concealment, adequate blinding, no risk of

selective outcome reporting, low risk of incomplete outcome data bias, no risk of 'for profit' bias), which was the primary strength of this randomised clinical trial. "

Revised manuscript, page 18-19:

#### "Strengths

First of all, the trial was conducted with an overall high level of methodological quality and we assessed the validity of the trial results according to the procedure proposed by Jakobsen et al., including adjusting the thresholds for significance according to the number of randomised participants and the planned sample size.<sup>35</sup> We also proved the feasibility of our trial design, which can be used for larger trials provided that funding can be raised. Our trial has a number of additional strengths: (1) The trial protocol was registered before randomisation began at ClinicalTrials.gov. In this protocol the outcome hierarchy and plans for analyses were presented. Our trial was altogether conducted according to good clinical research practice, with low risk of bias (adequate allocation sequence generation, adequate allocation concealment, adequate blinding, no risk of selective outcome reporting, low risk of incomplete outcome data bias, no risk of 'for profit' bias), and a high degree of external validity.<sup>41-45</sup>"

We fully agree with the reviewer that it is a major limitation that we did not reach our sample size and this is clarified in the revised manuscript:

Revised manuscript, page 4:

- "The primary limitation of this randomised clinical trial was that only 44 out of the planned 84 participants were randomised in this small-scale trial."

Revised manuscript, page 20:

#### "Limitations

Our trial has a number of limitations. This small-scale trial was in essence failed because we only included 44 out of the planned 84 participants. The trial inclusion lasted for about two years as planned but we had problems with recruiting participants. Basically, not enough eligible depressed patients were referred to the clinic within the planned trial period. The great advantage of the randomised clinical trial in general is that all known and unknown participant characteristics will be similar at baseline in compared intervention groups.<sup>40</sup> However, even though our baseline characteristics indicate similarity between the two groups on assessed baseline characteristics, it is unlikely that all baseline characteristics will be similar when only 44 participants are randomised. The low number of randomised participants in this small-scale trial increases the risks of wrong results due to type I errors, and type II errors,<sup>49, 50</sup> and our adequate trial methodology cannot necessarily compensate for these increased risks."

Reviewer 1: b) the underlying manuals are not accessible to the international psychotherapeutic community. This creates problems with possible differences of the quality of the underlying manual, therapist training and allegiance. These limitations are still not sufficiently discussed.

Our response: We thank the reviewer for this important comment. It is a clear limitation that the treatment manuals are only available in Danish and we have now discussed these limitations in our revised manuscript.

Revised manuscript, page 19:

"Both of the trial interventions were conducted using manuals (available at <http://ctu.dk/publications/supplementary-material.aspx>) and adherence to the manuals was assessed as relatively high by an independent Danish psychologist trained both in mentalization-based therapy and third wave cognitive therapy. The manualization of the trial interventions makes it possible, to

some extent, to implement the two trial interventions in clinical practice and to replicate or refute our results in future trials, but both treatment manuals are currently only available in Danish, which limits the possibility for non-Danish speakers to assess the quality of the treatment manuals. We are in the process of translating the third wave cognitive manual, which will be published at a later time point. The mentalization-based treatment is described thoroughly elsewhere.<sup>14-17</sup> Nevertheless, it is a clear limitation that the manuals are not currently available in English.”

Reviewer 1: The paper is considerably improved but still needs major revision.

Our response: We thank the reviewer for this positive comment and we now hope that the above amendments to our manuscript represent sufficient revisions.

Reviewer: 3

Reviewer Name Andrew Mackinnon

Reviewer 3: The authors have very diligently addressed the issues I raised, as well as those of the other reviewers.

Our response: We thank the reviewer for this very positive comment.

Reviewer 3: I make only one suggestion. Rather than referring to ‘regression to the mean’ as a possible explanation for the changes observed (p. 18), I believe it would be more accurate and comprehensive to refer to the nature history of the disorder. For example:

However, we did not include a control group receiving no intervention in this head-to-head trial so it is unclear whether it was trial intervention effects or the nature progress of the disorder in this sample is responsible for these changes. (Regression to the mean embraces only measurement effects.)

We agree with the reviewer that natural history of the disorder may represent one additional explanation to the course of the patients. We have now revised the manuscript accordingly.

Revised manuscript, page 18:

“Compared with the baseline scores, both intervention groups improved during the trial period on all continuous outcomes. We did not include a control group receiving no intervention in this head-to-head trial so it is unclear whether it was trial intervention effects, regression towards the mean, or the natural progression of the disorder in this sample or regression towards the mean which was responsible for these changes.<sup>40</sup> More randomised clinical trials are needed to assess the effects of third wave cognitive therapy and mentalization-based treatment for major depressive disorder.”

I agree with the authors that it would not be ethically acceptable to undertake their trial with an arm in which patients were offered no or placebo-only treatment. It would also not answer any clinically relevant question. If they proceed to further trials, I suggest they consider current treatment or a standardized form of conventional CBT as the comparator. In this context, undertaking a comparison of CBT3 and mentalization within an equivalence/non-inferiority framework may be appropriate.

Our response: We are very happy that the reviewer agrees that it would not be ethically justifiable to conduct this trial with a ‘no treatment’ group. The reviewer suggests an equivalence/non-inferiority framework design, but such designs have a number of limitations (see e.g., Non-inferiority trials are unethical because they disregard patients' interests’, by Garattini and Bertele, *The Lancet* 2007 & Noninferiority trials: clinical understandings and misunderstandings, Powers et al., *PMC* Feb 19, 2014.). The equivalence/non-inferiority framework design might in some circumstances be the optimal design for a future trial, on the other hand if less severely depressed participants are included in

future trials a placebo controlled superiority trial might be the optimal trial design. The specific trial design will depend on the choice of the trial intervention, the trial populations, and methodological issues. Our recommendation is basically that more trials, in general, are needed to assess the effects of third wave cognitive therapy and mentalization-based treatment for major depressive disorder because evidence is lacking.

Revised manuscript, 'Discussion' page 18:

“Compared with the baseline scores, both intervention groups improved during the trial period on all continuous outcomes. We did not include a control group receiving no intervention in this head-to-head trial so it is unclear whether it was trial intervention effects, regression towards the mean, or the natural progression of the disorder in this sample or regression towards the mean which was responsible for these changes.<sup>40</sup> More randomised clinical trials are needed to assess the effects of third wave cognitive therapy and mentalization-based treatment for major depressive disorder.”

Revised manuscript, page 26:

“Our preliminary results show that third wave cognitive therapy compared with mentalization-based treatment may be a more effective intervention for depressive symptoms measured on the HDRS. The effects of the two interventions did not seem to differ significantly regarding BDI II, SCL 90-R, and WHO 5. More randomised clinical trials are needed to assess the effects of third wave cognitive therapy and mentalization-based treatment.”