## PEER REVIEW HISTORY



## ARTICLE DETAILS

| | |
|---|---|
| **TITLE (PROVISIONAL)** | Enhancing Social Interaction in Depression (SIDE Study) – Protocol of a randomized controlled trial on the effects of a Cognitively-Based-Compassion-Training (CBCT®) for Couples |
| **AUTHORS** | Aguilar-Raab, Corina; Jarczok, Marc; Warth, Marco; Stoffel, Martin; Tieck, Maria; Berg, Judith; Negi, Lobsang; Harrison, Tim; Pace, Thaddeus; Ditzen, Beate |

## VERSION 1 – REVIEW

| | |
|---|---|
| **REVIEWER** | Steven D Barger<br>Professor, Psychological Sciences Northern Arizona University, Flagstaff, Arizona US |
| **REVIEW RETURNED** | 26-Nov-2017 |

| | |
|---|---|
| **GENERAL COMMENTS** | Review of BMJOpen Submission - Enhancing Social Interaction in Depression (SIDE Study) – Protocol of a randomized controlled trial on the effects of a Cognitively-Based-Compassion-Training (CBCT®) for Couples<br><br>The authors are to be commended for preregistration and full disclosure of their proposed trial. The study design figures are helpful, the data monitoring plan appears rigorous and the authors provide a data sharing plan (p. 19). The ethical obligations of the trial seem robustly addressed.<br><br>Generally the manuscript has sufficient detail and the analytic plan is clear. Additional suggestions and concerns are described below.<br><br>The power analysis is reasonable for the primary outcome measure, the Hamilton depression rating scale (p. 17). However, there are a large number of secondary analyses which are likely to increase the likelihood of effect vibration and spurious statistical significance, particularly for biomarkers (Nature Reviews Neuroscience 2013; 14, 365–376). The number of secondary comparisons in the abstract numbers 11 and this is not the full set of secondary analyses. I see around 20 or more secondary outcome measures in a proposed sample size of 50. This ratio strongly undermines the scientific credibility of any one of these secondary analyses. Greater selectivity for secondary outcomes would improve the paper, particularly in light of inconsistencies in the descriptions of primary citation sources in the introduction (see below).<br><br>There are more than 15 psychosocial assessments listed in Table 1. I did not see a rationale for these measures or how they will be analyzed.<br><br>Study objective #2 (p. 8) is to improve social cognition process and |

| | |
|---|---|
| | interpersonal skills of the female partner. This key outcome is measured indirectly with a single measure, eye tracking. Why not assess social interaction directly? Several times potentially relevant outcome measures are listed (increased caring, kindness, levels of relatedness-p. 8 line 3) but these are not included as outcome measures. In contrast, there are a large number of biomarker assessments that have much weaker links with the underlying therapeutic processes. Thus, the outcome measures do not correspond well with the study objectives. That is, the study manipulates social interaction processes but only examines those processes via eye tracking. The couple interaction videotapes seem a more appropriate and sensitive context in which to assess the efficacy of the treatment.<br><br>In this regard, will there be any manipulation checks to determine whether the compassion training altered the processes described on pp. 12-13?<br><br>Those in the control group who enroll in psychotherapy will be excluded (p. 14) but what about those who initiate pharmacotherapy? It seems important to track prescription medication use across all study groups.<br><br>I did not see a plan for missing data analysis (p 11) although intention to treat and per protocol approaches are listed. Describing how adherence will be modeled in the primary analysis would be helpful.<br><br>Please provide specific inclusion criteria (i.e., cut scores) for the Hamilton and Beck scales (p. 9 lines 26-28).<br><br>Additional details of block randomization would be helpful-what criteria are used for blocking?<br><br>Biomarker assessments are collected on “Day 1” but there are two “Day 1” sessions listed in the CONSORT flow diagram. I would be helpful to distinguish pre- and post-test assessment days in the text and in the flow diagram.<br><br>Issues regarding presentation and interpretation of the literature<br>The rationale for including the serotonin transporter gene was that it is sensitive to stress (p. 5). However, the study cited in support of this measure showed only a baseline difference and no sensitivity of this marker to depression treatment. Thus, reference # 15 would argue against including this marker in a depression treatment trial.<br><br>As part of the rationale to examine cortisol the authors claim “intensified cortisol reaction in social stress tests” (p. 5, citations 20 and 21). Citation 20 is a meta-analysis with a median study N of 15 participants (196 total). This study found substantial heterogeneity in cortisol patterns and made 21 statistical comparisons across both stress reactivity and stress recovery. Thus there were more statistical comparisons than studies (21 vs 8 studies in the meta-analysis). In combination with the very small samples in this analysis there is little foundation for inference regarding these patterns.<br><br>The second citation in support of cortisol reaction to social stress (#21) was a study of hamsters, ground squirrels and bears. Neither social stress (it examined hibernation) nor cortisol were evaluated. It seems that the cost and effort required to assess cortisol exceeds |

| | |
|---|---|
| | the scientific rationale for its utility.<br><br>Alpha-amylase is also multiply determined and has an ambiguous relation with sympathetic activity that is glossed over in the introduction. This ambiguous association is clearly described in reference #23. As noted above, why not measure stress and distress in social interaction directly rather than rely upon indirect and possibly invalid biomarkers? |

| | |
|---|---|
| **REVIEWER** | Mario Miniati, M.D., Ph.D.<br>University of Pisa, Italy |
| **REVIEW RETURNED** | 16-Dec-2017 |

| | |
|---|---|
| **GENERAL COMMENTS** | The protocol is on an interesting issue. The study aims to investigate whether a CBCT program adapted for couples (CBCT-fC) can improve depressive symptoms, distress, social interaction skills, and the neurobiological regulation of stress, considering that relationship quality and the partners' health might be negatively affected by depressive symptomatology, which may result in overall impairments in social functioning of a romantic couple.<br>The paper certainly is falling within the scope of the Journal.<br>However, some point of discussion, in my opinion, should be addressed:<br><br>A. Inclusion criteria section<br>1. Male partners are included if they 'are free of mental disorders, such as depression or anxiety', as detected by SCID. Is this meaning that they are included if free of mental disorders with a cross-sectional assessment or even with a longitudinal one? Are mental disorders excluded in the entire lifespan or only at baseline assessment?<br>2. Are male partners included even if heavy smokers?<br>3. What the authors mean for 'severe' metabolic, endocrinological, neurological etc. conditions? Are they meaning 'not stabilized' conditions?<br>4. Are medical treatments that might interfere with mood and anxiety levels (for example, thyroid hormones, corticosteroids or interferon) allowed?<br>5. No formal exclusion criteria list is provided. No lifetime or cross-sectional comorbidity issue for mental disorders is raised.<br><br>B. Please, specify more in detail the characteristics of TAU.<br>1. Is it encompassing a psychopharmacological treatment the 'individual therapy' of TAU?<br>2. Subjects in 'regular psychotherapy' will be excluded from the analysis. Giving that, what the Authors mean for individual therapy? Only pharmacological treatments are allowed?<br>3. Is a psycho-educational support allowed?<br><br>C. The definition of the 'primary outcome' is unclear.<br>The Authors stated that 'the primary study outcome is the reduction of depressive symptoms from baseline to post-intervention as measured by the HDRS observer-based rating'.<br>1. Is it defined a HAM-D cut-off score at baseline for the inclusion in the study? It is different to consider a sample affected with a severe depression or a sample with a mild/moderate depression.<br>2. The HAM-D score 'reduction' is unclear. Are the Authors meaning a 50% reduction of baseline HAM-D scores for the definition of 'clinical response'?<br>3. Are they differentiating 'response' form 'remission' (defined with a |

| | |
|---|---|
| | cut-off score, for example HAM-D<7) ?<br>4. Is it considered the definition of 'partial response'?<br>5. No triage points are included in the study protocol<br><br>D. Research ethics (e.g. participant consent, ethics approval) are in the original language (German). Only a native German-speaking reviewer can address them appropriately. |

| | |
|---|---|
| **REVIEWER** | Stefan Schmidt<br>Department of Psychosomatic Medicine and Psychotherapy, Medical Faculty, University of Freiburg, Germanny |
| **REVIEW RETURNED** | 21-Dec-2017 |

| | |
|---|---|
| **GENERAL COMMENTS** | This is an excellent, well-written and very detailed study protocol. I have some general remarks and some minor suggestions for improvement.<br>General Remarks: This is a very ambitious studies in terms of participant recruitment and in terms of outcome variables. There is multitude of outcomes on different levels which is excellent from a methodological perspective but also on the brink of a) being too much burden for the (depressive) participants and b) getting lost in a multitude of analyses. Especially regarding b) I recommend also within the secondary analyses to clearly pre-define the variable within each measurement and to bring them also in a hierachial order regarding importance. This is especially true for HRV, circadiane variation and eye-gaze. Here variables are somehow fuzzy in description.<br>The other general concern is to describe the control condition as TAU. In my understanding TAU needs to secure a standard treatment for all patients in that group, which is not the case. Also there is no monitoring of standard treatment as far as I can see. Another relevant point is that patients in the control group are offered a treatment with CBCT later on, which will create expectancies. Thus, the control group is best characterized as a wait-list condition. I strongly recommend to montitor concurrent treatments in both groups!<br>Minor Remarks<br>On page 7 you introduce the variable vmHRV. Please explain the abbreviation and introduce (the group of) variable(s) appropriately.<br>On page 8 the intervention is introduced. The basic ideas and approach of the intervention should be described in one or two sentences<br>Page 9 Please provide a justification why the study is limited to female patients only.<br>Page 9 The design and the measurement points described the lower paragraph are bit fuzzy. Please name all measurement points and assign based on weeks/months after BL assessment. This is also true regarding figure 1, where BL takes place at ten weeks including 8 weeks of recruitment. This does not make sense at all.<br>Page 9/10 I suggest separate subheadings for Design and Participants (incl. recruitment).<br>page 10 It was not enirely clear to me how the camera for gaze detection is mounted ('next to head/eyes'). Is it mounted to head?<br>Page 10: Please give the rationale for only talking about positive conntent.Please describe in brief the instruction on how to change topics if this doesn't work.<br>Page 13. You mention that the instructor is certified. Please give a short description what is meant by that.<br>Page 14 The first and the third module of the course will be repeated. Please give the rationale for that. |

| | |
|---|---|
| | Page 14 Textmessages: how can the text message capture exercise frequency? I also would be careful not to put to much pressure and demands regarding exercise on patients. This will be perceived as control and missing exercises might be attributed personally (depressive participants!). So this might easily backfire.<br>Page 16 HRV variables need to be more precise in description. How exactly will you make inferences regarding vagal activity? How will you aggregate the segments. etc. For the circadiane analyses variables and data segmentation need to be speificed. Please report which behavioral parameters regarding situations are availabe to segment the data.<br>Page 17 Stastics. Group assignment as level 2 predictor refers to experimental vs. controll group or the different courses (which need also to be modelled)? |

**VERSION 1 – AUTHOR RESPONSE**

Reviewer(s)' Comments to Author:

**Reviewer: 1**

Reviewer Name: Steven D Barger

Institution and Country: Professor, Psychological Sciences, Northern Arizona University, Flagstaff, Arizona USA

Please state any competing interests: None declared

The authors are to be commended for pre-registration and full disclosure of their proposed trial. The study design figures are helpful, the data monitoring plan appears rigorous and the authors provide a data sharing plan (p. 19). The ethical obligations of the trial seem robustly addressed.

1. Generally the manuscript has sufficient detail and the analytic plan is clear. Additional suggestions and concerns are described below.

We thank the reviewer.

2. The power analysis is reasonable for the primary outcome measure, the Hamilton depression rating scale (p. 17). However, there are a large number of secondary analyses which are likely to increase the likelihood of effect vibration and spurious statistical significance, particularly for biomarkers (Nature Reviews Neuroscience 2013; 14, 365–376). The number of secondary comparisons in the abstract numbers 11 and this is not the full set of secondary analyses. I see around 20 or more secondary outcome measures in a proposed sample size of 50. This ratio strongly undermines the scientific credibility of any one of these secondary analyses. Greater selectivity for

secondary outcomes would improve the paper, particularly in light of inconsistencies in the descriptions of primary citation sources in the introduction (see below).

We appreciate that the Reviewer 1 mentioned this very important issue. The group involved in planning this multidisciplinary trial had similar discussions on inflated likelihood of false-positive results, that led to the careful selection of 13 secondary indicators [see Table X below, under Reviewer 3], which are all strong components of the conceptual model. In addition, and as recommended in the referenced article by Reviewer 1, because we planned to publish our study protocol we pre-registered our study, which included the specific hypothesis for each outcome. We also have a data sharing plan with a local service of the Ruprecht-Karls-University for published articles of this trial, as well as the full data set, after an embargo period after finishing the study. The Nature review article refers to neuroimaging studies of which most are cross sectional, between-subjects comparisons. The currently proposed block-randomized controlled trial has a stronger design that further decreases the likelihood of false positive results. We are aware of the dangers associated with underpowered studies.

3. There are more than 15 psychosocial assessments listed in Table 1. I did not see a rationale for these measures or how they will be analyzed.

We thank the reviewer for pointing this out. We organized Table 1 in alphabetical order of the instruments. However, this table includes items like standard demographic assessments, checklists during the experimental procedure, control items, and screening tools. We have rearranged Table 1 "*Overview of measures*" (p.26) of the manuscript in order to clarify the rationale for each of the mentioned measures.

4. Study objective #2 (p. 8) is to improve social cognition process and interpersonal skills of the female partner. This key outcome is measured indirectly with a single measure, eye tracking. Why not assess social interaction directly? Several times potentially relevant outcome measures are listed (increased caring, kindness, levels of relatedness-p. 8 line 3) but these are not included as outcome measures. In contrast, there are a large number of biomarker assessments that have much weaker links with the underlying therapeutic processes. Thus, the outcome measures do not correspond well with the study objectives. That is, the study manipulates social interaction processes but only examines those processes via eye tracking. The couple interaction videotapes seem a more appropriate and sensitive context in which to assess the efficacy of the treatment.

We thank the reviewer for this for this important perspective. We choose our focus to be the actual translation into the interaction behavior of the training content, as well as changes in biomarkers as a result of the intervention. As Reviewer 1 pointed out above, with each additional questionnaire, there is increasing risk of a false positive result. We kindly disagree that the outcome measures do not correspond to the study objectives and our study model. As stated on p. 9f (page number of revised manuscript), the major goals of the study are to determine whether or not the intervention:

A) reduces depressive symptoms in the female partner, B) improves social cognition processes and interpersonal skills of the female partner, and C) reduces stress reactivity in social interactions in both partners associated to alterations in methylation patterns of SLC6A4 and OXTR. Further, we will video-record the positive social interaction in the lab for control and validation reasons, as described on p. 12 (revised manuscript). First and foremost, we are interested in determining if the benefits of

the training will be reflected in change patterns of the biomarkers or mediated by psychobiological processes.

5. In this regard, will there be any manipulation checks to determine whether the compassion training altered the processes described on pp. 12-13?

Indeed, we implemented several manipulation checks and control assessments of adherence and patient compliance:

1) On p. 13 (revised manuscript) we note that during positive social interaction (PSI):

"Conversations will be videotaped and checked for protocol adherence."

2) On p. 23 (revised manuscript) section "Data monitoring" we indicate that during the training session:

"All training sessions will be videotaped and a selection will be reviewed by senior researchers. All PSI sessions will be videotaped and reviewed to ensure couples compliance to the PSI protocol."

3) And on p. 18 (revised manuscript, section "*adherence*") we mention that:

"Couples' compliance with the CBCT®-fC meditations and exercises will be assessed with pen and paper questionnaires after each weekly in-house session using Therapeutic Presence Inventory[89] (TPI, modified for the study), Practice Quality[90](PQ-M, adapted version of the Practice Quality-Mindfulness) and the Scale for the Multiperspective Assessment of General Change Mechanisms in Psychotherapy (SACiP).[91] The CBCT®-fC group will receive one text message four times weekly during the ten-week period. This message will serve as a reminder for the meditation practice and home-based exercises, and will also capture which meditations or practices participants completed, and how long they practiced. Participants will be asked to indicate reasons for non-performance. Participants' emotional status will also be assessed via the Positive and Negative Affect Schedule (PANAS).[92]

The TAU control group will receive one text message per week during the same time period asking to indicate their emotional status."

6. Those in the control group who enroll in psychotherapy will be excluded (p. 14) but what about those who initiate pharmacotherapy? It seems important to track prescription medication use across all study groups.

We thank the reviewer for this question, as this information was missing before. Medication use will be monitored in all groups. We have added this detail accordingly on p. 14 (revised manuscript, subheading "intervention": *"All treatments, including prescriptions, will be monitored for both groups."*).

7. I did not see a plan for missing data analysis (p. 11) although intention to treat and per protocol approaches are listed. Describing how adherence will be modeled in the primary analysis would be helpful.

Thank you for bringing up this important issue. We have added our plan for missing data to the statistical analysis section p. 22 (revised manuscript):

*"Several strategies will be used to minimize the likelihood of missing data. We will clearly explain to couples what their efforts will be for the study before they agree to participate. We will also support couples by proactively manage their appointments of the pre-post assessments and other study requirements by sending reminder emails, texts, and by making reminder phone calls. Training sessions will be offered in the evenings with a flexible start (between groups), to meet the needs of each training group. Completeness of questionnaire data will be closely monitored to avoid missing questionnaires. Participants will fill in their questionnaires at the IMP using a tablet PC. Study personnel will be available in the event of any questions or technical issues. Biomarker sampling will be monitored using standardized checklists, and we will note missing specimens as well as the reasons for their not being available.*

*Patients drop-outs from the training will be closely monitored and reasons for leaving the study before the end of the protocol will be assessed to distinguish study -related reasons for attrition compared to other reasons, and to improve study procedures"*

8. Please provide specific inclusion criteria (i.e., cut scores) for the Hamilton and Beck scales (p. 9 lines 26-28).

We now mention on p. 11 (revised manuscript, section "Participants") that the inclusion criteria / cut-off scores for the HDRS is ≥12, and for the BDI-II ≥ 16.

9. Additional details of block randomization would be helpful-what criteria are used for blocking?

We have updated the section "Blinding and randomization" with additional details (see p. 14 of the revised manuscript).

*"…Blocks for each study cohort (consisting of at least 8 couples) will be randomized after recruitment and baseline assessments are completed."*

10. Biomarker assessments are collected on "Day 1" but there are two "Day 1" sessions listed in the

CONSORT flow diagram. I would be helpful to distinguish pre- and post-test assessment days in the text and in the flow diagram.

A similar point was raised by Reviewer 3. To make the schedule for the study clearer we have re-labeled the measurement points with the recruitment start as reference throughout the manuscript and graphs as follows (p. 10 “Study protocol” and see also figure 3: “Flow chart”):

*“Changes in outcome measures from baseline (T1 at week 8-10 after recruitment start); to post-intervention assessment (T2 week 20-21 after recruitment start); to follow up 4-week post Intervention (FU1 week 24-25) and to 12-week follow up post intervention (FU2 week 33-34) will be compared between the two study arms (Figure 1).”*

Issues regarding presentation and interpretation of the literature:

11. The rationale for including the serotonin transporter gene was that it is sensitive to stress (p. 5). However, the study cited in support of this measure showed only a baseline difference and no sensitivity of this marker to depression treatment. Thus, reference # 15 would argue against including this marker in a depression treatment trial.

We thank the reviewer for this comment. We added additional information, which shows a sensitivity of SLC6A4 methylation to pharmacologic anti-depressive treatment, to cognitive behavioral therapy, as well as to the severity of depressive symptoms. This suggests a susceptibility of serotonin transporter gene promoter methylation to treatments targeting mental health. Thus, we hypothesize that alterations in serotonin transporter methylation may be a mechanism mediating the health promoting effects of the intervention explored by this study.

12. As part of the rationale to examine cortisol the authors claim “intensified cortisol reaction in social stress tests” (p. 5, citations 20 and 21). Citation 20 is a meta-analysis with a median study N of 15 participants (196 total). This study found substantial heterogeneity in cortisol patterns and made 21 statistical comparisons across both stress reactivity and stress recovery. Thus, there were more statistical comparisons than studies (21 vs 8 studies in the meta-analysis). In combination with the very small samples in this analysis there is little foundation for inference regarding these patterns. The second citation in support of cortisol reaction to social stress (#21) was a study of hamsters, ground squirrels and bears. Neither social stress (it examined hibernation) nor cortisol were evaluated. It seems that the cost and effort required to assess cortisol exceeds the scientific rationale for its utility.

We thank the reviewer for his very careful attention to our citations, and we must also profusely apologize. The second citation was a mistake that happened when we merged our Endnote files. We were actually referring to “Stetler & Miller (2011): Depression and Hypothalamic -Pituitary-Adrenal Activation: A Quantitative Summary of Four Decades of Research”. This extensive meta-analysis synthesizes data from over 350 studies and more than 18,000 participants for the comparison of differences in cortisol levels between healthy and depressed individuals. The authors found a medium-sized increase in cortisol in depression, which was in part dependent on moderating factors including timing of sampling and methodological quality of the primary studies. Together with the earlier meta-analysis by Burke and colleagues (2005) on cortisol response to psychological stress, there is sufficient evidence to assume that HPA-axis regulation is altered in major depression, although the mechanisms involved are not yet fully understood and require more research. We discuss the results and limitations of the two studies more clearly in the present revision on p. 5 (revised manuscript):

*"Moreover, a meta-analysis by Stetler and Miller found both higher basal cortisol levels[26] and elevated cortisol release in response to acute psychosocial stress challenge in depressed versus healthy individuals[17]. The magnitude and direction of these cortisol effects, however, were found to be strongly dependent on moderating variables such as methodological quality, type of cortisol assessment, sex, measurement time, or type of stress task[17, 26]. In particular, women with remitted depression showed a blunted cortisol response to a psychologically stressful task compared to healthy controls, while cortisol reactivity was increased in depressed men.[27, 28] Together these findings suggest that more research is needed, although there is evidence for HPA-axis dysregulation in major depression."*

13. Alpha-amylase is also multiply determined and has an ambiguous relation with sympathetic activity that is glossed over in the introduction. This ambiguous association is clearly described in reference #23. As noted above, why not measure stress and distress in social interaction directly rather than rely upon indirect and possibly invalid biomarkers?

This section presents both critical and encouraging views on the validity of saliva alpha amylase as an indicator of autonomic activity. We would, however, disagree that mixed findings for a certain biomarker make should automatically make it an invalid measure that should no longer be used in research. In contrast, the majority of studies cited in our manuscript show strong potential for saliva alpha amylase, and suggest this marker should be investigated further. Nevertheless, in the present study we will also measure stress/distress directly; in the checklist during the laboratory assessments we will assess for current stressful events. We will also assess for self-reported stress before and after the PSI, and in the battery of questionnaires (with the Trier Inventory for chronic Stress [TICS]).

**Reviewer: 2**

The protocol is on an interesting issue. The study aims to investigate whether a CBCT program adapted for couples (CBCT-fC) can improve depressive symptoms, distress, social interaction skills, and the neurobiological regulation of stress, considering that relationship quality and the partners' health might be negatively affected by depressive symptomatology, which may result in overall impairments in social functioning of a romantic couple.

The paper certainly is falling within the scope of the Journal.

We sincerely thank the reviewer for this comment about our manuscript and study.

However, some point of discussion, in my opinion, should be addressed:

A. Inclusion criteria section

1. Male partners are included if they 'are free of mental disorders, such as depression or anxiety', as detected by SCID. Is this meaning that they are included if free of mental disorders with a cross - sectional assessment or even with a longitudinal one? Are mental disorders excluded in the entire lifespan or only at baseline assessment?

We will exclude male participants/partners with acute mental disorders at baseline assessment, not life span (see p. 11 of revised manuscript):

*"Male partners are included if they 1) are at least 20 years of age, 2) have a romantic relationship for at least 2 years (with the targeted depressed female partner), and 3) are not diagnosed with a primary recurrent depressive disorder (ICD-10: F33.0 or F33.1) by the Structured Clinical Interview (SCID) of the Diagnostic and Statistical Manual of Mental Disorders (DSM-V).[73] General exclusion criteria for the male patients are primarily with regard to acute mental disorders at baseline not life span: 1) other severe mental disorders (such as psychotic symptoms; bipolar disorders, acute suicidal tendency), 2) altered physical condition (chronic or severe metabolic, endocrinological, neurological, nephrological, cardiac or hepatic conditions), 3) heavy smoking (≥ 20 cigarettes/ day), substance abuse or acute addiction, 4) enrolling in couple therapy; and 5) current participation in a mindfulness- or compassion-based group training."*

2. Are male partners included even if heavy smokers?

We thank the reviewer for noticing this point and apologize for not making this clear. No, males (as well as females) who are heavy smokers will not be included. We have updated the section on p. 11 (revised manuscript) accordingly (see just quotation of the revised text above).

3. What the authors mean for 'severe' metabolic, endocrinological, neurological etc. conditions? Are they meaning 'not stabilized' conditions?

We again apologize for not making this clear. No, by severe metabolic conditions we were referring to diagnosed conditions *("chronic or severe metabolic, endocrinological, neurological, nephrological, cardiac or hepatic conditions"),* such as dementia, Parkinson, heart failure, acute kidney injury, etc. These are criteria for exclusion; medication will be monitored accordingly.

4. Are medical treatments that might interfere with mood and anxiety levels (for example, thyroid hormones, corticosteroids or interferon) allowed?

Yes, participants will not be excluded for those conditions, but we will monitor any medication (see above).

5. No formal exclusion criteria list is provided. No lifetime or cross -sectional comorbidity issue for mental disorders is raised.

On p. 11 of the revised manuscript we did list before, and continue to include, the following:

*"General exclusion criteria for the female patients are: 1) other severe mental disorders (such as psychotic symptoms; bipolar disorders, acute suicidal tendency), 2) altered physical condition (pregnancy, chronic or severe metabolic, endocrinological, neurological, nephrological, cardiac or hepatic conditions), 3) heavy smoking (≥ 20 cigarettes/ day), substance abuse or acute addiction, 4) enrolling in psychological therapies (couple therapy; individual psychotherapy – except for probatory phase including a maximum of the first six sessions), and 5) current participation in a mindfulness - or compassion-based group training."*

B. Please, specify more in detail the characteristics of TAU.

1. Is it encompassing a psychopharmacological treatment the 'individual therapy' of TAU?

We thank the reviewer for this and the following questions about the Treatment as Usual (TAU) condition. Participants of the control group (TAU) will be allowed to undergo or continue pharmacological treatment. This is what we meant by 'individual therapy' in addition to medical check-ups or short interviews with a physician (p. 14 of the revised manuscript):

*"All participants (including those randomized to CBCT®-fC) will receive the standard treatment (TAU) for depression. All treatments, including prescriptions, will be monitored for both groups. At a minimum this monitoring will include general practitioner care and contact with community mental health providers."*

And p. 17 of the revised manuscript "Treatment as usual (TAU)":

*"Couples assigned to the TAU group will start (or continue) with TAU which will consist of receiving medical treatment and primary health care, or waiting for starting individual therapy (average waiting time 3 months."*

2. Subjects in 'regular psychotherapy' will be excluded from the analysis. Giving that, what the

Authors mean for individual therapy? Only pharmacological treatments are allowed?

Thank you for this comment. We have changed wording/terms in order to clarify what is meant by individual treatment allowed in the TAU condition. Regular psychotherapy is understood as a short-/long-term psychotherapy that is provided by a licensed psychotherapist. In the German health care system, licensed psychotherapists are able to provide up to 25/50/100 (or even more) psychotherapy sessions, depending on the therapeutic approach. A psychodynamic psychotherapist/psychoanalyst can provide up to 100/300 sessions that are paid for by the health insurances. Participants of the TAU conditions will be included in the analysis only if they are undergoing medical check-ups or short interviews with their physician, and/or are being treated with medication. However, they will be

excluded when they are being treated by a Behavioral, Cognitive-Behavioral-, Psychodynamic- or other psychotherapist providing at least 12-24 therapy sessions, at 50-minute sessions every week.

3. Is a psycho-educational support allowed?

Yes, psycho-educational support is allowed for participants in the TAU group. But this is not a common practice/treatment in Germany. Usually this treatment takes place during medical check-ups or sessions with the physician.

C. The definition of the 'primary outcome' is unclear. The Authors stated that 'the primary study outcome is the reduction of depressive symptoms from baseline to post-intervention as measured by the HDRS observer-based rating'.

1. Is it defined a HAM-D cut-off score at baseline for the inclusion in the study? It is different to consider a sample affected with a severe depression or a sample with a mild/moderate depression.

Yes, for inclusion in the study we defined baseline cut-off scores of ≥12 for HDRS, and ≥ 16 for BDI-

II. We have added this information to the manuscript, and we thank the reviewer for pointing out that this information was not included before (p. 11 revised manuscript):

*"(…) in combination with the observer-based Hamilton Depression Rating Scale (HDRS ≥ 12; within the range of a mild depression)[74] and the Beck Depression Inventory (BDI-II ≥ 16; within the range of mild depression) self-rating. [76]"*

2. The HAM-D score 'reduction' is unclear. Are the authors meaning a 50% reduction of baseline HAM-D scores for the definition of 'clinical response'?

Overall, the treatment is offered to any kind of (recurrently) depressed patients; patients who are looking for a treatment taking the partner and relational aspects into account; or patients who just left a psychosomatic/psychiatric clinic as an in-patient or who are waiting for an outpatient psychotherapeutic treatment (24 sessions up to 100/300).

Based on the current S3-guidelines of the Working Group of the Scientific Medical Societies of Germany (AWMF; second edition, 2015; register-nr. nvl-005) the scores for the HDRS German version are defined as follows: ≤ 8 = normal/remitted; 9-16 = mild; 17-24 = moderate; ≥25 severe depression.

Although we are predominantly interested in symptom reduction taking into account that some of the included patients left inpatient treatment some weeks before (illness might not be too severe, and treatment is an *add-on* to TAU), we might investigate the rates of clinical response following the

guidelines of the AWMF. That is, non-improvement is defined as < 20% reduction from baseline HDRS scores, clinical response is defined as 20-50% reduction from baseline HDRS scores (minimal effect); < 50% reduction from baseline HDRS scores (strong effect); whereas remission will be defined as reaching the cut-off score of the HDRS (≤ 8) at post-treatment.

3. Are they differentiating 'response' form 'remission' (defined with a cut-off score, for example HAM-D<7)?

Yes – please see comments above (point 2).

4. Is it considered the definition of 'partial response'?

No, we did not plan to consider a definition of a partial response, and instead will focus on the above-mentioned definitions of responses "only".

5. No triage points are included in the study protocol

First, one of the exclusion criteria is acute suicidal tendencies . We feel that patients with suicidal tendencies would not benefit from and would not be appropriate for CBCT®-fC. In any case, if patients do develop suicidal tendencies during training we will to refer them for individual psychotherapy session(s)/crisis intervention. These sessions will be conducted by a psychotherapist on the study team - and if necessary, an inpatient treatment will be provided in our psychiatric or psychosomatic departments/clinic.

D. Research ethics (e.g. participant consent, ethics approval) are in the original language (German).

Only a native German-speaking reviewer can address them appropriately.

We apologize for this, but the ethical committee preferred that these study documents be written in German. We decided to include these original forms as part of the submission.

**Reviewer: 3**

Reviewer Name: Stefan Schmidt

Institution and Country: Department of Psychosomatic Medicine and Psychotherapy, Medical Faculty, University of Freiburg, Germany

Please state any competing interests: None declared

Please leave your comments for the authors below

This is an excellent, well-written and very detailed study protocol. I have some general remarks and some minor suggestions for improvement.

Thank you so much for these kind comments about our manuscript.

**General Remarks:** This is a very ambitious study in terms of participant recruitment and in terms of outcome variables. There is multitude of outcomes on different levels which is excellent from a methodological perspective but also on the brink of a) being too much burden for the (depressive) participants and b) getting lost in a multitude of analyses. Especially regarding b) I recommend also within the secondary analyses to clearly pre-define the variable within each measurement and to bring them also in a hierarchical order regarding importance. This is especially true for HRV, circadian variation and eye-gaze. Here variables are somehow fuzzy in description.

These are valuable points, thank you for bringing them up. The multidisciplinary study group had intense discussion on the burden of measures for study participants. However, experience from a pilot study as well as from partners having the same target group (depressed females with similar depression level) gave no reason for concern about the amount of burden imposed on participants for the measures conducted. Furthermore, compared to the pilot trial, we have reduced the time and psychological effort for the participants by curtailing some of the outcomes assessed. Furthermore, if indicated, participants can take a break during parts of the assessments or are free to stop participation withdrawing their consent to the study.

We thank the reviewer for this comment and in response have generated a table with a hierarchical order of importance of the outcome variables for a better overview:

Table X: Outcome variables in hierarchical order of importance

| | | Instrument | Unit of analysis pre-post training |
|---|---|---|---|
| Reactivity on training | Depressive symptom severity expert rating | HDRS | Score change |

| | | | |
|---|---|---|---|
| Reactivity on training | Depressive symptom severity self-rating | BDI-II | Score change |
| Reactivity on PSI | Eyetracking | total gaze duration toward partner | Milliseconds |

| | | | |
|---|---|---|---|
| Reactivity on PSI | Stress sensitive biomarkers | saliva cortisol | Area under the curve |
| Reactivity on PSI | Stress sensitive biomarkers | saliva alpha amylase | Area under the curve |
| Reactivity on PSI | Stress sensitive biomarkers | RMSSD during PSI | Millisecond change |
| Reactivity on training | Stress sensitive biomarkers | Circadian Amplitude RMSSD | Millisecond change |
| Reactivity on training | Stress sensitive inflammatory markers | Cytokines/interleukins: IL-1ß & IL-6 | Concentration change (pg/ml) |
| Reactivity on training | Polymorphism as moderator | SLC6A4 and OXTR | Serotonin: LL vs. SL or SS<br>Oxytocin: GG vs. GA vs. AA |
| Reactivity on training | Self-rating of different psychological | Questionnaires (see list in Table 1: Overview of | Score change |

| | | | |
|---|---|---|---|
| | constructs regarding perceived distress/stress, relationship quality, mindfulness, compassion etc. | measures; see in the manuscript) | |

1. The other general concern is to describe the control condition as TAU. In my understanding TAU needs to secure a standard treatment for all patients in that group, which is not the case. Also, there is no monitoring of standard treatment as far as I can see. Another relevant point is that patients in the control group are offered a treatment with CBCT later on, which will create expectancies. Thus, the control group is best characterized as a wait-list condition. I strongly recommend to monitor concurrent treatments in both groups!

Thank you for this advice! Indeed, we will monitor/ask for details of current treatment in the first telephone screening, where patients will be asked to indicate their current circumstances. We will also discuss these details again at the diagnostic interviews (with individuals, and couples). Furthermore, each couple/individual will be asked to indicate if their treatment or degree of burden / distress had changed. If indicated, therapy sessions will be provided for individuals or couples by team psychotherapists or transferred to our psychosomatic/psychiatric departments.

With regard to the second point, we will offer CBCT®-fC to the TAU participants later for ethical reasons (because we think there is a strong likelihood of CBCT® efficacy to ameliorate depression), and to avoid higher dropout rates. Couples who are allocated to the TAU condition will be tested at the same time as those randomly assigned to the treatment condition (pre T1, post T2, FU1 4 weeks later, and FU2 10-12 weeks later). Participants in the TAU will be offered CBCT®, but only after all study assessment have taken place (after FU2).

**Minor Remarks**

2. On page 7 you introduce the variable vmHRV. Please explain the abbreviation and introduce (the group of) variable(s) appropriately.

We kindly refer the reviewer to p. 6 of the revised manuscript (subheading "*Psychobiological underpinnings"*), where vmHRV is introduced as well as p. 28 "*Abbreviations*".

3. On page 8 the intervention is introduced. The basic ideas and approach of the intervention should be described in one or two sentences.

Thank you for this advice. We have added information on ideas and approach of the intervention in this section (p. 15 revised manuscript):

*"For CBCT®, all human beings are considered to be endowed with a biologically -based compassion that is focused on those close to us. But as research has shown, compassion as an inner quality or skill can be trained based on the enlargement of feelings of endearment towards others and comes along with a variety of health benefits.[79] The change process described here is rooted in Tibetan-Buddhist understanding - as the training itself is based on the so called lo-jong tradition, which means "mind training" and is taught to cultivate compassion towards others not only close to us.*

*The model of how the intervention works is that change takes place in three steps: (1) starting at changing views by getting to know and to familiarize new perspectives receiving information, (2) based on that changing behavior through conviction that comes about through critical and reflective thinking, and then (3) embodied understanding takes place through thorough and repetitive, contemplative practice."*

During the training process, several prerequisites and primary conditions are introduced in a stepwise manner, ultimately leading to compassion.

4. Page 9 Please provide a justification why the study is limited to female patients only.

Our study focuses primarily on females due to several reasons. Current prevalence rates of major depression indicate that females are affected twice as often as men; there also seem to be some differences in the phenomenological symptomatology. Additionally, as the study focuses on psychobiological markers and as there are hormonal differences in women and men, and for financial and infrastructural (feasibility) reasons, we had to keep the sample size manageable. We therefore decided to focus on females first. We have added the rationale for focusing on women to the manuscript accordingly, on p. 10 (revised manuscript "*Participants*").

5. Page 9 The design and the measurement points described the lower paragraph are bit fuzzy. Please name all measurement points and assign based on weeks/months after BL assessment. This is also true regarding figure 1, where BL takes place at ten weeks including 8 weeks of recruitment. This does not make sense at all.

We thank the reviewer for picking up on this issue. We have relabeled the measurement points with the recruitment start as reference throughout the manuscript and graphs (p. 10 revised manuscript *"Study design*" as well as in the flow chart figure 3:

*“Changes in outcome measures from baseline (T1 at week 8-10 after recruitment start); to post-intervention assessment (T2 week 20-21 after recruitment start); to follow up 4-week post Intervention (FU1 week 24-25) and to 12-week follow up post intervention (FU2 week 33-34) will be compared between the two study arms (Figure 1).”*

6. Page 9/10 I suggest separate subheadings for Design and Participants (incl. recruitment).

Thank you for this advice - we have added separate subheadings to this section.

7. page 10 It was not entirely clear to me how the camera for gaze detection is mounted ('next to head/eyes'). Is it mounted to head?

Thank you for this question. No, the small webcam (Microsoft LifeCam) will be mounted on the arm of a microphone-like stand and positioned on the horizontal line of women’s eyes (y-axis) in the plane of eyes position (z-axis) but shifted to the left next to the head (x-axis). Technically, the position between both eyes would be best, but this is not feasible. An x-axis shift provided best results in pretests (e.g. compared to a y-axis shift with the camera over the head of the female (y-axis shift). We have added a picture of the lab set-up to the appendix 1.

8. Page 10: Please give the rationale for only talking about positive content. Please describe in brief the instruction on how to change topics if this doesn't work.

First, each person will receive a list of 27 standardized conversation topics and will be asked to choose topics they wish to talk about in whatever order is fitting for a particular couple. Topics were adopted from the Problem List (PL) used in research on relationship conflict (Hahlweg, Kraemer, Schindler, & Revenstorf, 1980), but modified so that the conversational focus is on positive instead of negative issues (e.g. trust, financial security, affection). Couples will be verbally instructed to exclusively talk about positive content, to be supportive, and if authentic, to express compliments towards each other. As soon as they are noticed to fall of track, they will be asked to continue with another topic (p. 13 revised manuscript):

*“Conversations will be videotaped and checked for protocol adherence.”*

Based on a video-analysis of the conversations, a manipulation check will take place.

9. Page 13. You mention that the instructor is certified. Please give a short description what is meant by that.

We added a short description on page 15 (revised manuscript). For the reviewers information: Dr. Aguilar-Raab is the primary trainer in this study: Besides being a couple and a family therapist, Dr. Aguilar-Raab is in an advanced stage of individual psychotherapy training currently conducting individual psychotherapy under supervision and has deepened clinical experience with treating

psychiatric and psychosomatic disorders during her practical training (she will receive certification/approbation for individual therapy in fall 2018). Fulfilling several requirements such as long-term practice of mindfulness- and compassion based meditation under guidance of a recognized supervisor and meditation teachers (Prof. Lobsang Tenzin Negi, Loden Sherab Dagyab Rinpoche and others), Dr. Aguilar-Raab attended the CBCT® Teacher certification program at Emory University (Atlanta, USA). Through the Francisco Varela Award of the Mind and Life Organization, USA, she initiated the modification process of the CBCT® protocol in close collaboration with Emory University (Atlanta, USA) and Stanford University (Palo Alto, USA) in 2015, experts in couple and family therapy.

10. Page 14. The first and the third module of the course will be repeated. Please give the rationale for that.

The first and the third module will be repeated for several reasons. The first module teaches the basic skill of attuning one's attention towards the present moment anchoring it at the breath (traditionally described as focused attention/awareness, or mindfulness on the breath). It functions as a basic or fundamental quality which is necessary for all following modules and practices. It needs a deepened understanding in order to prevent misunderstanding - in terms of a formal and informal practices. On top of that, in the second round, mindful communication and relational aspects where mindfulness can be a helpful skill will be introduced and guided dyadic exercises are taught. The rationale behind teaching module III on self-compassion is that former research with depressed individuals has shown that self-compassion needs to be addressed more elaborately. From a practical viewpoint for many, it is difficult to focus on self-compassion as several misunderstandings of "self-" related issues seem to be associated with an egoistic and self-focused approach. Furthermore, one part in this module is to deal with aspects in life that are difficult and hurtful, and how to change perspective in a way that means to change our relationship towards our self towards a more warm-hearted sense (which is an opposite approach of what many depressed individuals are used to - self-critical thinking seems to be more prone to (depressed) individuals in western countries).

11. Page 14 Text Messages: how can the text message capture exercise frequency? I also would be careful not to put too much pressure and demands regarding exercise on patients. This will be perceived as control and missing exercises might be attributed personally (depressive participants!). So this might easily backfire.

We thank the reviewer for pointing out this important issue. Participants will receive a link with a questionnaire on exercise adherence and about how they felt at that day (20 items). They will have to indicate whether they practiced that day or not, and if not, for what reasons. They will also have to indicate the practice they did, and how long and at what time they practiced. Our colleagues' previous experience suggests there is little or no concern that this puts too much pressure on participants. However, this issue is also closely addressed during training, i.e. how to deal with resistances when it arises, and how to deal with self-made criticisms and pressure.

12. Page 16 HRV variables need to be more precise in description. How exactly will you make inferences regarding vagal activity? How will you aggregate the segments etc. For the circadiane analyses variables and data segmentation need to be specified. Please report which behavioral parameters regarding situations are available to segment the data.

We extended the description of vagal activity inference from HR time series on page 20 (revised manuscript). We added a more detailed description of circadian analysis of rmssd. We do not plan to segment circadian analysis based on behavior. Here, the cosine method fits a single cosine cycle (period = 24h) into the data of each individual using trigonometric regression based on ordinary least squares. In healthy humans, with a standard diurnal trajectory, vagal activity increases during sleep and is lower during day (showing up in the AMPLITUDE parameter), which is expected to be blunted in depression. So far, previous research has shown an overall decrease in HRV in depressed vs. non-depressed patients, but this is indicative of the overall 24h mean (=MESOR Parameter).

13. Page 17 Statistics. Group assignment as level 2 predictor refers to experimental vs. control group or the different courses (which need also to be modelled)?

Group assignment refers to experimental vs. control group. We have edited this section accordingly (p. 21 revised manuscript “Statistical analysis and missing data”).

**VERSION 2 – REVIEW**

| | |
|---|---|
| **REVIEWER** | Steven Barger<br>Northern Arizona University, USA |
| **REVIEW RETURNED** | 13-Mar-2018 |

| | |
|---|---|
| **GENERAL COMMENTS** | Review of BMJOpen Resubmission - Enhancing Social Interaction in Depression (SIDE Study) - Protocol of a randomized controlled trial on the effects of a Cognitively-Based-Compassion-Training (CBCT®) for Couples<br><br>The authors graciously and effectively responded to the editorial feedback. Adding the hierarchy of outcome measures is a welcome addition to the manuscript. The editorial efforts have clarified and improved the flow of the manuscript.<br><br>Thank you for fixing the reference error. In that corrected passage (line 51 p 77) - it appears that the Stetler and Miller meta-analysis is being cited for cortisol responses to stress but cite #17, which examines stress responses, is Burke et al. 2005.<br><br>In the introduction eye tracking is noted as a dependent measure for objective #2. Later (untracked revision section, p 18 line 43) it is stated "In this study, we interpret eye-tracking data as an objective proxy for social cognition processes with regard to emotion-related information processing biases." The former is reasonable whereas the latter claim presumes a 1:1 relationship between eye movement and these psychosocial processes. This assertion requires strong evidence or the claim should be revised.<br><br>It would also be helpful to state how eye tracking data is to be aggregated. If this is restricted to off-face/on face (rather than eyes vs other face regions) this should be stated.<br><br>What is the partner burden in Depression Scale (Table 1)?<br><br>Viel glück mit Ihre Arbeit. |

| | |
|---|---|
| **REVIEWER** | Stefan Schmidt<br>Department of Psychosomatic Medicine and Psychotherapy, Medical Center - University of Freiburg, Freiburg, Germany |
| **REVIEW RETURNED** | 21-Mar-2018 |

| | |
|---|---|
| **GENERAL COMMENTS** | Paper has been approved substantially. I recommend acceptance. |

| | |
|---|---|
| **REVIEWER** | Mario Miniati, M.D., Ph.D.<br>Department of Clinical and Experimental Medicine, University of Pisa, Italy |
| **REVIEW RETURNED** | 29-Mar-2018 |

| | |
|---|---|
| **GENERAL COMMENTS** | The authors addressed the reviewers' comments. |

**VERSION 2 – AUTHOR RESPONSE**

Reviewer(s)' Comments to Author:

Reviewer: 1

Reviewer Name: Steven Barger

Institution and Country: Northern Arizona University, USA

Please state any competing interests: None declared

Please leave your comments for the authors below

Review of BMJOpen Resubmission - Enhancing Social Interaction in Depression (SIDE Study) - Protocol of a randomized controlled trial on the effects of a Cognitively-Based-Compassion-Training (CBCT®) for Couples

The authors graciously and effectively responded to the editorial feedback. Adding the hierarchy of outcome measures is a welcome addition to the manuscript. The editorial efforts have clarified and improved the flow of the manuscript.

We thank the reviewer.

Thank you for fixing the reference error. In that corrected passage (line 51 p 77) - it appears that the Stetler and Miller meta-analysis is being cited for cortisol responses to stress but cite #17, which examines stress responses, is Burke et al. 2005.

We thank the reviewer and fixed the error by clarifying the wording.

… Moreover, a meta-analysis by Stetler and Miller found higher basal cortisol levels in depressed individuals[26]. Similar, depressed versus healthy individuals showed elevated cortisol release in response to acute psychosocial stress challenge.[17]

In the introduction eye tracking is noted as a dependent measure for objective #2. Later (untracked revision section, p 18 line 43) it is stated "In this study, we interpret eye-tracking data as an objective proxy for social cognition processes with regard to emotion-related information processing biases." The former is reasonable whereas the latter claim presumes a 1:1 relationship between eye movement and these psychosocial processes. This assertion requires strong evidence or the claim should be revised.

Thank you for identifying this potential misunderstanding. We do not presume a 1:1 relationship, however, by parsing out the different patterns of orientation and engagement of attention (locations of initial fixations indicate orientation i.e. where one looks first, and duration of these fixations indicate the engagement of attention i.e. how long one looks), Eye tracking provides a rich source for the analysis of emotion-related information processing biases.

…eye-tracking data as a more objective proxy measure for social cognition processes with regard to emotion-related information processing biases.

It would also be helpful to state how eye tracking data is to be aggregated. If this is restricted to off-face/on face (rather than eyes vs other face regions) this should be stated.

We added the requested information on p18:

… Raw coordinates of gaze and fixation will be mapped to the following predefined areas of interest: 1. eyes, 2. mouth, 3. other regions of the face, and 4. off face. The areas of interest will be compared pre- to post-intervention.

What is the partner burden in Depression Scale (Table 1)?

This is a questionnaire generated by our group. It is based on the PHQ9 but referring to the burden that the depressed partner may cause. We figured that there is no appropriate instrument to assess partners' burden in depression. In the current stage, this instrument is exploratory.

Viel glück mit Ihre Arbeit.

Reviewer: 2

Reviewer Name: Mario Miniati, M.D., Ph.D.

Institution and Country: Department of Clinical and Experimental Medicine, University of Pisa, Italy

Please state any competing interests: None declared

Please leave your comments for the authors below

The authors addressed the reviewers' comments.

Reviewer: 3

Reviewer Name: Stefan Schmidt

Institution and Country: Department of Psychosomatic Medicine and Psychotherapy, Medical Center - University of Freiburg, Freiburg, Germany

Please state any competing interests: none

Please leave your comments for the authors below

Paper has been approved substantially. I recommend acceptance.