Reviewer's report

Title: The Working Alliance in Online vs. Face-to-Face Cognitive-Behavioral Therapy for Depression: A Randomized Controlled Trial

Version: 1 Date: 13 July 2011

Reviewer: Brjánn Ljótsson

Reviewer's report:

This study highlights an important area of research within internet-delivered psychological therapy, which is the role of therapeutic alliance. The success of online therapies challenges the view that the therapeutic relationship is the most important factor influencing outcome in all psychological treatment. Unfortunately, this study cannot shed light on this issue, since it measured the alliance at post-treatment. Although it is interesting to note that the WAI scores were similar between the two groups, the study has little more to say. If this is a secondary article I seriously question the value of reporting these data separately.

- MAJOR COMPULSORY REVISIONS

1. Abstract. “The therapeutic implications of these findings are discussed.”

Where are these therapeutic implications discussed?

2. p12. "(b) the residual gain score (difference between depression score at baseline and posttreatment)"

Computing residual gain score is more complicated than just subtracting the post-treatment score from the pre-treatment score (raw gain score). Please explain how the residual gain score was computed (or rename it to “change score” if it was indeed calculated as the difference between pre- and post-treatment).

3. p13. “To quantify the magnitude of differences between the two groups (online versus face-to-face), we used Cohen’s d as a measure of effect size. Cohen [44] distinguished between small (d = .20), medium (d = .50) and large (d = .80) effect sizes.”

This seems reasonable. But on p14 Cohen’s d is given for correlations and coefficients between WAI subscales and BDI post score, reported separately for each group. This is extremely confusing - to what do the “large” (between-groups?) effect sizes reported on p14 refer?

4. p14. The total amount of variance explained when regressing BDI baseline score (35.1%) and WAI-C composite score (44.5%) on the post-treatment BDI score seems to be about 80%. This is an unusually large number and a bit strange since the correlation coefficient between WAI-C and BDI post is only .40
Please recheck the calculations. Adding a couple of tables showing the regression analyses and coefficients for all entered independent variables would aid the reader in interpreting the data.

5. p15. “This finding is in line with previous studies reporting that a strong working alliance, comparable to that formed in face-to-face settings, can also be established in online settings”

The authors need to qualify this statement (that the working alliance was "strong") by comparing the scores they got with other studies or normative data. A point relating to this is that fact that all WAI scores were between 5.77 and 6.22 (of a maximum of 7, indicating that scores were indeed high). Is there any risk of ceiling effects in this study, making it hard to differentiate between the two conditions? It would be valuable if this were discussed in the manuscript.

6. p16. "In our sample, the attrition rates in the online group (22%) versus the face-to-face group (7%) differed widely.”

Was this "wide" difference really significant? If not, the authors should be more conservative in their description of the magnitude of the difference and consider if it is really worth discussing.

7. p17. “However, our results are in line with previous research by Knaevelsrud and Maercker [24] who found correlations between working alliance early in the treatment (measured at the fourth treatment session) and treatment outcome in an online setting, but no significant effects in the regression analysis. Working alliance and treatment outcome do seem to be connected somehow; further research is needed to cast more light on the relationship between these variables.”

These sentences are confusing. In the regression analyses, the authors showed that the post-treatment status was fully explained by the pre-treatment status, with the WAI not adding significant explanatory value. So why do they claim that the working alliance influences outcome? Isn’t the best explanation about this mysterious “connection” which disappears in the multiple regression analysis a simple correlation between pre-treatment BDI and WAI scores? I think adding the correlation between the pre-treatment BDI value and the WAI scales to Tables 3+4 (or a combined table as I suggest below) would help the reader in interpreting the data.

8. None of the limitations that are mentioned in the Discussion are as important as the fact that the WAI was administered at post-treatment. This makes the results very hard to interpret, as the WAI scores may simply be a function of post-treatment depressive status. This may also explain the fact that the BDI pre-treatment score was the best predictor of post-treatment status. This limitation should be highlighted and it should be stated that this study can hardly say anything about the influence of alliance on outcome considering that they were measured simultaneously. The authors should also include references that
state that therapeutic alliance should be measured before the outcome to have explanatory value.

9. p18. “Our study contributes to a better understanding of the mechanisms underlying internet-supported therapeutic interventions, replicating previous findings [19, 21, 23-26] showing that a strong working alliance can be established in an online setting and that the underlying mechanisms seem, to some extent, to be comparable with those operating in traditional face-to-face interventions.”

It is clear that there was no meaningful difference in perceived alliance at post-treatment between the two conditions. However, the fact that the WAI was assessed at post-treatment precludes any conclusions to be drawn about possible mechanisms of online treatment and how it compares to face-to-face treatment. The authors should be much more modest about the value of this study. It shows that at post-treatment, patients and therapists in both groups were equally satisfied with the therapeutic alliance, but is says very little, if anything, about treatment mechanisms.

- MINOR ESSENTIAL REVISIONS

10. Abstract. "Results from multiple regression analysis showed a significant influence of the pre-treatment depression score in both groups and a significant influence of clients’ ratings of the working alliance in the face-to-face group."

It is not clear what the dependent variable was.

11. p4. "In view of these findings, the therapeutic alliance has traditionally been seen as a key element adding to the treatment success of face-to-face psychotherapy."

Please provide a reference to this statement.

12. p5. "Therefore, it is no surprise that critics of internet-based interventions are concerned that therapeutic contact is reduced, not only time-wise, but also to purely text-based, computer-mediated communication."

Please provide a reference to such criticism, and if such a reference does not exists, please rephrase (e.g. "...internet-based interventions could be criticized...")

13. p6. “Further, the composite scores of both the therapists’ and the clients’ ratings of the therapeutic alliance late in treatment were moderately correlated with treatment outcome. In contrast, the correlations between ratings of the therapeutic alliance early in treatment were low to modest [24].”

The latter sentence doesn’t make sense. It sounds like a low correlation between client’s and therapist’s ratings early in treatment is contrasted with moderate correlation between these measures and treatment outcome later in treatment.

15. p6. “Beside these studies of internet-supported therapeutic interventions with intense therapist support, Klein and colleagues [25] and Kiropoulos and
colleagues [26] have reported positive results on the therapeutic alliance in therapist-assisted internet programs."

It’s difficult to understand what the author’s mean with "intense therapist support". The previous paragraph has not explicitly stated that the studies described in there used "intense therapist support". Perhaps the sentence "Patients in the internet-based group had significantly less therapist contact than those in the face-to-face group," is supposed to explain this difference, but you would always expect face-to-face therapy to have more intensive therapist support than internet-based treatment. Please clarify how the referenced studies differ in intensity of therapist support.

16. p6. "These authors also reported low therapist time across the 10-week intervention (194.5 min)"

194.5 minutes is almost double the amount of time that Palmqvist et al. found to be sufficient for acceptable treatment effect (also referenced in this manuscript on p3). Again, it is not clear why this is considered to be less intensive therapist support.

17. p7. “To our knowledge, this is the first randomized controlled trial to compare the therapeutic alliance between patient and therapist in the two settings in an experimental design.”

To the reader it would seem that the study by Kiropoulos et al (ref 26) did exactly this. Please clarify.

18. p8. Demographic characteristics.

Move all in-text information about demographics to Table 1 (add a “Both groups” column).

19. p8. "Participants who indicated that they met and were comfortable with the requirements entered an online screening procedure, data from which were later used as pretest measures"

Please give the average, min, and max if the number of days between screening and treatment start – so the reader can judge how "old" the pre-treatment assessments are.

20. p9. "Participants were then randomly assigned to one of the two conditions"

Were all participants randomized simultaneously?

21. p10. Exclusion criteria

Figure 1 shows that applicants were excluded because of “Other psychiatric disorder”, “Bipolar disorder”, and “Low symptom severity”, but these criteria are not given in the text and there is no explanation of how they were assessed.

22. p11. “A life-review module was added to the standard CBT treatment manual"
Please repeat the reference (32) from the introduction after this sentence. Since the reference is in German it would be valuable for the reader to get a brief explanation of what a “life-review module” is.

23. p12. “They were also set additional homework assignments taking approximately one hour per week.”

One hour of homework assignment per week seems very low in a treatment that (traditionally) includes daily behavioral activation. Please clarify.

24. p12. “The treatment protocol contained instructions on the writing assignments, which were personalized by the therapist for each patient.”

Were these “writing assignments” simply homework reports or were they treatment interventions (like expressive writing)? Please clarify.

25. p12. “Multiple regression analyses were then used to calculate the effect of the working alliance on therapy outcome. Further, t tests were used to compare the therapeutic alliance in the two intervention groups. In addition, Pearson’s correlations were calculated to examine the relationship between the working alliance and therapy outcome.”

This is not the order in which the results are presented.

26. Please report the post-treatment scores on BDI for both groups and associated test statistics and also the amount of therapist time spent per participant in each group. If these data are part of a different report then this should be stated and referenced to if published.

27. p13. “With the exception of the WAI-T tasks score, which was significantly higher in the online condition, t(47) = 2.360; p < 0.05, Cohen’s d = 0.66”

These test-statistics are already given in Table 2 and need not be repeated. Adding the df=47 of the t-tests as a footnote of Table 2 would be helpful. Please also report Cohen’s d for all subscales and composite scores.

28. In Tables 3 and 4 it is unnecessary to report the correlations within the subscales of WAI-C and WAI-T (and they are never referred to in the manuscript). The reader is interested in the correlations between the 8 subscales and two outcome scores on the BDI in the two groups. Combining the tables into one and only reporting these 32 correlations in a 4x8 table would make the relevant correlations much easier to read and interpret.

29. p15. “In the same vein, [21] reported ratings of the therapeutic alliance in an online setting to be similar to existing data from a face-to-face group.”

Please add the name of the first author before ref 21.

30. p16. “for the composite score and the tasks subscale in the online group and for the composite score and the tasks and goals subscales in the face-to-face group.”
Typos: “sore”.

31. p17. “Working alliance ratings were only associated with treatment outcome in the face-to-face group, indicating that this treatment form may be the better predictor of the working alliance.”

The last part of the sentence is incomprehensible. I suggest removing.

- DISCRETIONARY REVISIONS

32. p6. “The program combines standardized instructions and information with e-mail contact to a therapist.”

Perhaps “with” is a better preposition than “to” in this sentence.

33. p8. “No significant differences in baseline characteristics were noted between groups, apart from gender—there were significantly more female participants in the online group.”

This reviewer does not believe in using significance tests as a measure of pre-treatment differences (Senn. Testing for baseline balance in clinical trials. Statistics in Medicine (1994) vol. 13 (17) pp. 1715-26). However, this is common practice. If the Editor agrees, I recommend removing these tests.

34. p9. “Demographic data were assessed at baseline only”

This information is superfluous.

35. p10. “age younger than 18 years”

A person is younger than 18 years while he as an age below 18 years.

36. p11. “Both treatment conditions were based on an evidence-based short-term CBT treatment manual for depression [30].”

Perhaps it could be stated here that the treatments were equal in length (8 weeks).

37. p12. “They were also set additional homework assignments taking approximately one hour per week.”

Perhaps rephrase to “They were also given approximately one hour of homework assignments per week”. (no need to use “additional” since other homework assignments were not given).

38. p12. “Patients were set two weekly 45-minute writing assignments over an eight-week period.”

Perhaps use “given” instead of “set”.

**Level of interest:** An article whose findings are important to those with closely related research interests
Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests