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.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Long-term effectiveness and outcome predictors of therapist-guided internet-based cognitive behaviour therapy for social anxiety disorder in routine psychiatric care
AUTHORS	El Alaoui, Samir; Hedman, Erik; Ljótsson, Brjánn; Lindefors, Nils

VERSION 1 - REVIEW

REVIEWER	Nickolai Titov Macquarie University, Australia
REVIEW RETURNED	29-Mar-2015

GENERAL COMMENTS	This paper reports the long-term effectiveness of internet-delivered CBT (ICBT) for social anxiety disorder in a cohort of consecutively recruited patients, treated as part of routine psychiatric care. Using a large sample (N=446) the authors administered measures at several times points, up to 4 years post-treatment to determine the stability of gains following ICBT for SAD and to explore characteristics of responders and non-responders. The results indicate that gains were sustained and indeed appear to improve over time, with rapid improvements in those with higher illness severity, but worse response from those with a family history of social anxiety disorder. These results replicate and extend a growing body of work indicating the stability and magnitude of gains following ICBT The strengths of this paper include the large sample, high levels of questionnaire completion, a naturalistic clinical setting, and reporting of adverse effects of treatment. The paper is also well written. The main limitation is the absence of a control group. The results of this paper are likely to be of considerable interest to planners and funders of mental health services who are interested in new models of service provision which are clinically and cost-effective, and accessible to consumers. I hope the following comments assist the authors in improving an already strong paper. - It would be helpful if the authors provided more detail about the baseline diagnostic procedures to help the reader better characterise the sample. Did these patients have SAD as a principal diagnosis? What was the extent of comorbidity with other anxiety and depressive disorders? Were suicidal patients excluded or allowed to participate, and if the latter, how were they managed? Were substance abuse or psychotic syndromes the only diagnostic exclusion criteria, and does the latter include Bipolar Disorder? - Many readers may not have used the LSAR-SR or MADRS-S. Perhaps the authors could describe the correlations between these measures and other common measures of SAD and depression,

	such as the SIAS/SPS and BDI-II? In addition, information about the cut-offs for these measures would also help characterise this sample. - The authors report the proportion of patients who received psychological or pharmacological treatment for SAD post ICBT (Table 1). Although the proportion receiving such treatments is low, it would be helpful if the authors reported whether such treatment was associated with outcomes on the main measures. - I regret that I am not clear about the mean duration of long-term follow-up, which appears to range between 1 and 4 years post-treatment. Perhaps the authors could describe the frequency of long-term follow-up using quintiles or some similar indicator, and also provide analyses to indicate whether those who completed the long-term follow up at a short (e.g., 1 year) duration differed from those who completed the follow up after a longer period (e.g., 4 year)? - Finally, although I appreciate the limited word count, I encourage the authors to briefly discuss the reasons why family history may be an important factor, as this speaks to the potential need to trial alternative treatment models, including possibly engaging more with family members during treatment, where this is relevant.

REVIEWER	Louise Mewton University of New South Wales
REVIEW RETURNED	08-Apr-2015

GENERAL COMMENTS	This is an interesting study investigating the longer term effects of ICBT on the symptoms of social anxiety disorder. The study is conducted within a naturalistic setting, and whilst this presents some limitations, effeciveness studies such as this are an important addition to the literature. The methods are sound and well described, given their complexity. My main concern is how the longer term follow up "time point" was treated. The authors treat it as a single time point but it ranges over a period of several years. I believe that this can be accomodated by the types of modelling used (I know it can in mixed modelling approaches, anyway) but a more detailed discussion of how this was dealt with is missing from the current manuscript. I have some other small comments that the authors might want to address: Abstract Last sentence - I'm not sure you can say there is "strong" evidence presented here - maybe consider toning this down? Introduction First paragraph - maybe also worth mentioning that because of the nature of SAD, ICBT may be particularly advantageous. The authors state we need more details on the longer term effectiveness of ICBT for SAD in naturalistic settings. What do we

	know about the shorter term effects of ICBT for SAD in naturalistic studies? Is there any research looking at this? Methods What were the qualifications of the therapists? Were they all psychologists? Was there some testing of the optimal covariance structure? Also it might help to explicitly state that you are talking about the "covariance structure" rather than just "structure". How did the models fit in terms of other indices? You talk about one model being superior to another in terms of -2LL (relative fit) but what about overall goodness of fit of the models (absolute fit)? Discussion First paragraph, last sentence - on what outcomes did patients continue to improve? This sentence needs to be more explicit. I'm also wary of saying that improvements continued up to four years after treatment beacuse of the loose way in which that longer term follow up is defined. I can't think of a better way to frame this, but it doesn't seem accurate to me given that the longer term follow up occasion can't really be termed 4 year follow up. This needs to be dealt with throughout the manuscript. Limitations - the lack of a control group is a serious limitation and could possibly be dealt with further in the limitations. I'm not sure I understand the sentence "since this was not a comparative study, it might be unclear as to how comparable this sample is to a typical SAD sample". Given that this is a naturalistic study, wouldn't you expect this sample to be well representative? Wouldn't that be the argument for conducting a study in a naturalistic setting? Representativeness data could maybe be framed in this way instead?

VERSION 1 – AUTHOR RESPONSE

REVIEWER 1:
COMMENT #1: It would be helpful if the authors provided more detail about the baseline diagnostic procedures to help the reader better characterise the sample. Did these patients have SAD as a principal diagnosis? What was the extent of comorbidity with other anxiety and depressive disorders? Were suicidal patients excluded or allowed to participate, and if the latter, how were they managed? Were substance abuse or psychotic syndromes the only diagnostic exclusion criteria, and does the latter include Bipolar Disorder?

RESPONSE TO COMMENT #1: We agree with the reviewer and have now expanded the description of the diagnostic procedures in the manuscript. For example, we have included a more detailed description of the exclusion criteria applied by the physicians during the pre-treatment diagnostic assessment. Page 6, first paragraph:

“In general, exclusion from ICBT was based on the following criteria: a) patients with severe depression (clinician rated MADRS ≥ 35) and/or moderate to high risk of suicide where monitoring is required, b) patients with low motivation, severe apathy or difficulty concentrating, c) patients with

psychosis, d) patients with untreated drug or alcohol problems that may pose barriers to treatment, e) reading and/or writing difficulties, including language difficulties, and f) patients residing outside the city county."

Also, we have added that included patients had SAD as their principal diagnosis (page 6, first paragraph):
"This was a longitudinal study investigating adult patients (N = 446) who had a principal diagnosis of SAD and had been treated for SAD as part of routine care."
Finally, we have expanded the description of the diagnostic procedures and the intervention (both in terms of administration and content). See page 6 paragraph 2 (section "Procedure and design"). Table 1 also presents data on the level of comorbid depressive symptoms (presented as mean MADRS scores, both clinician-administered and self-assessed).

COMMENT #2: Many readers may not have used the LSAR-SR or MADRS-S. Perhaps the authors could describe the correlations between these measures and other common measures of SAD and depression, such as the SIAS/SPS and BDI-II? In addition, information about the cut-offs for these measures would also help characterise this sample.

RESPONSE TO COMMENT #2: We thank the reviewer for this suggestion, and believe the inclusion of such comparison to other common measures of SAD would improve the manuscript further, as well as providing established cut-off scores to facilitate interpreting the results. Consequently, we have addressed this in the text, referring to previous research validating the use of LSAS in relation to SIAS and SPS (page 8, second paragraph):
"The LSAS-SR has showed strong convergent validity with other common self-report measures of social anxiety such as the Social Interaction Anxiety Scale (SIAS) and the Social Phobia Scale (SPS) [24] with correlation coefficients of 0.71 and 0.61 respectively [25]. A cut-off score of 30 or less on the LSAS-SR has been suggested as indicative of probable absence from social anxiety disorder [26]."
Also, in regard to measuring symptoms of depression and the MADRS-S scale, we have included the following paragraph (page 9, first paragraph):
"A cut-off score of 13 or higher has been used to distinguish depressed from non-depressed patients [31]. The MADRS-S has been shown to have a high correlation with the commonly used Beck Depression Inventory (BDI) (r = .87) [32]."

COMMENT #3: The authors report the proportion of patients who received psychological or pharmacological treatment for SAD post ICBT (Table 1). Although the proportion receiving such treatments is low, it would be helpful if the authors reported whether such treatment was associated with outcomes on the main measures.

RESPONSE TO COMMENT #3: This is an interesting addition to the analyses. We have investigated the proportions of patients in remission at the time of long-term follow up, and we observed a difference in the rate of patients in remission depending on the use of psychotropic medication. We have therefore included a new paragraph in the Results section of the manuscript (page 14, last paragraph):
"At post-treatment, 20.7% of patients who provided data had achieved remission from SAD (LSAS-SR score ≤ 30), 34.1% at six-month follow-up and 35.2% at the time of long-term follow-up. The rate of achieved remission at long-term follow-up was found to be independent of when the follow-up assessment was completed; $\chi 2$ (1, N = 141) = 1.20, p = .27. However, there was a significant difference in the proportion of patients in remission at follow-up depending on the use of psychotropic medication; $\chi 2$ (2, N = 267) = 12.41, p < .01, indicating a moderate association (Cramer's V = 0.22, p < .02). Among those who reported no use of medication either during treatment or at the time of long-term follow up, 41.8% (79 patients out of 189) had achieved remission, as opposed to only 15.7% (8 out of 51) among those who reported using medication during treatment and at the time of long-term

follow up."

In addition, we have commented on the interpretation of these data in the discussion (page 21, first paragraph):
"Finally, although we observed a difference in proportions of patients in remission from SAD at long-term follow-up depending on the use of medication during and after treatment (41.8% of patients with no medication during or after ICBT had achieved remission at follow-up as opposed to only 15.7% among those who reported using medication during treatment and at the time of follow-up), these subgroups of patients may be difficult to compare because we do not how they differ in terms of illness history."

COMMENT #4: I regret that I am not clear about the mean duration of long-term follow-up, which appears to range between 1 and 4 years post-treatment. Perhaps the authors could describe the frequency of long-term follow-up using quintiles or some similar indicator, and also provide analyses to indicate whether those who completed the long-term follow up at a short (e.g., 1 year) duration differed from those who completed the follow up after a longer period (e.g., 4 year)?

RESPONSE TO COMMENT #4: We have included information regarding the distribution of the duration between baseline measurement and long-term follow-up assessment in Table 1, including mean, minimum and maximum number of days to follow-up, and also percentiles.
To analyse whether those who completed the long-term follow up at a short duration differed from those who completed the follow up after a longer period we have now made additional analyses. The method is described on page 11, first paragraph:
"However, due to the relatively wide distribution of duration to long-term follow-up assessment (i.e. between 1-4 years), we tested whether those who completed the long-term follow up at a short duration differed from those who completed the follow up after a longer period in terms of the number of patients in remission at the time of follow-up. Consequently, patients were categorized into two groups ("short" or "long" duration), operationalised as either 1 SD below or above the mean duration. The data were analysed with a Chi-square design."

Also, the result of this analysis is presented in the Results section (page 11, second paragraph):

"The rate of achieved remission (LSAS score ≤ 30) at the time of long-term follow-up was found to be independent of when the follow-up assessment was completed; $\chi 2$ (1, N = 141) = 1.20, p = .27."

COMMENT #5: Finally, although I appreciate the limited word count, I encourage the authors to briefly discuss the reasons why family history may be an important factor, as this speaks to the potential need to trial alternative treatment models, including possibly engaging more with family members during treatment, where this is relevant.

RESPONSE TO COMMENT #5: We thank the reviewer for this suggestion as it fills a gap on this topic in the discussion. We have therefore extended the discussion of the role of family history (page 20, last paragraph):
"We also identified family history of social anxiety as a predictor of poorer outcome. This relationship has been observed in both pharmacological treatment for SAD [41] and in an RCT on ICBT for SAD [7]. Although family history has been shown to be associated with a two- to three-fold risk for developing SAD [42] it is unclear which mediating mechanisms of genetic and environmental factors may be responsible for the variance in treatment effects. Still, the identification of family history as a long-term outcome predictor may have implications for the need to monitor this patient group more carefully due to an increased risk of lower response rate and possibly also a higher level of therapist guidance during exposure training."

REVIEWER 2:

COMMENT #6: My main concern is how the longer term follow up "time point" was treated. The authors treat it as a single time point but it ranges over a period of several years. I believe that this can be accomodated by the types of modelling used (I know it can in mixed modelling approaches, anyway) but a more detailed discussion of how this was dealt with is missing from the current manuscript.

RESPONSE TO COMMENT #6: We agree with the reviewer that the manuscript would benefit from discussing this methodological choice in more detail. Therefore, we have addressed this issue in the following paragraph (page 22, first paragraph):
“Third, a note on the statistical approach of handling the relatively wide time-frame of the long-term follow-up “time point” is warranted, since these de facto ranges over a period of several years. As described in the methods section, there were mainly two alternative approaches to model how the outcome variables evolve over time. We could either treat time as discrete or continuous. Within a discrete framework, time values are defined as distinct time periods or measurement occasions; in the present study, pre-treatment, post-treatment, 6-month-follow-up and long-term follow-up measurements were categorised into four separate assessment occasions irrespective of the actual duration since baseline. The alternative approach, treating time as continuous, would have meant that the time variable would range over the complete duration of the study (i.e. 1492 days). We acknowledge that there are benefits with both approaches. For example, it could be argued that developing a model using a continuous time variable might be more exact and mathematically tractable. On the other hand, it might be easier to understand and interpret a model that corresponds to how the theoretical hypothesis to be tested is expressed (e.g. to understand the degree of symptomatic improvement at long-term follow-up). Ultimately, we chose a discrete time framework in order to facilitate the development and interpretation of the multilevel model.”

COMMENT #7: Abstract. Last sentence - I'm not sure you can say there is "strong" evidence presented here - maybe consider toning this down?

RESPONSE TO COMMENT #7: We have removed the word “strong” from the abstract, and simply stated that “These findings provide evidence for the long-term effectiveness of ICBT for SAD in routine clinical practice, even for more severe cases.”

COMMENT #8: Introduction. First paragraph - maybe also worth mentioning that because of the nature of SAD, ICBT may be particularly advantageous.

RESPONSE TO COMMENT #8: We thank the reviewer for this suggestion, and agree that this is an important aspect of the benefits of internet-based intervention. We have now included a brief paragraph on this (page 4, second paragraph):
“To increase accessibility, therapist-guided internet-based CBT (ICBT) has been shown to be a promising method of delivering evidence-based psychological treatments. This format of delivering CBT may also be particularly advantageous in the treatment of SAD because of the nature of the disorder, especially for patients presenting with greater illness severity who might fear face-to-face interactions with a therapist. [5 6] ICBT may in such cases be preferable for those who might otherwise avoid seeking help.”

COMMENT #9: The authors state we need more details on the longer term effectiveness of ICBT for SAD in naturalistic settings. What do we know about the shorter term effects of ICBT for SAD in naturalistic studies? Is there any research looking at this?

RESPONSE TO COMMENT #9: We agree that it would strengthen the manuscript by including

references to such research. We have therefore included the three references: (page 4, last paragraph):
“Although there is evidence that ICBT may be equally effective as group CBT for SAD also under clinically representative conditions [12 13], more knowledge on the long-term effectiveness of ICBT for SAD in naturalistic settings […]”

COMMENT #10: What were the qualifications of the therapists? Were they all psychologists?

RESPONSE TO COMMENT #10: We have included a phrase clarifying the qualifications of the therapists (page 7, first paragraph):
“The duration of the active psychotherapeutic intervention was twelve weeks, during which patients were guided online by licensed psychologists who had training in CBT.”

COMMENT #11: Was there some testing of the optimal covariance structure? Also it might help to explicitly state that you are talking about the "covariance structure" rather than just "structure".

RESPONSE TO COMMENT #11: We have now clarified in the methods section that it is a first-order autoregressive covariance structure (page 10, second paragraph). There were no testing of alternative covariance structures; rather, the basis for choosing the autoregressive covariance structure for this study was that this is generally the recommended covariance structure for analysing longitudinal data with repeated measurements where correlations between measurements tend to weaken as time between them increases.

COMMENT #12: How did the models fit in terms of other indices? You talk about one model being superior to another in terms of -2LL (relative fit) but what about overall goodness of fit of the models (absolute fit)?

RESPONSE TO COMMENT #12: Since comparing the two approaches to modelling the time variable, in this case based on -2LL, may not be the optimal due to difficulties in interpreting model fit, we argue that the choice of using a multilevel model where assessments are grouped is based on facilitating interpretation of the model. We have now revised the paragraph accordingly (page 10, last paragraph):
“Two approaches to treating time were considered; (i) either using a two-piece model with discrete time, grouping measurements into four measurement occasions (i.e. pre-treatment, post-treatment, 6-month-follow-up and long-term follow-up) or (ii) a model with a continuous time variable using the number of days to follow-up. Since grouping measurements facilitates the interpretation of the model, this approach was used.”

COMMENT #13: Discussion. First paragraph, last sentence - on what outcomes did patients continue to improve? This sentence needs to be more explicit.

RESPONSE TO COMMENT #13: We have clarified on which outcomes patients improved (page 19, last paragraph):
“In line with these reports, we have observed similar results in a naturalistic setting; a large treatment effect achieved at post-treatment and that patients, overall, continued to improve in social anxiety symptoms between 1-4 years after treatment. In addition, continued improvements in health-related quality of life were observed during the follow-up period.”
COMMENT #14:

I'm also wary of saying that improvements continued up to four years after treatment beacuse of the loose way in which that longer term follow up is defined. I can't think of a better way to frame this, but

it doesn't seem accurate to me given that the longer term follow up occasion can't really be termed 4 year follow up. This needs to be dealt with throughout the manuscript.

RESPONSE TO COMMENT #14: We have made the following changes throughout the manuscript:
Abstract: “for up to four years” have been changed to “long-term”.
Discussion (page 19, last paragraph): “for up to four years” have been changed to “between 1-4 years”.

COMMENT #15: Limitations - the lack of a control group is a serious limitation and could possibly be dealt with further in the limitations.

RESPONSE TO COMMENT #15: We agree with the reviewer and have now commented further on the issue of having a control group (page 21, second paragraph):
“As this was a cohort study conducted within a naturalistic setting, this was an unavoidable limitation. However, available empirical data suggest a low probability of spontaneous recovery from SAD. For example, an eight-year study following the long-term course of SAD reported that only 13-14% had achieved remission after one year [Yonkers, 2001], which can be compared to 34.1% at six-month follow-up in the present study. Also, ICBT for SAD has demonstrated superiority to waiting list control [Furmark, 2009]. Taken together, it would seem unlikely that the observed improvements among patients in the present study might be attributed to the effect of time alone.”

COMMENT #16: I'm not sure I understand the sentence "since this was not a comparative study, it might be unclear as to how comparable this sample is to a typical SAD sample". Given that this is a naturalistic study, wouldn't you expect this sample to be well representative? Wouldn't that be the argument for conducting a study in a naturalistic setting? Representativeness data could maybe be framed in this way instead?

RESPONSE TO COMMENT #16: We agree, and have decided to remove this sentence.

VERSION 2 – REVIEW

REVIEWER	Nickolai Titov Macquarie University, Australia
REVIEW RETURNED	01-May-2015

GENERAL COMMENTS	Thank you for fully addressing my questions. I believe the authors have strengthened this manuscript.

REVIEWER	Louise Mewton University of New South Wales, Australia
REVIEW RETURNED	28-Apr-2015

GENERAL COMMENTS	The authors have satisfactorily addressed all of my comments and concerns.