Author's response to reviews

Title: Psychosocial functioning in patients with treatment-resistant depression after group cognitive behavioral therapy

Authors:

Miki Matsunaga (mikim112@hotmail.com)
Yasumasa Okamoto (oy@hiroshima-u.ac.jp)
Shin-ichi Suzuki (ssuzuki@waseda.jp)
Akiko Kinoshita (akino0707@hiroshima-u.ac.jp)
Shinpei Yoshimura (xanth@hiroshima-u.ac.jp)
Atsuo Yoshino (a-yoshi0011@hiroshima-u.ac.jp)
Yoshihiko Kunisato (ykunisato@hiroshima-u.ac.jp)
Shigeto Yamawaki (yamawaki@hiroshima-u.ac.jp)

Version: 2 Date: 3 December 2009

Author's response to reviews: see over
We are grateful to the editor for this comment and useful suggestions that have helped to improve our paper. Thank you for your interest and patience.

**Editor’s comment**

Ethics - Experimental research that is reported in the manuscript must have been performed with the approval of an appropriate ethics committee. Research carried out on humans must be in compliance with the Helsinki Declaration ([http://www.wma.net/e/policy/b3.htm](http://www.wma.net/e/policy/b3.htm)), and any experimental research on animals must follow internationally recognized guidelines. A statement to this effect must appear in the Methods section of the manuscript, including the name of the body which gave approval, with a reference number where appropriate.

**Response**

We are in complete agreement with the Editor’s comment. We have added the following sentences to the methods section.

“The study protocol was approved by the Ethics Committee of the Hiroshima University Graduate School of Medical Sciences (Reference number: 628). Written informed consent was obtained from all patients.” (Methods, Page 7, L14–16).
Response to Reviewers

We are grateful to Reviewer 1 for their comments and useful suggestions that have helped to improve our paper. These comments gave us a better perspective on our work. Thank you for your interest and patience.

Reviewer 1’s comment

Comment #1
I have some issue with the overall reasoning behind this study. The background section lacks focus (this may be due to the writing style of the manuscript though). In the first paragraph the authors define TRD. In the second, they review articles suggesting a relationship between depressive symptom severity and poor psychosocial functioning. They conclude saying ?Such results suggest that combining a psychosocial approach with drug therapy could be important in the management of TRD?. Does this imply that CBT may improve TRD by reducing poor psychosocial functioning, regardless of its effect on depressive symptoms or, alternatively, that CBT can improve TRD by reducing depressive symptoms and as a consequence, reduce psychosocial dysfunction? The third paragraph consists in a review of the efficacy of CBT in TRD (the Fava article being more pertinent here, although it is an open trial. The Moore article is, from reading the abstract, a pilot study of a small number of patients receiving CT sequential to their antidepressant treatment. In this light, perhaps the authors should also consider the Thase study, which is similar but 10 years later, Am J Psychiatry, 2007, 164:739). Finally, the fourth paragraph is on chronic, but not necessarily TR depression. In the fifth, they state the goals of this study, elaborating only on social functioning and not at all on depressive symptomatology, although they collect such data throughout this study.

If the overall philosophy behind this study is?.

A- some patients are treatment refractory, defined by. B- These patients, because of the sustained depressive symptoms, have reduced psychosocial functioning?. C- Some studies show that sequentially administering CBT to antidepressant therapy in TRD may be of benefit in reducing depressive symptoms? D- To date, few studies have examined the effect of the sequential administration of CBT in TRD on both depressive symptoms
and psychosocial functioning? E- We intend to do just that...

**Response**

We are in agreement with this comment. We have revised the background section as follows;

“About 20 to 40% of depressed patients do not respond satisfactorily to treatment with only antidepressant medications [1-3]. These patients are defined as having treatment-resistant depression (TRD) when they fail to respond to at least two adequate trials of antidepressant medications from different classes [3-4].

TRD patients frequently have impaired social functioning because of sustained depressive symptoms [5]. The impairments affect marriages, cause interpersonal problems, and difficulty in work environments [6]. Continued depression and psychosocial impairment may induce social isolation, loneliness, and interpersonal difficulties that also interfere with the improvement of depressive symptoms [7]. TRD patients who received treatment as usual (TAU) with only medication continued to have functional disability [8].

Cognitive behavioral therapy (CBT) has been shown to be effective in the treatment of major depressive disorder. DeRubeis et al. [9] suggested that CBT can be as effective as medication for the initial treatment of moderate to severe major depression. Other studies have shown that adding CBT to medication for TRD may be beneficial in reducing depressive symptoms. For example, Thase et al. [10] compared the effectiveness of CBT and medication as second-step strategies for the patients with an unsatisfactory response to an initial trial of medication (citalopram). They reported that those patients who received CBT (either alone or in combination with citalopram) had similar response and remission rates to those who received only medication.

However, these studies have mainly investigated the short-term effects of CBT on depressive symptoms. Several studies investigated whether CBT improved social functioning in individuals with chronic depression. Scott et al. [11] assessed psychological and social functioning, and compared medication management alone to CBT plus medication management. They reported that patients receiving cognitive therapy plus medication management had better psychosocial functioning than those who receiving medication management alone. Hirschfeld et al. [12] studied patients who underwent a cognitive behavioral analysis system of psychotherapy (CBASP) as CBT for chronic depression, and compared the efficacy of (1) CBASP, (2) nefazodone,
or (3) CBASP combined with nefazodone for improving psychosocial functioning. They reported that the combined therapy had greater effects than either monotherapy. These studies have been limited to consideration of the short-term effectiveness of CBT for social functioning, and they did not necessarily meet criteria for treatment resistant.

Impaired social functioning may be a contributing cause as well as an effect of depression in individuals with TRD. Studies have not examined the effectiveness of CBT, along with medication, for patients with TRD with regard to both depressive symptoms and psychosocial functioning, particularly with longer-term follow-up. Therefore, we examined the short-term effectiveness of combined therapy (group-CBT and medication) on not only the depressive symptoms but the social functioning of TRD patients. Moreover, we studied these long-term effects (12 months) after the termination of group-CBT. We addressed the following questions:

1. Is the combined therapy (group-CBT and medication) effective in improving not only the depressive symptoms but the social functioning of patients with treatment-resistant depression?

2. Are these effects of the combined treatment for TRD maintained 12 months after termination of the group-CBT?

Comment #2

One of the great advantages of online journals is the lack (within reason) of a space requirement. Word counts worries are part of the past and authors can concentrate on fully explaining their data. The authors of this study should take maximum advantage of this possibility. I will give a few examples of what I mean in several of the manuscript’s sections.

First, the abstract does not fully reflect the content of this manuscript. In the methods subsection of the abstract the authors mention all rating instruments used. In the results section, data for only 2 of the 5 rating scales is given.

Second, in the methods section, care should be taken to fully document treatment refractoriness. Although I find it great that they are using a well-defined staging system for TRD, they still have to explain how they got there. They might include a paragraph (or table) stating the antidepressants used (at least the most frequently used ones), the dosages they consider to represent an adequate trial of each and, the duration of treatment. For instance, citalopram was taken at an average of 40 mg per day (range 30
to 60 mg, N=?), for a minimum of 8 weeks.

Third, again in the methods section, were any modifications of the current medication allowed in this study. Were there any concomitant drugs allowed in this study and / or did they vary in dose?

Fourth, in the results section, ranges should be given for the rating scale averages. For instance, for the HRSD and GAF scales, we should know what the ranges are. Given that the averages are, at best, modest, a distribution would be appropriate. For example, of the 38 patients, N=7 had HRSD between 9 and 14, N=20 had scores of between 15 and 18 and N=11 had scores 18 and above?..

Fifth, also in the results section, more detail should also be given for the HRDS end points. In table 3, it is stated that 55% of patients were in remission. However, it is much easier to be in remission if patients started out the trial with a HRSD of, say, 9. The drop would be a scant 2 rating points. So in essence, patients could be in remission and not fulfill the criteria for 50% improvement here.

Sixth, again in the results section, the ever-dwindling number of patients on follow-up is simply not explained. We go from 38 at end point, to 28 without explanation (although the 28 to 20 drop is explained).

There are several other examples although at this point I will conclude this section.

Response
Referring to these comments, we have modified the manuscript.

First, we have revised in the abstract under results and conclusions as follows;

Results
Thirty-eight patients completed treatment; five dropped out. For the patients who completed treatment, post-treatment scores on the GAF and SF-36 were significantly higher than baseline scores. Scores on the HRSD, DAS, and ATQ-R were significantly lower after the treatment. Thus patients improved on all measurements of psychosocial functioning and mood symptoms. Twenty patients participated in the 12-month follow-up. Their improvements for psychosocial functioning, depressive symptoms, and dysfunctional cognitions were sustained at 12 months following the completion of
Conclusions
These findings suggest a positive effect that the addition of cognitive behavioral group therapy to medication improved not only depressive symptoms but social functioning of patients with TRD.

Second, we have mentioned the past antidepressants treatment in the methods section as follows:

“All patients had previously taken two different classes of antidepressant medications for a minimum of 8 weeks without remission of symptoms: clomipramine (n=13, average 146 mg per day), paroxetine (n=13, average 29 mg per day), milnacipran (n=13, average 103 mg per day), or others.” (Methods; Page 7, L17-20)

Third, we also described the current medication of the patients in the methods section.

“During the group-CBT treatment, the drug type and dose was maintained, except for one patient whose medication was changed when his condition deteriorated rapidly. In addition, the patients did not take any other forms of treatment except medication for the 12 months after the group-CBT.” (Methods; Page 7, L20-24)

Fourth, we have provided the ranges of the rating scale averages, especially HRSD and GAF scores in the results section.

“The baseline scores on the Hamilton Rating Scale for Depression (HRSD) indicated mild to moderate levels of depression among the patients (Mean = 14.4, SD = 4.4). Seven patients (16%) had HRSD scores between 8 and 10, 13 (30%) had scores between 11 and 14, 13 (30%) had scores of between 15 and 18, and 10 (23%) had scores between 18 and 27. The baseline GAF scores indicated a poor level of social functioning. 31 patients (72%) had scores between 40 and 60, and the other 12 (28%) had scores between 61 and 70.” (Results; page 12, L6-12)

“For both the ITT and Completer analyses, the GAF scores increased significantly (ITT: F(1, 42) =36.58, p < .001, partial$\eta^2$ = 0.47; Completer: F(1, 37) = 41.19, p < .001, partial$\eta^2$ = 0.53 ). For the ITT sample, the number of patients who were rated as showing mild functional impairment (defined as GAF scores over 60) improved from 12 (28%) at baseline to 30 (70%) at post-treatment; 7 (16%) of these patients were rated as having minimal impairment (GAF > 70). As expected, the Completer sample
comprised those 30 patients who had a post-treatment GAF score over 60 (79%), and the 7 patients (18.4%) who were rated as having minimal impairment (GAF > 70).” (Results; Page 12, L23-Page 13, L6)

“Among the Completers, 21 (55%) of the patients had scores of 7 or less on the HRSD at post-treatment. 12 had scores between 8 and 14, and 5 had scores between 15 and 21.” (Results; Page 13, L22-24)

Fifth, we agree this comment that added the remission and response rates for ITT samples in the results section and Table 3.

“Table 3 shows the remission and response rates at post-treatment for the ITT and Completer sample analyses. Twenty-one participants (ITT: 49%; Completer: 55%) met criteria for remission (HRSD score of 7 or less), and 18 participants (ITT: 42%; Completer: 47%) showed at least a 50% reduction of their scores on the HRSD from the pre-treatment score. The number of participants who met criteria both for remission and 50% reduction of were 17 (ITT: 40%; Completer: 45%).” (Results; Page 14, L1-6)

Table 3- Remission and response rates for patients who completed group-CBT
Table 3 summarizes the remission and response rates at post-treatment for ITT and Completer samples.

Six, we should have explained that the patients of the follow-up study was much smaller than these of the baseline. Of the 38 patients who completed the group-CBT, those who ended a one-year follow-up after group-CBT was 28 patients. The remaining 10 persons had not passed for one year after group-CBT. They are ongoing of the follow-up study. We have modified the following sentence in the result section.
“Of the 38 patients who completed the group-CBT, a total of 28 patients had completed the treatment more than one year previously at the time of our follow-up. The remaining 10 persons had completed the group-CBT and were less than one year previously at the time of our follow-up. Twenty of the 28 patients (71.4%) completed all measurements one year after finishing the group-CBT; the other 8 refused to participate in the follow-up (4 refused the participation in the follow-up study, and 4 refused accesses to contact for follow-up).” (Results; Page 15, L7-13)

Comment #3
What was the justification for using such a low cut off point for the HRSD. I think there would be general agreement that patients with a HRSD score of 8 to 12 would be considered, at best, very mildly depressed. If patients entered into this study had a HRSD of 8 or 10 then how do the authors should explain this with regards their own entrance criteria (of stage II or greater depression)? Also, how do they explain this with regards their triage system (the SCID mentioned in the Methods section)?

Response
We agree that we should have explained this issue. In this study, we intended to assess whether patients reached remission or not using a HRSD score of 7 or less after group-CBT. Therefore, we included the patients with a HRSD of 8 or greater as the inclusion criteria.

Actually, in the present study, the baseline scores on the Hamilton Rating Scale for Depression (HRSD) indicated mild to moderate levels of depression among the patients (Mean = 14.4, SD = 4.4). Seven patients (16%) had HRSD scores between 8 and 10, 13 (30%) had scores between 11 and 14, 13 (30%) had scores of between 15 and 18, and 10 (23%) had scores between 18 and 27. The baseline GAF scores indicated a poor level of social functioning. 31 patients (72%) had scores between 40 and 60, and the other 12 (28%) had scores between 61 and 70.

Furthermore, we recruited the outpatients who could participate in the group-CBT for 12 weeks. However, we did not describe this entrance criterion in our original paper. Therefore, we have revised these descriptions of the entrance criteria in the methods section, and added the following sentences in the results and discussion section.

“Criteria for inclusion in the treatment study were: (a) outpatients who could
participate in the group-CBT for 12 weeks, (b) a diagnosis of major depressive disorder for the current episode established by a psychiatrist or a clinical psychologist using the Structured Clinical Interview for DSM-IV (SCID) [13-14], (c) Hamilton Rating Scale for Depression (HRSD) [15] score of 8 or greater, and (d) patients being defined as the treatment resistant according to the staging system of antidepressant resistance, [4] with the level of the treatment resistance at stage 2 or greater. Exclusion criteria were: current or previous diagnosis of a psychotic spectrum disorder, evidence of organic brain disorder, mental retardation, personality disorder, current high risk of suicide, substance abuse, or serious somatic disease. All patients were evaluated by a psychiatrist or a clinical psychologist using the Structured Clinical Interview for Axis I (SCID-I) [16] and the Structured Clinical Interview for Axis II (SCID-II) [17]. 

"The baseline scores on the Hamilton Rating Scale for Depression (HRSD) indicated mild to moderate levels of depression among the patients (Mean = 14.4, SD = 4.4). Seven patients (16%) had HRSD scores between 8 and 10, 13 (30%) had scores between 11 and 14, 13 (30%) had scores of between 15 and 18, and 10 (23%) had scores between 18 and 27. The baseline GAF scores indicated a poor level of social functioning. 31 patients (72%) had scores between 40 and 60, and the other 12 (28%) had scores between 61 and 70." (Results; Page 12, L6-12)

Comment #4
Why was group, rather than individual CBT chosen for this study? Although the jury is not at all definitive on this point it is possible that CBT in the individual setting be slightly more efficacious for depressive symptoms on the median term (6 months post treatment). Cost effectiveness is also debatable (Tucker, Beh Cog Psychotherapy 2007: 35:77).

Response
We agree that, in depressive patients with non TRD, individual CBT is slightly more efficacious for depressive symptoms at follow-up than group-CBT. However, there are few evidences whether individual CBT is more effective than group-CBT in the patients with TRD. However, a group-CBT is also expected not only the support effect but also the modeling effect (Zettle & Rains, J of Clin Psychology, 1989; 45(3): 436-445). Probably the group format provided patients opportunities for practicing new
skills, and the view that the social function has improved will also be possible (Oei, Bullbeck, & Campbell, J Affect Disord, 2006; 92(2-3): 231-241). Because these effects are more important to the treatment for TRD, we chose the group format. We added the following sentences in the discussion sections.

“Our protocol used basic CBT strategies, and did not include social skills training or stress management. However, the patients learned appropriate cognitive and behavioral coping strategies for increasing meaningful activity and managing interpersonal stress [32, 37]. The group-CBT provided both social support and also modeling effect [38-39]. The group format provided patients with opportunities for practicing new cognitive and behavioral skills, which they could apply in their lives after completion of the group-CBT [39]. These cognitive and behavioral skills may have influenced the improvement of social functioning.” (Discussion; Page 18, L18-25)

Comment #5
I also have an issue with how the authors are analyzing their data. We go from intent to treat (ITT) on 43 patients and end up with 20 at the 12-month follow-up. Statistics are only run on these latter. Statistics are only run on the 38 completers of the active phase of the study also. Where is the concept of LOCF here (last observation carried forward). Did the protocol not include a LOCF at the time of withdrawal from this study or were endpoint ratings only done on those who actually reached endpoint? If so, then the authors should be especially careful in their interpretation of their results. For instance, take the GAF scores at endpoint that increase from 60 to 67. Not a huge increase but again, it is clear that some of these patients are but very mildly depressed. The problem here is that N=38, so that the 5 drop outs are not included. As drop outs frequently do so due to lack of efficacy, then what would the inclusion of a 12% of reduced GAF scores have on this average (indeed, on all of the other averages of the other rating scales also)? Would the conclusions of this manuscript be the same? This may not be a huge problem for the active phase of the study if the response in the 38 patients is consistent, but it might be for the follow-up phase, where the drop out rate is much higher.

Response
We are in agreement with this comment. We have conducted not only the completers-only analyses but also the intent-to-treatment (ITT) analyses for the
treatment phase and the follow-up phase, as indicated in the methods section. The results of ITT have been added to Table 2 and Table 4 in revised version, and we have explained these results in the results section.

“All analyses were conducted on intent-to-treat (ITT) and completed treatment (Completer) samples. In the ITT analyses, the missing post-treatment or follow-up data were considered to be non-responders or adverse events, and their last available observations were carried forward (LOCF: last observation carried forward).”

Methods; Page 11, L15- L18)

“Acute treatment outcomes

a) Functional status

Table 2 displays the results of repeated measures ANOVAs for the GAF and Short-Form Health Survey (SF-36) scores from pre- to post-treatment. For both the ITT and Completer analyses, the GAF scores increased significantly (ITT: $F(1, 42) = 36.58$, $p < .001$, partial $\eta^2 = 0.47$; Completer: $F(1, 37) = 41.19$, $p < .001$, partial $\eta^2 = 0.53$). For the ITT sample, the number of patients who were rated as showing mild functional impairment (defined as GAF scores over 60) improved from 12 (28%) at baseline to 30 (70%) at post-treatment; 7 (16%) of these patients were rated as having minimal impairment (GAF > 70). As expected, the Completer sample comprised those 30 patients who had a post-treatment GAF score over 60 (79%), and the 7 patients (18.4%) who were rated as having minimal impairment (GAF > 70).

On the SF-36, the physical health (PCS) and mental health (MCS) scores at post-treatment were higher than at baseline, for both the ITT and Completer samples (all $p$ values < 0.01). The effect sizes of the MCS improvement were greater than these for the PCS, indicating that the group-CBT was more strongly associated with improvement in mental health than physical health. Seven of the 8 subscale scores were improved significantly (bodily pain was not). The pre-post effect sizes (Cohen’s $d$) for the vitality (ITT: 0.90; Completer: 1.08) and mental health (ITT: 0.83; Completer: 0.95) subscales were especially larger than for the other subscales.

b) Depressive symptoms

The repeated measures ANOVAs for the Hamilton Rating Scale for Depression (HRSD) scores showed a highly significant time effect. For the ITT sample, the mean HRSD scores decreased from 14.7 at pre-treatment to 9.2 at post-treatment ($F(1, 42) = 40.86$, $p < .001$, partial $\eta^2 = 0.49$, Cohen’s $d = 1.09$). For the Completer, the mean...
HRSD scores decreased from 14.2 at pre-treatment to 8.2 at post-treatment (F(1, 37) = 48.88, p < .001, partial $\eta^2 = 0.57$, Cohen’s $d = 1.30$). Among the Completers, 21 (55%) of the patients had scores of 7 or less on the HRSD at post-treatment. 12 had scores between 8 and 14, and 5 had scores between 15 and 21.

Table 3 shows the remission and response rates at post-treatment for the ITT and Completer sample analyses. Twenty-one participants (ITT: 49%; Completer: 55%) met criteria for remission (HRSD score of 7 or less), and 18 participants (ITT: 42%; Completer: 47%) showed at least a 50% reduction of their scores on the HRSD from the pre-treatment score. The number of participants who met criteria both for remission and 50% reduction of were 17 (ITT: 40%; Completer: 45%).

In addition, we calculated the reliable change and clinically significant change using Jacobson and Truax’s formula [29]. For the ITT sample, the cutoff point on the HRSD was 5. The criteria for “recovered” were fulfilled by 9 (21%) participants, 10 (23%) were “improved”, and 24 (56%) were “unchanged” or “deteriorated”. Among the 38 patients who completed the treatment, the cutoff point on the HRSD was 6. Nineteen (50 %) met criteria for “recovered” or “improved”, and the other 19 (50%) were classified as “unchanged or deteriorated”.

c) Dysfunctional cognitions

The score on the Dysfunctional Attitude Scale (DAS) decreased significantly from pre-treatment to post-treatment for both the ITT and Completer samples. The mean of the DAS scores changed using the LOCF (last observation carried forward) method from 161.3 to 147.6 (F(1, 42) = 17.24, p < .001, partial $\eta^2 = 0.29$, Cohen’s $d = 0.39$). The mean for the 38 in the Completer sample decreased from 156.3 to 140.9 (F(1,37) = 18.16, p < .001, partial $\eta^2 = 0.33$, Cohen’s $d = 0.48$).

In addition, the means on the ATQ-R negative scale at post-treatment were significantly lower than the means at pre-treatment using the same two methods of analysis. The mean of the ATQ-R negative scale scores changed using the LOCF method from 90.0 to 70.8 (F(1, 42) = 35.22, p < .001, partial $\eta^2 = 0.46$, Cohen’s $d = 0.76$). The mean for the 38 in the Completer sample decreased from 87.2 to 65.5 (F(1,37) = 39.46, p < .001, partial $\eta^2 = 0.52$, Cohen’s $d =1.00$). However, there was no significant difference on the ATQ-R positive scale between pre-treatment and post-treatment in the ITT or Completer sample analyses.” (Results; Page 12, L20- Page 15, L4)

“We analyzed the follow-up data using both the ITT and Completer samples. For
the ITT analysis which included 13 dropouts (5 who did not complete the treatment, and 8 who refused the follow-up study), the last observation values were carried forward (LOCF). Table 4 shows the changes in functional status measured by the GAF and SF-36 for the ITT and Completer samples. The repeated measures ANOVAs for GAF revealed a significant time effect for the both the ITT sample and Completer samples (both p values < 0.01). Post hoc paired t-tests with a Bonferroni correction showed that the score at post-treatment was higher than the score at baseline, and the score at the 12-month follow-up was also higher than at the post-treatment (p < 0.001). For the ITT sample, including the 13 dropouts, 27(84%) met criteria for the mild-minimal impairment (GAF < 60), and 12 patients (38%) reached the level of functioning well. For the Completer sample, except for one patient, all patients (95%) were at the mild-minimal impairment level (GAF < 60), and 10 patients (50%) were functioning well (GAF > 70) at the 12-month follow-up.

The repeated measures ANOVAs for the SF-36 (both PCS and MCS) also showed significant time effects for both the ITT and the Completer samples (all p values < 0.01). Post hoc paired t-tests with a Bonferroni correction demonstrated that MCS scores at the post-treatment and the 12-month follow-up were higher than the baseline score in both analyses (p < 0.001). In contrast, for the PCS score, the 12-month follow-up scores were not significantly different from the pre-treatment scores. Regarding the subscale scores, 7 of the 8 subscale scores at the 12-month follow-up were significant higher than the baseline scores (bodily pain was the exception).

Regarding depressive symptoms, in both the ITT and Completer analyses, there was a significant change in the Hamilton Rating Scale for Depression (HRSD) score during the 12-month follow-up ( both p values < 0.001), and the score at the follow-up was lower than not only the score at baseline but also the score at the end of the treatment (ps < 0.01). For the ITT sample, 22 patients (69%) had scores of 7 or less on the HRSD at the 12-month follow-up. 8(25%) had scores between 8 and 14, and other two had scores between 15 and 22. Of the 20 in the Completer sample, 14 (70%) had scores of 7 or less on the HRSD at the 12-month follow-up. Five (25%) had scores between 8 and 14, and one scored 22.

Additionally, dysfunctional cognitions measured by the DAS and the ATQ-R (both negative and positive scales) showed sustained improvements in both the ITT and Completer analyses (all p values < 0.05). For the Completer sample, the ATQ-R negative score at the 12-month follow-up was lower than the score at post-treatment (p < 0.01). ” (Results; Page 15, L14- Page 16, L24)
Table 2 - Repeated measures ANOVAs of treatment outcome as measured by GAF and SF-36

Table 2 shows the results of repeated measures ANOVAs for the GAF and SF-36 from pre- to post-treatment, for both ITT and Completer analyses.

<table>
<thead>
<tr>
<th></th>
<th>ITT (N=43)</th>
<th>Completers (N=38)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
</tr>
<tr>
<td></td>
<td>pre</td>
<td>post</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAF</td>
<td>59.49 (6.10)</td>
<td>65.51 (6.68)</td>
</tr>
<tr>
<td></td>
<td>60.18 (5.79)</td>
<td>67.00 (5.17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-36 PCS</td>
<td>41.00 (11.22)</td>
<td>46.78 (10.22)</td>
</tr>
<tr>
<td></td>
<td>41.18 (11.04)</td>
<td>47.72 (9.62)</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>76.33 (17.38)</td>
<td>84.77 (13.92)</td>
</tr>
<tr>
<td></td>
<td>77.43 (16.65)</td>
<td>86.97 (13.76)</td>
</tr>
<tr>
<td>Role functioning</td>
<td>27.33 (42.19)</td>
<td>58.14 (42.86)</td>
</tr>
<tr>
<td></td>
<td>23.68 (41.08)</td>
<td>58.55 (43.21)</td>
</tr>
<tr>
<td>Body pain</td>
<td>61.58 (25.68)</td>
<td>68.45 (26.94)</td>
</tr>
<tr>
<td></td>
<td>62.68 (26.63)</td>
<td>70.46 (27.55)</td>
</tr>
<tr>
<td>General health perception</td>
<td>39.66 (15.76)</td>
<td>48.74 (20.36)</td>
</tr>
<tr>
<td></td>
<td>38.62 (16.03)</td>
<td>50.03 (20.76)</td>
</tr>
<tr>
<td>SF-36 MCS</td>
<td>25.77 (8.70)</td>
<td>33.52 (10.97)</td>
</tr>
<tr>
<td></td>
<td>25.93 (8.95)</td>
<td>34.71 (10.90)</td>
</tr>
<tr>
<td>Vitality</td>
<td>24.65 (13.69)</td>
<td>39.65 (19.32)</td>
</tr>
<tr>
<td></td>
<td>24.87 (12.81)</td>
<td>41.84 (18.25)</td>
</tr>
<tr>
<td>Role functioning-emotional</td>
<td>11.63 (28.06)</td>
<td>34.88 (39.14)</td>
</tr>
<tr>
<td></td>
<td>11.40 (29.28)</td>
<td>37.72 (40.40)</td>
</tr>
<tr>
<td>Social functioning</td>
<td>43.84 (20.79)</td>
<td>59.59 (23.75)</td>
</tr>
<tr>
<td></td>
<td>44.34 (21.83)</td>
<td>62.17 (23.87)</td>
</tr>
<tr>
<td>Role functioning</td>
<td>11.63 (28.06)</td>
<td>34.88 (39.14)</td>
</tr>
<tr>
<td></td>
<td>11.40 (29.28)</td>
<td>37.72 (40.40)</td>
</tr>
<tr>
<td>Mental health</td>
<td>37.86 (15.82)</td>
<td>52.09 (18.34)</td>
</tr>
<tr>
<td></td>
<td>38.00 (16.10)</td>
<td>54.11 (17.93)</td>
</tr>
</tbody>
</table>

Table 4 - Repeated measure ANOVAs of 12-month follow-up measured by GAF and SF-36

Table 4 shows the results of repeated measures ANOVAs of 12-month follow-up data for the GAF and SF-36, for both ITT and Completer analyses.
Table 4 Repeated Measures ANOVAs of 12-month follow-up measured by GAF and SF-36

<table>
<thead>
<tr>
<th></th>
<th>Pre-treatment</th>
<th>Post-treatment</th>
<th>12 months</th>
<th>effect size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ITT (N=32)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAF</td>
<td>60.62 (6.36)</td>
<td>66.41 (6.74)</td>
<td>71.22 (9.16)</td>
<td>26.36³</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SF-36 PCS</td>
<td>42.23 (8.88)</td>
<td>48.21 (9.76)</td>
<td>46.97 (9.34)</td>
<td>9.76″</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>76.87 (16.84)</td>
<td>84.53 (14.67)</td>
<td>84.53 (14.11)</td>
<td>11.57″</td>
</tr>
<tr>
<td>Role functioning-physical</td>
<td>21.88 (39.53)</td>
<td>60.16 (45.73)</td>
<td>59.38 (40.54)</td>
<td>15.05″</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>65.84 (22.56)</td>
<td>70.73 (25.29)</td>
<td>74.06 (24.06)</td>
<td>2.30</td>
</tr>
<tr>
<td>General health perception</td>
<td>39.78 (14.92)</td>
<td>51.13 (20.59)</td>
<td>49.91 (19.64)</td>
<td>6.69</td>
</tr>
<tr>
<td><strong>SF-36 MCS</strong></td>
<td>26.16 (9.70)</td>
<td>33.59 (11.20)</td>
<td>38.34 (11.63)</td>
<td>19.34″</td>
</tr>
<tr>
<td>Vitality</td>
<td>26.41 (14.66)</td>
<td>40.31 (19.10)</td>
<td>43.75 (19.76)</td>
<td>18.48</td>
</tr>
<tr>
<td>Social functioning</td>
<td>44.84 (21.91)</td>
<td>60.55 (23.14)</td>
<td>68.75 (25.79)</td>
<td>10.77</td>
</tr>
<tr>
<td>Role functioning-emotional</td>
<td>13.54 (31.52)</td>
<td>37.50 (42.12)</td>
<td>54.17 (43.76)</td>
<td>16.43</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>65.84 (22.56)</td>
<td>70.73 (25.29)</td>
<td>74.06 (24.06)</td>
<td>2.30</td>
</tr>
<tr>
<td>General health perception</td>
<td>39.78 (14.92)</td>
<td>51.13 (20.59)</td>
<td>49.91 (19.64)</td>
<td>6.69</td>
</tr>
<tr>
<td><strong>Completers (N=20)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAF</td>
<td>60.24 (6.86)</td>
<td>66.48 (6.40)</td>
<td>73.81 (7.29)</td>
<td>26.61″</td>
</tr>
<tr>
<td>SF-36 PCS</td>
<td>43.80 (8.48)</td>
<td>50.32 (7.88)</td>
<td>48.18 (7.28)</td>
<td>5.31″</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>80.25 (13.13)</td>
<td>89.50 (9.02)</td>
<td>89.00 (7.88)</td>
<td>7.52</td>
</tr>
<tr>
<td>Role functioning-physical</td>
<td>21.25 (40.76)</td>
<td>57.50 (47.37)</td>
<td>66.25 (38.79)</td>
<td>6.75″</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>66.60 (24.30)</td>
<td>73.28 (27.40)</td>
<td>76.30 (26.75)</td>
<td>1.35</td>
</tr>
<tr>
<td>General health perception</td>
<td>39.55 (16.17)</td>
<td>55.90 (22.06)</td>
<td>54.10 (21.19)</td>
<td>6.94</td>
</tr>
<tr>
<td>SF-36 MCS</td>
<td>25.23 (9.77)</td>
<td>33.78 (11.06)</td>
<td>40.56 (10.71)</td>
<td>15.93″</td>
</tr>
<tr>
<td>Vitality</td>
<td>25.50 (13.66)</td>
<td>42.25 (16.58)</td>
<td>47.25 (17.66)</td>
<td>14.52</td>
</tr>
<tr>
<td>Social functioning</td>
<td>48.00 (21.86)</td>
<td>61.87 (26.12)</td>
<td>73.13 (29.32)</td>
<td>5.12</td>
</tr>
<tr>
<td>Role functioning-emotional</td>
<td>13.33 (33.16)</td>
<td>40.00 (45.37)</td>
<td>63.33 (45.76)</td>
<td>11.63″</td>
</tr>
<tr>
<td>Mental health</td>
<td>34.80 (16.68)</td>
<td>52.00 (16.57)</td>
<td>57.60 (17.38)</td>
<td>16.56″</td>
</tr>
</tbody>
</table>

³p<.001, †p<.01, *p<.05

a: significant difference between Pre- and Post-treatment (p<.05)
b: significant difference between Pre- and 12 months after treatment (p<.05)
c: significant difference between Post- and 12 months after treatment (p<.05)

Comment #6

Would a comparison group of TAU (treatment as usual) not have been warranted here, given the population studied and the modest level of psychopathology?

Response

We agree that we should mention the population and the level of psychopathology of patients. The mean patient’s depressive illness duration was 19.4 ± 15.6 months which indicated that their depressive symptoms had existed over one year. In addition, at the baseline score of Hamilton Rating Scale for Depression (HRSD) was 14.4 which indicated in the mild depression range. Seven patients had HRSD scores between 8 and 10, 13 had HRSD scores between 11 and 14, 13 (30.2%) had scores of between 15 and 18, and 10 (23.3%) had scores 18 and above. However, the baseline GAF scores indicated a poor level of the social functioning. 31 patients (72.1%) had scores between 40 and 60, and the other 12 (27.9%) had scores between 61 and 70.

After all, most patients in the present study were less severe than previous studies
(Moore & Blackburn; Scott et al.; Hirshfeld et al.), although they met diagnostic criteria from mild to moderate depression. So we do not know whether these treatments can be implemented or these results can be found for severe TRD. We have added the following sentence in the results and discussion section.

“The baseline scores on the Hamilton Rating Scale for Depression (HRSD) indicated mild to moderate levels of depression among the patients (Mean = 14.4, SD = 4.4). Seven patients (16%) had HRSD scores between 8 and 10, 13 (30%) had scores between 11 and 14, 13 (30%) had scores of between 15 and 18, and 10 (23%) had scores between 18 and 27. The baseline GAF scores indicated a poor level of social functioning. 31 patients (72%) had scores between 40 and 60, and the other 12 (28%) had scores between 61 and 70.” (Results; Page 12, L6-12)

“We examined the efficacy of the adding cognitive behavioral therapy to treatment with medication for improving both the depressive symptoms and the social functioning of TRD patients. The baseline scores on the HRSD in the present study were in the mild to moderate depression range. However, psychosocial functioning in the majority of patients was poor. The mean of the enrolled patients’ depressive episode at the baseline was 19.4 months, which indicated that their depressive symptoms and psychosocial functioning impairments had existed for long term. The cognitive behavioral therapy combined with medication for the patients with TRD resulted in significant improvement in both the depressive symptoms and the social functioning of the patients, and maintained improvement after a one year follow-up. As far as we know, this is the first study to investigate the long term effectiveness in of adding cognitive behavioral therapy to medication for improving both depressive symptoms and social functioning of patients defined as TRD.” (Discussion; Page 17, L2- L14)

Comment #7
I think there is a certain amount of confusion with regards reporting of the HRSD scores for the long-term portion of this study on page 14 (results section). They state that the HRDS scores at follow-up are higher than at baseline or at end of the active treatment. This would imply that patients are less well. This is not at all mentioned in the Discussion or Conclusion section, which is odd. In fact, on page 17 of the Discussion section active treatment gains are once again mentioned, but not the follow-up HRSD scores. Are we to assume that the authors made a typo in the results section and that the
final HRSD scores are also maintained at a lower level? If this is not the case then the authors have a bit of explaining to do.

Response
We agree that we made a typo in the results section. The HRSD score at follow-up were lower than at baseline or at the end of active treatment. We have corrected the sentence in the results section, and added the following sentences in the discussion section.

“Regarding depressive symptoms, in both the ITT and Completer analyses, there was a significant change in the Hamilton Rating Scale for Depression (HRSD) score during the 12-month follow-up (both p values < 0.001), and the score at the follow-up was lower than not only the score at baseline but also the score at the end of the treatment (ps < 0.01).” (Results; Page 16, L11-15)

Minor revision point 1;
Comment #8
(1) The authors should correct some obvious typos, such as on page 3, in the phrase ?Group-CBT involved of 12 weekly? and also on page 9, where it is written ?twelve 90-minite sessions? There are other examples.

Response
We have corrected these typos as follows;

“Forty-three patients with TRD were treated with 12 weekly sessions of group-CBT.” (Abstract; page 3, L11)

“Cognitive behavioral group therapy was conducted for 12 weekly 90-minute sessions;” (Methods; page 9, L18-19)

(2) Also, some phrases are just a bit difficult to understand largely due to sometimes odd word selections. For instance, the development of pharmacological strategies in the treatment of major depressive disorder is remarkable? on page 5. Remarkable is perhaps not the most appropriate choice of words, although one gets the message.

Response
We are in agreement with this comment. We have removed this sentence in the revised our paper.

(3) Likewise, on page 15 in the Discussion section, the phrase beginning with In the present study, these scores?? is difficult to understand which is unfortunate as they are making an important point (contrasting their results with those of Dunner et al, one of the few comparable studies).

Response
We are in agreement with this comment. We have made the modification in this point as follows;

“They reported that the scores on the PCS and MCS scales of the SF-36 did not change over the two years. In the present study, the PCS and MCS scores were similar at baseline to the results of Dunner et al. [8] , but these scores in our study showed sustained improvement, especially for the mental components (MCS), after CBT treatment and one year later.” (Discussion; page 17, L17-21)

(4) Another example is on page 16, in the phrase “It is possible that chronic depression in these studies might be involved with treatment-resistant??. The choice of the word involved? here is difficult to understand and again, the authors are making an important point.

Response
We corrected the errors in this sentence.

“It is likely that chronic depression in these studies included treatment-resistant depression.” (Discussion; Page18, L14-15)

Discretionary revision point 1;

Comment #9
An abbreviations section at the end of the manuscript would be useful.

Response
We have added the list of abbreviations at the end of the manuscript (Page 20).
List of abbreviation used

ATQ-R (Automatic Thought Questionnaire-Revised), CBT (cognitive behavioral therapy), DAS (Dysfunctional Attitude Scale), GAF (Global Assessment of Functioning), HRSD (Hamilton Rating Scale for Depression), ITT (intent-to-treat), LOCF (last observation carried forward), MCS (Mental Component Summary), PCS (Physical Component Summary), RCI (Reliable Change Index), SF-36 (the 36-item Short-Form Health Survey), TRD (treatment-resistant depression), TAU (treatment as usual).

Discretionary revision point 2;
Comment #10
I would add the time line to the flow chart (ie, Entered group-CBT (day 1), Completed ?(day X)?).

Response
We have added the time line to the follow chart (Page 27, Figure 1).
Figure 1: Flow chart of participants

Entered group-CBT week 1 (N=43)

Drop out (N=5)

Completed group-CBT week 12 (N=38)

Had not ended follow-up study yet (N=10)

Ended follow-up study week 60 (N=28)

Did not assess at 12-month follow-up (N=8)

Available 12-month follow-up assessment (N=20)
We are grateful to Reviewer 2 for their comments and useful suggestions that have helped to improve our paper. These comments gave us a better perspective on our work. Thank you for your interest and patience.

**Reviewer 2’s comment**

**Comment #1**
The small number of patients included in the study (N=38 for T0 and T1; N= 20 for the FU).
The small number of patients is especially problematic in the use of multiple hierarchical regression analysis (MHRA). In general the ratio variables: subjects needs to be at least 1:15. The number of predictors in the analyses was 10, whereas the number of subjects was 38, a very unfavourable ratio, precluding the use of MHRA.
Besides, since it is well known that the level of distress/impairment at the start of treatment is often the best predictor of level of distress/impairment after treatment, it would have been better to include GAF and SF-36 at T1 as the dependent variables (in stead of difference scores) and to enter GAF and SF-36 as measured at T0 in the first step of the HRMA, before entering the other predictors.
Also in the repeated measures ANOVA, one should control for baseline levels of impairment.
-Advise: leave regression analyses out, and use baseline levels of dependent variables as covariates in repeated measures ANOVA.

**Response**
We agree that there were little number for conducting multiple linear regression analysis. We have attempted to conduct the multiple hierarchical regression analysis, with GAF and SF-36 at T1 as the dependent variables, and to enter GAF and SF-36 as measured at T0 in the first step of the HRMA, before entering the other predictors.
However, we did not find any significant predictive variables after controlled the baseline scores. Therefore, we cut the result of the multiple hierarchical regression analysis.

Moreover, using the baseline score as a covariate in repeated ANOVA was not suitable for finding the prediction of the treatment effect. Thus, we removed the following sentences and Table 4 in revised version of our paper.

“3. What predicts the improvements of social functioning in patients with TRD
after treatment?" (Background)

“Next, in order to elucidate the baseline predictors of the treatment outcomes, multiple regression analyses were conducted with the difference scores between the baseline and post-treatment of the GAF and SF-36 subscale score as dependent variables and the baseline demographic and clinical variables (age at intake, gender, age of onset, number of depressive episodes, duration of current episode, duration of illness, marital status, employment status, years of school education and baseline HRSD score) as independent variables.” (Methods - Statistical Methods)

“Findings of the multiple regression analysis are summarized in Table 4.” (Results)

“We also investigated whether the baseline variables predicted the functional improvement from group-CBT.” (Discussion)

Comment #2
Authors excluded a.o. patients with comorbid personality disorder. Severe depression however often goes together with a personality disorder. So how many patients were excluded on the basis of the exclusion criteria, especially comorbid PD? Next, and more importantly, how were comorbid personality disorders assessed?

Response
We agree that we should mention this point. In our study, there was none of them who met the diagnosis of the personality disorder, although there were a few patients near diagnostic criteria. We gathered and selected patients who were satisfied with the entrance criteria. Therefore, we conducted the interviews using the Structures Clinical Interview for Axis II (SCID-II) to assessed comorbid personality disorders. We added the following sentences in the method section.

“All patients were evaluated by a psychiatrist or a clinical psychologist using the Structured Clinical Interview for Axis I (SCID-I) [16] and the Structured Clinical Interview for Axis II (SCID-II) [17].” (Methods; page 7, L14-16)
Comment #3
Authors need to provide reliability coefficients (Cronbach’s alpha) for the DAS, the ATQ-R and the SF-36.

Response
We are agreement with this comment. We have added following sentences in the results section.

[SF-36] “The Cronbach’s alpha reliability estimates for the Japanese SF-36 are 0.71-0.87 for the subscales, indicating good test-retest reliability [20].”
(Methods; Page 8, L21-23)

[DAS] “The Cronbach's alpha reliability estimate for the Japanese version is 0.86, consistent with good test-retest reliability [24].”
(Methods; Page 9, L9-11)

[ATQ-R] “The Japanese version has been tested in university students, and satisfactory reliability and construct validity has been reported [26].”
(Methods; Page 9, L14-15)

Comment #4
The most important limitation concerns the lack of a control group. In fact, the study shows that treatment resistant depressed patients show an improvement in social functioning after adding short term group CBT to medication, and that those who were willing to participate in the FU, maintain improvement. This holds for a specific group of depressed patients without comorbidity with axis II disorders and without high suicidal risk. However, it remains unknown whether the change in scores is related to natural course, to having had more attention, to other, unknown, factors or to CBT in particular. Without a control group that received only medication, and a control group that received another type of treatment in addition to medication, the main questions of the study remain unanswered, unfortunately.

Although this limitation of the study is mentioned in the discussion, the discussion says that adding CBT to medication enhanced social functioning, speaks of a study on the long term effectiveness of CBT, mentions as reason for improvement of patients that CBT provided patients with appropriate cognitive and behavioral coping strategies. In other words, the conclusions described in the discussion and the abstract are to my opinion overstated because the present study design does not allow for such conclusions.
I would suggest to be somewhat more modest in discussing the results of this study and start with the limitations. So I would not say: despite the findings, the study has several limitations..., but rather: despite several limitations, the present study suggests that adding CBT to medication might have a positive effect, a suggestion that needs to be confirmed in larger samples using randomized controlled trials?

Response

We agree that we have overstated the results without a comparison of a sufficient control group. We modified the limitation and the conclusion of this study in the revised version.

“This study has several limitations. First, the lack of a control group limits the interpretation of the results. It remains unknown whether the improvement in social functioning with TRD is related to natural course of depression. In addition, it is not clear whether the group affiliation or the CBT strategy is the active factor accounting for the improvements. More research using a TAU (treatment as usual) control group or different treatment groups is needed. Second, most TRD patients in the present study were less severely depressed than previous studies of patients with TRD [11-12, 40], although they met diagnostic criteria for mild to moderate depression. So we do not know whether the combined treatment can be implemented or these results would be found for severe TRD. Third, there were missing data from people who did not complete treatment. We used not only the completer analyses but also the ITT analyses. Although the results did not differ much between the dropouts (included in the ITT sample) and the treatment completers, there were some patients who did not complete group-CBT because they got worse. Also, control of the specific antidepressants could not be implemented in our longitudinal study. Therefore, the results of maintaining improvement may include some effects of medication.

Despite these limitations, the present study suggests that using group-CBT along with medication has a positive effect on both depressive symptoms and psychosocial functioning, a suggestion that needs to be confirmed in larger samples using randomized controlled trials.” (Discussion; page 19, L8-23)

Conclusions

“This study suggests a positive effect that combining cognitive behavioral therapy with medications improves both depressive symptoms and social functioning with TRD.
Moreover, these improvements in both depressive symptoms and social functioning were maintained over one year following completion of CBT while continuing on medication. ” (Conclusions; Page 20; L5-9)

Discretionary Revisions ;

Comment #5

Effect sizes were expressed as partial eta squared (partial $\eta^2$). According to conventional criteria a partial $\eta^2$ of 0.01 is small; 0.06 moderate; 0.14 large. As such all effects found, except for one, are (very) large. However, a more appropriate ES measure is Cohen’s $d$. According to conventional criteria, $d' .20$ is considered a small ES; $d' .50$ a medium ES; and $d' .80$ a large ES.

Expressed in terms of Cohen’s $d$, four of the ESs found in this study are medium (SF-36, SF-36 physical functioning, SF-36 PCS general health perception and SF-36 MCS role functioning-emotional) and not large, as is suggested by the partial $\eta^2$.

Response

We agree that the effect size of pre-post treatment is often used the Cohen's $d$. However, we considered that the appropriate effect size of the repeated measure design was partial $\eta^2$. Therefore, the effect sizes of ANOVAs expressed by partial $\eta^2$. If there were statistically significant differences with paired $t$-tests (using a Bonferroni correction), we also computed Cohen’s $d$. As a result, we agree that we found four of the ESs in this study are medium (SF-36, SF-36 physical functioning, SF-36 PCS general health perception and SF-36 MCS role functioning-emotional) and not large by the Cohen’s $d$

We have added the following sentences in the methods and the results sections. Table 2 has been modified to add the effect size of the Cohen's $d$.

“If there were statistically significant differences, we also computed Cohen's $d$ as a measure of the pre-post effect size. According to the criteria of Cohen's classification a $d$ of 0.2 is small, 0.5 is medium, and 0.8 is large [29] .” (Methods; Page 10, L19-21)

“The pre-post effect sizes (Cohen's $d$) for the vitality (ITT: 0.90; Completer: 1.08) and mental health (ITT: 0.83; Completer: 0.95) subscales were especially larger than for the other subscales.” (Results; Page 13, L12-14)
Table 2 - Repeated measures ANOVAs of treatment outcome as measured by GAF and SF-36

Table 2 shows the results of repeated measures ANOVAs for the GAF and SF-36 from pre- to post-treatment, for both ITT and Completer analyses.

<table>
<thead>
<tr>
<th></th>
<th>ITT (N=43)</th>
<th></th>
<th></th>
<th></th>
<th>Completers (N=38)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pre</td>
<td>post</td>
<td>F value</td>
<td>effect size</td>
<td>pre</td>
<td>post</td>
<td>F value</td>
</tr>
<tr>
<td>GAF</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>59.49 (6.10)</td>
<td>65.51 (6.68)</td>
<td>36.58***</td>
<td>0.47</td>
<td>0.94</td>
<td>60.18 (5.79)</td>
<td>67.00 (5.17)</td>
</tr>
<tr>
<td>SF-36 PCS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical functioning</td>
<td>41.00 (11.12)</td>
<td>46.78 (10.22)</td>
<td>14.10**</td>
<td>0.25</td>
<td>0.04</td>
<td>41.18 (11.04)</td>
<td>47.72 (9.62)</td>
</tr>
<tr>
<td>Role functioning-physical</td>
<td>76.33 (17.38)</td>
<td>84.77 (15.92)</td>
<td>20.78***</td>
<td>0.33</td>
<td>0.51</td>
<td>77.43 (16.65)</td>
<td>86.97 (13.78)</td>
</tr>
<tr>
<td>Body pain</td>
<td>61.58 (25.68)</td>
<td>68.45 (26.94)</td>
<td>3.81</td>
<td>0.08</td>
<td>0.26</td>
<td>62.68 (26.60)</td>
<td>70.46 (27.55)</td>
</tr>
<tr>
<td>General health perception</td>
<td>38.66 (15.76)</td>
<td>48.74 (20.36)</td>
<td>12.18**</td>
<td>0.23</td>
<td>0.00</td>
<td>38.62 (16.03)</td>
<td>50.03 (20.76)</td>
</tr>
<tr>
<td>SF-36 MCS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Physical functioning</td>
<td>25.77 (6.70)</td>
<td>33.52 (10.97)</td>
<td>23.91***</td>
<td>0.36</td>
<td>0.78</td>
<td>25.93 (8.89)</td>
<td>34.71 (10.99)</td>
</tr>
<tr>
<td>Role functioning-emotional</td>
<td>24.65 (13.69)</td>
<td>39.65 (19.32)</td>
<td>30.21***</td>
<td>0.42</td>
<td>0.90</td>
<td>24.87 (12.81)</td>
<td>41.84 (18.25)</td>
</tr>
<tr>
<td>Vitality</td>
<td>43.84 (20.79)</td>
<td>59.59 (23.75)</td>
<td>13.84**</td>
<td>0.25</td>
<td>0.71</td>
<td>44.34 (21.83)</td>
<td>62.17 (23.87)</td>
</tr>
<tr>
<td>Social functioning</td>
<td>11.63 (28.06)</td>
<td>34.88 (39.14)</td>
<td>18.68***</td>
<td>0.31</td>
<td>0.68</td>
<td>11.40 (29.26)</td>
<td>37.72 (40.40)</td>
</tr>
<tr>
<td>Role functioning-emotional</td>
<td>37.86 (15.82)</td>
<td>52.09 (18.34)</td>
<td>28.14***</td>
<td>0.40</td>
<td>0.83</td>
<td>38.00 (16.10)</td>
<td>54.11 (17.93)</td>
</tr>
</tbody>
</table>

*p<.001, **p<.01, *p<.05
GAF: Global Assessment of Functioning Scale
SF-36-36: 36-item Short-Form Health Survey
PCS: Physical Component Summary
MCS: Mental Component Summary

Comment #6

Treatment response was defined as a 50% or greater reduction on the HRSD, compared to pretreatment score. However this way of calculating response may not imply a reliable change. Why did not the authors use Jacobson and Truax formula to calculate reliable change and clinical significant change?

Response

We are grateful to this comment. We have calculated the reliable change and clinically significant change using Jacobson and Truax’s formula, and added the following sentence in the revised manuscript.

“We also calculated the reliable change and clinically significant change of depressive symptoms using Jacobson and Truax’s (JT) method, which uses two steps [30-31]. The first step is to define a cutoff point that separates the functional population from the dysfunctional population. The cutoff we used point was ±2 SD from the pre-treatment mean. The second step compares an individual’s change from pre- to post-treatment to a standard error of measurement of the outcome, referred to as the Reliable Change Index (RCI). If the RCI is higher than 1.96, the probability that the pre-post treatment difference occurred by chance is less than 5%. Using the results of
these two steps, we classified patients into three categories: recovered (passed cutoff point and $RCI > 1.96$), improved (did not pass cutoff point but $RCI > 1.96$), or unchanged or deteriorated (passed neither criterion).” (Methods; Page 10, L25- Page 11, L10)

“In addition, we calculated the reliable change and clinically significant change using Jacobson and Truax’s formula [29]. For the ITT sample, the cutoff point on the HRSD was 5. The criteria for “recovered” were fulfilled by 9 (21%) participants, 10 (23%) were “improved”, and 24 (56%) were “unchanged” or “deteriorated”. Among the 38 patients who completed the treatment, the cutoff point on the HRSD was 6. Nineteen (50%) met criteria for “recovered” or “improved”, and the other 19 (50%) were classified as “unchanged or deteriorated.” (Results; Page 14, L7-13)

“In terms of the clinical significant change [30], half of the patients showed recovery or improvement.” (Discussion; Page 19, L4-7)

Comment #7
The fact that only 20 patients participated in the follow up measurement, may constitute a serious bias. In addition, it is unknown whether these 20 patients continued to use medication or took other forms of treatment. It is a pity that an intention-to-treatment analysis could not be done, due to the small number of patients included in the study.

Response
We agree that we should mention this issue. This issue is also pointed by reviewer 1. We did not provide the sufficient explanation in the smaller sample of the follow-up. Of the 38 patients who completed the group-CBT, a total of 28 patients had completed the treatment more than one year ago at the time of our research follow-up. The remaining 10 persons had completed the group-CBT less than one year at the time of follow-up. Twenty of the 28 patients (71.4%) completed all the self-report questionnaires one year after finishing group-CBT; the other 8 refused to participate in the follow-up.

We have conducted the intent-to-treatment analysis (ITT) with not only the active treatment but also the follow-up data. We have added the results of ITT analyses in the results section, and discussed them.

Moreover, we have added the explanations in the methods section that the patients had continued to treatment with medications and did not take any other forms of
treatment except medication for 12 months after the group-CBT. In the limitations under the discussion section, we have indicated that it was the small number of patients included the 12-month follow-up study, and they had medications which were not controlled during follow-up period.

“All analyses were conducted on intent-to-treat (ITT) and completed treatment (Completer) samples. In the ITT analyses, the missing post-treatment or follow-up data were considered to be non-responders or adverse events, and their last available observations were carried forward (LOCF: last observation carried forward).” (Methods; Page 11, L15-L18)

“Acute treatment outcomes

a) Functional status

Table 2 displays the results of repeated measures ANOVAs for the GAF and Short-Form Health Survey (SF-36) scores from pre- to post-treatment. For both the ITT and Completer analyses, the GAF scores increased significantly (ITT: $F(1, 42) = 36.58, p < .001$, partial $\eta^2 = 0.47$; Completer: $F(1, 37) = 41.19, p < .001$, partial $\eta^2 = 0.53$).

For the ITT sample, the number of patients who were rated as showing mild functional impairment (defined as GAF scores over 60) improved from 12 (28%) at baseline to 30 (70%) at post-treatment; 7 (16%) of these patients were rated as having minimal impairment (GAF > 70). As expected, the Completer sample comprised those 30 patients who had a post-treatment GAF score over 60 (79%), and the 7 patients (18.4%) who were rated as having minimal impairment (GAF > 70).

On the SF-36, the physical health (PCS) and mental health (MCS) scores at post-treatment were higher than at baseline, for both the ITT and Completer samples (all $p$ values < 0.01). The effect sizes of the MCS improvement were greater than these for the PCS, indicating that the group-CBT was more strongly associated with improvement in mental health than physical health. Seven of the 8 subscale scores were improved significantly (bodily pain was not). The pre-post effect sizes (Cohen’s $d$) for the vitality (ITT: 0.90; Completer: 1.08) and mental health (ITT: 0.83; Completer: 0.95) subscales were especially larger than for the other subscales.

b) Depressive symptoms

The repeated measures ANOVAs for the Hamilton Rating Scale for Depression (HRSD) scores showed a highly significant time effect. For the ITT sample, the mean HRSD scores decreased from 14.7 at pre-treatment to 9.2 at post-treatment ($F(1,42) =$
40.86, \( p < .001 \), partial \( \eta^2 = 0.49 \), Cohen’s \( d = 1.09 \)). For the Completer, the mean HRSD scores decreased from 14.2 at pre-treatment to 8.2 at post-treatment (\( F(1, 37) = 48.88, p < .001 \), partial \( \eta^2 = 0.57 \), Cohen’s \( d = 1.30 \)). Among the Completers, 21 (55%) of the patients had scores of 7 or less on the HRSD at post-treatment. 12 had scores between 8 and 14, and 5 had scores between 15 and 21.

Table 3 shows the remission and response rates at post-treatment for the ITT and Completer sample analyses. Twenty-one participants (ITT: 49%; Completer: 55%) met criteria for remission (HRSD score of 7 or less), and 18 participants (ITT: 42%; Completer: 47%) showed at least a 50% reduction of their scores on the HRSD from the pre-treatment score. The number of participants who met criteria both for remission and 50% reduction of were 17 (ITT: 40%; Completer: 45%).

In addition, we calculated the reliable change and clinically significant change using Jacobson and Truax’s formula [29]. For the ITT sample, the cutoff point on the HRSD was 5. The criteria for “recovered” were fulfilled by 9 (21%) participants, 10 (23%) were “improved”, and 24 (56%) were “unchanged” or “deteriorated”. Among the 38 patients who completed the treatment, the cutoff point on the HRSD was 6. Nineteen (50%) met criteria for “recovered” or “improved”, and the other 19 (50%) were classified as “unchanged or deteriorated”.

c) Dysfunctional cognitions

The score on the Dysfunctional Attitude Scale (DAS) decreased significantly from pre-treatment to post-treatment for both the ITT and Completer samples. The mean of the DAS scores changed using the LOCF (last observation carried forward) method from 161.3 to 147.6 (\( F(1, 42) = 17.24, p < .001 \), partial \( \eta^2 = 0.29 \), Cohen’s \( d = 0.39 \)). The mean for the 38 in the Completer sample decreased from 156.3 to 140.9 (\( F(1,37) = 18.16, p < .001 \), partial \( \eta^2 = 0.33 \), Cohen’s \( d = 0.48 \)).

In addition, the means on the ATQ-R negative scale at post-treatment were significantly lower than the means at pre-treatment using the same two methods of analysis. The mean of the ATQ-R negative scale scores changed using the LOCF method from 90.0 to 70.8 (\( F(1, 42) = 35.22, p < .001 \), partial \( \eta^2 = 0.46 \), Cohen’s \( d = 0.76 \)). The mean for the 38 in the Completer sample decreased from 87.2 to 65.5 (\( F(1,37) = 39.46, p < .001 \), partial \( \eta^2 = 0.52 \), Cohen’s \( d = 1.00 \)). However, there was no significant difference on the ATQ-R positive scale between pre-treatment and post-treatment in the ITT or Completer sample analyses.” (Results; Page 12, L20- Page 15, L4)
We analyzed the follow-up data using both the ITT and Completer samples. For the ITT analysis which included 13 dropouts (5 who did not complete the treatment, and 8 who refused the follow-up study), the last observation values were carried forward (LOCF). Table 4 shows the changes in functional status measured by the GAF and SF-36 for the ITT and Completer samples. The repeated measures ANOVAs for GAF revealed a significant time effect for the both the ITT sample and Completer samples (both p values < 0.01). Post hoc paired t-tests with a Bonferroni correction showed that the score at post-treatment was higher than the score at baseline, and the score at the 12-month follow-up was also higher than at the post-treatment (p < 0.001). For the ITT sample, including the 13 dropouts, 27(84%) met criteria for the mild-minimal impairment (GAF < 60), and 12 patients (38%) reached the level of functioning well. For the Completer sample, except for one patient, all patients (95%) were at the mild-minimal impairment level (GAF < 60), and 10 patients (50%) were functioning well (GAF > 70) at the 12-month follow-up.

The repeated measures ANOVAs for the SF-36 (both PCS and MCS) also showed significant time effects for both the ITT and the Completer samples (all p values < 0.01). Post hoc paired t-tests with a Bonferroni correction demonstrated that MCS scores at the post-treatment and the 12-month follow-up were higher than the baseline score in both analyses (p < 0.001). In contrast, for the PCS score, the 12-month follow-up scores were not significantly different from the pre-treatment scores. Regarding the subscale scores, 7 of the 8 subscale scores at the 12-month follow-up were significant higher than the baseline scores (bodily pain was the exception).

Regarding depressive symptoms, in both the ITT and Completer analyses, there was a significant change in the Hamilton Rating Scale for Depression (HRSD) score during the 12-month follow-up (both p values < 0.001), and the score at the follow-up was lower than not only the score at baseline but also the score at the end of the treatment (ps < 0.01). For the ITT sample, 22 patients (69%) had scores of 7 or less on the HRSD at the 12-month follow-up. 8(25%) had scores between 8 and 14, and other two had scores between 15 and 22. Of the 20 in the Completer sample, 14 (70%) had scores of 7 or less on the HRSD at the 12-month follow-up. Five (25%) had scores between 8 and 14, and one scored 22.

Additionally, dysfunctional cognitions measured by the DAS and the ATQ-R (both negative and positive scales) showed sustained improvements in both the ITT and Completer analyses (all p values < 0.05). For the Completer sample, the ATQ-R negative score at the 12-month follow-up was lower than the score at post-treatment (p < 0.01).  

(Results; Page 15, L14- Page 16, L24)
Table 4 shows the results of repeated measures ANOVAs of 12-month follow-up data for the GAF and SF-36, for both ITT and Completer analyses. (Page, 30)

<table>
<thead>
<tr>
<th></th>
<th>Pre-treatment</th>
<th>Post-treatment</th>
<th>12 months</th>
<th>F</th>
<th>partial Eta²</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td>Mean (SD)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ITT (N=32)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAF</td>
<td>60.62 (6.36)</td>
<td>66.41 (6.74)</td>
<td>71.22 (9.16)</td>
<td>26.36**</td>
<td>0.46</td>
</tr>
<tr>
<td>SF-36 PCS</td>
<td>42.23 (8.88)</td>
<td>48.21 (9.76)</td>
<td>46.97 (9.34)</td>
<td>9.76**</td>
<td>0.24</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>76.87 (16.84)</td>
<td>84.53 (14.67)</td>
<td>84.53 (14.11)</td>
<td>11.57**</td>
<td>0.27</td>
</tr>
<tr>
<td>Role functioning-physical</td>
<td>21.88 (39.53)</td>
<td>60.16 (45.73)</td>
<td>59.38 (40.54)</td>
<td>15.05**</td>
<td>0.33</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>65.84 (22.56)</td>
<td>70.73 (25.29)</td>
<td>74.06 (24.06)</td>
<td>2.30</td>
<td>0.07</td>
</tr>
<tr>
<td>General health perception</td>
<td>39.78 (14.92)</td>
<td>51.13 (20.59)</td>
<td>49.91 (19.64)</td>
<td>6.69**</td>
<td>0.18</td>
</tr>
<tr>
<td><strong>Completers (N=20)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAF</td>
<td>60.24 (6.86)</td>
<td>66.48 (6.40)</td>
<td>73.81 (7.29)</td>
<td>26.61**</td>
<td>0.57</td>
</tr>
<tr>
<td>SF-36 PCS</td>
<td>43.80 (8.48)</td>
<td>50.32 (7.88)</td>
<td>48.18 (7.28)</td>
<td>5.31**</td>
<td>0.22</td>
</tr>
<tr>
<td>Physical functioning</td>
<td>80.25 (13.13)</td>
<td>95.50 (9.02)</td>
<td>89.00 (7.88)</td>
<td>7.52**</td>
<td>0.28</td>
</tr>
<tr>
<td>Role functioning-physical</td>
<td>21.25 (40.78)</td>
<td>75.70 (47.37)</td>
<td>56.25 (38.79)</td>
<td>6.75</td>
<td>0.26</td>
</tr>
<tr>
<td>Bodily pain</td>
<td>66.60 (24.30)</td>
<td>73.28 (27.40)</td>
<td>76.30 (26.75)</td>
<td>1.35</td>
<td>0.07</td>
</tr>
<tr>
<td>General health perception</td>
<td>39.55 (16.17)</td>
<td>55.90 (22.06)</td>
<td>54.10 (21.19)</td>
<td>6.94**</td>
<td>0.27</td>
</tr>
<tr>
<td>SF-36 MCS</td>
<td>25.23 (9.77)</td>
<td>33.78 (11.06)</td>
<td>40.56 (10.71)</td>
<td>15.93**</td>
<td>0.46</td>
</tr>
<tr>
<td>Vitality</td>
<td>25.50 (13.66)</td>
<td>42.25 (16.58)</td>
<td>47.25 (17.66)</td>
<td>14.52**</td>
<td>0.43</td>
</tr>
<tr>
<td>Social functioning</td>
<td>48.00 (21.86)</td>
<td>61.87 (26.12)</td>
<td>73.13 (29.32)</td>
<td>5.12</td>
<td>0.21</td>
</tr>
<tr>
<td>Role functioning-emotional</td>
<td>13.33 (33.16)</td>
<td>40.00 (45.37)</td>
<td>63.33 (45.76)</td>
<td>11.63**</td>
<td>0.38</td>
</tr>
<tr>
<td>Mental health</td>
<td>34.86 (16.68)</td>
<td>52.00 (16.57)</td>
<td>57.80 (17.36)</td>
<td>16.56**</td>
<td>0.47</td>
</tr>
</tbody>
</table>

*p<.001,  **p<.01, *p<.05

a: significant difference between Pre- and Post-treatment (p<.05)
b: significant difference between Pre- and 12 months after treatment (p<.05)
c: significant difference between Post- and 12 months after treatment (p<.05)

**Comment #8**

**Minor points:** a typo in abstract under Methods (number of patients included) and unclear wording on page 14, third line from below: "higher than these scores).Which scores?

**Response**

We have corrected these typos and sentences as follows;

"Forty-three patients with TRD were treated with 12 weekly sessions of group-CBT." (Abstract; page 3, L11)

"Regarding depressive symptoms, in both the ITT and Completer analyses,
there was a significant change in the Hamilton Rating Scale for Depression (HRSD) score during the 12-month follow-up (both p values < 0.001), and the score at the follow-up was lower than not only the score at baseline but also the score at the end of the treatment (ps < 0.01).” (Results; Page 16, L11-15)

We are grateful to Reviewer 3 for their comments and useful suggestions that have helped to improve our paper. These comments gave us a better perspective on our work. Thank you for your interest and patience.

Reviewer 3’s comment

Comment #1
It provides an important contribution to the literature on the effectiveness of psychological interventions for depression. The introduction is missing references to other key studies in the area of augmenting antidepressant medication treatment with CBT. Although the focus is on treatment resistant depression, the work of, for example, DeRubeis et al. Archives of General Psychiatry, Vol. 62, April 2005 seems relevant showing that cognitive therapy can be as effective as medications for the initial treatment of moderate to severe major depression. About 40% of the patients in the study under review had a first episode. The DuRubeis et al study had initial Hamilton Rating Scale for Depression mean (HRSD) score of about 23 whereas the study under review had a mean HRSD score of 14.2, which is in the milder range and surprisingly low for a treatment resistant population. The authors may wish to comment on that. Are their patients more impacted by poor coping skills in various areas of life as opposed to actual symptoms of depression?

Response
We agree that we have added the reference to DuRubeis et al study of augmenting antidepressant medication treatment with CBT. In addition, we should have explained the patients in this study who were included mildly depressive patients. This point is also commented by other reviewers.

The baseline scores on the Hamilton Rating Scale for Depression (HRSD) indicated mild to moderate levels of depression among the patients (Mean = 14.4, SD = 4.4).
Seven patients (16%) had HRSD scores between 8 and 10, 13 (30%) had scores between 11 and 14, 13 (30%) had scores of between 15 and 18, and 10 (23%) had scores between 18 and 27. The baseline GAF scores indicated a poor level of social functioning. 31 patients (72%) had scores between 40 and 60, and the other 12 (28%) had scores between 61 and 70.

Although the mildly depressed patients were diagnosed partial remission, they had suffered from the long-term depressive episode in the past. The previous researches indicate functioning impairment persists beyond symptomatic improvement (Miller et al., J Clin Psychiatry 2002; 59:608-619). Moreover, combined medication and CBASP (as a type of CBT) may have an independent effect on psychosocial functioning above and beyond its effect on depressive symptoms per se (Hirschfeld et al., Biol Psychiatry 2002; 51:123-133). Therefore, we used the following criteria; (a) outpatients who could participate in the group-CBT for 12 weeks, (b) a diagnosis of major depressive disorder (including partial remission) established by a psychiatrist or a clinical psychologist using the Structured Clinical Interview for DSM-IV [13], (c) Hamilton Rating Scale for Depression (HRSD) [17] score of 8 or greater, because we used a HRSD of 7 or less as the criteria for remission after the group-CBT; and (d) patients being defined as the treatment resistant according to the staging system of antidepressant resistance, [4] with the level of the treatment resistance at stage 2 or greater.

“Cognitive behavioral therapy (CBT) has been shown to be effective in the treatment of major depressive disorder. DeRubeis et al. [9] suggested that CBT can be as effective as medication for the initial treatment of moderate to severe major depression.” (Background; Page 5, L13-16)

“Criteria for inclusion in the treatment study were: (a) outpatients who could participate in the group-CBT for 12 weeks, (b) a diagnosis of major depressive disorder for the current episode established by a psychiatrist or a clinical psychologist using the Structured Clinical Interview for DSM-IV(SCID) [13-14], (c) Hamilton Rating Scale for Depression (HRSD) [15] score of 8 or greater, and (d) patients being defined as the treatment resistant according to the staging system of antidepressant resistance, [4] with the level of the treatment resistance at stage 2 or greater. Exclusion criteria were: current or previous diagnosis of a psychotic spectrum disorder, evidence of organic brain disorder, mental retardation, personality disorder, current high risk of suicide, substance abuse, or serious somatic disease. All patients were evaluated by a
psychiatrist or a clinical psychologist using the Structured Clinical Interview for Axis I (SCID-I) [16] and the Structured Clinical Interview for Axis II (SCID-II) [17].” (Methods; Page 7, L4-16)

“The baseline scores on the Hamilton Rating Scale for Depression (HRSD) indicated mild to moderate levels of depression among the patients (Mean = 14.4, SD = 4.4). Seven patients (16%) had HRSD scores between 8 and 10, 13 (30%) had scores between 11 and 14, 13 (30%) had scores of between 15 and 18, and 10 (23%) had scores between 18 and 27. The baseline GAF scores indicated a poor level of social functioning. 31 patients (72%) had scores between 40 and 60, and the other 12 (28%) had scores between 61 and 70.” (Results; Page 12, L6-12)

“We examined the efficacy of the adding cognitive behavioral therapy to treatment with medication for improving both the depressive symptoms and the social functioning of TRD patients. The baseline scores on the HRSD in the present study were in the mild to moderate depression range. However, psychosocial functioning in the majority of patients was poor. The mean of the enrolled patients' depressive episode at the baseline was 19.4 months, which indicated that their depressive symptoms and psychosocial functioning impairments had existed for long term. The cognitive behavioral therapy combined with medication for the patients with TRD resulted in significant improvement in both the depressive symptoms and the social functioning of the patients, and maintained improvement after a one year follow-up. As far as we know, this is the first study to investigate the long term effectiveness in of adding cognitive behavioral therapy to medication for improving both depressive symptoms and social functioning of patients defined as TRD.” (Discussion; Page 17, L2- L14)

Comment #2
The standard CBT Depression protocol used in this study is primarily focused on symptoms and in particular cognitive and behavioural. Whilt patients do get opportunities to set goals and practice new ways of thinking in the real world? as part of their homework, the CBT protocol does not include stress management, communication skills, interpersonal functioning. Other studies showing that integrated therapies as opposed to monotherapies had greater effects on improving psychosocial functioning. This aim is to show that a combined approach, CBT with medication, can improve symptoms and psychosocial functioning at end of treatment and one year later. How do
the authors explain the improvement in psychosocial functioning when their CBT protocol was not integrated with another psychological therapy? Did the group format perhaps provide opportunities for practicing new skills? Did they deviate from the CBT protocol at times and allowed for a more interpersonal component to take place as group members supported each other?

Response

We would like to touch a few about the present condition of our country. In Japan, there had not been an effective protocol of group-CBT with not only TRD but also major depressive disorder. Thus, first of all, we conducted the basic protocol, and examined it important to evaluate an effect. Therefore, our protocol was a basic CBT, and that did not include social skills training and stress management. However, the patients learned the skills which notices and restructures the negative thinking in a problem of interpersonal relations. Moreover, they trained to look for new ideas, and practiced them in daily life. The new ideas were not only positive thinking replaced with negative thinking but also coping strategies to manage interpersonal stress (Patelis-Siotis et al., J Affect Disord, 2001; 65(2): 145-153).

In addition, a group-CBT could also expect not only the support effect but also the modeling effect (Zettle & Rains, J of Clin Psychology, 1989;45(3): 436-445). Probably the group format provided patients opportunities for practicing new skills, and the view that the social function has improved will also be possible (Oei, Bullbeck, & Campbell, J Affect Disord, 2006; 92(2-3): 231-241). Therefore, we agree that we mentioned that the group format provided patients opportunities for practicing and modeling new skills. We have added the following sentence in the discussion section.

“Our protocol used basic CBT strategies, and did not include social skills training or stress management. However, the patients learned appropriate cognitive and behavioral coping strategies for increasing meaningful activity and managing interpersonal stress [32, 37]. The group-CBT provided both social support and also modeling effect [38-39]. The group format provided patients with opportunities for practicing new cognitive and behavioral skills, which they could apply in their lives after completion of the group-CBT [39]. These cognitive and behavioral skills may have influenced the improvement of social functioning.” (Discussion; Page 18, L18-25)
Comment #3
Interesting that more men than women enrolled in the therapy and study considering that depression is twice as prevalent among women compared to men. Also, women still to seek depression treatment more often than men. Was this issue discussed in the groups? Is this a more typical scenario in the country where the study took place? What may account for this gender balance?

Response
We agree that a woman tends to suffer from the depressive symptoms than a male. However, we did not know why male patients were more than female patients in our study. The data of the present study gathered from the patients who ended the group-CBT by 2008. At the November in 2009, those who ended the group CBT was about 50. They were male 27, and female 23. The proportion of female has been increased gradually. These days, the big difference in gender balance is not found.

Comment #4
It would be helpful for the authors to comment on level of depression and take that into consideration when drawing conclusions. The discussion section needs to be softened?It seems reasonable to conclude that combining cognitive behavioral group therapy with medications could improve social functioning more than medication alone.? This may be limited to patients who may be mild to moderately depressed at least as measured by the HRSD.

Response
We agree that we have better interpreted more carefully the results in our study. We modified the discussion section as follows:

“This study has several limitations. First, the lack of a control group limits the interpretation of the results. It remains unknown whether the improvement in social functioning with TRD is related to natural course of depression. In addition, it is not clear whether the group affiliation or the CBT strategy is the active factor accounting for the improvements. More research using a TAU (treatment as usual) control group or different treatment groups is needed. Second, most TRD patients in the present study were less severely depressed than previous studies of patients with TRD [11-12, 40], although they met diagnostic criteria for mild to moderate depression. So we do not know whether the combined treatment can be implemented or these results would be
found for severe TRD. Third, there were missing data from people who did not complete treatment. We used not only the completer analyses but also the ITT analyses. Although the results did not differ much between the dropouts (included in the ITT sample) and the treatment completers, there were some patients who did not complete group-CBT because they got worse. Also, control of the specific antidepressants could not be implemented in our longitudinal study. Therefore, the results of maintaining improvement may include some effects of medication.

Despite these limitations, the present study suggests that using group-CBT along with medication has a positive effect on both depressive symptoms and psychosocial functioning, a suggestion that needs to be confirmed in larger samples using randomized controlled trials.” (Discussion; page 19, L8-23)

Comment #5
Other clinically relevant questions to be addressed. 1. Why were the groups so small in size like 5-6 when about 8 is the more usual in the group therapy literature? Was there a therapeutic rationale for this small group size and the proportionally large therapist number, like some groups may have had three therapists and five patients. That also allows for much individual attention perhaps at the expense of the group climate factors. 2. The drop-out rate of only five of 43 (about 11%) is very low and commendable for group therapy. Can the authors explain this? Was there any group therapy preparation or prior experience with group? Cultural factors? Once you commit, you stay? Pleasing/fearing the therapists (authorities)?

Response
In the limited time of 90 minutes, in order to get all the patients to speak, we thought that about five to six persons were appropriate. It might be said that it can pay the attention to each and related the low dropout rate. Furthermore, we thought that also the patients who had been gathered from its hospital and included mildly depressions relate to the lowness of a dropout rate.

Comment #6
Limitations section. How do the authors suggest a controlled study be designed?

Response
We suggest a controlled study which uses a TAU group, or difference processing groups to reveal the more clearly efficacy of the addition of group-CBT to medication. We have added the following sentences in discussion section.

“It remains unknown whether the improvement in social functioning with TRD is related to natural course of depression. In addition, it is not clear whether the group affiliation or the CBT strategy is the active factor accounting for the improvements. More research using a TAU (treatment as usual) control group or different treatment groups is needed.” (Discussion; Page 19, L9-13)

Comment #7
Proof read the manuscript. There are several typos and a few sentences that do not work well in English.

Response
We have corrected typos and sentences as follows;

“Forty-three patients with TRD were treated with 12 weekly sessions of group-CBT.” (Abstract; page 3, L11)

“Cognitive behavioral group therapy was conducted for 12 weekly 90-minute sessions;” (Methods; page 9, L18-19)

“It is likely that chronic depression in these studies included treatment-resistant depression.” (Discussion; Page 18, L14-15)