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)	Does recruitment source moderate treatment effectiveness? A subgroup analysis from the EVIDENT study, a randomised controlled trial of an internet intervention for depressive symptoms.
AUTHORS	Klein, Jan; Gamon, Carla; Späth, Christina; Berger, Thomas; Meyer, Björn; Hohagen, Fritz; Hautzinger, Martin; Lutz, Wolfgang; Vettorazzi, Eik; Moritz, Steffen; Schröder, Johanna

VERSION 1 - REVIEW

REVIEWER	Margaret Anne Hurley College of Health and Wellbeing University of Central Lancashire PRESTON Lancashire PR12HE UK
REVIEW RETURNED	16-Dec-2016

GENERAL COMMENTS	Major comment: On the whole the statistical analyses described and carried out in this study look valid and convincing. However, there are some areas of ambiguity and the paper would benefit from a number of revisions and amendments for the readers' benefit. 1. My main comment regards the Statistical analysis section on page 10. For the linear mixed models, adjustment for baseline measure was chosen. I interpret this to mean that the outcome was PHQ-9 at post and follow-up assessment and that baseline PHQ-9 was included as a covariate (lines 24 and 25). However, on lines 26 and 27, the authors say that the outcome was change from baseline. This I interpret to mean the outcome was PHQ-9 at post or follow-up minus PHQ-9 at baseline. This sounds somewhat contradictory and needs to be resolved. Also, apart from the random intercept for each participant and the random error term, all the included terms are fixed effects, so the model does not have the full complexity of a linear mixed model. It is rather more a regression model with two error terms, one for participants and one for observations; a multilevel model with two levels. I hope that the model did not use change from the baseline as the outcome and then include an adjustment for baseline (as stated on page 11, line 1) since this would not be correct. I feel that this statistical analysis section could be usefully rewritten to improve the clarity of exposition.

2. My second main concern is Table 1. I do not see where the chi-squared degrees of freedom come from. Take for example Marital Status, which has 3 levels and there are 4 recruitment sources. The degrees of freedom are 15. Compared this to 18 for Educational Status, also with 3 levels, and 9 for Employment Status, also with 3 levels. Further, is there a level missing? Since 36, 24 and 29 does not sum to 105 (the clinical recruitment source) and the same comment applies to the other recruitment sources. I can see a rationale for 9 degrees of freedom but not for 15 or 18. Table 1 needs some attention.

My other concerns are more minor

3. When describing range of scores, I suggest you use the notation ‘from 10 to 14’ rather than between 10 and 14 since it is not clear whether 10 and 14 are included in the range. See page 9, line 14 and elsewhere.

4. The measure PHQ-9 often has the ‘-9’ omitted or the ‘9’ omitted so I suggest you use the full PHQ-9 throughout text and tables.

5. Regarding the APOI, the four subscales are stated to have range 4 to 16 but the total 16 to 80. There is a typo here I think since 4 times 16 is 64.

6. Where the paper says ‘group’, it refers to the treatment group. The paper concerns a subgroup analysis where the subgroups are recruitment source. It would help the reader if the word ‘treatment’ was inserted in front of ‘group’ throughout to avoid ambiguity.

7. It would also help if the word ‘effect’ was removed from ‘interaction effect’ and just use ‘interaction’ since this is more common usage.

8. In the Results section, under Participant characteristics (Page 11, line16) Table 1 and 2 are referenced. I had a little difficulty matching the text to the tables. For example, which rows in Table 2 are ‘self and clinician rated depression’? What is SF-12 PH, SF-12 MH and FEP-2. I am not sure if these are described in the measures.

9. Page 12 line 7 mentions ‘technological disadvantages’. Is this the same as ‘Technologization Threat’ in Table 3 – if so then please standardize on terminology.

10. Page 12, line 19 mentions ‘CAU’; is this ‘care as usual’, in which case please define on first use.

11. Page 12, line 23 mentions ‘treatment response’ and by this I think the authors must mean difference between change in outcome for treatment minus change in outcome for care as usual (known as the difference in differences). It would help the reader for this terminology to be defined somewhere.

12. In the Discussion section, page 14 lines 1 and 2, the text reads ‘While we have found that the between-groups effect is the same

	across recruitment sources, we did find that the within-group effect is smaller in those recruited via internet forums than in those recruited via other settings.' The first part of the sentence is expressing the result that the interaction was not significant and this is fine. So, if we look in Table 4 at the difference between 2.900 and 1.994, the difference between 1.790 and 0.428, the difference between 2.827 and 1.135 and the difference between 2.584 and 0.942 then all these differences could be estimates of the same quantity. The second part of the sentence present some difficulty since it refers to the 'within group effect' and this is rather ambiguous terminology since, previously, the exposition has implicitly defined the treatment effect/response to be the difference in differences (see comment 11 above). I think the authors are saying that the intervention outcome difference for the internet forums subgroup, 1.794 is lower than that for the other sub-groups with values 2.900, 2.827 and 2.584. The same comment applies to the care as usual treatment group, 0.428 for internet forums versus 1.994, 1.135 and 0.942 for the other subgroups. This ambiguity needs to be resolved to help the reader accept the results. 13. In Table 1, the samples sizes (N) are replicated for Educational status, so this needs to be made consistent with the other characteristics in Table 1. 14. In Table 4, the Clinical subgroup has n=105, with n=42 for intervention and n=38 for CAU. I have not understood why the subgroup sample size is more than the sum of 42 and 38. Is this because the 105 refers to all data points in the model (so participants have more than one data point if they have been measured at more than one time point) whereas the intervention and control n values are for numbers of participants? This inconsistency needs to be explained. 15. A final comment is that the authors should take a critical look at the numbers of decimal places in the four tables. Readers, including myself, can get a kind of number blindness when confronted with so many numbers. The last decimal place could be helpfully removed in some of the tables.

REVIEWER	Lasse Sander Department of Rehabilitation Psychology and Psychotherapy, Institute of Psychology, University of Freiburg, Germany
REVIEW RETURNED	21-Feb-2017

GENERAL COMMENTS	Dear editor and authors, thank you very much for the chance to review this paper. It presents new insights into an important research question. The research methodology is sound and applied properly. I only have the following minor comments/questions:

	1. In the abstract on page 3 in line 13, there is a little typo ("internet interventions"). 2. Is there any description of care as usual? CAU might differ between recruitment sources which leads to a potential thread to validity and should be discussed. 3. Have any missing data operations been performed and, if so, which?. 4. With regard to the CONSORT statement (Moher et al., 2010), this subgroup analysis was not prespecified in the protocol. However, this creates a risk of bias. This should be discussed in the limitations section. 5. The CONSORT statement further stipulates that "if adjustment was made for baseline variables, both unadjusted and adjusted analysis should be reported" (Moher et al., 2010). As adjustment for baseline measure was done in your study, unadjusted analysis should be reported or this discrepancy from the CONSORT statement should be discussed. 6. How was the mean total usage time measured? Is there any chance to observe persons, who only entered the programme without working on it or who did not log off from the programme after use? 7. The concept of „marginally significance" is dubious and may be misleading (Pritschet et al., 2016). Hence, the p value should either be labeled as not significant or "marginally significance" should be clarified. Again, thank you very much for inviting me as reviewer to this paper. I am looking forward to your reply and the discussion of my comments and questions.

VERSION 1 – AUTHOR RESPONSE

Reviewer # 1

"1. My main comment regards the Statistical analysis section on page 10. For the linear mixed models, adjustment for baseline measure was chosen. I interpret this to mean that the outcome was PHQ-9 at post and follow-up assessment and that baseline PHQ-9 was included as a covariate (lines 24 and 25). However, on lines 26 and 27, the authors say that the outcome was change from baseline. This I interpret to mean the outcome was PHQ-9 at post or follow-up minus PHQ-9 at baseline. This sounds somewhat contradictory and needs to be resolved. Also, apart from the random intercept for each participant and the random error term, all the included terms are fixed effects, so the model does not have the full complexity of a linear mixed model. It is rather more a regression model with two error terms, one for participants and one for observations; a multilevel model with two levels. I hope that the model did not use change from the baseline as the outcome and then include an adjustment for baseline (as stated on page 11, line 1) since this would not be correct. I feel that this statistical analysis section could be usefully rewritten to improve the clarity of exposition."

It is true that important details were missing from the description of the statistical analysis. One core feature of LMMs is the opportunity to choose an appropriate covariance structure reflecting the potential dependence due to repeated measurements. Therefore we have now added that we "chose an autoregressive covariance structure and allowed variances to vary between assessment points. The choice was based on Akaike's Information Criterion (AIC) from a fixed set of candidate structures, namely a first order autoregressive (AR1), or scaled identity structure or heterogeneous versions thereof."

In keeping with the reviewers suggestion we have re-analysed the data without baseline correction. The results are essentially the same: main effect of group: $F_{1,827}$ = 20.47, p < .001; main effect of recruitment source on PHQ change: ($F_{3,827}$ = 2.28, *p* = .078) and interaction effect (group assignment by recruitment source): $F_{3,827}$ = 0.18, *p* = .91). Still, we would like to defend the correction for baseline PHQ-9 and now also provide a reference to justify this choice: "Adjustment for baseline measure was chosen as this accounts for regression to the mean (Vickers et al., 2001)". We believe that analysing change alone would not control for baseline imbalance.

One further reason for retaining the current analysis is that it is based on the analysis we used for our main results (Klein et al., 2016). We are willing to add the analysis without baseline correction as a sensitivity analysis or use it as the main analysis if you do not find our reasons for retaining the current analysis convincing.

"2. My second main concern is Table 1. I do not see where the chi-squared degrees of freedom come from. Take for example Marital Status, which has 3 levels and there are 4 recruitment sources. The degrees of freedom are 15. Compared this to 18 for Educational Status, also with 3 levels, and 9 for Employment Status, also with 3 levels. Further, is there a level missing? Since 36, 24 and 29 does not sum to 105 (the clinical recruitment source) and the same comment applies to the other recruitment sources. I can see a rationale for 9 degrees of freedom but not for 15 or 18. Table 1 needs some attention."

We mistakenly inserted an abbreviated version of the full table. We had prepared this abbreviated table for as slide set and apologise for our error. We are grateful that the reviewer spotted this error and have now included the correct table.

"3. When describing range of scores, I suggest you use the notation 'from 10 to 14' rather than between 10 and 14 since it is not clear whether 10 and 14 are included in the range. See page 9, line 14 and elsewhere."

Thank you. We have carefully gone over the entire manuscript and changed this.

"4. The measure PHQ-9 often has the '-9' omitted or the '9' omitted so I suggest you use the full PHQ-9 throughout text and tables."

Thank you. We are now referring to this measure as the PHQ-9 throughout the manuscript.

"5. Regarding the APOI, the four subscales are stated to have range 4 to 16 but the total 16 to 80. There is a typo here I think since 4 times 16 is 64."

Thank you for pointing out this inaccuracy. The paragraph should have stated that the subscales range from 4 to 20 and that the total score ranges from 16 to 80.

"6. Where the paper says 'group', it refers to the treatment group. The paper concerns a subgroup analysis where the subgroups are recruitment source. It would help the reader if the word 'treatment' was inserted in front of 'group' throughout to avoid ambiguity."

This is indeed confusing, thank you. We have therefore replaced the word "group" with the word "treatment" wherever it refers to the main effect of treatment or the treatment by recruitment source interaction.

"7. It would also help if the word 'effect' was removed from 'interaction effect' and just use 'interaction' since this is more common usage."

This has also been corrected.

"8. In the Results section, under Participant characteristics (Page 11, line16) Table 1 and 2 are referenced. I had a little difficulty matching the text to the tables. For example, which rows in Table 2 are 'self and clinician rated depression'? What is SF-12 PH, SF-12 MH and FEP-2. I am not sure if these are described in the measures."

Thank you for pointing out that the description of these measures in the methods section was too brief. We have now expanded the description: "We also employed the following self-rating scales: a measure of health-related quality of life (Short-Form Health Survey: SF-12) that covers physical health related qualitiy of life (SF-12 PH) and mental health related quality of life (SF-12 MH); a broad self-rated symptom measure covering dimensions ranging from general well-being to interpersonal relationships (Questionnaire for the Evaluation of Psychotherapeutic Processes - FEP-2)…". We also provide an explanation for the measures alongside the table now.

"9. Page 12 line 7 mentions 'technological disadvantages'. Is this the same as 'Technologization Threat' in Table 3 – if so then please standardize on terminology."

Thank you for noticing this inconsistency, which has now been corrected.

"10. Page 12, line 19 mentions 'CAU'; is this 'care as usual', in which case please define on first use."

Again, thank you for pointing this out – the abbreviation “CAU” is now defined under “interventions” in the methods section.

“11. Page 12, line 23 mentions ‘treatment response’ and by this I think the authors must mean difference between change in outcome for treatment minus change in outcome for care as usual (known as the difference in differences). It would help the reader for this terminology to be defined somewhere.”

We have indeed used the term “treatment response” loosely here and replaced it with the more precise term “treatment effect”. Also, we have added a definition of what we mean by “treatment effect” at the end of the description of statistical analysis in the methods section.

“12. In the Discussion section, page 14 lines 1 and 2, the text reads ‘While we have found that the between-groups effect is the same across recruitment sources, we did find that the within-group effect is smaller in those recruited via internet forums than in those recruited via other settings.’ The first part of the sentence is expressing the result that the interaction was not significant and this is fine. So, if we look in Table 4 at the difference between 2.900 and 1.994, the difference between 1.790 and 0.428, the difference between 2.827 and 1.135 and the difference between 2.584 and 0.942 then all these differences could be estimates of the same quantity. The second part of the sentence present some difficulty since it refers to the ‘within group effect’ and this is rather ambiguous terminology since, previously, the exposition has implicitly defined the treatment effect/response to be the difference in differences (see comment 11 above). I think the authors are saying that the intervention outcome difference for the internet forums subgroup, 1.794 is lower than that for the other sub-groups with values 2.900, 2.827 and 2.584. The same comment applies to the care as usual treatment group, 0.428 for internet forums versus 1.994, 1.135 and 0.942 for the other subgroups. This ambiguity needs to be resolved to help the reader accept the results.”

We have carefully changed the wording in this paragraph and use the terms “moderator” and “predictor” in the summary of findings to avoid confusion with the term “treatment effect” as defined above.

“13. In Table 1, the samples sizes (N) are replicated for Educational status, so this needs to be made consistent with the other characteristics in Table 1.”

This has been corrected; thank you for your attention to detail here and throughout the manuscript!

“14. In Table 4, the Clinical subgroup has n=105, with n=42 for intervention and n=38 for CAU. I have not understood why the subgroup sample size is more than the sum of 42 and

38. Is this because the 105 refers to all data points in the model (so participants have more than one data point if they have been measured at more than one time point) whereas the intervention and control n values are for numbers of participants? This inconsistency needs to be explained."

The sum of participants in the treatment groups (intervention plus CAU) is smaller than the number of participants in each subgroup because some participants did not complete the post or the follow-up assessment and could therefore not be included in the main analysis because the outcome for the main analysis (change from baseline) could not be calculated for those without any assessment beyond baseline. A sensitivity analysis using multiple imputation to replace missing values yielded essentially the same results as the main analysis. This fact has now been added to the results section (sensitivity analysis using multiple imputation with 50 imputations) and the caption for table 4. We have decided to use the analysis without replacement of missing values as the main analysis as this is in keeping with published study protocol (Klein et al., 2013), the published main results (Klein et al., 2016) and the statistical analysis plan that we wrote before starting the analyses.

"15. A final comment is that the authors should take a critical look at the numbers of decimal places in the four tables. Readers, including myself, can get a kind of number blindness when confronted with so many numbers. The last decimal place could be helpfully removed in some of the tables."

In keeping with your suggestion and to improve readability, only the p-values are presented with three decimal places now.

Reviewer # 2

"1. In the abstract on page 3 in line 13, there is a little typo ("internet interventions")."

Thank you. We have also gone over the manuscript carefully again and hope that we have now eliminated all typos.

"2. Is there any description of care as usual? CAU might differ between recruitment sources which leads to a potential thread to validity and should be discussed."

We have described in the methods section that "Following a naturalistic and pragmatic design approach, care-as-usual was not influenced by the investigators. All participants were permitted to use any form of treatment, including antidepressant medication and psychotherapy."

It is also correct that the treatment received at baseline differed between recruitment sources: "Compared to participants recruited through insurance companies and other sources, participants recruited in clinical settings were more likely to be in psychiatric treatment, psychotherapy and inpatient psychiatric treatment." This is almost tautological, of course, because participants recruited from treatment settings would, by definition, be more likely to receive treatment than those recruited from non-treatment settings. In the discussion section we also cite indirect evidence that this concomitant psychiatric and psychotherapeutic treatment might reduce the effectiveness of the internet intervention even though the treatment assignment by recruitment source interaction was not statistically significant: “We have previously reported that the internet intervention was less effective for patients with mild to moderate depressive symptoms who received concurrent psychiatric or psychotherapeutic treatment [3]. Internet interventions may therefore confer the greatest benefit for individuals who are not in specialized psychiatric or psychotherapeutic care. However, this difference may also depend on symptom severity, as we have previously observed strong effects among severely depressed individuals who used an internet intervention and received concurrent antidepressant medication (Meyer et al., 2015).”

As you can see in the following table, treatments received after randomization did not differ between the intervention group and the CAU group, but that is beyond the scope of this manuscript, we believe. We can add this table as a supplement though, should you wish to see it included.

	Intervention		Care as usual			
	n = 509 (50.1%)		*n* = 504 (49.9%)		chi^2	*p*
Treatment between baseline and three months assessment						
Psychotherapy	127	32.3	140	35.1	0.767	.381
Outpatient psychiatric treatment	106	27.0	108	27.1	0.005	.941
Inpatient psychiatric treatment	5	1.3	9	2.3	1.123	.289
Antidepressant medication	193	48.9	204	51.3	0.455	.500
Treatment between three months and six months assessment						
Psychotherapy	118	31.3	119	31.7	0.016	.898
Outpatient psychiatric treatment	94	24.9	96	25.6	0.044	.833
Inpatient psychiatric treatment	6	1.6	9	2.4	0.629	.428
Antidepressant medication	192	50.8	192	51.2	0.012	.911

“3. Have any missing data operations been performed and, if so, which?”

In keeping with our protocol and the statistical analysis plan we had drafted before starting the analyses, we have not substituted missing data. This has now been specified in the methods section: “No missing values were substituted as LMMs based on all observed data are valid and unbiased methods for missing at random (MAR) data [37].” Following up on an observation of the statistical reviewer we have also added a sensitivity analysis were we replaced missing values using multiple imputations. This analysis yielded the same results as the main analysis.

"4. With regard to the CONSORT statement (Moher et al., 2010), this subgroup analysis was not prespecified in the protocol. However, this creates a risk of bias. This should be discussed in the limitations section."

We would argue that this is a pre-specified subgroup analysis. The study protocol states that "One subgroup analysis will concern the influence of referral source (medical and psychological services versus other) and primary motivation for study participation (self versus other) on the main outcome as many trials of online-based psychological interventions have been conducted outside routine clinical practice and the applicability of the results to routine clinical practice has therefore been debated." We have added this to the description of methods: "This subgroup analysis had been pre-specified in the study protocol [23]."

Technically we should also have analysed whether motivation influences the effectiveness of the intervention. We have not done that since the manuscript is already quite long. For completeness, these are the results of the analysis with self- vs. other motivation: main effect of treatment $p < .001$, main effect of motivation (self vs. other) $p = .221$, interaction treatment*motivation $p = .129$. If anything, those who had been motivated by others to participate in the study benefited more from deprexis. But with $p = .129$, this interaction is not statistically significant.

"5. The CONSORT statement further stipulates that "if adjustment was made for baseline variables, both unadjusted and adjusted analysis should be reported" (Moher et al., 2010). As adjustment for baseline measure was done in your study, unadjusted analysis should be reported or this discrepancy from the CONSORT statement should be discussed."

This issue has also been raised by the statistical reviewer. We have argued that we would like to defend the correction for baseline PHQ-9 because this accounts for regression to the mean (Vickers et al., 2001). We believe that analysing change alone would not control for baseline imbalance. One further reason for retaining the current analysis is that it is based on the analysis we used for our main results (Klein et al., 2016). In keeping with your suggestion we have re-analysed the data without baseline correction. The results are essentially the same: main effect of group: $F_{1,827} = 20.47$, $p < .001$; main effect of recruitment source on PHQ change: ($F_{3,827} = 2.28$, $p = .078$) and interaction effect (group assignment by recruitment source): $F_{3,827} = 0.18$, $p = .91$). We are willing to add the analysis without baseline correction as a sensitivity analysis or use it as the main analysis if you do not find our reasons for retaining the current analysis convincing.

"6. How was the mean total usage time measured? Is there any chance to observe persons, who only entered the programme without working on it or who did not log off from the programme after use?"

We have now specified that “Periods of inactivity of 5 min or longer were subtracted in the computation of the total usage time.” This is the same algorithm we have previously used and reported to compute usage time (Meyer et al., 2015). This method ensures that periods without active use (e.g., forgetting to log off, or logging on but then attending to another website) are not added to usage time.

“7. The concept of „marginally significance“ is dubious and may be misleading (Pritschet et al., 2016). Hence, the p value should either be labeled as not significant or “marginally significance” should be clarified.”

In keeping with your suggestion we have relabelled these as “not significant”.

We believe our manuscript has benefited substantially from the review process. We would like to thank the editor and the reviewers again for their comments. Please do not hesitate to contact me should you have any further questions.

VERSION 2 – REVIEW

REVIEWER	Margaret Anne Hurley University of Central Lancashire UK
REVIEW RETURNED	04-Apr-2017

GENERAL COMMENTS	The authors have adequately responded to the issues raised in the statistical review.

REVIEWER	Dipl. Psych. Lasse Sander University of Freiburg, Germany
REVIEW RETURNED	07-Apr-2017

GENERAL COMMENTS	I thank the authors for their detailed replies to my questions and comments. The authors clarified most issues. Pleas find my minor additional comments below. **#Reviewer 2 comments:** I thank the authors for their detailed replies to my questions and comments. The authors clarified most issues. Pleas find my minor additional comments underneath. **On Comment 2:** **#R2:** Thank you for your reply. I agree, that you described that

	participants were permitted to use whatever additional treatment they like. I also agree that the presented table is somewhat beyond the scope of your article. But it is still nice to have to lower readers concerns on the potential validity thread. Hence, an addition as supplement is desirable. **On comment 4:** **#R2:** Sorry, I totally agree. I expected to find this in the protocol´s method section where a rather undefined handling of subgroup analysis was stated. **On comment 5:** **#R2:** Concerning baseline correction, I agree with the author's argumentation. Unadjusted analysis may be added as sensitivity analysis or as a supplement to go along with the CONSORT statement. Finally, as some authors declared their competing interests due to funding by Servier and GAIA, the authors should not specify the products name (Deprexis) in the conclusion section again, to avoid any impression of surreptitious advertising. Again, thank you very much for the opportunity to review this manuscript, which substantially improved even further after the review-process.

VERSION 2 – AUTHOR RESPONSE

We were glad to hear that the reviewers have now recommended publication and are happy to send you the final minor suggestions today. In the following we provide point-by-point responses to the issues raised reviewer # 2. All major changes are highlighted in the revised version of the manuscript.

"On Comment 2: #R2: Thank you for your reply. I agree, that you described that participants were permitted to use whatever additional treatment they like. I also agree that the presented table is somewhat beyond the scope of your article. But it is still nice to have to lower readers concerns on the potential validity thread. Hence, an addition as supplement is desirable."

The table has now been included as a supplementary table. The section "Intervention usage" has now relabelled "Intervention usage and utilization of other treatments" and the following sentence has been added: "The intervention and the CAU group did not differ with respect to the use of concomitant treatments (e.g. psychotherapy, psychotropic medication) during the study period (Supplemental Table 1)."

On comment 4: #R2: Sorry, I totally agree. I expected to find this in the protocol´s method section where a rather undefined handling of subgroup analysis was stated.

Thank you. We hope that we have clarified this in the manuscript now.

"On comment 5: #R2: Concerning baseline correction, I agree with the author's argumentation. Unadjusted analysis may be added as sensitivity analysis or as a supplement to go along with the CONSORT statement."

As per your suggestion we provide this analysis as one of the sensitivity analyses now: Rerunning the primary analysis without baseline correction did not alter our results substantially (main effect of group: $F_{1,827} = 20.47$, $p < .001$; main effect of recruitment source on PHQ change: $F_{3,827} = 2.28$, $p = .078$; treatment assignment by recruitment source interaction: $F_{3,827} = 0.18$, $p = .91$)."

"Finally, as some authors declared their competing interests due to funding by Servier and GAIA, the authors should not specify the products name (Deprexis) in the conclusion section again, to avoid any impression of surreptitious advertising."

I totally agree, the product name should only be mentioned where strictly necessary. The name of the product has therefore been deleted from the conclusion.

We would like to thank the reviewer for his comments. We would also like to thank the editor again for orchestrating a smooth and efficient review process.