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)	Individual Patient Data Meta-analysis of combined treatments versus psychotherapy (with or without pill placebo), pharmacotherapy, or pill placebo for adult depression: A protocol
AUTHORS	Weitz, Erica; Kleiboer, Annet; van Straten, Annemieke; Hollon, Steven; Cuijpers, Pim

VERSION 1 - REVIEW

REVIEWER	Peter Bower University of Manchester
REVIEW RETURNED	12-Aug-2016

GENERAL COMMENTS	In the strengths and limitations section, it might be worth noting that IPD not only maximises power, but also provides protection against ecological fallacies which can be a problem when moderating factors are assessed using aggregate data and meta-regression. The introduction might also highlight that better targeting of combination treatments also has really important economic significance as such treatments are more expensive and should be provided only when they are likely to lead to significant benefit. The moderators listed are varied in scope, but obviously assessing very large numbers of moderators could lead to accusations of 'fishing'. There are some published guidelines for quality in moderation analysis of individual trials, and I wonder if these might be considered here? Are there a maximum number of moderators that they feel it is useful to assess? If so, how should they be prioritised? A short discussion of this issue might be useful Finally, did the authors have any thoughts about the 'clinical significance' criterion that should be used when assessing moderation? Does it make sense to use the standard conventions around main effects?

REVIEWER	Sarah Liebherz University Medical Center Hamburg-Eppendorf, Department of Medical Psychology, Hamburg, Germany
REVIEW RETURNED	30-Aug-2016

GENERAL COMMENTS	This paper describes a protocol on an individual patient data meta-analysis of combined vs. other treatments for adult depression. Overall, the methodology is sound, the study's aim is relevant and the manuscript is well written. However, in my opinion, some changes need to be made:

	Language Some sentences are extremely long and hard to understand. Please check the whole manuscript for language clarity and use short and precise formulations, e.g. page 3, line 12-15, page 4, line 34-41. Some formulations are very vague and need to be specified: - Page 2, line 39: “provides important information” ◊ this is very vague, please specify - Page 4, line 12: “many of these” ◊ these treatments? - Page 4, line 27-30 ◊ “yet different patients…” this is hard to understand, please clarify - Page 7: 40/41: “other relevant measures” sounds very vague. As you already mention several moderator variables, I think you can delete this or – if you think of other specific moderators, please specify. - Page 8, line 48-49: Please clarify. Aim of the study In my opinion, the aim of the study needs to be specified more clearly. I think the study focuses the question: What kind of patients really need combined treatments (because they benefit more from this kind of treatments than from psychotherapy or pharmacotherapy alone). If this is the main questions, it needs to be focused more clearly in the abstract, the introduction and the discussion, e.g.: - Page 2, line11-12: little is known… ◊ I think this sentence should focus combination treatment instead of the question which patient benefits from which treatment in general - Are there any hypotheses on relevant predictors? Of course, there may be little research on the correlation of patient characteristics and outcome of specific treatments, but – as you describe – there are – for example – some studies/guidelines recommending combined treatments for severe and chronic depression. I wonder if you have any hypothesis on these predictor variables (severity, chronicity). - The pros and cons of IPD-meta-analyses are described in detail, but I think more information on possible clinical implications are necessary. Methodology / General study approach: - Page 6, line 20/21: Please specify how you select the moderator variables. - Page 7, line 36: How will you operationalize “comorbid psychopathological disorders”: Number of comorbid disorders? Specific diagnosis? - Page 7: Timing of outcome assessments: What kind of variability do you mean? The variability in the length of the intervention? In this case, I think, the length should be a moderator (instead of addressing it in sensitivity analyses). Or do you mean the time between end of intervention and the assessment? I think, it should be specified a priori which period between end of intervention and assessment you accept as “post intervention”. However, addressing this period is not a sensitivity analysis, unless you exclude studies with a predefined period longer than xxx. - Page 8, line 11-12: I think the literature search needs to be described more clearly: As you use a database of studies already included in other meta-analyses: Will you update the search? 1,756 full-text articles were identified for possible inclusion. But you already performed some meta-analyses on adult depression? Will you screen the 1,756 full-texts again for inclusion criteria in this study? Or did you already pre-select some studies? Please explain

	this procedure. - Page 8, line 23-27: Why will you exclude the two items on other sources of bias and selective outcome reporting? I did not understand why this is not possible especially in this meta-analysis. - Page 9, line 14-17: Please specify: What do you mean by "necessary data" and "when missing data are not excessive"? How many missing data (and in which variables) are allowed? - Page 10, line 26/27: other study characteristics? Please specify. - Page 10, line 43-45: If you use predictor variables to impute missing outcome scores, the correlation between predictor variables and outcome scores will increase, please discuss this fact. - Page 11, line 32-35: In my opinion, these analyses are subgroup and not sensitivity analyses or will you exclude specific kinds of treatments? Further points - Page 2, line 22: The searches were already conducted, it would be better to write "searches were conducted…" - Page 7, line 28: I think "maximum absolute score reflecting normalization" is not clear enough to define remission, I would prefer explanations such as "no longer fulfilling diagnostic criteria for depression" or a score on depression outcome measures below a cut-off. - Page 7, line 29/30: What does "extreme response" mean? Please specify. - References: Reference 3: The title of the journal is missing. Reference 29: There is a new version of the Cochrane Handbook, please cite the current version.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Peter Bower
Institution and Country: University of Manchester
Please state any competing interests or state 'None declared': None declared'

Please leave your comments for the authors below
Reviewer comment:
In the strengths and limitations section, it might be worth noting that IPD not only maximises power, but also provides protection against ecological fallacies which can be a problem when moderating factors are assessed using aggregate data and meta-regression.

Response:
Thank you for this suggestion. We have added the following to the strengths and limitations section, "Utilizing IPD meta-analysis methods will allow for examination of individual patient clinical and demographic characteristics as moderators between combined treatment and comparator treatments for depression by maximizing statistical power while protecting against ecological fallacies that present problems when examining aggregate data using conventional meta-analysis techniques."

Reviewer Comment:
The introduction might also highlight that better targeting of combination treatments also has really important economic significance as such treatments are more expensive and should be provided only when they are likely to lead to significant benefit.

Response:

We agree with this point and have added the following into the introduction, “Similarly, many patients may do as well on a specific monotherapy as they do in combined treatment, however, utilizing combined treatment for these patients would waste valuable economic resources given that combined treatments are much more costly to provide.” (page 4, line 21).

Reviewer Comment:
The moderators listed are varied in scope, but obviously assessing very large numbers of moderators could lead to accusations of ‘fishing’. There are some published guidelines for quality in moderation analysis of individual trials, and I wonder if these might be considered here? Are there a maximum number of moderators that they feel it is useful to assess? If so, how should they be prioritised? A short discussion of this issue might be useful

Response:
We agree with the reviewer’s comments, and to avoid fishing expeditions have chosen our moderators based on previous research and not on chance. We have clarified which moderators we have chosen and how we have included them by better defining the clinical correlates that we will be examining as moderators and citing the particular research that supports examining the moderator next to each variable of interest. However, given that our analysis depends on the availability of these moderators from a large majority of the trials included, it is possible that not all of these moderators will be able to be analyzed (depending on what authors are able to share with us).
In the paper we stated, “Published papers will be examined to determine valid predictors reported across studies. This project will focus on clinical and demographic moderators of treatment outcomes including correlates of depression severity. Treatment guidelines recommend combined treatment for patients with severe depression, thus suggesting that there is a differential effect of treatment (combined versus monotherapy) as a function of depression severity. In addition, each of the particular moderator variables selected has been examined in previous RCTs and has been found to predict or moderate treatment outcomes in depression. The clinical predictors that will be examined in this study are: baseline depression severity [33–35] measured on the measures outlined above, having a comorbid mental health diagnosis[15,23] , anxiety symptoms [23], number of previous episodes (recurrence)[26,36] length of current episode (chronicity)[17], global assessment of functioning (GAF)[24], and previous exposure to depression treatments[15], Demographic moderators that will be examined in this study are: marital status[15,22,37], employment[15,22], education[26], and age[17]. Other baseline demographic characteristics will be gathered in order to adjust the analysis for these baseline characteristics. In addition, previous literature has found that social adjustment[24] predicted outcomes, and thus will be included when available as it is in a majority of trials. . It is expected that not all studies will assessed will be able to contribute all variables, and thus, indices will be selected when they uniquely examine a clinical correlate of interest (ie are not similar to another variable included) and when a majority of studies have provided this particular data.” (page 7 line 27 – page 8 line 11).

Reviewer Comment:
Finally, did the authors have any thoughts about the ‘clinical significance’ criterion that should be used when assessing moderation? Does it make sense to use the standard conventions around main effects?

Response:
Moderators that are found to be significant will be further assessed with subgroup analyses that standardize effect sizes. Effect sizes of d= .24 or above are considered to be clinically relevant (Cuijpers, et al, 2014). (page 11, lines 33-35).

Reviewer: 2
Reviewer Name: Sarah Liebherz

Institution and Country: University Medical Center Hamburg-Eppendorf, Department of Medical Psychology, Hamburg, Germany
Please state any competing interests or state ‘None declared’: none declared

Please leave your comments for the authors below

Reviewer comment:
This paper describes a protocol on an individual patient data meta-analysis of combined vs. other treatments for adult depression. Overall, the methodology is sound, the study's aim is relevant and the manuscript is well written.

Response:
Thank you for your comments.
Reviewer Comment A:

Language
Some sentences are extremely long and hard to understand. Please check the whole manuscript for language clarity and use short and precise formulations, e.g. page 3, line 12-15, page 4, line 34-41.
Some formulations are very vague and need to be specified:
1. Page 2, line 39: “provides important information” ◊ this is very vague, please specify
2. Page 4, line 12: “many of these” ◊ these treatments?
3. Page 4, line 27-30 ◊ “yet different patients…” this is hard to understand, please clarify
4. Page 7: 40/41: “other relevant measures” sounds very vague. As you already mention several moderator variables, I think you can delete this or – if you think of other specific moderators, please specify.
5. Page 8, line 48-49: Please clarify.

Response A:
We have edited the manuscript for clarity and have specifically clarified the concerns on page 2,4, 7, and 8.
1. On page 2, line 28, we state, “provides new information on moderators of treatment outcome that can be utilized by patients, clinicians, and researchers.”
2. On page 4, line 7, we state, “many of these treatments…”
3. On page 4, line 17, we state, “For instance, different treatments may be comparably effective for the average patient, yet some patients may improve more on a combination of treatments than a certain monotherapy”
4. This section has been edited and that sentence deleted.
5. These sentences were determined to not be necessary, and thus were deleted.

Reviewer Comment B:
Aim of the study
6. In my opinion, the aim of the study needs to be specified more clearly. I think the study focuses the question: What kind of patients really need combined treatments (because they benefit more from this kind of treatments than from psychotherapy or pharmacotherapy alone). If this is the main questions, it needs to be focused more clearly in the abstract, the introduction and the discussion, e.g.:

7. Page 2, line11-12: little is known… ◊ I think this sentence should focus combination treatment instead of the question which patient benefits from which treatment in general

8. Are there any hypotheses on relevant predictors? Of course, there may be little research on the correlation of patient characteristics and outcome of specific treatments, but – as you describe – there are – for example – some studies/guidelines recommending combined treatments for severe and

chronic depression. I wonder if you have any hypothesis on these predictor variables (severity, chronicity).

9. The pros and cons of IPD-meta-analyses are described in detail, but I think more information on possible clinical implications are necessary.

Response:
6. We have clarified our statement about our objective in the introduction to state, “The main objective of this meta-analysis is to determine which patients respond better to combined treatment (psychotherapy + pharmacotherapy) compared with monotherapies (pharmacotherapy, psychotherapy, or pill placebo monotherapy or psychotherapy versus pill placebo combination treatment).” (page 6 line 3). In the abstract, we state, “Therefore, this protocol outlines an individual patient data (IPD) meta-analysis to explore which patients, with which clinical characteristics, have better outcomes in combined treatment compared to psychotherapy (alone or with pill placebo), pharmacotherapy, and pill placebo.” (page 2, line 10). In the discussion, we state, “This project aims to contribute this knowledge of which patients respond best to which treatments, t to clinicians and researchers in the field of depression treatment.” (page 12, line 35).

7. This has been changed in the manuscript to state, “little is known about which specific groups of patients may respond best to combined treatment versus monotherapy.”(page 2).

8. This is now addressed under the section on moderators – page 7 line 27 – page 8 line 11. We explain the interest in clinical correlates of baseline severity and highlight for each individual moderator the previous literature that supports examining it.

9. We have added the following to the discussion, “Previous RCTs have not had sufficient power to thoroughly examine moderators. Thus, although we know depression treatments are equally effective, we do not know whether certain kinds of patients (for example, those who are older or more severe patients) will respond better to a specific type of treatment than another. Knowing which types of patients benefit more from combined treatment than monotherapy can both ensure that patients get the optimal treatment and relieve clinicians of the burden to choose the best treatment option for a given patient with very little information to inform that decision.” (page 13 line 9-16)

Reviewer Comment:
Methodology / General study approach:
10. Page 6, line 20/21: Please specify how you select the moderator variables.
11. Page 7, line 36: How will you operationalize “comorbid psychopathological disorders”: Number of comorbid disorders? Specific diagnosis?
Response:
10. This is also similar to the first reviewer’s comments and we have now specified more clearly how we choose the moderator variables to be tested on page 7 line 27 – page 8 line 11. Also see above for further response to reviewer 1.

11. This will be operationalized as having a comorbid diagnosis (and which one). It is possible that studies only examine Axis II disorders or comorbid anxiety disorders, and in this case, only these comorbid diagnoses will be included. The definition has been clarified in the paper as, “having a comorbid mental health diagnosis”.
Reviewer Comment:
12. Page 7: Timing of outcome assessments: What kind of variability do you mean? The variability in the length of the intervention? In this case, I think, the length should be a moderator (instead of addressing it in sensitivity analyses). Or do you mean the time between end of intervention and the assessment? I think, it should be specified a priori which period between end of intervention and

assessment you accept as “post intervention”. However, addressing this period is not a sensitivity analysis, unless you exclude studies with a predefined period longer than xxx.

13. Page 8, line 11-12: I think the literature search needs to be described more clearly: As you use a database of studies already included in other meta-analyses: Will you update the search? 1,756 full-text articles were identified for possible inclusion. But you already performed some meta-analyses on adult depression? Will you screen the 1,756 full-texts again for inclusion criteria in this study? Or did you already pre-select some studies? Please explain this procedure.
Response:
12. We meant the variability in the length of treatments or the time elapsed from baseline to post-treatment assessments. Most studies will lie within the range of 8-24 weeks, but studies that include acute phase follow-ups up to 36 weeks will be included. For this reason, we have added that we will accept trials with follow-up lengths between 5 and 36 weeks and that the length of follow-up will be controlled for in the regression analyses. We have also added that separate analyses will be conducted for acute phase versus extended follow-ups. (page 8, lines 13-18).

13. We have clarified that the articles that are included in the database were screened for inclusion in this meta-analysis. (page 8 line 31 -32).

Reviewer comment:
14. Page 8, line 23-27: Why will you exclude the two items on other sources of bias and selective outcome reporting? I did not understand why this is not possible especially in this meta-analysis.
15. Page 9, line 14-17: Please specify: What do you mean by “necessary data” and “when missing data are not excessive”? How many missing data (and in which variables) are allowed?

Response:
14. We have included these two items of the risk of bias assessment.

15. We mean that we are checking to make sure that there is not more missing data in the database received than in the published paper, as this can be a sign that the data has been corrupted. We have clarified this in the paper as, “…missing data are not excessive (relative to what is reported in the paper)” (page 9 line 33-34).

Reviewer comment:
16. Page 10, line 26/27: other study characteristics? Please specify.

17. Page 10, line 43-45: If you use predictor variables to impute missing outcome scores, the correlation between predictor variables and outcome scores will increase, please discuss this fact.

18. Page 11, line 32-35: In my opinion, these analyses are subgroup and not sensitivity analyses or will you exclude specific kinds of treatments?

Response:
16. This was changed to state that study characteristics such as bias score, type of recruitment, and other characteristics of the interventions will be entered as independent variables (ie – controlled for in the metaregression analysis) (page 11, line 4-5).

17. We follow recommended procedures for imputation. Experts have stated that if you do not include the predictor variables in your model, than the relationship between the variables imputed may be biased (White, Royston, Wood, 2010), therefore we need to include these predictor variables in the imputation. However, to be sure that this does not increase correlations, we will also run analysis on only the complete cases as a sensitivity analysis to the model utilizing imputation.

18. If there are only very few (1-3) studies utilizing certain types of psychotherapy then these may be excluded from this analysis and only the main types of psychotherapy included in the meta-analytic database will be examined (studies of 4 or greater). The data that are provided will dictate how these sensitivity or subgroup analyses are conducted. This has been changed in the manuscript to state, "Sensitivity analysis using individual types of psychotherapy alone will be conducted when there are at least 4 studies utilizing a particular psychotherapy. This analysis will explore whether moderators are specific to certain types of psychotherapies." (page 12, line 9-11)

Reviewer comment:
Further points
19. Page 2, line 22: The searches were already conducted, it would be better to write "searches were conducted…"
20. Page 7, line 28: I think "maximum absolute score reflecting normalization" is not clear enough to define remission, I would prefer explanations such as "no longer fulfilling diagnostic criteria for depression" or a score on depression outcome measures below a cut-off.
21. Page 7, line 29/30: What does "extreme response" mean? Please specify.
22. References: Reference 3: The title of the journal is missing. Reference 29: There is a new version of the Cochrane Handbook, please cite the current version.

Response:
19. We have changed this in the manuscript.

20. Our official definition of remission exists on page 11, and states, "If a sufficient number of trials incorporate HAM-D-17 scores, than a dichotomous variable indicating remission, defined as a HAM-D-17 score of ≤ 7, will be calculated and analyzed as an outcome." (page 12, line 3-5).

21. Extreme response is defined by the paper cited in the text. We will follow the previously defined definitions.

22. Thank you, we have corrected these references.

VERSION 2 – REVIEW

REVIEWER	Sarah Liebherz University Medical Center Hamburg-Eppendorf, Hamburg, Germany
REVIEW RETURNED	25-Oct-2016

GENERAL COMMENTS	The authors now have adressed all the reviewers' concerns.