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)	Meta-analysis of the effectiveness of psychological and medical treatments for binge-eating disorder (MetaBED): Study protocol
AUTHORS	Hilbert, Anja; Petroff, David; Herpertz, Stephan; Kersting, Anette; Pietrowsky, Reinhard; Tuschen-Caffier, Brunna; Vocks, Silja; Schmidt, Ricarda

VERSION 1 - REVIEW

REVIEWER	Michele Fornaro New York State Psychiatric Institute, USA
REVIEW RETURNED	29-Aug-2016

GENERAL COMMENTS	Title: “Meta-analysis of the effectiveness of psychological and medical treatments for binge-eating disorder (MetaBED): Study protocol” Type: (secondary [meta-analytic]) research) protocol Preliminary and specific considerations: 1) Language proficiency need proofreading by mother-tongue. Some minor twists and punctuation error all over the text. 2) References 20-22 (in the introduction): those references advocated in support of “rationale and background” are quite biased towards the psychological treatment of BED rather than the pharmacological treatment. Please add-up some more references and try to discern between those focusing on either (varying) pharmacological vs. (varying) psychological approaches, and/or side-by-side comparison treatment interventions. 3) It is recommended that the authors edit the introduction and background to provide a concise, yet accurate, overview of the essential changes introduced by the DSM-5 against the DSM-IV with reference to BED. Indeed, the shift of BED as a full-threshold category is a main nosological event having also major implications in terms of clinicians’ diagnostic sensitivity and subsequent treatment intervention (promptness to, and chances of future guidelines revisions as well as insurance coverage). Nonetheless, the changes about the actual duration and overall diagnostic criteria of BED in the DSM-5 over the DSM-IV need additional comment. Quoting the (open-access) passage from the introduction by Fornaro M. et al., 2016 about the impact of the changes introduced by the DSM-5, […”These higher figures would ultimately reflect a higher diagnostic sensitivity developed by the clinicians over the time, as well as the impact of more permissive criteria introduced by the Fifth Edition of the DSM (DSM-5),9 finally acknowledging BED as a distinct diagnostic category now characterized by reduced “frequency” (“once per week” instead of “twice per week”) and

	"duration" ("three months" vs "six months") criteria compared to the DSM-IV ones.9,12,13.]. I purposely kept the "referenced numbers" to ease you with the punctual, most updated, referenced to support how and why the updated DSM-5 codes would lead to much more permissive diagnostic criteria for BED over the DSM-IV (provisional category) codes. This would ultimately inflate or at least significantly increase the number of both psychological and psychopharmacological interventions (reimbursed and approved by the insurance companies…). 4) Is the planned study going to assess all the medication with evidence in support in the treatment of BED? If so, please bear in mind that the only FDA-approved drug to date for BED is lisdexamfetamine. And that a very updated meta-analysis about it (focusing of the extended approval) for BED already exists ("Lisdexamfetamine in the treatment of moderate-to-severe binge eating disorder in adults: systematic review and exploratory meta-analysis of publicly available placebo-controlled, randomized clinical trials". Fornaro M. et al., 2016). A rational extension and update of this latter meta-analysis specifically focusing on lisdexamfetamine for moderate-to-severe BED in adults should therefore either include additional moderators to be extracted from the original trials at pooling, and/or be able to include further original studies to pool, which are not disclosed at present time. The (open-access) meta-analytic report by Fornaro M. et al., 2016 also includes a quick overview in the introduction about the impact of the changes introduced by the DSM-5 over the DSM-IV, and I strongly encourage the authors to review it and reference for rationale and background. This latter meta-analytic report is also much accurate in terms of inquired databases, which is a issue I will expand further in the upcoming remarks for your convenience. 5) If going to include other pharmacological treatments with some sort of evidence for BED, e.g. some other stimulants, mood stabilizers or antidepressants, please bear in mind that (due to a diagnostic shift introduced by the DSM-5 updated codes for BED), none of these latter would be FDA-approved. This should be accounted when setting-up your dataset and multiple sheets for meta-analysis, ideally including an additional moderator (FDA vs. non-FDA approved treatment; which would obviously include lisdexamfetamine as the only positive coding). Adding "year of publication" and/or "year of trial registration" will also contribute to explain why a concrete publication bias may exist on the matter. 6) As per DSM-5 codes, BED can occur either in adult and non-adult population. You should therefore set a specific population target. Likewise, varying degree of severity of BED have been coded. Please bear in mind that moderate-to-severe BED (in adults) is the population who received the approval and is going to received most attention from future trials leading to extended approval of previous or ex-novo approval for any eventual novel compound to come. 7) As per FDA recommendation, the pharmacological treatment of BED should be part of a multidisciplinary treatment approach. In this view, your meta-analysis is going to be a very relevant one, ideally including moderators to compare univocal vs. integrate treatment modalities. Yet again, current publication bias about approved medication due to recent DSM-5 diagnostic shift would preclude inclusion of a representative number of alternative combinations to test. 8) About the database you are going to inquire, please consider that stating the novel full-threshold category of BED in the DSM-5, your search strategy should be significantly expanded. Yet again, the method and the adapted PRISMA flowchart from the pivotal (open-

	access) article by Fornaro M., 2016 should be referenced and replicated. You may nonetheless add-up some more psychologically-oriented databases since the goal of your planned study is to expand the work beyond the sole lisdexamfetamine (or alternative drugs, if ever going to come any approved one any time soon). Negative results indexes must be considered, though at least for the only FDA-approved drug for BED, a serious sponsorship bias exists… 9) PRISMA guidelines are fine, yet I would encourage the authors to merge with the MOOSE guidelines too. Please refer to: Stroup DF, Berlin JA, Morton SC, et al, for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group. Meta-analysis of Observational Studies in Epidemiology. A Proposal for Reporting. JAMA. 2000;283(15):2008-2012. doi: 10.1001/jama.283.15.2008. Also note that P-PRISMA (2015) is more updated than PRISMA (2009), yet the latter is still more user-friendly and comprehensive when it comes to set-up of the planned systematic review (and meta-analysis). 10) About the planned effect sizes to extract from the original studies: your premises are find, yet you may want to consider how and what specific effect size to compute then. Indeed, many measures of effect size have been proposed and the most common are Pearson's correlation coefficient, r, Cohen's d, and the odds ratio (OR). However, there may be reasons to prefer unstandardized effect size measures. No matter what, it is the clinical question and study outcome which should be better expanded beyond "pre- and post-treatment" variables in my opinion. 11) Which method(s) are you going to use and why? Basically, the fixed-effect model assumes that studies in the meta-analysis are sampled from a population in which the average effect size is fixed or can be predicted from a few predictors. The alternative assumption is that the average effect size in the population varies randomly from study to study: studies in a meta-analysis come from populations that have different average effect sizes, so population effect sizes can be thought of as being sampled from a "super-population" (Hedges, 1992). I assume a random-model would better fit in this case, yet you should better discuss this prior starting your meta-analytic report. 12) About participants: if going to include only cases with a DSM-5 diagnosis of BED, you will end in a more homogeneous, yet too tiny set. I would therefore consider including also DSM-IV ones and properly discriminate the adopted diagnostic codes using an ad-hoc moderator. ICD cases should also be accounted. 13) Duplicate records should be checked owing to the PRISMA and/or automatic screening (endnote method) along with manual screening. 14) About the language restriction: I would strongly encourage to either screen English only or, alternatively, any language ("no language restriction"). German literature on the topic could be covered, but unless accounted for, systematic exclusion of alternative languages would introduce a systematic bias too. "Fortunately" enough, I do not expect any high number of hits in languages other than English on this topic, so you should easily screen that handful or records at the abstract (English language anyway) level. In case any suggestive records would deserve further assessment, you may seek for support for an international colleague of expert reader. 15) Similarly, despite the relative (appendix) introduction of BED by the DSM (Fourth edition, and its subsequent text revision), no date of publication restriction should apply.

	Finally, I took the liberty to attach a concise synthesis of general, yet accurate, guidance for either planning or reviewing (into advance) systematic review and meta-analysis. I am confident the authors would find some additional hints beyond the afore mentioned ones: 16) “How to do a meta-analysis” Field A. and Gillett R., 2010 (British Journal of Mathematical and Statistical Psychology). 17) “How to review a meta-analysis” Russo M., 2007 (Gastroenterology and Hepatology). 18) “Ho to read a Systematic Review and Meta-analysis and Apply the Results to Patient Care – User’s Guides to the Medical Literature” by Murad M.H. et al., 2014 (Clinical Review & Education). 19) “Systematic reviews and meta-analyses: An illustrated, step-by-step guide” Madhukar P. et al., 2004 (The National Medical Journal of India). To sum-up, this is going to be an extremely relevant contribute, with considerable impact on the clinical practice and fair citation potential if properly run. I did my best to provide the authors with the most exhaustive guidance I could, including some pivotal reference they should integrate and owe for their research project. Please bear in mind that a strong publication bias still exists at this time due to the recent introduction of BED as a full-threshold category (and therefore either psychological or pharmacological approved treatments are virtually none – with the exception of lisdexamfetamine). Unless properly discriminated (multiple Excel sheets with many moderators), this would otherwise introduce a major apple and orange bias, which the authors are prompted to avoid in order to provide a clinically informative and sound report. Thank you for the opportunity of assisting in this.

REVIEWER	Gianluca Castelnuovo 1 Department of Psychology, Catholic University of Milan, Italy 2 Istituto Auxologico Italiano IRCCS, Psychology Research Laboratory, Ospedale San Giuseppe, Verbania, Italy
REVIEW RETURNED	14-Oct-2016

GENERAL COMMENTS	The manuscript is well written and the protocol is of good quality. I suggest to specify in the methods section how the information about methodological quality and equity will be treated during data analysis, since this is stated only in the abstract. It could be useful to clarify why you are planning to use two tools for the quality assessment. Moreover I do not agree about pooling the data of non controlled trials with the data of RCTs and not randomized trials. Indeed, not controlled trials may be severely biased and both the Cochrane and the EPOC tools are not appropriate for their evaluation. As a consequence, the power analysis is insufficient. We suggest to separate these data in the analysis. Finally, it is theoretically questionable to create an index of general pathology summing different conditions such as anxiety, depression and various other symptoms.

REVIEWER	Nancy D Berkman RTI International USA

REVIEW RETURNED	21-Oct-2016

GENERAL COMMENTS	MA of the effectiveness of psychological and medical treatments for binge-eating disorder (MetaBED): study protocol Overall, I found the protocol to be straightforward and complete. I have several comments intended to clarify several points. 1. Pg. 3 of 23, line 45. The authors state that they will conduct a moderator analysis with consideration of equity aspects. Please further describe and clarify what that analysis would entail. 2. Pg. 8 of 23, line 28. Participants include those with DSM-IV or DSM-5 diagnoses of BED. A parenthetical states (including BED of low frequency and/or limited duration). I found this unclear. Are the authors intending to include participants who do not meet DSM criteria? If not, I would remove the parenthetical. If they are, the more relaxed criteria should be more clearly specified. 3. Pg. 8 of 23, line 45. Please further clarify whether you will limit the analysis of bariatric surgery just to BED patients--those who received a diagnosis of BED prior to the surgery or if you will be also/rather examining loss-of-control eating post-surgery. 4. Pg. 10 or 23, line 32. If you are just at the protocol stage, I would expect the final search to be later than January 2016. This date would likely make the search “cold” by the date of final result/publication. 5. Pg. 12 of 23. Because you are planning to include non-RCT analyses, risk of bias will need to include control for confounding and issues of increased vulnerability to various types of selection bias. I recommend that you include a risk of bias tool that is designed to evaluate these non-RCT designs. Consider using the new Cochrane ROBINS-I tool.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Michele Fornaro
Institution and Country: New York State Psychiatric Institute, USA Competing Interests: None declared

Preliminary and specific considerations:

1) Language proficiency need proofreading by mother-tongue. Some minor twists and punctuation error all over the text.
Response: Thank you very much for this comment. We had the manuscript proofread by a native English speaker again before resubmission.

2) References 20-22 (in the introduction): those references advocated in support of “rationale and background” are quite biased towards the psychological treatment of BED rather than the pharmacological treatment. Please add-up some more references and try to discern between those focusing on either (varying) pharmacological vs. (varying) psychological approaches, and/or side-by-side comparison treatment interventions.
Response: As recommended, we updated our reference list with recent references for pharmacological or combined treatments (p. 7). The study is intended as a meta-analysis comparing diverse approaches to the treatment of BED. Therefore we kept the description of pharmacological treatment fairly short. Treatment category (e.g., pharmacotherapy, combined treatment) or the impact of a single treatment (e.g., specific drugs) will be addressed in the moderator analysis (p. 15).

3) It is recommended that the authors edit the introduction and background to provide a concise, yet accurate, overview of the essential changes introduced by the DSM-5 against the DSM-IV with reference to BED. Indeed, the shift of BED as a full-threshold category is a main nosological event having also major implications in terms of clinicians' diagnostic sensitivity and subsequent treatment intervention (promptness to, and chances of future guidelines revisions as well as insurance coverage). Nonetheless, the changes about the actual duration and overall diagnostic criteria of BED in the DSM-5 over the DSM-IV need additional comment.
Quoting the (open-access) passage from the introduction by Fornaro M. et al., 2016 about the impact of the changes introduced by the DSM-5, […"These higher figures would ultimately reflect a higher diagnostic sensitivity developed by the clinicians over the time, as well as the impact of more permissive criteria introduced by the Fifth Edition of the DSM (DSM-5),9 finally acknowledging BED as a distinct diagnostic category now characterized by reduced "frequency" ("once per week" instead of "twice per week") and "duration" ("three months" vs "six months") criteria compared to the DSM-IV ones.9,12,13.]. I purposely kept the "referenced numbers" to ease you with the punctual, most updated, referenced to support how and why the updated DSM-5 codes would lead to much more permissive diagnostic criteria for BED over the DSM-IV (provisional category) codes. This would ultimately inflate or at least significantly increase the number of both psychological and psychopharmacological interventions (reimbursed and approved by the insurance companies…).
Response: As suggested, we included a description of the changes of the diagnostic criteria according to DSM-IV and DSM-5 in the Introduction section (p. 5). Thank you very much for your support.

4) Is the planned study going to assess all the medication with evidence in support in the treatment of BED? If so, please bear in mind that the only FDA-approved drug to date for BED is lisdexamfetamine. And that a very updated meta-analysis about it (focusing of the extended approval) for BED already exists ("Lisdexamfetamine in the treatment of moderate-to-severe binge eating disorder in adults: systematic review and exploratory meta-analysis of publicly available placebo-controlled, randomized clinical trials". Fornaro M. et al., 2016). A rational extension and update of this latter meta-analysis specifically focusing on lisdexamfetamine for moderate-to-severe BED in adults should therefore either include additional moderators to be extracted from the original trials at pooling, and/or be able to include further original studies to pool, which are not disclosed at present time.
The (open-access) meta-analytic report by Fornaro M. et al., 2016 also includes a quick overview in the introduction about the impact of the changes introduced by the DSM-5 over the DSM-IV, and I strongly encourage the authors to review it and reference for rationale and background. This latter meta-analytic report is also much accurate in terms of inquired databases, which is a issue I will expand further in the upcoming remarks for your convenience.
Response: Thank you very much for your comments. The publication of your meta-analysis fell in the period of the last turnaround to the submission of our study protocol. Of course, we included it in the revised manuscript. As described on pp. 3, 7, 8, the planned meta-analysis will assess randomized, non-randomized controlled, and uncontrolled trials on psychological and medical treatments that have been used in BED, including all pharmacological agents, not only lisdexamfetamine. The study focus is thus different from your meta-analysis so that the overlap with your study should not limit the value of the planned meta-analysis. Moderators of outcome were specified on p. 15.

5) If going to include other pharmacological treatments with some sort of evidence for BED, e.g. some other stimulants, mood stabilizers or antidepressants, please bear in mind that (due to a diagnostic shift introduced by the DSM-5 updated codes for BED), none of these latter would be FDA-approved. This should be accounted when setting-up your dataset and multiple sheets for meta-analysis, ideally including an additional moderator (FDA vs. non-FDA approved treatment; which would obviously include lisdexamfetamine as the only positive coding). Adding "year of publication" and/or "year of trial registration" will also contribute to explain why a concrete publication bias may exist on the matter.

Response: Please note that single medications will be addressed in the moderator analysis (p. 15). Reviewer #1 is correct that a specific emphasis should be placed on medications with versus without approval by the FDA, the European Medicines Agency, etc.

6) As per DSM-5 codes, BED can occur either in adult and non-adult population. You should therefore set a specific population target. Likewise, varying degree of severity of BED have been coded. Please bear in mind that moderate-to-severe BED (in adults) is the population who received the approval and is going to received most attention from future trials leading to extended approval of previous or ex-novo approval for any eventual novel compound to come.
Response: Reviewer #1 is correct that BED can occur in youth, and a few treatment trials in young populations have been published. We addressed this issue by including age as a moderator (p. 15). Likewise, the severity of BED is addressed using baseline number of binge-eating episodes as a moderator (p. 15).

7) As per FDA recommendation, the pharmacological treatment of BED should be part of a multidisciplinary treatment approach. In this view, your meta-analysis is going to be a very relevant one, ideally including moderators to compare univocal vs. integrate treatment modalities. Yet again, current publication bias about approved medication due to recent DSM-5 diagnostic shift would preclude inclusion of a representative number of alternative combinations to test.
Response: The question of single versus combined treatment is addressed in the moderator analysis, as described on p. 15.

8) About the database you are going to inquire, please consider that stating the novel full-threshold category of BED in the DSM-5, your search strategy should be significantly expanded. Yet again, the method and the adapted PRISMA flowchart from the pivotal (open-access) article by Fornaro M., 2016 should be referenced and replicated. You may nonetheless add-up some more psychologically-oriented databases since the goal of your planned study is to expand the work beyond the sole lisdexamfetamine (or alternative drugs, if ever going to come any approved one any time soon). Negative results indexes must be considered, though at least for the only FDA-approved drug for BED, a serious sponsorship bias exists…
Response: We cited your meta-analysis in the revision of our manuscript (p. 7). As described on p. 7, a PRISMA flowchart will be completed in our planned meta-analysis. We do not see a specific need to replicate your PRISMA flowchart as your study included a part of the studies that our more comprehensive meta-analysis will include. The databases for our search do include psychological databases, as described on p. 11. Of course, all results, positive and negative, will be coded. We will consider potential publication biases as described (see Assessment of Reporting Biases, p. 15).

9) PRISMA guidelines are fine, yet I would encourage the authors to merge with the MOOSE guidelines too. Please refer to: Stroup DF, Berlin JA, Morton SC, et al, for the Meta-analysis Of Observational Studies in Epidemiology (MOOSE) Group. Meta-analysis of Observational Studies in Epidemiology. A Proposal for Reporting. JAMA. 2000;283(15):2008-2012. doi: 10.1001/jama.283.15.2008.
Also note that P-PRISMA (2015) is more updated than PRISMA (2009), yet the latter is still more user-friendly and comprehensive when it comes to set-up of the planned systematic review (and meta-analysis).
Response: Thank you very much for your recommendation. We included the MOOSE guidelines in the Introduction section and corrected our reference to PRISMA-P instead of that to PRISMA in the text (p. 7). In fact, we already had used the PRISMA-P checklist for our submission, not the PRISMA checklist.

10) About the planned effect sizes to extract from the original studies: your premises are find, yet you may want to consider how and what specific effect size to compute then. Indeed, many measures of

effect size have been proposed and the most common are Pearson's correlation coefficient, r, Cohen's d, and the odds ratio (OR). However, there may be reasons to prefer unstandardized effect size measures. No matter what, it is the clinical question and study outcome which should be better expanded beyond "pre- and post-treatment" variables in my opinion.
Response: Thank you very much. On p. 14, it is now more clearly explained that our choice of difference in means and not standardized means was made since the latter can be difficult to interpret and since outcomes are generally measured on the same scale or in very similar ways. The pooled result will then also be translated into Cohen's d for comparative purposes.

11) Which method(s) are you going to use and why? Basically, the fixed-effect model assumes that studies in the meta-analysis are sampled from a population in which the average effect size is fixed or can be predicted from a few predictors. The alternative assumption is that the average effect size in the population varies randomly from study to study: studies in a meta-analysis come from populations that have different average effect sizes, so population effect sizes can be thought of as being sampled from a "super-population" (Hedges, 1992). I assume a random-model would better fit in this case, yet you should better discuss this prior starting your meta-analytic report.
Response: Under the heading "Measures of Treatment Effect" we had specified that "Random effects models (general and generalized linear mixed models) will be used and compared with fixed effects models as a sensitivity analysis." To make it easier to locate this specification of methods, we have now changed the sub-heading to read "Statistical Models and Measures of Treatment Effect" (p. 13).

12) About participants: if going to include only cases with a DSM-5 diagnosis of BED, you will end in a more homogeneous, yet too tiny set. I would therefore consider including also DSM-IV ones and properly discriminate the adopted diagnostic codes using an ad-hoc moderator. ICD cases should also be accounted.
Response: Thank you. This is actually what we are planning to do; please see the moderator variables (p. 15). We agree that an exclusive focus on DSM-5 diagnosis of BED would result in a database that is too small for valid meta-analytical estimation of effects.

13) Duplicate records should be checked owing to the PRISMA and/or automatic screening (endnote method) along with manual screening.
Response: Duplicates were checked both automatically (using endnote) and manually. This was now more clearly stated in the text (p. 11). The number of duplicates will be included in the PRISMA flowchart.

14) About the language restriction: I would strongly encourage to either screen English only or, alternatively, any language ("no language restriction"). German literature on the topic could be covered, but unless accounted for, systematic exclusion of alternative languages would introduce a systematic bias too. "Fortunately" enough, I do not expect any high number of hits in languages other than English on this topic, so you should easily screen that handful or records at the abstract (English language anyway) level. In case any suggestive records would deserve further assessment, you may seek for support for an international colleague of expert reader.
Response: As described on p. 8, our meta-analysis will form the basis of the renewal of the German evidence-based S3 Guidelines of Diagnosis and Treatment of Eating Disorders, specifically BED. This is the reason why we initially searched for studies in both English and German. German-language studies are important because there might be treatment approaches with evidence published in German only, but which are relevant to the care of BED in German-speaking countries. However, considering your valuable comment, we decided to describe the German studies in the guidelines text only, not in the meta-analysis, and changed the inclusion criteria for the meta-analysis, now including studies published in English only (p. 10).

15) Similarly, despite the relative (appendix) introduction of BED by the DSM (Fourth edition, and its

subsequent text revision), no date of publication restriction should apply.
Response: As now more clearly specified on p. 10, there is no lower limit of study inclusion.

Finally, I took the liberty to attach a concise synthesis of general, yet accurate, guidance for either planning or reviewing (into advance) systematic review and meta-analysis. I am confident the authors would find some additional hints beyond the aforementioned ones:
16) “How to do a meta-analysis” Field A. and Gillett R., 2010 (British Journal of Mathematical and Statistical Psychology).
17) “How to review a meta-analysis” Russo M., 2007 (Gastroenterology and Hepatology).
18) “Ho to read a Systematic Review and Meta-analysis and Apply the Results to Patient Care – User’s Guides to the Medical Literature” by Murad M.H. et al., 2014 (Clinical Review & Education).
19) “Systematic reviews and meta-analyses: An illustrated, step-by-step guide” Madhukar P. et al., 2004 (The National Medical Journal of India).
Response: Thank you very much for these very helpful references!

To sum-up, this is going to be an extremely relevant contribute, with considerable impact on the clinical practice and fair citation potential if properly run. I did my best to provide the authors with the most exhaustive guidance I could, including some pivotal reference they should integrate and owe for their research project. Please bear in mind that a strong publication bias still exists at this time due to the recent introduction of BED as a full-threshold category (and therefore either psychological or pharmacological approved treatments are virtually none – with the exception of lisdexamfetamine). Unless properly discriminated (multiple Excel sheets with many moderators), this would otherwise introduce a major apple and orange bias, which the authors are prompted to avoid in order to provide a clinically informative and sound report.
Thank you for the opportunity of assisting in this.
Response: We would like to thank Reviewer #1 for the very thorough and comprehensive review!

Reviewer: 2
Reviewer Name: Gianluca Castelnuovo
Institution and Country: 1 Department of Psychology, Catholic University of Milan, Italy; 2 Istituto Auxologico Italiano IRCCS, Psychology Research Laboratory, Ospedale San Giuseppe, Verbania, Italy Competing Interests: No

The manuscript is well written and the protocol is of good quality. I suggest to specify in the methods section how the information about methodological quality and equity will be treated during data analysis, since this is stated only in the abstract. It could be useful to clarify why you are planning to use two tools for the quality assessment.
Response: Thank you very much for your comment. We corrected the abstract, now saying that “Study search, selection, and data extraction, including risk of bias assessment, will be independently performed by two reviewers and consensus will be sought. Moderator analyses will be conducted, and equity aspects will be considered.” In the Methods section, it was specified that we are planning to use the Effective Practice and Organization of Care’s definition of risk of bias for non-randomized and uncontrolled studies in our meta-analysis. The Cochrane Collaboration’s Risk of Bias Tool addresses randomized-controlled studies only (p. 12). Further, it is now specified that equity aspects will be considered in the moderator analyses if indicated (p. 13).

Moreover I do not agree about pooling the data of non controlled trials with the data of RCTs and not randomized trials. Indeed, not controlled trials may be severely biased and both the Cochrane and the EPOC tools are not appropriate for their evaluation. As a consequence, the power analysis is insufficient. We suggest to separate these data in the analysis.
Response: We agree with you that it is difficult to pool randomized-controlled designs with other trial

designs. We now re-organized the Statistical Models and Measures of Treatment Effect section (p. 13), specifying that “two types of analyses will be performed. First, the treatment effect will be compared for treatment versus non-interventional and placebo controls. These analyses will be performed for each treatment category separately (see below), and only RCTs will be included. Second, the treatment effect will be estimated within each treatment category, where all types of studies will be included.” Because the preceding German clinical guidelines meta-analysis by Vocks et al. (2010) compiled randomized, non-randomized controlled, and uncontrolled designs in second within-treatment analyses and because of a potential gain in external validity associated with uncontrolled designs (despite a lowered internal validity), we would like to maintain our strategy to estimate the treatment effect in within-treatment analyses, based on all study designs. A moderator analysis will address study designs separately (p. 15), and the power analysis planned in conjunction will shed light onto how feasible it is to take into account putative confounders.

Finally, it is theoretically questionable to create an index of general pathology summing different conditions such as anxiety, depression and various other symptoms.
Response: After considering this very helpful comment, we decided to focus on depression instead of a composite of general psychopathology, and specified this in the text (p. 9).

Reviewer: 3
Reviewer Name: Nancy D Berkman
Institution and Country: RTI International, USA Competing Interests: None declared

Overall, I found the protocol to be straightforward and complete. I have several comments intended to clarify several points.

1. Pg. 3 of 23, line 45. The authors state that they will conduct a moderator analysis with consideration of equity aspects. Please further describe and clarify what that analysis would entail.
Response: Thank you. We corrected the abstract, now stating that “Moderator analyses will be conducted, and equity aspects will be considered.” (p. 3) Equity aspects will be addressed in the moderator analysis, for example, age, sex, or ethnicity, which is now more clearly indicated in the text (p. 13).

2. Pg. 8 of 23, line 28. Participants include those with DSM-IV or DSM-5 diagnoses of BED. A parenthetical states (including BED of low frequency and/or limited duration). I found this unclear. Are the authors intending to include participants who do not meet DSM criteria? If not, I would remove the parenthetical. If they are, the more relaxed criteria should be more clearly specified.
Response: Thank you very much for this important comment. We aim to include participants with DSM-IV or DSM-5 diagnosis of BED and with DSM-5 diagnosis of BED of low frequency and/or limited duration. The text was modified according to your suggestion (p. 8).

3. Pg. 8 of 23, line 45. Please further clarify whether you will limit the analysis of bariatric surgery just to BED patients--those who received a diagnosis of BED prior to the surgery or if you will be also/rather examining loss-of-control eating post-surgery.
Response: We would like to thank you for this very helpful comment. We intend to include patients with a diagnosis of BED prior to treatment, including bariatric surgery only. This was more clearly described on p. 8. On p. 9, we wrote that as bariatric surgery limits the possible amount of food intake, episodes of loss of control eating will be considered among the primary outcomes where appropriate.

4. Pg. 10 or 23, line 32. If you are just at the protocol stage, I would expect the final search to be later than January 2016. This date would likely make the search “cold” by the date of final

result/publication.
Response: We fully agree with you. We clarified on p. 10 that the searches will be re-run before the final analyses and further studies will be retrieved for inclusion prior to submission of the meta-analytical report. January 2016 was the inclusion deadline of our first search.

5. Pg. 12 of 23. Because you are planning to include non-RCT analyses, risk of bias will need to include control for confounding and issues of increased vulnerability to various types of selection bias. I recommend that you include a risk of bias tool that is designed to evaluate these non-RCT designs. Consider using the new Cochrane ROBINS-I tool.
Response: Thank you very much for referring us to the new Cochrane ROBINS-I tool which was published after submission of our study protocol to BMJ Open. We are already using the Effective Practice and Organization of Care's definition of risk of bias for non-randomized and uncontrolled studies in our meta-analysis, in addition to the Cochrane Collaboration's Risk of Bias Tool for randomized-controlled studies. As the ROBINS-I tool raises a number of additional issues, we specified in the text that we will also consider this tool (p. 12).

VERSION 2 – REVIEW

REVIEWER	Gianluca Castelnuovo Catholic University of Milan IRCCS Istituto Auxologico Italiano
REVIEW RETURNED	15-Dec-2016

GENERAL COMMENTS	Unfortunately, as we have already pointed out, none of quality assessment tools you are planning to use is suitable for the evaluation of non-controlled trials. The EPOC is appropriate for controlled studies or for interrupted - series studies and would reject these trials. Therefore, you will not be able to assess if the methodological quality affects the results across study designs. A possible partial solution is to modify your tools adding other items, as other authors did (e.g. doi:10.1186/s12913-016-1816-5).

VERSION 2 – AUTHOR RESPONSE

Reviewer: 2
Reviewer Name: Gianluca Castelnuovo
Institution and Country: Catholic University of Milan, IRCCS Istituto Auxologico Italiano, Italy
Competing Interests: None declared

Unfortunately, as we have already pointed out, none of quality assessment tools you are planning to use is suitable for the evaluation of non-controlled trials. The EPOC is appropriate for controlled studies or for interrupted - series studies and would reject these trials. Therefore, you will not be able to assess if the methodological quality affects the results across study designs. A possible partial solution is to modify your tools adding other items, as other authors did (e.g. doi:10.1186/s12913-016-1816-5).

Response: Thank you very much for your comment which adds to the previous comment #5 of Reviewer #3 in the first revision. In the Methods section, it is now more clearly specified that we are planning to use the Cochrane Collaboration's Risk of Bias Tool[33] and the Effective Practice and Organization of Care Risk of Bias Tool[34] to assess the risk of bias in randomized-controlled studies as well as non-randomized studies, with additional items for assessing the risk of bias in uncontrolled studies, for example, considering the newly published Risk of Bias in Non-randomized Studies of Interventions (ROBINS-I) tool.[35] (page 12)