Author’s response to reviews

Title: Prospective, randomized, double-blind, placebo-controlled phase IIa clinical trial on the effects of an estrogen-progestin combination as add-on to inpatient psychotherapy in adult female patients suffering from anorexia nervosa

Authors:
georgios paslakis (georgios.paslakis@uk-erlangen.de)
stefanie maas (Stefanie.Maas@uk-erlangen.de)
bernd gebhardt (bernd.gebhardt.extern@uk-erlangen.de)
manfred rauh (Manfred.Rauh@uk-erlangen.de)
Erim Yesim (Yesim.Erim@uk-erlangen.de)
Andreas Mayr (andreas.mayr@fau.de)

Version: 2 Date: 22 Dec 2017

Author’s response to reviews:

John Reece (Reviewer 1): This trial protocol describes a clinically important, scientifically interesting, and well-design randomised controlled trial to assess the effect of hormonal treatment on a range of psychological, physiological, and neuropsychological outcomes in women with a diagnosis of anorexia nervosa. This is an important trial that has the potential to inform a new branch of clinical intervention for sufferers of this disorder. I applaud the researchers on their efforts. This is a very challenging trial, and the authors have shown great care in presenting such a rigorous trial protocol. I look forward to the publication of the results of this trial. I am happy to recommend that this protocol be accepted for publication with the following minor required amendments:

Thank you very much! We appreciate these encouraging words.

1) Line 44-45. I don’t know the literature personally, but I was very surprised to read that there are no evidence-based psychopharmacological interventions for the treatment of AN. I'm not questioning the authors assertion; nor am I requiring an amendment. But I am very surprised to read this. Given the prevalence of AN, it surprises me that drug treatment has not been investigated.

There have been studies with regard to psychopharmacological interventions in AN, however, none of the studies has shown that psychopharmacological treatments (e.g. using antidepressants or second generation antipsychotics etc.) are associated with significant benefits for patients
suffering from AN. Thus, there are indeed no evidence-based psychopharmacological interventions for the treatment of AN.

2) Lines 61 - 63: I accept the outcomes and summary of the Watson review, but surely there have been other relevant meta-analyses that can be briefly reported here. These should be briefly summarised in this section. If no such publications exist, then this should be noted.

There are just few empirically-supported treatments for AN. International clinical guidelines recommend psychological treatments showing strong empirical support, although some recommendations are non-evidence based and likely reflect the particularities in healthcare systems (Hilbert, Hoek, & Schmidt, 2017: Evidence-based clinical guidelines for eating disorders: International comparison. Current Opinion in Psychiatry, 30(6), 423–437). Cognitive behavioral therapy (CBT) has been demonstrated as efficacious (National Institute of Clinical Excellence, 2017: Eating disorders: Recognition and treatment. Available at: www.nice.org.uk/guidance/ng69).

We have now added the following information (please see “introduction”):

“There are just few empirically-supported treatments for AN.”

“According to a recent systematic review of third-wave behavior therapies for the treatment of AN, none of the third-wave therapies meet criteria for an empirically-supported treatment for AN; thus, patients with AN should be treated with cognitive behavioral therapy, with interpersonal therapy considered an alternative (Linardon et al. 2017).”

3) Line 246-267: The exclusion criteria, while well-justified, are very restrictive! How can the sample size be guaranteed with such extensive exclusion criteria?

Exclusion criteria are dictated by the contraceptive “pill” itself and are mostly related to conditions that would predispose for occurrence of thrombosis or liver function failure. Besides, this is the first study examining the application of the “pill” for this new indication (anorexia nervosa), thus setting high requirements for safety and tolerability (which have also been defined as outcomes).

As the study is ongoing, we can actually say that it is not the extensive exclusion criteria making recruitment more difficult than expected, but rather the unwillingness of patients towards intake of a drug that would potentially be associated with weight gain.

4) Section 285: Who will govern the randomisation and how will it be done? Computer-based random numbers? Coin toss? And who will control this?
Prior to study initiation, a randomization list was generated by an external independent person (staff member of the Department of Clinical Pharmacology) as a block randomization (block size n = 6) using the nQuery software.

This information has been added to the manuscript.

“Prior to study initiation, a randomization list was generated by an external independent person using the nQuery software (Statistical Solutions Ltd., Cork, Ireland) and a block randomization design (block size n = 6). The assignment of medication to study participants is carried out by two persons, one acting as the operator and the second acting as the checker.”

5) Line 241: I have a number of suggested minor amendments in the statistics section.

- No mention is made of the calculation and reporting of effect size measures and measures of clinical significance. Both are relevant to a trial of this type and should be included.

We have added the following (please see “statistics”):

“Effect sizes will be reported based on Cohen’s d for group differences for primary and secondary outcomes. Clinical improvements, amelioration in neuropsychological performance and restitution of hormonal aberrancies are all to be considered of clinical significance for patients with AN.”

- Why has an a priori decision been made not to consider any form of data imputation as part of the intention to treat approach? If the amount of missing data is low and meets the necessary conditions, why would some form of imputation not be considered? Please explain this approach.

We thank the reviewer for this important comment. We re-thought our planned dealing with missing data and now consider the imputation of missing values for the cross-sectional analyses following the ITT principle. We plan to impute missing outcome variables for the cross-sectional analysis at end of treatment with the corresponding baseline value, which can be considered as a rather conservative approach. For the longitudinal analyses, however, we already had planned to apply mixed effect regression analysis, which is particularly well-suited to deal with single missing outcome measures. As a consequence of this modelling strategy, single missing values for a patient do not lead to a removal of this patient from the analyses, but all available data from each randomized patient enters in the final model.

We have changed the corresponding paragraph in the revised manuscript (please see “statistics”).

- Why has the decision been made to use non-parametrics to compare baseline values? If the data meet the assumptions, I would have thought that parametric procedures would be more appropriate. Please explain this approach.

Thank you for bringing up this issue. We are planning to apply Fisher’s exact test for categorical variables and non-parametric Mann-Whitney tests for metric variables to assess the differences
between the two treatment groups at baseline, because we feel that the relatively small sample size (n = 25 in each group) does not really justify the use of statistical tests, where the test-statistic follows only asymptotically the corresponding distribution (e.g., chi^2 test or t-tests). We agree that in cases where the necessary assumptions are truly met (e.g., the outcome follows exactly a Gaussian distribution) the parametric tests could provide a higher statistical power. However, we do not see that for our variables and therefore prefer to stick to the non-parametric or exact tests. Additionally, please note that applying statistical tests to compare baseline differences in RCTs in general has a rather exploratory character, as we know for sure that both groups belong to the same population which corresponds to the null hypothesis. The tests outcome (e.g., the corresponding p-values) hence only have a descriptive meaning, to check if the randomization led to the desired results.

Following the reviewers comment, we now justify our decision in the revised manuscript (please see “statistics”).

- The power analysis is unconvincing. The use of a trait anxiety measure as the basis for the sample size estimation seems inappropriate given the primary outcomes of the study. A further attempt should be made to drive the power analysis from the perspective of the primary outcomes, even if it means extrapolating from another clinical sample. One needs to be creative when conducting a power analysis in an area where there is little research in the specific area.

Another approach is to drive the power analysis from the perspective of the available sample size. If the maximum possible sample is fixed and known, which is often the case, conduct the power analysis with the sample size as a fixed parameter and provide the power of reliably identifying effects of varying magnitude; for example, "with a sample of this size, small effects will be observed with a power of X, moderate effects with a power of Y, and large effects with a power of Z". Then make every reasonable and justifiable attempt to argue for why you will observe effects that have a reasonable power of being identified.

In addition, provide more detail around the power analysis: the type of test being used, is it one- or two-tailed, at what time point in the trial are these power calculations related to?

I accept the challenge of this exercise, and I understand the Phase IIa limitations in conducting sample size estimation, but I would contend that a little more can be done in this part of the protocol.

We thank the reviewer for this important comment and the constructive suggestions. We re-arranged the power analysis following the suggestions of the reviewer and now computed the statistical power under different conditions (different effect sizes for a fixed n = 25 in both groups). Additionally, we provide more details about the calculations (please see “statistics”).

With these suggested amendments addressed appropriately, I am happy to recommend acceptance of this protocol for publication.

Jennifer Jordan (Reviewer 2):
1. **Will the study design adequately test the hypothesis?**

This is an ambitious study examining the potential impact of an oestrogen replacement treatment being trialed as a novel treatment for broad range of outcomes in 50 women with anorexia nervosa. The authors do not specify hypotheses, but rather exploratory aims.

The sample size is on the small size but justified by a power calculation taken from the Misra et al (2013) study which found that trait anxiety (but not eating disorder related variables) was improved after an oestrogen replacement therapy. In the current protocol, however anxiety is a secondary outcome and the study is not powered for neuropsychological changes (the primary outcome) - these authors state that there are no existing studies on which to base such a power calculation. I wonder if it might have been possible to look at clinically meaningful neuropsychological changes following other treatments in anorexia nervosa e.g. for cognitive remediation (e.g. see Tchanturia, K., Lloyd, S., & Lang, K. (2013). Cognitive remediation therapy for anorexia nervosa: current evidence and future research directions. International Journal of Eating Disorders, 46(5), 492-495)?

Thank you for this comment. Indeed, we have spent some time thinking about which parameter to choose and base our sample size calculation upon.

The suggestion of the reviewer is certainly justified and would be an option; however, we decided to design our protocol according to a study that has also applied estrogen (which is the study by Misra et al. 2013).

That aside, I consider the study design is suitable to achieve the aims of the study; to establish the magnitude of any effects on a range of outcomes being examined in this study.

2. **Are sufficient details provided to allow replication of the work or comparison with related analyses: if not, what is missing?**

The protocol provides sufficient detail to allow replication. The authors do not specifically mention this but I presume that the neuropsychological outcome will be based on examining change on each test rather than on a composite neuropsychological measure.

Yes. We have constructed a neuropsychological “battery” consisting of several tests. Changes on each test will be reported. We have now added this information (please see “methods”).

I note that St John's Wort is an exclusion criterion but I don't think that current antidepressant use was. I see that change in antidepressant use is listed as an outcome. An explicit statement about whether this is a contraindication or not at baseline would be useful in the inclusion/exclusion criteria.

St John’s Wort is excluded as it interacts with the CYP-metabolization of contraceptives. Stable antidepressant use is not an exclusion criterion.
An explicit statement that the intake of antidepressants at a stable dose for a minimum of 14 days prior to study inclusion does not constitute an exclusion criterion has now been added (please see “methods”).

The authors state that the protocol follows the SPIRIT guidelines however there are a number of details in the SPIRIT guidelines that are missing from this protocol. Please review and add details where appropriate in the manuscript or as a supplementary checklist.

We have once more reviewed the SPIRIT guidelines and believe to have included information to all relevant points, as far as applicable:


[6-8] Introduction: 6: Background and rationale, 7: Objectives, 8: Trial design


[32-33] Appendices: 32: Informed consent materials, 33: Biological specimens

We have now added information on monitoring of the study by a member of the Center for Clinical Studies according to the GCP guidelines.

3. Is the planned statistical analysis appropriate?

The statistics appear to be appropriate for this preliminary study, which aims to provide effect sizes data to inform future studies. I suspect this may end up being under-powered for variables other than anxiety but the data provided will still be useful.

4. Is the writing acceptable?
The protocol is well written overall.

General comments:

The authors clearly make the case for the need to trial novel biological treatments for anorexia nervosa, given the well documented biological dysregulation, morbidity and mortality associated with this serious condition. Traditionally oral contraceptives (OC) were prescribed with the purpose of ameliorating bone loss but impacts on other areas examined here (e.g. neuropsychological performance) have not been considered until very recently.

This is the first study trialing oestrogen replacement for AN for the purpose of comprehensively assessing improvement in 1) neuropsychological functioning 2) ED psychopathology and 3) HPA including stress hormones and appetite elated gut peptides. It seems that there have been at least three studies cited by these authors reporting a positive impact of OC on specific aspects such as verbal memory and other cognitive domains (refs 80-84), however some studies have produced contradictory results (their refs 83, 89, 90) which the authors suggest may be due to methodological differences. This novelty in this study is that it examines all these key but likely inter-related variables, addressing gaps in the literature not addressed by previous studies which just looked at one or two of these outcome variables. Examining the influence of gut peptides is a hot topic these days and this methodology seems to be widely used for assessing HPA axis functioning.

The authors' overview of the impacts of AN on biological systems is thorough and informative for the general reader. The choice of measures overall is appropriate. The choice of neuropsychological tests seems reasonable as these are well established tests and executive function deficits (especially mental flexibility) are well established. There are other tests that might have been more sensitive (see cognitive remediation studies again) however I realise that the assessment paradigm needs to be targeted in the context of the total assessment burden for participants.

Thank you very much for these encouraging remarks!

Please check that the manuscript complies with the BMC Psychiatry submission guidelines for protocols e.g. in relation to the prescribed order of sections of the manuscript need for an abbreviation list, and the detail in the referencing style.

The BMC submission guidelines are fulfilled.

Small points: The wording of the ethics statement could do with rephrasing (p 9 line 237) consider replacing "favourable opinion" with "approval". Similarly, "Next to female sex" would read better as something like "Other inclusion criteria were… (P9 line 242).

Corrections made as suggested.