PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (see an example) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below. Some articles will have been accepted based in part or entirely on reviews undertaken for other BMJ Group journals. These will be reproduced where possible.

ARTICLE DETAILS

TITLE (PROVISIONAL)	MINOCYCLINE AND ASPIRIN IN THE TREATMENT OF BIPOLAR DEPRESSION: a randomized, double-blind, placebo-controlled, parallel-group clinical trial
AUTHORS	Jonathan Savitz, Sheldon H. Preskorn, T. Kent Teague, Douglas A. Drevets, William Yates and Wayne C. Drevets

VERSION 1 - REVIEW

REVIEWER	Andrew H. Miller, M.D. Professor of Psychiatry and Behavioral Sciences Emory University School of Medicine United States
REVIEW RETURNED	09/12/2011

GENERAL COMMENTS	Savitz and colleagues propose a clinical trial with minocycline and aspirin in patients with Bipolar depression. The trial is novel and timely and will fit nicely into an emerging literature examining the potential efficacy of these and other anti-inflammatory strategies in affective disorders. The primary weakness of the trial is the lack of a more concerted effort in the design to directly address whether this approach is more successful in subjects with increased inflammation versus those without. On page 4, line 53, “significant minority” seems awkward to refer to the significant percentage of patients who develop depression in this context. Some mention of the impact of minocycline on the gut microflora and its potential impact on inflammation and mood should be included. There should be some discussion of the lack of using a stratified design based on inflammatory status (possible CRP). Balancing inflammation across the groups and ensuring a significant sample of subjects with increased inflammation (the basis of the mechanism of interest) is highly desirable. Stratification on sex might also be useful. NSAIDS should be disallowed during the study

REVIEWER	Andrea Roalfe Senior Lecturer in Medical Statistics Primary Care Clinical Sciences University of Birmingham UK
REVIEW RETURNED	12/12/2011

GENERAL COMMENTS	This is an interesting and worthwhile study, however the manuscript is difficult to follow in parts and may be better placed in a specialist journal. There are a number of issues that I consider need addressing and/or clarification in particular the statistical analysis plan. This study is described as both a parallel and 2x2 design, the methodology described infers a factorial rather than a parallel design however the authors have not discussed the issue of interaction between the two drugs and whether the study was powered to detect this. More details of the randomisation method need to be included eg blocking. The timepoint for the primary outcome has not been clearly identified. This trial is described as a pilot proof-of-concept study within the methods section but not in the title or abstract. Sample size is inadequately described. The authors should identify what an effect size of 0.81 means in relation to the MADRS score. The methods of analysis need to be improved whether the data are normal or skewed. Regression based methods are best conducted for this type of design. If there is imbalance in the numbers in each treatment arm during the study then a more appropriate method such as multilevel or mixed modelling method should be used rather than repeated measures ANOVA. The authors should clarify how they propose to compare the rate of completion in the two cells and what these 'cells' refer to? I have some concerns relating to persons not complying with treatment being excluded from data analysis if the analysis is ITT. Furthermore the authors should consider a less biased method of imputation such as multiple imputation or perhaps a sensitivity analysis comparing several imputation methods.

REVIEWER	John M. Davis University of IL at Chicago
REVIEW RETURNED	14/12/2011

GENERAL COMMENTS	This is a useful protocol for those designing exploratory studies of psychopharmacological agents in that it can serve as a model indicating how one research group does this. The 2x2 design, assuming they neither augment each other nor interfere with each other, essentially doubles the statistical power by testing the two drugs individually. If two drugs augment each other, it is of great interest because then you have an even more effective therapeutic effect than either one alone. This is worth knowing. If two drugs interfere with each other and if the individual drugs are used or may be used together, this is valuable information. Under these two conditions, there would be no advantage of power, but the additional research is important clinical information. If the individual drugs are well studied with respect to toxicity and, in this case, as they are both widely used agents known to be safe, the hazards are well

	known. Of course, drugs that are generally regarded as safe do have “known” adverse effects. In this protocol, both experimental agents are used to augment a standard treatment and, as a consequence, in evaluating potential risks, we are dealing with the combined toxicity of three drugs. This issue needs to be thoroughly evaluated. Experimental agents can have unknown hazards. This design would be particularly problematic when using an experimental drug. I think this type of design should be used much more frequently.

VERSION 1 – AUTHOR RESPONSE

Reviewer: Andrew H. Miller, M.D.
Professor of Psychiatry and Behavioral Sciences Emory University
School of Medicine United States

Savitz and colleagues propose a clinical trial with minocycline and aspirin in patients with Bipolar depression. The trial is novel and timely and will fit nicely into an emerging literature examining the potential efficacy of these and other anti-inflammatory strategies in affective disorders. The primary weakness of the trial is the lack of a more concerted effort in the design to directly address whether this approach is more successful in subjects with increased inflammation versus those without.

We acknowledge the reviewer’s point and plan to conduct a post-hoc analysis in order to test whether baseline level of inflammation affects treatment response (pp. 22, para. 2). We note that this trial is an exploratory investigation of whether an anti-inflammatory strategy shows treatment benefits in individuals with bipolar disorder, and thus secondary, post-hoc, analyses are described in relatively limited detail in the protocol. We are happy to provide additional detail, however, if the editor deems such information important for publication.

On page 4, line 53, ³significant minority² seems awkward to refer to the significant percentage of patients who develop depression in this context.

This phrase has been altered so that it instead states the actual proportion of cases reported in the relevant literature.

Some mention of the impact of minocycline on the gut microflora and its potential impact on inflammation and mood should be included.

We have now mentioned that minocycline may potentially exert anti-inflammatory effects by modulating the behavior of the gut microflora (pp. 8, para. 1).

There should be some discussion of the lack of using a stratified design based on inflammatory status (possible CRP). Balancing inflammation across the groups and ensuring a significant sample of subjects with increased inflammation (the basis of the mechanism of interest) is highly desirable. Stratification on sex might also be useful.

In order to conduct a stratified trial with 8 experimental groups (4 x high versus low inflammation, we would have to essentially double the sample size of the study, which would significantly increase costs and decrease feasibility. This problem would be additionally compounded by stratifying for sex. Nevertheless, such a stratification approach certainly would be desirable if promising results are obtained in this clinical trial. In the current trial, the effect of gender and baseline inflammation on treatment outcome will be assessed post-hoc. Given the absence of formal stratification in this study, there is a risk that an imbalance in the inflammatory status of subjects across groups could confound

our results. Nevertheless, given a predicted sample size of 30 patients per group, the risk that the four experimental groups would have clinically meaningful differences in inflammation appears to be small enough to justify an initial phase study. Moreover, the issue of stratification is complicated by the fact that at study entry type I BD subjects must have been taking a stable dose of a mood-stabilizing medication (lithium, valproate, carbamazepine, lamotrigine, antipsychotic agents), for at least 4 weeks since a number of these medications also exert anti-inflammatory effects. A comment about stratifying for inflammation has now been added (pp. 22, para. 2).

NSAIDS should be disallowed during the study.

The use of other NSAIDS indeed was listed as an exclusion criterion. Please see pp. 24, point 10 under the paragraph headed, “exclusion criteria”. In addition, the disallowance of NSAIDS during the study was added to the “Concurrent Medications” Section (p. 14).

Reviewer: Andrea Roalfe
Senior Lecturer in Medical Statistics
Primary Care Clinical Sciences
University of Birmingham
UK

This is an interesting and worthwhile study, however the manuscript is difficult to follow in parts and may be better placed in a specialist journal. There are a number of issues that I consider need addressing and/or clarification in particular the statistical analysis plan.

This study is described as both a parallel and 2x2 design, the methodology described infers a factorial rather than a parallel design however the authors have not discussed the issue of interaction between the two drugs and whether the study was powered to detect this.

We have changed the title to reflect the factorial nature of the design. The reviewer is correct that our study likely does not have the statistical power needed to adequately test for aspirin-minocycline interaction effects. This is a pilot or early phase study designed to generate as well as test hypotheses. If the results suggest the potential for an interaction between the two drugs then a subsequent study can be designed with adequate power to detect the interaction effect suggested by this study. Reviewer 3 has explicitly made reference to the early-phase nature of this study: “This is a useful protocol for those designing exploratory studies…”.

More details of the randomisation method need to be included (e.g., blocking).

We will be using a restricted randomization method, i.e. a permuted block randomization where subjects are randomly allocated to each block (n=30) to ensure that equal numbers of participants receive each drug combination. In order to ensure that experimental group assignment is not skewed across the two trial sites, the study progress will be monitored by individuals who are not involved in the data collection, and in the case of “drift”, adjustments will be made as necessary (pp. 15, para. 2).

The timepoint for the primary outcome has not been clearly identified.

The time-point for the primary outcome variable is now specified as the last visit of the study (pp. 16, para. 1, page 20, para. 2).

This trial is described as a pilot proof-of-concept study within the methods section but not in the title or abstract.

We have now added the term “proof of concept” to both the abstract and title.

Sample size is inadequately described. The authors should identify what an effect size of 0.81 means in relation to the MADRS score.

An effect size of 0.81 would correspond to approximately 3 points difference between the active medication and placebo arms of the study.

The methods of analysis need to be improved whether the data are normal or skewed. Regression based methods are best conducted for this type of design. If there is imbalance in the numbers in each treatment arm during the study then a more appropriate method such as multilevel or mixed modelling method should be used rather than repeated measures ANOVA.

Thank you to the reviewer for these helpful comments. We agree with this point and have clarified our analysis plan in the text of the paper (pp. 20, para. 2).

The authors should clarify how they propose to compare the rate of completion in the two cells and what these 'cells' refer to?

Information about assessment of completion rates has now been presented (pp. 21, para. 3). Our comment about completion rate in two cells is an error and should read 4 cells.

I have some concerns relating to persons not complying with treatment being excluded from data analysis if the analysis is ITT. Furthermore the authors should consider a less biased method of imputation such as multiple imputation or perhaps a sensitivity analysis comparing several imputation methods.

We agree with the reviewer and have revised to the protocol to address this point (pp. 20, para. 3). Briefly, we plan to conduct a mixed effect repeated measures model (MMRM). A last observation carried forward (LOCF) analysis and an observed cases (OC) analysis will be conducted post-hoc for the purposes of confirming the results obtained under the MMRM analysis.

Reviewer: John M. Davis
University of IL at Chicago

This is a useful protocol for those designing exploratory studies of psychopharmacological agents in that it can serve as a model indicating how one research group does this. The 2x2 design, assuming they neither augment each other nor interfere with each other, essentially doubles the statistical power by testing the two drugs individually. If two drugs augment each other, it is of great interest because then you have an even more effective therapeutic effect than either one alone. This is worth knowing. If two drugs interfere with each other and if the individual drugs are used or may be used together, this is valuable information. Under these two conditions, there would be no advantage of power, but the additional research is important clinical information. If the individual drugs are well studied with respect to toxicity and, in this case, as they are both widely used agents known to be safe, the hazards are well known. Of course, drugs that are generally regarded as safe do have [3]known[2] adverse effects. In this protocol, both experimental agents are used to augment a standard treatment and, as a consequence, in evaluating potential risks, we are dealing with the combined toxicity of three drugs. This issue needs to be thoroughly evaluated. Experimental agents can have unknown hazards. This design would be particularly problematic when using an experimental drug. I think this type of design should be used much more frequently.

The reviewer’s comments nicely summarize the value of the two-by-two design in exploratory, proof-

of-concept, trials involving experimental drugs that already have been well-studied with respect to toxicity, as is the case with aspirin and minocycline. We also concur that a parallel arm design, as opposed to a 2 x 2 factorial design, would be more clearly informative in the case of an experimental drug for which the toxicity and drug interaction potential have not been thoroughly studied in human subjects (see pp. 29).

VERSION 2 – REVIEW

REVIEWER	Andrea Roalfe Senior Lecturer in Medical Statistics Primary Care Clinical Sciences University of Birmingham UK
REVIEW RETURNED	20/01/2012

THE STUDY	I have a few minor concerns with some of the statistical methodology proposed: The MMRM is incorrectly described as an imputation method (2nd para of Outcome measures and data analysis). Inappropriate method to compare completion rates is suggested, the authors should consider the chi-squared test or perhaps logistic regression. It would be useful to state that effect size relates to a 3 unit difference in MADRS score (1st para of Statistical Power).