Author’s response to reviews

Title: Mirroring everyday clinical practice in clinical trial design: a new concept to improve the external validity of randomized double-blind placebo-controlled trials in the pharmacological treatment of major depression

Authors:

Emanuel Severus (Emanuel.Severus@uniklinikum-dresden.de)
Florian Seemüller (Florian.Seemueller@med.uni-muenchen.de)
Michael Berger (M.Berger@med.uni-muenchen.de)
Sandra Dittmann (Sandra.Dittmann@med.uni-muenchen.de)
Michael Obermeier (Michael.Obermeier@med.uni-muenchen.de)
Andrea Pfennig (Andrea.Pfennig@uniklinikum-dresden.de)
Michael Riedel (Riedel@med.uni-muenchen.de)
Sophia Frangou (sophia.frangou@kcl.ac.uk)
Hans-Juergen Möller (Hans-Juergen.Moeller@med.uni-muenchen.de)
Michael Bauer (Michael.Bauer@uniklinikum-dresden.de)

Version: 2 Date: 1 May 2012

Author’s response to reviews: see over
1st of May, 2012

Dear Dr. D’Souza,

attached please find the substantially revised version of our manuscript “Introducing a new Concept to assess the Efficacy of psychopharmacological and psychotherapeutic Interventions in Patients with Major Depression” which we hereby submit for publication to your journal.
In the following please find our responses (set in italics, red colour) to the thoughtful comments made by both reviewers.

Sincerely yours,

W. Emanuel Severus, M.D.
Reviewer's report

Title: Introducing a new Concept to assess the Efficacy of psychopharmacological and psychotherapeutic Interventions in Patients with Major Depression

Version: 1 Date: 6 March 2012
Reviewer: Jonathan W. Stewart

Reviewer's report:

There are many reasons to doubt the generalizability of RCT’s to clinical practice. These authors propose a remedy by recommending a study that randomizes patients to typical RCT conditions vs. several alternatives including one that mimics clinical treatment. The beauty of their proposal is it directly compares a typical RCT to studies having various degrees of deviation from a typical RCT potentially allowing a determinations of how closely each iteration comes to what is seen in general practice. Thereby, the proposed randomization of patients to various trial designs will allow clinicians to judge what they might expect in their clinical practices based on the findings of an RTC.

1) My first problem is they are too wordy and convoluted, undermining clarity. The manuscript needs English-speaking editor.

We thank Prof. Stewart for pointing at this issue. Consequently, to improve clarity, we have completely revised the entire manuscript. In addition an English- speaking editor has gone through the manuscript.

2) Their message is simple and straightforward. Why obfuscate? They should limit their comments to the problem generalizing from RTC’s to clinical practice and their proposed remedy. In discussing their proposal, they need to mention its negatives, including requiring a huge sample size and risking being unable to reach conclusions due to differential attrition. They should omit irrelevancies, such as psychotherapy and how to maximize placebo effects.
We agree with Prof. Stewart and have substantially revised the manuscript according to his comments.

For example, we omitted the section on psychotherapy and are now exclusively focusing on the different problems when it comes to generalize from double-blind RCTs to clinical practice. We cut down on the number of treatment trial arms in order not to require an unrealistically large sample size. We propose to compare outcome-relevant demographic and clinical characteristics of enrolled patients in each design (enrolment fraction) to the respective eligible population to deal with the problem of differential attrition.

Substantive issues:

3) If it is “difficult for the placebo effect . . . to take place in these trials” (i.e., RTC’s), how do they explain that alarming increase in placebo response rates over the recent years? They acknowledge the recent increase in RTC placebo response rates a page or two later, so I either misunderstand their point here or their assertions are inconsistent.

We believe that the increase in placebo response rates is the result of a complex interaction between a variety of factors, some of which foster placebo response, such as initial overrating or response bias while others, such as severity of the illness, or the probability of receiving placebo might mitigate it. However we agree with Prof. Stewart that our statement in its present form is unclear, so we decided to eliminate it.

4) How does limiting RCT’s to arbitrary durations, such as 8 weeks, limit inferences? 8 weeks certainly rarely mimics clinical practice, but should RCT’s mimic clinical practice on this one, or should clinical practice mimic RCT’s?

We agree with Prof. Stewart that 8 weeks certainly rarely mimics clinical practice. While data suggest that a period of less than 4 weeks is too short due to the increased risk of erroneously concluding that an effective treatment is ineffective (Tedeschini et al., 2011) less is known about the impact of study periods longer than 8 weeks. Therefore, in the “enhanced RCT” which we propose, we stick to the 8 week’s study period, with the option for study
participants to continue with the study medication (or the active medication) for an extended period of time, thereby mimicking continuation treatment.

5) They assert that clinical practice treats more severely ill patients than do RCT’s. This does not jibe with RCT’s having increasing requirements for minimal severity while in practice patients below these limits receive these medications. Certainly, the exclusion of suicidal and psychotic patients may equate to excluding the more severely ill, so the main point is not so much a severity one as one of how broad the population is that is treated with these agents. But, this broadness of allowed population is inherent in clinical practice viz a viz RCT. Unless one has no entry criteria, one cannot mimic a clinician’s choice of to whom to give a given medication, since someone somewhere will use my favorite antidepressant for whoever walks in the door. But, without entry criteria, it seems difficult to draw any kind of logical inferences beyond “this works/does not work for whoever walks in the door.” And, the likelihood of such an “all comers” study demonstrating efficacy seems slim unless the sample size is huge, while who would you generalize it to even if successful? As a “whoever walks in the door” study would have to be generalized to “whoever walks in the door”, how is that an improvement over what we have now?

Prof. Stewart’s points are very well taken. In particular we agree with his statement that unless one has no entry criteria one cannot mimic a clinician’s choice of to whom to give a given medication. However what we aim at with the “enhanced RCT” which we propose is to increase the enrolment fraction (proportion of people who are eligible for participation and who actually enroll in the RCT) – and to have the sample of patients recruited being as representative as possible of the eligible population in terms of outcome-relevant demographic and clinical characteristics. One of the main reasons why patients who are eligible for participation (i.e. fulfilling all of the inclusion criteria and meeting none of the exclusion criteria) do not get enrolled in a double-blind, placebo-controlled trial is that they decline to participate in it. One of the main reasons for their refusal is the probability of being randomized to placebo. And placebos, as Prof. Stewart correctly pointed out at 9), are not used in clinical practice.
6) So, their BACKGROUND section needs to be stated more succinctly and more clearly. It seems to me they make two interlocking points. First, RTC’s do not mimic clinical practice because they X, Y and Z. Second, drugs marketed due to RTC demonstrations of efficacy are used in a much broader population of subjects than allowed into RTC’s. We need RTC’s to more closely mimic clinical practice both in terms of who gets treated and how the treatment is conducted.

*We are very grateful to Prof. Stewart for this “roadmap” and have substantially revised the background section according to his suggestions.*

**DISCUSSION**

7) I do not follow “1)”. In any case, I doubt recommendationis 1) and 2) are followed in clinical practice so I am unsure how they help the generalization from RTC’s to clinical practice.

*In short, in clinical practice the combination of using an approved antidepressant, along with verbal suggestions of benefit, as well as the patients’ active participation in treatment choices (shared decision making) creates positive expectations regarding the effectiveness of the treatment agreed upon and may positively affect the overall outcome. In contrast, in double-blind RCTs the expectation of a positive outcome is lessened by the knowledge that participants may be randomized to the placebo arm, with placebo often referred to as “a dummy or pretend medicine” which has “no effect on a person” (WHO ERC). Therefore a detailed, balanced and truthful description of potential benefits associated with being randomized to the placebo arm may redress patients’ fears associated with being randomized to the placebo arm – and thereby mimic clinical practice with regard to creating positive expectations. For a more complete elaboration on this issue we would like to refer to the revised “Discussion” section of our manuscript.*

8) “3)” makes a lot of humanitarian sense and ought to be done in all studies. However, at least in the US, the FDA has rules about exposing patients to safety risks prior to efficacy having been at least presumptively demonstrated. That is, I do not see how the FDA would allow use beyond a short-term study of an experimental agent in the participants in the initial efficacy studies. I doubt the
FDA would alter its rules on this one, but I suppose these authors could try. Short of that, I see their proposed study a Phase IV trial rather than premarketing.

We are very grateful to Prof. Stewart for pointing to this important issue. Therefore we would recommend starting with Phase IV trials (as now being stated in the revised research agenda). In case our proposed “enhanced RCT” yields promising results with regard to external validity one might implement this study design at a late stage of Phase III trials, after efficacy and safety have been at least presumptively demonstrated.

9) I am unclear how presenting positive expectations of placebo effects attracts “patients more representative of clinical practice”. Since placebos are not used (or, at least, are not labeled as such) in clinical practice, the presentation of positive expectations of placebo effects is also not used in clinical practice. So, how is it that presenting positive expectations of placebo effects attracts “patients more representative of clinical practice”?

In major depression both positive and negative expectations regarding treatment outcome may affect the outcome itself. Expectations responsible for placebo effects are not limited to the administration of placebo interventions, but occur in everyday clinical practice. In clinical practice, they are created by the combination of shared decision making, the choice between a variety of approved antidepressant drugs and verbal suggestions of benefit given along with the active treatments. In contrast, in double-blind, placebo-controlled RCTs the probability of receiving a placebo, “a dummy or pretend medicine” which has “no effect on a person” (WHO ERC) makes it harder to give rise to positive expectations regarding treatment outcome. Therefore patients representative of clinical practice may decline to take part in such a trial even if they fulfill all inclusion criteria and meet none of the exclusion criteria. Consequently one of the outcome criteria of our proposed set of studies will be to compare the enrolment fraction between the traditional RCT and the “enhanced RCT”.

10) Informed consent is also not a tool used in clinical practice. Are they recommending its use? Else, how is the informed consent process helpful in mimicking clinical practice?
Research indicates, that both the content and manner in which information is shared with the patient, and the patient’s experience of being involved in the decision, can directly alter therapeutic outcomes via placebo responses (Brody et al., 2012). In a double-blind, placebo-controlled trial, the informed consent process allows to share information with eligible patients regarding what the study is about in a manner which may give rise to positive expectations similar to those occurring in clinical practice.

11) I’d delete the section on psychotherapy. If retained, they need to acknowledge that control conditions besides “wait lists” have been used, for example, supportive psychotherapy or an exercise program in order to control for attention and time, differences demonstrating something “specific” about the experimental condition beyond that the patient is “doing something” or “receiving a kind person’s attention”.

We agree and have deleted the section on psychotherapy.

12) Indeed, my impression is that “doing something” alternatives have largely supplanted “wait lists” in psychotherapy research. In any case, this section gets in the way without helping.

We agree and have deleted the section on psychotherapy.

13) While the four studies they present have intuitive merit, such studies suffer from attrition. That is, I may initially agree hoping I will be assigned to Study #4, then be disappointed to be assigned to Study #1 and not agree to continue. Should sufficient numbers not agree to their assigned study, that would invalidate comparisons across studies. The only conclusion would be “there is differential acceptability” but I submit that is not a very interesting conclusion. They indeed state that one of their intended analyses would be to determine whether there was differential nonacceptance but they do not indicate whether differential nonacceptance would affect other between study comparisons, e.g., of response/remission rates.
To avoid a scenario like the one proposed above by Prof. Stewart the first informed consent form (ICF #1) will just outline the general purpose of the trial but will not include any details regarding the different study designs. Nevertheless there is a high probability of different attrition as the percentage of eligible patients which may decline to be randomised to the conventional RCT may be higher than the percentage of eligible patients which may decline to be randomised to the “enhanced RCT”. Therefore comparing the enrolment fraction (proportion of people who are eligible for participation and who actually enrol in the RCT) as well as outcome-relevant demographic and clinical characteristics of the enrolment fractions between the two RCTs is one of the outcome criteria of our proposed study design.

SUMMARY

14) “Modifying the meaning of what a placebo constitutes . . . seems essential”.

Perhaps, but I do not see that they have demonstrated this.

We are very grateful to Prof. Stewart for having pointed to the lack of clarity in our original manuscript. We sincerely hope to have clarified this issue by the explanations given above as well as the information provided in our substantially revised manuscript.

15) It seems to me that if they can avoid differential non-acceptance, then their study design (i.e., randomizing patients to one of several study designs) is both novel and will allow an estimate of the relationship, if any, between the usual RTC (Study #1) and usual clinical practice (Study #4). That is, their design promises to tell us what the generalizability of RTC’s might be.

As outlined in our revised background section the external validity of today’s RCT’s may be compromised, among other factors, by the blinded use of placebo, and possibly connected with this, the fact that patients recruited may not be representative of the eligible population in terms of demographic and clinical characteristics. Therefore, instead of trying to completely avoid differential non-acceptance (which will be hard to achieve, if not impossible to attain), we decided to evaluate the impact of the two study designs (traditional RCT, “enhanced RCT”) not only on the efficacy but also on the enrolment fraction. Comparisons between the two study designs can be straightforward performed using ordinary linear
models where treatment success is explained by study arm, study design and their interaction.

16) I wonder, though, what the sample size would have to be to have reasonable power. They do not address power/sample size. I suspect they would need a huge sample to be able to detect between study differences and/or suggest that failure to demonstrate differences is more likely due to lack of difference than lack of the power to detect.

We are very grateful to Prof. Stewart for pointing to this important issue. In order not to require an unrealistically large sample size we decided to cut down on the number of treatment trial arms. Furthermore by choosing sertraline as the medication of choice we will be able to use existing literature for power calculations for the traditional and “enhanced RCT's” we propose to conduct.

Quality of written English: Not suitable for publication unless extensively edited

The manuscript has been extensively edited.

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I do not see how any organization could gain (or lose) financially from publication of the manuscript except or unless they use it to design a study using the ideas these authors present.

I do not own any patents and am unaware of any patents relating to the content of this manuscript; therefore, I am unaware of whether any organization with which I have done business holds or has applied for patents relating to the content of this manuscript.

I am unaware of any competing financial or nonfinancial interests. Thus, to the best of my knowledge, I have no competing interests, financial or otherwise.
Reviewer's report

Title: Introducing a new Concept to assess the Efficacy of psychopharmacological and psychotherapeutic Interventions in Patients with Major Depression

Version: 1 Date: 14 March 2012

Reviewer: Bruno Falissard

Reviewer's report:

This is a short conceptual paper which does not present some study results but instead suggestions to improve the usefulness of randomized controlled trials (RCT) in the field of depression (but these suggestions could be in fact applied also to many other therapeutic areas, included outside the field of psychiatry). The subject is all the more important as we are indeed at a turning point in the process of evaluation of treatments, pharmacological or not. The points raised by the authors are definitely relevant and their suggestions for improvement are simple, easy to implement in practice and could improve the interest of RCT results.

Major compulsory revision:

My main reproach against the paper is that it is indeed too short for a conceptual paper. It touches the subject but do not tackle it.

We are very grateful to Prof. Falissard for making this point. Therefore we decided to eliminate the section on psychotherapy and are now focusing exclusively, but more thoroughly, on the external validity of randomized double-blind placebo-controlled trials in the pharmacological treatment of major depression.

A paper which deals with the question of the incompatibility of double blinded/placebo RCTs with real life clinical practice has first to explain why these double blinded/placebo RCTs are supposed to be inescapable. There are of course methodological reasons for that, but also sociological ones and even anthropological ones (for instance RCTs can be considered as a ritual which goal is to pacify the relationships between health authorities and pharmaceutic firms).
These reasons make that it can be (at least this can be discussed) simply absurd to transpose straightforwardly these double blinded/placebo RCTs from the field of development of medications to the field of development of psychological treatments.

We agree with Prof. Falissard that transferring methods developed and established in the field of psychopharmacological treatment straightforwardly to the field of psychological treatments may be too simplistic. Therefore we decided to eliminate the section on psychotherapy. In return, in the “Background” section, we are explaining in more detail why these RCTs are supposed to be inescapable, and why at the same time, the very strength of this type of trial, i.e. the blinded use of placebo, compromises the external validity of results obtained from these trials.

Moreover there is not only the questions of double blind and placebo which make problem. These trials are done in particular countries and this raises transcultural problems. They are done in selected centers and this hampers the generalisabilty of results. The randomization itself can interfere with the level of efficacy of a treatment (a patient which receives a psychoanalytic treatment after a randomization could be less involved in this treatment than the same patient who would have chosen it deliberately). And since there are nowadays methodological approaches that are serious alternative to randomization, RCT themselves are questionable, especially in the field of psychological treatment.

We agree with Prof. Falissard that there are other issues than those we discussed in our article which may limit the degree to which the results of double-blind placebo-controlled trials can be generalized to clinical practice, in particular in the field of psychological treatment. We are particularly grateful for pointing to the impact randomization may have on the efficacy on a treatment and included this topic in our background section. In addition Prof. Falissard pointed out that there are nowadays methodological approaches that are serious alternatives to randomization. In the field of psychopharmacological treatments one of these methods are Marginal Structural Models, which the authors of this paper themselves have dealt with (e.g. see reference #33). However these methods as well do have problems, in particular with type I error rates. Therefore, at least for the time being, randomized,
Double-blind placebo-controlled trials will remain irreplaceable for testing the efficacy for new pharmacological intervention in major depression and other psychiatric as well as medical disorders in which expectations regarding treatment outcome may affect the outcome itself.

To sum up, the paper begins a promising discussion, proposes some interesting elements of answer. But the reader has the unpleasant feeling that (perhaps) the most important problems are elsewhere.

While we agree that there are other important problems associated with randomized, double-blind, placebo-controlled trial, in particular in the field of psychological treatments, we feel that the problem we are addressing in our article is of substantial clinical importance, not only in the field of psychopharmacological research, but also in the pharmacological treatment for other psychiatric as well as medical disorders in which expectations regarding treatment outcome may affect the outcome itself.

Quality of written English: Acceptable

An English-speaking editor has gone through the revised manuscript

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests