Reviewer's report

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

Version: 2 Date: 18 May 2012

Reviewer: Jonathan W. Stewart

Reviewer's report:

Cover letter addressing reviews:

2) differential attrition - I am unclear how comparing demographics and clinical characteristics deals with differential attrition. Certainly, demographics and clinical characteristics might well influence who does not complete a treatment trial. But, it also seems that within given demographics and clinical characteristics one might also anticipate that different clinical settings and circumstances might result in differential likelihood of attritting. Just a simple example is patient A has been seeing his psychiatrist for years and is started on a newly marketed antidepressant vs. A’s twin brother (B) enrolls in a clinical trial of this same antidepressant with a clinician he has never met before entering the study. My bet is A is more likely to complete the trial than B. How do comparing demographics and clinical characteristics address this likely differential setting effect?

13) a 2nd stab at differential attrition - they essentially give the same answer which at best is a partial answer but does not address whether demonstration of differential acceptability and/or attrition would affect the inferences they might draw. That is, their purpose in proposing their multilayered study is to assess how adequately RTC’s mimic clinical practice. That is, how translatable are RTC’s into clinical practice? My question is whether they can answer their question if there is significantly higher non-acceptance/attrition in the RTC group than in the usual practice group? Will they still be able to assert that the RTC response rates are the same as vs. different from the responses I can expect in my own clinical practice?

16) sample size - OK, but does the manuscript state what the sample size would have to be given the literature’s estimate of sertraline’s effect size? A problem with using the literatures estimate of sertraline’s effect size is based on RTCs. This drug’s effect size in a nonresearch population must be unknown. And, because their proposed study design has not been utilized, they have no idea what the effect size is of sertraline as usual treatment vs. sertraline as RTC treatment. The problem with assessing sertraline as usual treatment vs. sertraline as RTC treatment is that demonstrating equivalency (i.e., we should not reject the null hypothesis of no difference) requires a much larger sample size than demonstrating nonequivalency (i.e., we ought to reject the null). If they
think the RTC will demonstrate difference response rates the required sample size will be smaller than if they think the RTC can be translated into clinical practice. But, they need to calculate the numbers and show and discuss them to convince us whether to think they have a viable idea vs. it is nice in theory but too impractical to actually implement. If it would require a large sample, would they still recommend that, say, a national health service or WHO or other very large organization take it on anyway?

Revised manuscript
1) the English is very much improved rendering a much more readable manuscript.

2) I am unfamiliar with this use of “remit” (1st sentence of “Discussion”). Perhaps they could switch to a more commonly used word. I suspect they mean “responsibility” or some such.

3) I might add that the point of Phase III trials is to demonstrate whether there is efficacy so patients and doctors ought to have greater skepticism than with a marketed drug which has at least already been shown to “work” for someone among its Phase III participants.

4) I like their proposed way of explaining placebos to study participants.

5) As I see it their simplified study design will:

a) answer the simple question of whether it makes a difference to better explain placebo effects and add a maintenance period to the usual RTC
b) require many fewer subjects than the original study design; nevertheless, they still need to give an estimate of how many that would be
c) NOT all an answer to their original issue of how well RTCs translate into clinical practice

6) Thus, they have thrown the baby out with the bathwater. It seems to me their proposed design might better be described as “Will improved study design increase RTC retention?” rather than suggesting it mirrors everyday practice. If they want to do the latter, I’d suggest they either add a treatment as usual group or make it a two stage process. That is, Stage 1: two parallel RTCs having differing in explanation and whether there is a maintenance phase; Stage 2: comparing the winner of Stage 1 with TAU. Or, run the studies in reverse order. Doing it in stages would aid sample size so cost less.

BOTTOM LINE: Much improved readability and over-all manuscript but they redid their design without bringing the introductory remarks into line with the redesign. That is, they removed the potential to determine how well an RTC mirrors every day clinical practice which is what the manuscript title advertises.