Author’s response to reviews

Title: The effectiveness of joint crisis plans for people with borderline personality disorder: protocol for an exploratory randomised controlled trial.

Authors:

Paul Moran (paul.moran@iop.kcl.ac.uk)
Rohan Borschmann (rohan.borschmann@kcl.ac.uk)
Clare Flach (clare.flach@kcl.ac.uk)
Barbara Barrett (b.barrett@kcl.ac.uk)
Sarah Byford (s.byford@kcl.ac.uk)
Joanna Hogg (j.hogg@kcl.ac.uk)
Morven Leese (m.leese@kcl.ac.uk)
Kim Sutherby (k.sutherby@kcl.ac.uk)
Claire Henderson (c.henderson@kcl.ac.uk)
Diana Rose (k.sutherby@kcl.ac.uk)
Mike Slade (m.slade@kcl.ac.uk)
George Szmukler (g.szmukler@kcl.ac.uk)
Graham Thornicroft (g.thornicroft@kcl.ac.uk)

Version: 2 Date: 5 February 2010

Author’s response to reviews: see over
Editors-in-chief  
Trials  

5th February 2010  

RE. MS: 1164263216326050: The effectiveness of joint crisis plans for people with borderline personality disorder: protocol for an exploratory randomised controlled trial.  

Thank you for giving us the opportunity to re-submit this paper. We are grateful to the peer-reviewer for the very helpful comments on our original submission. The paper has been revised taking account of these comments.  

Please find below a list of the reviewer’s comments, with our responses and changes to the manuscript inserted below each comment.  

If you require further revisions we will be happy to consider them.  

We look forward to hearing from you.  

Yours sincerely,  

Paul Moran  
Clinical Senior Lecturer
1. Will the study design adequately test the hypothesis?
The study as reported should collect valid evidence to test the effectiveness of a JCP vs. TAU. Comments regarding this issue are:

- As the study will recruit over approximately 1.5 years, is it possible for the TAU to change in that period? Would be useful to document what TAU was actually delivered during the trial. If this is already planned, please report how it will be achieved.

OUR RESPONSE:
We do not anticipate that the TAU will change during the course of the trial. We will be documenting what TAU consists of through our collection of service use data using the ADSUS, which records in detail the use of all hospital and community services and any treatments provided. In our revised manuscript, we have stated this in our description of the control intervention (page 10):

“We do not anticipate that TAU will change during the course of the trial and will be documenting what TAU consists of through our collection of service use data.”

------------------------------------------------------------------------------------------------------

- The proposed assessment will be “standardized and objective”. Most of the outcome measures will be obtained through interviews (in the form of a questionnaire). Therefore please provide details about how researchers will be blinded during this phase. Which methods will be used to keep the allocation concealed?

OUR RESPONSE:
We apologise for the lack of clarity about this issue in our original submission. We have revised the section on “Randomisation procedure and methods to minimise bias” (page 8) in order to make this clearer and added the following description:

“There will be specific instructions to participants and clinical teams not to disclose the treatment details of any study participant. Every effort will be made to maximise the single blindness of research workers, whose ‘best guesses’ of treatment arm status will be assessed after the final assessment in order to test whether their blindness was successfully maintained.”

-------------------------------------------------------------------------------------------------------

- The baseline assessment and follow-ups have an extensive list of outcomes which in their majority will be collected via questionnaire (please clarify in text if this is incorrect). Please provide information on the length (duration) of this “interview” and if this level of measurements have already been piloted (or if this is the first time). Would be useful to have information as to whether you expect potential drop-outs due to the length of data collection.

OUR RESPONSE:
We piloted the assessment battery on a small sample (N= 13) of participants and found that the average duration of the baseline interview was 45 minutes, with no drop-outs. On the basis of this information, we do not anticipate drop-outs due to
length of data collection. We have added this information to our description of “Recruitment and baseline procedures” (page 7):

“In a pilot study of 13 participants (described further below) in which feasibility of administering the intervention and collecting data were examined, the average duration of the baseline interview was 45 minutes, with no drop-outs.”

• Regarding the primary outcome, self-harm, is there any literature regarding the reliability of self-reporting of this outcome for this group? If so, please provide. One further comment is that although this is reported as a pilot trial, the sample size calculation is based on showing effectiveness (although a large effect size is assumed). There are also costs recorded and a proposal to have an economic analysis. Not clear from this perspective why this is a “pilot” study instead of a “full” study. Could you please clarify?

OUR RESPONSE:
Within the field of self harm research, the use of self-report data is standard. However, knowledge about the reliability of this approach is limited and we accept that self-reporting can be susceptible to information bias. It is for these reasons that we have chosen ‘any self-harm’ as the most robust primary outcome. We will be gathering self-harm data from two sources: i) self-report (retrospective self-report at follow-up and also contemporaneous diaries which we will be asking participants to keep) and ii) staff report (from electronic case records). One of the objectives of the study will be to examine the various methods of recording self-harm data. We have added this point to our revised ‘Statistical Analysis Plan’ (page 14). Since there is some doubt about how best to record these outcomes and how reliable they are, and also because we have assumed a relatively large effect (as the reviewer has pointed out), this trial cannot be called definitive. This was the reason it was initially called a ‘pilot’ but on reflection, we concede that ‘exploratory’ would be a better description of the trial, as we do hope to achieve some useful conclusions about the likely range of effectiveness estimates. We have amended the title of the paper and the protocol accordingly.

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

• In the protocol, there is a hint that the authors have some previous results of a case series; it would be useful to show the results of the series in terms of the outcomes that will be obtained/reported from the pilot and if these are consistent with the assumptions made (sample calculation).

OUR RESPONSE:
The purpose of the pilot case series was purely to examine the feasibility of conducting JCP planning meetings and data collection methods. In the paper, we report that the process of developing JCPs proved to be feasible and that the assessment battery was acceptable to participants. When we followed them up, a number of participants also reported that the JCPs were potentially useful. However,
the numbers involved are too small to generate statistically meaningful results and have not been used to inform the power calculation.

• Please expand which “Outcome data [will be] obtained from electronic patient records” Are these only the self-harm data as reported further in the protocol or are there other data as well?

OUR RESPONSE:
We apologise for the ambiguous wording of this sentence and have amended it accordingly (page 8):

“Some outcome data will be obtained from electronic patient records, even for participants lost to follow-up interview. This will include any episodes of self-harm, hospital admissions and service use recorded in the electronic records.”

• The effect size reported as target is large. Please give information regarding the lower limits (closer to the null hypothesis) for the effect size in the “green cards” trial. This should allow the reader to get an idea of the precision in the quoted estimate. Also, it would be useful to report the similarity (or lack of) of the population recruited for the “green cards” trial and this one.

OUR RESPONSE:
The ‘Green Card Study’ (Morgan et al, 1993) was a trial of 212 patients who had been admitted to hospital following a first episode of deliberate self harm. From a socio-demographic perspective, the patient population in the Green Card Study is similar to the one proposed in our trial. The only anticipated difference in the population is that the JOSHUA trial will include individuals with a history of previous self-harm (as opposed to the Green Card Study which only included first episodes of self-harm). The risk of self-harm after randomisation in the green card group was 37% (95%CI 14% to 97%) of the risk in the TAU group. In our revised description of ‘Sample Size’, at the reviewer’s suggestion, we have now added the confidence interval around the reduction in risk.

• The sample size calculation does not provide accrual time or median survival time. If these are not required for the calculation, please provide a formula used or reference to the method used for this calculation. The data reported would be sufficient for a sample calculation based on a dichotomous outcome (difference in proportions) but not for a log-rank test (survival data).

OUR RESPONSE:
Thank you. We have reconsidered the primary outcome and for a number of reasons (to do with data completeness and integration with the cost effectiveness analysis) have decided to power the trial on the binary outcome of ‘any self-harm’. We have explained these reasons in the revised section on ‘Sample size’ (page 13): We will also analyse the time to first event and number of events as other potential outcomes.
However, the sample size needed for a survival analysis can be calculated using the probability of survival at a particular follow-up time (in this case 6-months) using the calculation by Freedman (Statistics in Medicine 1: 121-129.). This sample will have 80% power to detect a constant hazard ratio of 0.29 which is now stated in the protocol.

In order to account for the above points, we have completely revised the section on ‘Power Calculation’ (pages 13 and 14), which now reads as follows:

“On the basis of these predictions, an overall sample of 114 (randomised 1:1 to TAU: JCPs) would provide 80% power to detect an observed difference between TAU and JCPs based on a 2-sided test at the 5% significance level. This will be increased to 120 to allow a small loss in the administrative data on self-harm. This sample is also large enough to provide 80% power to detect a constant hazard ratio between the groups of 0.29 with proportions of events in the two groups as stated above, based on the log-rank statistic assuming no accrual rate, a fixed time of follow-up and an estimated 10% dropout.”

--------------------------------------------------------------------------------------

3. Is the planned statistical analysis appropriate?

- There will be three types of self-harm data collected and some comparison between them made. For the final analysis, if these appear to be in conflict, how will these sources be integrated? One will have preference over the others? This needs defining.

OUR RESPONSE:
If conclusions differ across the measurements used, we will investigate reasons for the discrepancy since this will be of interest for future trials. We have now specified this in the revised ‘Statistical analysis plan’ (page 14).

--------------------------------------------------------------------------------------

- The methods to be used for the analysis are rather vague For example, is there good evidence that the outcomes are normally distributed (and hence the use of t-tests)? Also, are dichotomous outcomes relatively common, none less than 2%? Otherwise simple Chi-square statistics might not be appropriate. Finally, although survival methods are stated as used for survival data, could you please be more explicit as to which methods will be used or are likely to be used (e.g. Cox regression).

OUR RESPONSE:
We will assess the assumptions of normality in the first exploration of the data and adjust the analyses accordingly using nonparametric tests where necessary. We do not expect small samples in dichotomous outcomes to be a problem but if so we will use Fisher’s exact test rather than chi-square. We intend to use logistic and cox regression to analyse self harm outcomes in order to allow for adjustment for potential confounding variables. We have added these specific analyses into the revised statistical analysis plan (page 14).

--------------------------------------------------------------------------------------

4. Is the writing acceptable?
• Yes, the protocol is generally well written. There is only one sentence in the page13 that could be made clearer. “Data collection methods for “time to fist self harm” and “number of self harm episodes” will be compared on attrition rates using …” Does this mean that the attrition will be compared or that the actual data?

OUR RESPONSE:
We apologise for the confusing wording of this sentence and have amended it accordingly (page 14):

“The three data collection methods for the outcome of self-harm (i.e., participant interviews, diaries and electronic case records) are likely to differ on the amount of missing data. We will compare proportions of missing data using chi-square tests and agreement between data collection methods using Lin’s Concordance Correlation Coefficient.”

• There also seems to be a verb missing in page 7 - “ A baseline interview will then be [missing] conducted”

OUR RESPONSE:
Thank you for spotting this. We have amended the sentence, which now reads:

“A baseline interview will then be conducted by the research worker, either at this point or at a subsequent meeting depending on the wishes of the participant.”

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.