Author’s response to reviews

Title: Feasibility of a randomized single-blind cross-over trial to assess the effects of the second generation slow-release dopamine agonists pramipexole and ropinirole on cued-recall memory in idiopathic mild or moderate Parkinson's disease without cognitive impairment.

Authors:

Nicola Edelstyn (n.edelstyn@keele.ac.uk)
Laura Longshaw (aura.longshaw@uhnm.nhs.uk)
Julius Sim (j.sim@keele.ac.uk)
Keira Watts (Keira.Watts@uhns.nhs.uk)
Andrew Mayes (Andrew.Mayes@manchester.ac.uk)
Michael Murray (m.murray@keele.ac.uk)
Simon Ellis (simon.ellis@uhns.nhs.uk)
Nicola Edelstyn (n.edelstyn@keele.ac.uk)

Version: 1 Date: 02 May 2017

Author’s response to reviews:

Dear Editor,

We would like to take this opportunity of thanking the reviewers and yourself for the time and attention given to our paper. A point-point response to both reviewers' reports is provided below.

Reviewer #1:

I read with interest this article, and have the following issues for you to consider:-

[Page 6]: Requirement to align the article and make reference to the recent CONSORT extension to pilot trials (Eldridge et al., Pilot and Feasibility Studies 2016; 2:64) as opposed to reference 28 (Schulz et al 2010)

We have updated this reference accordingly (page 25).
Randomisation is specified as 1:1; I think this was simple randomisation as there was a 12:10 ratio in 1st-order designation between arms; perhaps in a main trial there would be some blocking in the procedure if there is concern that there may be an order effect?

We have now calculated the period effect, and have noted that randomization in a full trial should be blocked (page 16).

Washout seems to be classed as 'satisfactory' if 1/16th of the elimination is met - is there any reference to support this?

Justification of the (4 half-lives) washout period and reference has been included on page 7.

'Outcomes' - It is stated, "The two primary outcomes were efficacy of processes and procedures used to manage symptoms during the washout period … and estimates of cued recall performance". The last of these corresponds with one of the two primary objectives for the pilot trial - as detailed on page 6 and in the abstract; however, it is not too clear that the efficacy of processes and procedures tallies with assessment of safety of switching (though I think this is probably what is being assessed here)?

To ensure consistency and avoid ambiguity, the objective described in the Abstract (page 2) “safety of switching” has been replaced by “examine the efficacy of processes and procedures used to manage symptoms during the washout period”.

Quantitative data - This section is brief, understandably because the focus is not on hypothesis testing. However, a primary aim was to provide estimates for informing a main trial. What I am not clear about is what is the intended main comparison in the main trial - the presentation of data and sample size precursor tend to suggest that the difference (OFF-ON) in cued recall between treatments is what is being targeted, but in the results section (page 16) the clinically important difference seems to be focused on actual cued-recall scores (as opposed to OFF-ON difference) - thus, further clarity around this would be helpful.

The main trial would indeed focus on the difference in cued recall between the treatments. We have now inserted the estimates related to this (page 16).
[Page 13]: Qualitative - There is limited information in this section (or it is not clear) on selection of the patients for interview. From the results section it is clear that 5 people were interviewed; yet, the pool of potential people would have been far greater (53 declined or didn't respond) - were the 5 interviewed all 5 who declined (48 being non-responders) or was there some selection around 5 out of xx decliners (and if so how selected)?

Also, there is no detail in this section on the other qualitative study which is reported in the results section on selection of 5 from 16 of the study completers in regards to the experience of study participation.

We have added details of both the end of study interviews and barriers to participation qualitative studies on page 13.

[Page 16]: The authors specify that 10% of the observed range of scores (in this case 0.070) qualifies as a clinically important difference - is there further justification or a reference in support of this, or was it an a priori decision?

In the absence of any clear guidance or published work that would indicate what the minimum clinically important difference should be, we took it that 10% of the observed range of scores would be a reasonable figure. As sample sizes are presented in Table 4 in relation to a range of effects, readers can determine the size of a future trial in terms of effects of alternative magnitudes.

[Page 18]: A large part of the discussion section focuses on initial problems with CTU set-up and logistical challenges. This may well have been the case, but identified numbers were quite large - at least relative to the number that actually decided to take part. The problem seemed to lie with eligibility (was this too restrictive?) and uptake of those deemed eligible for the study - two of the mentioned barriers were trial accessibility and fear of the research process - is it that a more developed CTU would have addressed these problems, and moreover contributed to greater recruitment numbers that may have changed the conclusion of the study in relation to its likely feasibility? Some clarity on this is provided in the Conclusion (on page 19) but additional thought and clarification could help clarify the discussion on page 18.

We have commented on these points in the Discussion on pages 19 and 20, and in the Conclusion, also on page 20.

[Page 19]: I am not sure if the interpretation in the Conclusion (of confidence in recruitment now the CTU is fully operational) is completely supported by the findings here - but certainly the
required sample size if small for the main trial as indicated is favourable, and any increase in efficiency of logistical processes can only help in making a case towards increased feasibility of conducting a trial.

The following text has been included in the Conclusion: Implementing these measures, together with integration of the QRI from the outset in order to prevent difficulties developing and optimize recruitment from the start, rather than applying it to an ongoing RCT experiencing recruitment shortfalls where time to rectify/address recruitment difficulties may be limited increase confidence that realistic recruitment targets, whilst still challenging, could be achieved. (page 20)

Reviewer #2: The authors have presented a comprehensive description of the conduct and research of a well-conducted trial. There are a number of minor comments:

1) The authors have provided a good description of the treatment patterns (discontinuations, switches and trial drop out). However it is not immediately clear how the 18% missing data rate was calculated: my estimation of missing data is based on completion rates (16/22 = 73%, ie missing data rate of 27%) but perhaps the authors can clarify here?

Thank you – we have recalculated the missing data rate on this basis.

2) Can I ask why the ON/OFF assessments were made prior to the 6 week stabilization period on each IMP (lines 5-7 on p11 it states "The ON-medication and OFF-medication sessions for each drug should have taken place during a 7-day window, on separate days separated by at least 3 days and followed by a 6 week stabilization period on each IMP")? If my understanding is correct, the ON/OFF assessments should take place after the stabilization treatment period?

The reviewer is correct, the ON/OFF sessions took place after stabilization on each IMP. This report has been amended to remove the ambiguity text. (page 11)

There are a few more important comments arising from the complexities introduced by the use of a crossover design:

3) I notice that results are split according to the pre-randomisation treatment. Did the authors consider the impact of carryover effect (from pre-randomisation treatment) - and perhaps whether a future phase III trial may benefit from a washout period prior to administration of the first treatment, in order to remove the potential effect of pre-randomisation treatments?
This is a good point. We have added a comment to this effect on page 16 (This also raises the issue of including a washout period prior to administration of the first treatment, in order to remove the potential effect of pre-randomisation IMP).

4) The use of a crossover design assumes that there is no "period effect" (which would occur if patients become more comfortable, for example (or more unsettled) during the course of the trial which in turn affected their outcome) in order to ensure that the two treatment periods are comparable. Might it be worth the authors discussing the whether such a period effect is likely in this clinical area?

An estimate of the period effect has now been inserted (page 16) and on the basis of this it is stated that the full trial should adopt an analysis that takes account of a period effect, and a method of randomization that ensures equal allocation to the two treatment sequences.

5) The advantage of using a crossover design is to permit within-patient differences (of the treatment effects from the two treatment periods) to be used rather than between-patient differences (which would be available from a parallel group design), which reduces the variability of the data and therefore increases power. Therefore, assuming that this pilot study aims to inform the sample size calculation of a future crossover trial, the authors need to provide an estimate of the SD of the within-patient differences (of the two ON/OFF treatment effects); however, this information has not been presented in the manuscript - instead, the authors have provided the individual SDs (of the ON/OFF differences) for each treatment period separately.

Thank you for pointing this out. The SD of the within-patients differences in the effects of the two drugs has now been presented (page 16).

As such, the projected sample sizes (in Table 4) are NOT accurate for a future crossover trial, as they have been calculated under the assumption that the SDs reflect the within-patient differences between the two treatment effects. The authors need to calculate the within-patient between-treatment SDs before calculating projected sample sizes.

Again, thank you for pointing this out. The sample sizes have been recalculated.

6) If the authors feel that some features of the crossover trial (e.g. the washout period) may be a barrier to recruitment (as evidenced by the low consent rate), they may want to suggest that a future trial should instead be based on a simpler parallel group design. In this case, the reported SD estimate (0.18) would be appropriate for the projected sample size calculations.
for a two-arm trial (ie the sample size projections in Table 4 would need to be amended to reflect results based on a two-group hypothesis test).

The recalculated sample sizes are on the assumption that a crossover design is still intended.

7) When discussing the trial results (at the top of p16), authors should stress the fact that the results are imprecise (as demonstrated by wide CIs which include 0) and therefore, even though the individual estimates of decrement in cued recall may be above or below 0, these could be chance findings rather than reflecting an indication of positive/negative treatment effects.

The imprecision of the estimates resulting from the small sample size has been alluded to on pages 15 and 19.