Author’s response to reviews

Title: Rethinking the treatment of chronic fatigue syndrome—A reanalysis and evaluation of findings from a recent major trial of graded exercise and CBT

Authors:

Carolyn Wilshire (Carolyn.Wilshire@vuw.ac.nz)
Tom Kindlon (tomkindlon@irishmecfs.org)
Robert Courtney (information785@gmail.com)
Alem Matthees (alem.matthees@gmail.com)
David Tuller (davetuller@yahoo.com)
Keith Geraghty (keith.geraghty@manchester.ac.uk)
Bruce Levin (Bruce.Levin@Columbia.edu)

Version: 1 Date: 02 Oct 2017

Author’s response to reviews:

Dr. Maria Pertl,

BMC Psychology

Re: PSYO-D-17-00052 “Rethinking the treatment of chronic fatigue syndrome—A reanalysis and evaluation of findings from a recent major trial of graded exercise and CBT”

Dear Dr. Pertl,

Thank you for giving us the opportunity to revise the above manuscript. Please find attached our completed revision, in which we have carefully addressed all of the editor’s and reviewers’ comments.
Herewith, I also attach our detailed responses to the editor’s and reviewers’ individual points.

Major changes to the content of the paper are highlighted in blue in the revised manuscript.

Thank you for your consideration.

Yours faithfully,

Carolyn Wilshire

Summary of Responses to Editor’s/Reviewers’ Comments

Editor Comments:

Two reviewers have considered your manuscript and, though their decisions were different, both highlight the need to include the original trial authors' rebuttals in defence of the trial and how it was analysed. Given the controversy surrounding the trial, I agree that it is prudent to present the original authors' perspective and rationale for their approach as appropriate throughout the manuscript.

We have now provided extensive discussion of the researchers’ original justifications for their changes. See Discussion of New Findings, lines 306-313, 327-357.

Reviewer 2's comments regarding the analysis should also be addressed.

We have carefully addressed Reviewer 2’s comments regarding the analysis. A full response. Please see detailed response to Reviewer 2 comments below.
Additional typos:

These have all been fixed.

Jim Coyne (Reviewer 1):

The manuscript italicizes controlled rather than putting in scare quotes, but a reader not thoroughly familiar with the PACE trial will wonder why the authors are confused. After all, the original trial protocol title refers to a randomized trial. The issue, of course, that the authors must be familiar with, is that it is not clear what exactly is controlled by comparison condition. In fact, the paper in The Lancet drops the designation. This has lots of implications, but the authors could simply note the discrepancy between the protocol and the published paper with primary outcomes and leave it at that. The advantage is that they would not look like they were simply confused, rather than the PACE investigators being confusing in their contradictory titles of their papers.

Actually, our original manuscript carefully avoided the term ‘controlled’ as a description of the PACE trial. However, we have now added an explicit comment on this issue. See lines 25-27, blue highlighted text.

Similarly, the authors still rather delicately with the switching of outcomes. The PACE investigators have defended their switching by indicating that it occurred before all the data were collected. I'm sure the authors know otherwise. I think briefly noting the conditions under which the switching occurred and why it is inappropriate would be worthwhile.

We have provided information about when the changes occurred, and noted that the timing is highly problematic. See Trial Summary, lines 67-70, 97-100; and Discussion of New Findings, lines 306-307, 327-328.

In general, the authors could be more explicit and cognizant of what the PACE investigators have said defending their trial so that there is no suggestion the authors are unaware. I understand their unwillingness to get embroiled in controversy, but I think some who are only familiar with the PACE investigators perspective might dismiss what is said in the current manuscript, just has the PACE investigators surely will. What the PACE investigators will say is quite predictable because they keep repeating themselves. However, the recent response to the
editorial commentary in Journal of Health Psychology is a rather succinct summary and perhaps there could be some siding of it and response to it. A little inoculation against their reaction might go a long way.

We have added a full discussion and evaluation of the PACE investigators’ justifications for their changes. See Discussion of New Findings, lines 306-313, 327-357.

Finbarr Leacy (Reviewer 2):

The appropriateness of last observation carried forward (LOCF) as the primary analysis under the intention-to-treat principle should be discussed. In particular, the assumptions underpinning and potential drawbacks of the LOCF approach and should be highlighted. Ideally, a sensitivity analysis e.g. best-case, worst-case analysis would be presented. See references [1] and [2] below for useful discussion.

Thank you for the references. We have specified exactly what we mean by the terms intention-to-treat and available-case in each context. See lines 148-152 174-175, 208-216.

Also in order to explore best-case/worst case scenarios, we have now presented descriptive data and inferential statistics for the primary outcome based on both an available case sample (counting only cases for which the key dependent variable was non-missing at 52 weeks), and an intention-to-treat sample (counting all cases with missing values at 52 weeks as not improved/not recovered). The conclusions are the same irrespective of which method is used (perhaps not surprisingly, since less than 5% of participants had missing outcome data). Again, see lines 148-152 174-175, 208-216.

We agree that it’s useful to be able to compare and contrast the results from both approaches.

In the FOI dataset, the investigators described the major outcome variables as “prorated” scales. Looking carefully at the statistical analysis plan [1], this would appear to refer to the procedures
that were adopted when there was missing item data for a particular scale. Contrary to what we wrote in the previous ms, it does not appear that the authors have used LOCF (measurements taken and 12 or 24 weeks) to replace missing data. So we have corrected this.

Unfortunately, some additional options for exploring the data were not available to us, because only a small subset of the measured variables was released to the public.

Please highlight rebuttals from the original study authors as appropriate.

We have included two new paragraphs explaining the investigators’ arguments in support of their changes. See Discussion of New Findings, lines 306-313, 327-357

Minor comments:

Abstract, RESULTS: Replace first sentence with: "We found no statistically significant differences on protocol specified primary outcome measures - improvement in self-reported fatigue and physical functioning - between the CBT and GET groups and conventional medical care following adjustment for multiple testing"

Done. Text now reads: “Scores on the original primary outcome measure – improvement in self-reported fatigue and physical function - were not significantly higher in the groups receiving CBT or GET than in the Control group after correcting for multiple comparisons.” See Abstract lines 11-14.

Abstract, CONCLUSIONS: Text insertion: " The results of PACE are [potentially] not as robust as claimed.

Agreed, our statement was too strong. We have instead opted for “These findings raise serious concerns about the robustness of the claims made about the efficacy of CBT and GET”. We think this avoids any vagueness that the word ‘potentially’ might introduce. See Abstract lines 18-19.

A brief summary of the trial entry criteria as they pertain to the primary outcome measure scores would aid the reader.
This has been added. See lines 29-34.

Line 47: Please highlight that the change in primary outcome has subsequently been defended by the original study authors, e.g. in [3].

This has been added. We have now included a full discussion of the authors’ published justifications for their various post-protocol changes (including [3]; thank you for the reference). See Discussion of New Findings, lines 306-313.

Line 48: Insert "Data for the protocol-specified improvement outcome was subsequently reported online by the PACE study authors [manuscript reference 19]"

We have added this information, together with an explanation of the timing of that release (only after the researchers had been directed by the Information Commissioner to release the relevant data). See footnote 2.

Line 72-4: Text insertion: "This small difference was reported as statistically significant, but [we conjecture] that it would very likely fail to do so using an intention-to-treat approach. There are a number of possible approaches consistent with the intention-to-treat principle. It would help the reader to clarify the particular approach to which the authors refer.

In this new ms, we have replaced the previous text with the following phrase “given the high and uneven drop-out rate, this result should be treated with caution.” Lines 111-113. There is a linking footnote which describes the results of our statistical exploration of this issue. We explain exactly how missing values were treated there (use of last observation carried forward, which is the same as baseline observation carried forward, since the outcome was only measured twice. We simply excluded participants with no baseline measure at all).

Line 77: Replace "finally" with "later".

This has been done.

Line 77: Text insertion: "there was no evidence that any yielded [statistically] significant treatment effects at 52 weeks".

This has been done.
Line 84-5: Provide numbers and/or incidence rates per group. Note that the number of serious adverse events in GET group was similar to the APT group [4, Table 4].

We have added incidence rates for the significant effects. Our reading of the adverse events results in Table 4 in the paper you refer to and its accompanying text is that the difference between APT and Control was not statistically significant. There are a lot of adverse events figures, and to keep our paper brief, we have reported only those that are significant and relevant to the discussion. We worry that it might just add confusion to report descriptive statistics for an effect that was non-significant, and that involved a treatment that is not discussed at all in this paper.

There is a lot to cover in this paper, and adverse effects outcomes are presented only for completeness. They are not a key part of our argument (indeed, we argue that the GET treatment was very conservative so it was unlikely to elicit adverse reactions in the first place).

Line 86: Provide number and percentage of respondents in each group.

Done. See lines 126-127, text highlighted in blue.

Line 135: Should read "[6, see Supplementary materials, Table C]"

We have added this.

Line 136: Please provide a brief description of how the relevant figures for the long-term follow up analysis excluding those receiving 10 or more sessions of CBT or GET post-trial were derived from Table C of the supplementary materials of [5].

Done. In fact, we had made an error: the results we had presented were actually from the subset of participants who did not receive any CBT or GET. In this revised ms, we have correctly described these data. We have also reported data for the larger subset of participants who received 9 or fewer sessions of post-trial CBT or GET. The outcome is the same irrespective of which subset is used. See lines 186-195, 236-258.

Thank you for helping us pick up that error.

Line 167: Provide number and percentage of study subjects receiving/not receiving a substantial dose of CBT/GET in each group.
Done. Raw Ns have been added to Figure 2 and percentages have been included in the text. Lines 236-239.

Line 185: Insert "after controlling for multiple testing" at end of this sentence.

We have added this. The text now reads: ‘However, these differences did not pass the threshold for statistical significance after corrections for multiple comparisons were applied’ See Discussion of New Results, lines 271-273.

Line 188-9: Remove one of the final two sentences in this paragraph for redundancy.

This has been done.

Line 204-5: Clarify that this choice has subsequently been defended by the original study authors, e.g. in [3].

This has been clarified. See Discussion of New Results, lines 307-313.

Line 206-7: Suggest splitting this sentence as follows: "Turning now to recovery rates, the definition of recovery used in published reports was changed substantially from that in the trial protocol. This made recovery much easier to achieve".

This sentence no longer appears in the revised paper.

Line 218: Insert "We suggest that" at the beginning of this sentence.

This sentence no longer appears in the revised paper.

Line 218: Replace "more" with "additional".

This sentence no longer appears in the revised paper.
Lines 218-9: Perhaps summarise comments as objection to use of term "recovery" for this outcome. Perhaps the authors would classify it as a more stringent definition of improvement?

Yes, the problem is that recovery is a word that readers take to mean a return to good health. Therefore, claims about recovery are strong claims indeed. We have commented on the definition issue in lines 344-349.

Line 319: Clarify that these changes were made prior to the examination of outcome data [4].

This has been done. Lines 67-70 (for timing of the primary outcome change), Lines 97-100 (for timing of the change to the recovery definition).

Line 357-8: I would replace "are likely" with "have the potential"

This has been done. Lines 508-9.

Figures 1 and 2: I think the data in these figures would be better presented in tables or as boxplots (Figure 1 only) with details provided on the number of subjects per group. The linear trend shown is potentially misleading to the reader. If figures are retained, the same symbol should be used for each group across figures.

We have considered this suggestion carefully, and decided that too much is lost if we change to boxplots or tables. The figure enables the reader to see not only group differences, but also the rate of change in the measures from trial commencement to the primary endpoint for each group separately. It also enables us to include normative data on the same scale, which we think is valuable.

However, we have made the symbols consistent across both figures, as you suggested. We have also included N’s in the figures themselves. In the revised ms, we have also included a bar graph depicting the protocol-specified rates of improvement, which provides some additional clarity.

Figure 2: Please verify that the means and confidence intervals presented in this figure are correct.

Checked, they are correct. But as noted above, we had erred in describing this subgroup as those having had no substantial post-trial CBT or GET. The group was in fact those who had had no post-trial CBT or GET at all.
Typos:

Page 7, footnote 2, use either "confidence interval" or "CI" consistently (same for Figure 1 caption)

Fixed. We have opted for “Confidence Intervals” throughout.

Line 36: replace "was" with "were"

I think this refers to Line 361 in the original ms (lines 389-90 in the revised ms): Text reads “a wide range of outcomes was measured”. “Was” is the correct form of the verb when the subject noun is the singular “range”.

We have now used plural throughout when referring to “data”

All of the other typos listed by Reviewer 2 have been fixed.