Formal Complaint re 'Exercise therapy for chronic fatigue syndrome'

Background

A formal complaint was submitted on 12th February 2018 from Robert Courtney regarding the review 'Exercise therapy for chronic fatigue syndrome'. This complaint was submitted as four separate documents, which each addressed the following concerns;

- Document 1: FINE trial unpublished data
- Document 2: PACE trial selective reporting bias
- Document 3: Misreporting of outcomes for physical function & overall health
- Document 4: Primary Outcome Switching

The [redacted] was asked to first conduct an independent CEU screening of the review, and then to investigate the specific issues raised in in the complaints.

The key issues raised and [redacted]'s responses to each concern are summarised below.

The initial independent CEU screening report by [redacted] is included as an appendix to this document, as it involves many of the same issues and concerns.

The proposed next steps are:

1. EMD (CEU) to share this report with the CMD
2. EMD to organise a conference call involving members of the EMD, CMD Co-Ed, and Senior Editor to discuss the next steps and agreed on how to address the issues raised in this document (both in the response to Robert Courtney's comments below, and any additional points raised by the CEU screening report).
3. CMD to communicate with the authors

Comments on feedback submitted by Robert Courtney (RC)

Document 1: FINE trial unpublished data

Summary: Regarding the use of data relating to the Wearden 2010 trial, the Cochrane review presented a Post-hoc informal analysis using unpublished data informally released by Wearden et al as a BMJ Rapid Response comment. This unpublished data used a difference scale (Likert 0,1,2,3 system with a scale of 0-33) to the one proposed in the Wearden protocol and published in the trial report (0,0,1,1 system with a scale of 0-11). This difference in scales is never acknowledged or justified in the text of the review. RC suggests that either;

- this data be replaced with the available published data using the protocol defined scales

  or

- the data clearly highlighted in the review as 'post-hoc', the risk of bias amended accordingly, and justification included in review explaining why this data is used in place of published data.

[redacted] *Comments:* I would **agree** with RC 's comments. The use of this post-hoc data must at a minimum be clearly acknowledged and justified.

Document 2: PACE trial selective reporting bias

Summary: The included study White 2011 is judged as having a low risk of bias under "Selective reporting (outcome bias)" domain. RC objects to this judgement for the following reasons;

- The trial protocol was submitted for publication after the trial had commenced (protocol submitted 2006, trial began 2005). It is questionable therefore whether it can truly be defined as a 'pre-trial report'.
- The statistical plan for the trial was submitted for publication in 2012, after the trial results had been published in 2011. The statistical analysis plan can therefore not be considered 'a priori'.
- The original protocol planned to use bimodal scoring system for a tool (Chalder Fatigue) which found no effect, but changed to Likert system in an informal post-hoc analysis which did find effect.

Courtney suggests that;

- since this is post-hoc data, a sensitivity analysis should be published (Cochrane Handbook 8.14.2)
- Study should be high risk of bias for selective outcome reporting.

[redacted] *comments:* I **agree** with RC's comments. Authors have justified their judgement of 'low risk' by stating "Our primary interest is the primary outcome reported in accordance with the protocol, so we do not believe that selective reporting is a problem". However, the current risk of bias assessment is based on the study as a whole, and this review used data for both primary and secondary outcomes in this meta-analysis. Therefore, unless they present an individual 'selective outcome reporting' judgement for each individual outcome, this judgement is not justifiable and should be changed to 'high risk', or at a minimum 'unclear risk'. Changing this judgement is unlikely to affect the overall conclusions of the review, as most outcomes are already downgraded for Risk of Bias.

Document 3: Misreporting of outcomes for physical function & overall health

Summary: RC notes that the abstract, conclusion and main discussion section indicate that there was positive treatment effect on both physical function and overall health. Yet for both of these outcomes, there was a 'significant' treatment effect only at end of treatment, but not at follow-up. These outcomes are not reflected accurately in the abstract, the main discussions or the conclusions of the review

[redacted] *comments:* I **agree** that the reporting of these outcomes is not fully reflective of all of the evidence available in this review. As previously discussed in point 2 of this screening report, it could be considered reasonable to focus attention on the post intervention evidence, as the pooled results at follow up are often based on fewer studies and participants, and demonstrate a higher inconsistency in their results. Therefore, it may be reasonable to state that results at long term follow up are 'low' or 'very low quality', and therefore the results at end of treatment are more 'certain' than those at follow up. However, I agree this uncertainty at follow up should still be explicitly stated in the Abstract, Discussion, and Conclusions of the review.

Document 4: Primary Outcome Switching

Summary (1/2): In the review protocol, authors planned to pool continuous data from all scales using SMDs. In the final version of the review, they changed this plan, and instead for the main analyses, pooled only data using the same scales using MDs, and presented SMDs in sensitivity analyses. RC also notes that "the prespecified primary analyses demonstrate that exercise therapy (versus passive control) had a significant pooled treatment effect on fatigue at end of treatment, but no significant effect at follow-up. Whereas the unplanned (revised) analyses demonstrate significant treatment effects at end of treatment but mixed outcomes at follow-up".

[redacted] *comments:* I **agree** this is a problem, particularly considering authors insufficient justification for this change in the differences between protocol and review section. Specifically: "We realise that the standardised mean difference (SMD) is much more difficult to conceptualise and interpret than the normal mean difference (MD); therefore, we decided to report both MDs and SMDs in the Results section. In general, MDs are reported in the main Results section, whereas SMDs are supplied under the "Sensitivity and subgroup analysis" subheading".

Summary (2/2): An additional issue raised by RC in this document was that he felt that authors did not sufficiently explore the possibility that "any initial positive treatment effects broadly seen in this review at end of treatment, may entirely, or to some degree, reflect biases inherent in trial methodologies that are unable to blind patients, therapists or trial investigators to the treatment arm.".

[redacted] *comments*: Authors did judge all open label studies to be at high risk of bias in the 'blinding' domain. In addition, all outcomes presented in the Summary of Findings table were downgraded due to "risk of performance bias, as they were unblinded". However, not all outcomes were assessed by GRADE. Therefore, I **agree** that this issue is not fully explored and acknowledged. To do so, at a minimum, authors should assess and present the quality of all outcomes, not just those in the Summary of Findings table.

Appendix 1 - CEU screening report

1. Review details

Title:	Exercise therapy for chronic fatigue syndrome
Authors:	Larun L, Brurberg KG, Odgaard-Jensen J, Price JR
CRG:	Common Mental Disorders
Review type:	Intervention
Archie version no.:	16.0
DOI:	10.1002/14651858.CD003200.pub7

2. Screening result

Further actions

The description of results in this review must reflect all the available evidence, the quality of the evidence, and subsequent certainty of the findings. Some details are missing from the Summary of Findings tables, and the GRADE considerations are somewhat inconsistent. Some changes to protocol methods have not been appropriately justified. Authors must either address all the points raised in this report, or provide a reasonable justification for why they are not necessary.

3. Screening findings

Implementation of protocol methods (Search date, inclusion decisions, differences between protocol & review, & additional considerations)

Differences between protocol and review

1. Authors have clearly outlined the changes made to the review since protocol. However, their justification for some of these changes remains unclear. Specifically, the following statement is problematic;

   "*We realise that the standardised mean difference (SMD) is much more difficult to conceptualise and interpret than the normal mean difference (MD); therefore we decided to report both MDs and SMDs in the Results section. In general, MDs are reported in the main Results section, whereas SMDs are supplied under the "Sensitivity and subgroup analysis" subheading*".

   This statement is not a sufficient justification for why it was more appropriate to present mean differences as the primary analyses, and SMD as a sensitivity analyses.

Interpretation (GRADE, SoF tables, full text discussion & conclusions)

Interpretation of Results

2. It is notable that authors have selectively focused their Discussion and Conclusions on the results of studies pooled at 'end of treatment', and have given much less focus to the results of studies pooled at longer term follow up. This may be justifiable to an extent, as the pooled results at follow up are often based on fewer studies and participants, and demonstrate a higher inconsistency in their results. Therefore, it may be reasonable to state that results at long term follow up are 'low' or 'very low quality', and therefore the results at end of treatment are more 'certain' than those at follow up. However, this should still be explicitly stated in the discussion and conclusion of the review. For example;

   "When exercise therapy was compared with 'passive control,' fatigue ~~was significantly~~may be reduced at end of treatment (Analysis 1.1), however we are uncertain if there is any difference between groups at follow up"

   All similar statements should be amended accordingly throughout the review.

Summary of Findings table

3. Authors must justify in the text of the review why this single comparison, and these outcomes were chosen for the Summary of Findings table.

4. Details regarding the 'setting' must be included in the PICO row of the Summary of Findings table.

5. The exact length of follow up should be included in the 'outcomes' column

6. The Summary of Findings table should only present at most seven outcomes, and each outcome should be presented once, either using the most appropriate measure, or merged as one row, and the multiple measures described narratively.

7. For the outcomes 'Self-perceived changes in overall health', and 'Drop-out', authors should only present either the 'study population' or the 'moderate' risk data.

GRADE Considerations

8. Some downgrading decisions seem inconsistent. For example, it is unclear why an outcome based on 504 participants (5 studies) was downgraded for imprecision, but an outcome based on 148 participants (1 study) was not downgraded for imprecision.

9. It is unclear how an outcome based on a single study can be downgraded for 'inconsistency'.

Authors Conclusions

10. Conclusions must reflect the quality and certainty of the evidence. Authors are advised to adhere to the language set out in Table 1 of the Cochrane Consumers and Communication guidance. For example;

"~~Encouraging~~ Low quality evidence suggests that exercise therapy ~~can~~ may contribute to alleviation of some symptoms of CFS, especially fatigue at the end of treatment, but the long term effects are uncertain. Exercise therapy ~~seems to~~may perform better than no intervention or pacing and ~~seems to~~may lead to results similar to those seen with cognitive behavioural therapy. Reported results were obtained from patients who were able to participate (not from those too disabled to attend clinics); these results were inconclusive as to type of exercise therapy and showed heterogeneity. ~~Few~~ There is probably no difference between groups regarding the number of serious adverse reactions ~~were~~ reported".

11. Authors must avoid giving direct recommendations, or referring to any intervention as 'safe' or 'effective'. Specifically, the following statement must be removed;

"*We think the evidence suggests that exercise therapy might be an effective and safe intervention for patients able to attend clinics as outpatients*."

Consistency (abstract, PLS, results in text & SoF tables)

Abstract

12. Findings for all important outcomes reported for main comparison must be described in full in the Abstract. Specifically, any outcome that was presented in the Summary of Findings table must be described in full in the Abstract, including direction, magnitude and confidence intervals of effects, and the individual quality rating of each outcome. (Statements such as "*It was not possible for review authors to draw conclusions regarding the remaining outcomes*" are insufficient).

13. Abstract conclusions must reflect the quality/certainty of the evidence. Specifically;

"*Patients with CFS may ~~generally benefit and~~ feel less fatigued following exercise therapy at the end of treatment, but we are uncertain with the long term effect. There is probably no difference between groups regarding adverse effects~~, and no evidence suggests that exercise therapy may worsen outcomes~~. Exercise may improve ~~A positive effect with respect to~~ sleep, physical function and self-perceived general health ~~has been observed~~, but ~~no conclusions for~~ we are uncertain regarding the outcomes of pain, quality of life, anxiety, depression, drop-out rate and health service resources ~~were possible~~. The effectiveness of exercise therapy ~~seems~~ may be greater than that of pacing but similar to that of CBT. Randomised trials with low risk of bias are needed to investigate the type, duration and intensity of the most beneficial exercise intervention*"

Additional comments/common errors/good practice

Date and CEU editor

Date: 10th April 2018