PEER REVIEW HISTORY

BMJ Paediatrics Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	Cognitive behavior therapy combined with music therapy for chronic fatigue following Epstein-Barr virus infection in adolescents: A feasibility study
AUTHORS	Malik, Sadaf; Asprusten, Tarjei; Pedersen, Maria; Mangersnes, Julie; Trondalen, Gro; van Roy, Betty; Skovlund, Eva; Wyller, Vegard

VERSION 1 – REVIEW

REVIEWER	Reviewer name: Maria Loades Institution and Country: University of Bath, UK Competing interests: None declared.
REVIEW RETURNED	18-Dec-2019

GENERAL COMMENTS	Overall, I commend the authors for conducting this study - and particularly for focusing on Chronic Fatigue more broadly, which is much needed. Also, for a novel intervention approach blending 2 therapies and using 2 therapists, which is quite unique in this field I think. Abstract Minor point – I'd suggest rewording the sentence “Endpoint evaluation was concealed for therapists and participants.” – I think this would read more easily as Therapists and participants were blinded to outcome evaluation. I would also suggest avoiding the idea of ‘statistically significant’ – see https://link.springer.com/article/10.1007/s10654-016-0149-3 for example. Could the sentence ‘At 15 months follow-up, there was a trend towards higher recovery rate in the intervention group’ be followed with statistics? Introduction This makes a good case for the intervention and this study specifically in the context of the wider literature. I struggle to understand from the aims of the study and the way the study is described whether this was intended as a feasibility study – i.e. to look at feasibility (can this be done?), acceptability (how do participants experience it?) and to give some indication of potential effect sizes to power a future larger scale trial, or whether this was intended as a fully powered trial. Throughout, I think this needs to be clarified for the reader and interpretations/conclusions drawn in light of what the aim was. Minor points: 1st paragraph – there is a typo in the acronym for CFS – it is written as CSF. Same paragraph, I'd suggest saying an estimated prevalence ‘of’ rather than ‘at’.

	2nd paragraph – 1st line – should read 'adults' rather than 'adult'. Methods Generally, well presented. The authors should be commended particularly for their young person participation (PPI) in developing the intervention and the study, which is impressive to see. Although I realise that the details about this cohort/the wider cohort have been published elsewhere, for the purposes of this as a standalone paper, a paragraph or two describing the participants, eligibility criteria, etc. would be useful I think. I also feel like a section on what TAU in Norway is would be helpful for the international reader particularly. And to know who the therapists were – how many, what their qualifications were, what training they were given for this study specifically, etc. Minor points - In the description of the physical activity measure, I think 'costume' made should be custom made. In the description of the intervention, there is a sentence which reads …tried seek to identify negative thoughts and feelings, and to motivate for…. I think this could be simplified grammatically as I struggle to follow what it means here. Questionnaire measures are described – my question would be, for each of these measures, is there evidence of validity and reliability specifically in adolescents? And what were the Cronbach's alpha's in the current study to give some indication of internal consistency here? The power calculation, if I understand it right, suggests that 50 participants were needed to detect a change of 2000 steps/day (which is deemed clinically significant?) at 80% power, alpha = 0.05. However, only 43 participants were randomised – so this study is underpowered? Again, I wonder about the intention of the study – was this a feasibility study or intended to be a fully powered RCT? Results I note a large attrition ('lost to follow-up') in the active intervention arm – 7/22 – 32%. Did these individuals complete therapy but not follow-up measures or did they drop out of therapy and if so, how many sessions did they complete? This is important in the context of understanding more about the acceptability of the intervention. I'd suggest moving away from the idea of 'statistical significance' and 'trends' – see my earlier comments in the abstract section, and also https://www.bmj.com/content/348/bmj.g2215.full for example. Discussion I think there are further strengths (e.g. use of objective activity measurement) and weaknesses (e.g. lack of fidelity checks) which should be highlighted. I feel that the conclusion paragraph is somewhat misleading as it overstates the conclusions that can really be drawn from 'trends' and from an underpowered exploratory study. I think this needs to be reworded.

REVIEWER	Reviewer name: Jeremy N Miles Institution and Country: Google Competing interests: None

REVIEW RETURNED	04-Jan-2020

GENERAL COMMENTS	The method section of the abstract tells us that there were 200 patients that were eligible or ineligible, but not the number in the trial, which is what we need to know. It then tells us that they were 'eligible' for a trial, but there is no mention of the methods of the trial. Finally when we get to the results, we learn about the number we care about (for the abstract): 43. All the other sample size numbers in the abstract are distracting and confusing. I haven't read beyond the abstract, but I don't understand how one can have 99% compliance among 43 patients. (In addition, this 99% is a proportion of some other, unknown, number - hence it is not meaningful.) The conclusion in the abstract is overstated, in my opinion. The authors failed to find evidence of its effectiveness, and failed to find evidence of harm. Neither of these are surprising in a study of this size - the effects would have to be extremely large to find such effects. The authors argue for interpreting a p-value of 0.12 as indicating evidence of effectiveness. By my calculation the difference in dropout across the two conditions has an OR of 0.15 and p-value of 0.08, indicating that patients found the intervention less acceptable than control. A power analysis would help us to understand the results.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1

We thank the reviewer for a thorough review and important comments.

Overall, I commend the authors for conducting this study - and particularly for focusing on Chronic Fatigue more broadly, which is much needed. Also, for a novel intervention approach blending 2 therapies and using 2 therapists, which is quite unique in this field I think.

Thank you!

Abstract

Minor point – I'd suggest rewording the sentence "Endpoint evaluation was concealed for therapists and participants." – I think this would read more easily as Therapists and participants were blinded to outcome evaluation.

We agree, the sentence has been revised accordingly

I would also suggest avoiding the idea of 'statistically significant' – see https://link.springer.com/article/10.1007/s10654-016-0149-3 for example.

We agree that “statistically significant” tend to be misleading, in particular when this study is rephrased as a feasibility study. The term has therefore been removed, as suggested, as well as the p-values.

Could the sentence ‘At 15 months follow-up, there was a trend towards higher recovery rate in the intervention group’ be followed with statistics?

We have inserted percentages in the revised version of the paper. However, we suggest not to report p-values for this trend in the abstract, in line with the considerations above.

Introduction

This makes a good case for the intervention and this study specifically in the context of the wider literature.

I struggle to understand from the aims of the study and the way the study is described whether this was intended as a feasibility study – i.e. to look at feasibility (can this be done?), acceptability (how do participants experience it?) and to give some indication of potential effect sizes to power a future larger scale trial, or whether this was intended as a fully powered trial. Throughout, I think this needs to be clarified for the reader and interpretations/conclusions drawn in light of what the aim was.

Thank you. We agree – this study should be regarded a feasibility study, and the manuscript has been rephrased accordingly.

Minor points: 1st paragraph – there is a typo in the acronym for CFS – it is written as CSF. Same paragraph, I’d suggest saying an estimated prevalence ‘of’ rather than ‘at’. 2nd paragraph – 1st line – should read ‘adults’ rather than ‘adult’.

Thanks, these typos have been corrected.

Methods

Generally, well presented. The authors should be commended particularly for their young person participation (PPI) in developing the intervention and the study, which is impressive to see.

Thank you!

Although I realise that the details about this cohort/the wider cohort have been published elsewhere, for the purposes of this as a standalone paper, a paragraph or two describing the participants, eligibility criteria, etc. would be useful I think. I also feel like a section on what TAU in Norway is would be helpful for the international reader particularly. And to know who the therapists were – how many, what their qualifications were, what training they were given for this study specifically, etc.

In the revised version of the paper, we have added information on the wider cohort, the treatment as usual, and the involved therapists.

Minor points - In the description of the physical activity measure, I think ‘costume’ made should be custom made. In the description of the intervention, there is a sentence which reads …tried seek to identify negative thoughts and feelings, and to motivate for…. I think this could be simplified grammatically as I struggle to follow what it means here.

Thanks. We have revised accordingly.

Questionnaire measures are described – my question would be, for each of these measures, is there evidence of validity and reliability specifically in adolescents? And what were the Cronbach's alpha's in the current study to give some indication of internal consistency here?

In the revised version of the manuscript, we have added 5 references that specifically address the psychometric properties of the different instruments in adolescents. Also, we have calculated Cronbach's alpha for all the applied instruments (range 0.75 to 0.94) based on the data from the entire post-infectious cohort, as stated in the revised version of the manuscript.

The power calculation, if I understand it right, suggests that 50 participants were needed to detect a change of 2000 steps/day (which is deemed clinically significant?) at 80% power, alpha = 0.05. However, only 43 participants were randomised – so this study is underpowered? Again, I wonder about the intention of the study – was this a feasibility study or intended to be a fully powered RCT?

We agree that the paragraph on power calculation appears somewhat confusing. In the revised version of the manuscript, the "feasibility character" of the study has been underlined throughout, and the paragraph on power considerations has been altered accordingly.

Results

I note a large attrition ('lost to follow-up') in the active intervention arm – 7/22 – 32%. Did these individuals complete therapy but not follow-up measures or did they drop out of therapy and if so, how many sessions did they complete? This is important in the context of understanding more about the acceptability of the intervention.

Thank you for pointing this out. The six participants lost to follow-up in the intervention group all left this group prior to or immediately after therapy startup. The primary reason, as reported by these individuals, was concern over school absence due to therapy sessions. In those participants that actually commenced therapy, the compliance was excellent. These points have been clarified in the revised version of the manuscript.

I'd suggest moving away from the idea of 'statistical significance' and 'trends' – see my earlier comments in the abstract section, and also https://www.bmj.com/content/348/bmj.g2215.full for example.

We agree, and have revised the paper accordingly

Discussion

I think there are further strengths (e.g. use of objective activity measurement) and weaknesses (e.g. lack of fidelity checks) which should be highlighted.

These points have been added to the relevant paragraph in the revised version of the manuscript.

I feel that the conclusion paragraph is somewhat misleading as it overstates the conclusions that can really be drawn from 'trends' and from an underpowered exploratory study. I think this needs to be reworded.

We agree, and have rephrased accordingly

Reviewer #2

We thank the reviewer for a thorough review and important comments.

The method section of the abstract tells us that there were 200 patients that were eligible or ineligible, but not the number in the trial, which is what we need to know. It then tells us that they were 'eligible' for a trial, but there is no mention of the methods of the trial.

We have clarified these points in the revised version of the abstract.

Finally when we get to the results, we learn about the number we care about (for the abstract): 43. All the other sample size numbers in the abstract are distracting and confusing.

We agree, and have revised accordingly.

I haven't read beyond the abstract, but I don't understand how one can have 99% compliance among 43 patients. (In addition, this 99% is a proportion of some other, unknown, number - hence it is not meaningful.)

We agree that this proportion is confusing in the context of the abstract, and have therefore removed it, just stating that compliance was high.

The conclusion in the abstract is overstated, in my opinion. The authors failed to find evidence of its effectiveness, and failed to find evidence of harm. Neither of these are surprising in a study of this size - the effects would have to be extremely large to find such effects.

We have rephrased the conclusion in the revised version of the manuscript; it should now be appropriate for a feasibility study, cf. responses to reviewer #1 and associate editor.

The authors argue for interpreting a p-value of 0.12 as indicating evidence of effectiveness. By my calculation the difference in dropout across the two conditions has an OR of 0.15 and p-value of 0.08, indicating that patients found the intervention less acceptable than control.

The drop-out in the intervention group actually occurred prior to or immediately after therapy startup, and concern over school absence was reported as the main reason for this, as stated above. That said, we agree that the study does not provide evidence of effectiveness, and have revised accordingly.

A power analysis would help us to understand the results.

The paragraph on power considerations has been rephrased in the revised version of the manuscript.