

Malik, Sadaf et al "Cognitive-behavioural therapy combined with music therapy for chronic fatigue following Epstein-Barr virus infection in adolescents: a randomised controlled trial." *BMJ Paediatrics Open* 4.1 (2020): e000797. Web. 27 Oct. 2020.
<https://bmjpaedsopen.bmj.com/content/4/1/e000797> doi: 10.1136/bmjpo-2020-000797

By Michiel Tack

3 November 2020 and 16 November 2020

This is a resubmitted version of a retracted paper and many issues remain unresolved

Inconsistencies in the retraction notice

It is unfortunate that the BMJ Paediatrics Open website doesn't make clear that this is an amended, republished version of a paper about which multiple methodological issues were raised [1] and that, eventually, was retracted [2]. The publication history, for example, does not mention the retracted version. Only at the very end of the article under "Provenance and peer review" is a link provided to the previous version. I suspect few readers will notice this. Hopefully, BMJ Paediatrics Open will place a more visible notification at the top of the page, for example the one it has put above the retraction notice [3].

Because this is a new publication that received its own DOI, the criticism of the retracted paper [1, 4] is no longer visible. This is unfortunate because, as I will try to clarify in this comment, many issues that were raised have not been addressed in this republished version.

In the retraction notice, BMJ Paediatrics Open explains that the retracted paper was misrepresented as a feasibility trial due to a mistake in the editorial process and that this was not due to error on behalf of the authors. As I will explain below, this is an implausible explanation given the information that is currently available. The review history shows that one of the reviewers was confused with how the authors originally presented their results. She stated: "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."

To this, the authors responded: "Thank you. We agree – this study should be regarded a feasibility study, and the manuscript has been rephrased accordingly." So it seems that the authors explicitly stated that this study should be regarded as a feasibility study when this was not the case. This information conflicts with the editorial statement that the authors were not in error. I think more information should be shared to clarify this contradiction.

It is unfortunate that the peer review history of this republished version has not been made public. On its website, BMJ Paediatrics Open claims it is committed to open peer review and states "as part of this commitment we make the peer review history of every article we publish publicly available." It is unclear why the peer review of this republished version became an exception to this commitment. In the peer review of the retracted version, one reviewer admitted that he hadn't read beyond the abstract. Considering the problems surrounding this work, it seems particularly important that readers are able to read through the review process and understand the decisions taken by the journal and authors.

The paper doesn't explain that the study was designed to test large effect sizes

A notable aspect of this study is that the number of participants the authors tried to recruit ($n = 60$) is much lower than the number the power analysis suggested was needed to detect a difference of moderate effect size ($n = 120$). In other words, the lack of power was not only due to recruitment problems or drop-outs; the study was designed as such.

The statistical analysis plan [5] shows that the study was powered to test a large effect size, namely 0.8 times the estimated standard deviation of the primary outcome measures (approximately 2000 steps per day). The authors justified this by saying that since "CBT alone is documented to have a moderate effect size in CFS/ME, only a substantial effect size is of direct clinical interest. Analogously, only a substantial treatment effect is of interest regarding markers of pathophysiology." [5] In the protocol, they also argued that "the FITNET study suggests that larger treatment effects might be assumed in adolescent CFS/ME patients as compared to adults." [6]

In this republished version of the manuscript, however, nothing suggests that the authors anticipated the intervention to produce a large effect size and that the intervention was designed to this hypothesis. The fact that the authors previously agreed to frame this study as a feasibility trial (when this was clearly not the case), suggests there was a willingness to disregard that the intervention failed to provide the anticipated effect sizes. Even though the republished version avoids the term 'feasibility trial', the issues remain the same.

There is little data to support the conclusion that the intervention is "feasible and acceptable"

As the intervention failed to provide the large effect sizes anticipated in the statistical analysis and protocol, the retracted paper focused on other aspects than the efficacy of the intervention, such as feasibility and acceptability. The same is true for this republished version which highlights that "combined CBT and music therapy is feasible and acceptable in adolescent postinfectious CF."

The data do not support this conclusion. More than half of the eligible individuals ($n=48$) did not consent to participation. Of those who started the program 6 out of 21 or 28% dropped out, compared to only 4.5% in the control group. The claim that the intervention is acceptable seems to be based only on a high attendance of the 15 patients who did start and stayed in the intervention arm of the study.

The authors also point to a lack of statistically significant differences of adverse events between the intervention and control group. But given that the trial was "strongly underpowered" to test the efficacy of the intervention, the same reasoning would apply to adverse events as well.

It should also be noted that the trial registration [7] lists approximately 20 different outcome measures for this trial, and with the exception of adverse effects, none of these focused on the acceptability of the intervention. Therefore, the main conclusion of this republished version of this paper remains unwarranted.

The outcome measure for post-exertional malaise was added post-hoc

In the discussion section, the authors state: "we observed a concurrent tendency of improvement of many symptom scores, including fatigue and postexertional malaise, in the intervention group." The authors fail to mention that approximately 20 outcomes measures were registered for this study but that the outcome 'postexertional malaise' was added post-hoc: it was not listed as an outcome in the protocol (1), statistical analysis plan (2), or trial registration (3).

The tendencies towards improvement were all quite small. The biggest difference found was for the primary outcome measure, where patients in the intervention group did worse than those in the control group. Patients in the intervention group had a mean of 6198 steps per day post-treatment, 2059 steps lower than in the control group. The text of the paper does not mention that this difference reached statistical significance in the per-protocol analysis.

The reported recovery rates risk misleading readers

Instead, the result section highlights a trend towards a higher recovery rate in the intervention group. The authors defined recovery as a score lower than 4 points on the Chalder Fatigue Scale using a dichotomous scoring method (range 0-11 points). A score of 4 or higher on the Chalder Fatigue Scale, however, was already used as an inclusion criterium. This means that participants could be classified as recovered as a result of reporting an improvement of just 1 point on the Chalder Fatigue Scale.

It should also be noted that the Chalder Fatigue Scale does not assess the intensity or impact of fatigue. Instead, it assesses whether participants experience fatigue-related symptoms such as having “problems starting things” or finding it “more difficult to find the right word” more than usual. Consequently, it seems inappropriate to use the term “recovery rate” for the percentage of participants who score lower than the threshold of 4 points on the Chalder Fatigue Scale. Post-treatment there was no difference in the percentage of patients meeting this 4-point threshold between the intervention and control group.

The intention-to-treat analysis of the Chalder Fatigue Scale ordinal scoring (range 0-33) was also reported, and this showed little difference between the two groups. A plausible explanation for what the authors describe as “a trend towards higher recovery rate in the intervention group” is the high drop-out rate. At follow-up, only 13 patients were in the intervention group. The analysis of recovery rates highlighted by the authors do not take into account the 8 persons who were in the intervention group but were lost to follow-up. There are little reasons to suggest that more patients in the intervention group recovered than in the control group. It is unfortunate that the authors have used this term in their manuscript.

References

- [1] Tuller D. Trial By Error: More on that Norwegian CBT/Music Therapy Study. Virology Blog. May 16, 2020. <https://www.virology.ws/2020/05/16/trial-by-error-more-on-that-norwegian-cbt-music-therapy-study/>
- [2] Marcus A. BMJ journal retracts, replaces study on chronic fatigue in children. Retraction Watch. October 21, 2020. <https://retractionwatch.com/2020/10...odBA9xf5vDnNPIIWayQ6mwr5JOVhslbloUjiRFmHwKl-k>
- [3] BMJ Paediatrics Open. Retraction: Cognitive-behavioural therapy combined with music therapy for chronic fatigue following Epstein-Barr virus infection in adolescents: a feasibility study. <http://dx.doi.org/10.1136/bmjpo-2019-000620ret>
- [4] Tack M. Inaccuracy in reporting CEBA part II. BMJ Paediatrics Open. <https://bmjpaedsopen.bmj.com/content/4/1/e000620.responses>
- [5] Statistical analysis plan – CEBA part 2. https://bmjpaedsopen.bmj.com/content/4/1/bmjpo-2020-000797supp002_data_supplement.pdf
- [6] Akershus University Hospital. Research Protocol - processing. Mental training for chronic fatigue syndrome (CFS/ME) following EBV infection in adolescents: a randomised controlled trial. Available from: https://www.ahus.no/seksjon/forskning/Documents/Forskningsgrupper/Barne-og_ungdomsklinikken/Paedia/Forskningsprotokoll_-_behandling.pdf
- [7] ClinicalTrials.gov Identifier: NCT02499302. Available from: <https://clinicaltrials.gov/ct2/show/NCT02499302>