

Draft Report for the Systematic Review\_ ME-CFS

Posted by the Centers for Disease Control and Prevention on May 17, 2021

Docket No. CDC-2021-0053

The following comments were submitted by Michiel Tack

### **The abstract should mention that the strength of evidence was insufficient or low**

Table 22. Summary of Evidence (starting on page 153) indicates that the strength of evidence was rated as ‘insufficient’ or ‘low’ for all outcomes and interventions. According to a 2014 report by the Agency for Healthcare Research and Quality (AHRQ), low strength of evidence means we have limited confidence in the estimates of effect and that the body of evidence has major or numerous deficiencies (or both). [1]

**Recommendation:** The strength of evidence grade should be made clear in the abstract. The current wording is too ambiguous.

#### **References**

[1] Agency for Healthcare Research and Quality (AHRQ). Methods Guide for Effectiveness and Comparative Effectiveness Reviews. 2014. [https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/cer-methods-guide\\_overview.pdf](https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/cer-methods-guide_overview.pdf)

### **“Inactive control therapies” requires a different formulation**

The statement that “*CBT and exercise therapy were associated with improved fatigue, function, and other outcomes versus inactive control therapies*” is problematic. In these trials, patients in the control group received usual care, specialist medical care or were put on a waiting list. The term ‘control therapies’ is inappropriate because in many of these trials patients in the control group did not receive an intervention. It is, for example, misleading to speak of ‘control therapy’ if patients are put on a waiting list for several weeks to get the treatment the other group received. More accurate would be to name these comparisons CBT or exercise therapy versus ‘no intervention’.

Because patients in the control group in these trials received no intervention, any improvement in symptom scores might be due to expectancy effects or unequal attention of therapists rather than the nature of the intervention itself. Despite this limitation, the review has not downgraded the strength of evidence of these trials for failing to use an active control intervention. Therefore, the control condition and its limitations need to be explicitly described in the text. It is important, for example, to avoid formulations that imply that CBT and exercise therapy lead to improved fatigue as it remains uncertain if improvements are due to the intervention itself or other factors such as expectancy effects or unequal attention of healthcare professionals.

**Recommendation:** A more cautious formulation is advised when discussing comparisons to control groups where no intervention was given, for example: “*Patients who received CBT and exercise therapy reported less fatigue and higher physical functioning than patients in the control group who received no intervention. It remains uncertain if these differences are due to the intervention itself or other factors such as expectancy effects and unequal attention of healthcare professionals.*”

## **The review underestimates bias due to lack of blinding on subjective outcomes**

In all trials of CBT and exercise therapy, neither patients nor therapists could be blinded to treatment allocation. Consequently, these trials are at high risk of bias when subjective outcomes are used such as the patient-reported symptom questionnaires that make up the bulk of evidence in this review. Savovic and colleagues found that lack of blinding was associated with an average 13% exaggeration of treatment effects (ratio of odds ratios: 0.87, 95% confidence interval 0.79 to 0.96). [1] A 2014 review by Hróbjartsson et al. on randomized trials that compared blinded and non-blinded groups found that the average difference in effect size for patient-reported outcome measures was 0.56 (95% confidence interval 0.41 to 0.71). In groups where patients were not blinded, the reported effect sizes were inflated by approximately half a standard deviation. [2]

In trials of CBT and graded exercise therapy, the risk of bias is particularly high because the intervention includes strong encouragements and promotion of positive expectations. A therapist manual on exercise therapy used in the PACE trial, for example, advised regular promotion of the belief that patients can improve. Therapists were instructed:

*“... it is important that you encourage optimism about the progress that they may make with this approach. You can explain the previous positive research findings of GET and show in the way you discuss goals and use language that you believe they can get better.”* [3]

The patient booklet used in the FINE trial informed trial participants how exercise therapy would make them feel:

*“You will experience a snowballing effect as increasing fitness leads to increasing confidence in your ability. You will have conquered CFS by your own effort and you will be back in control of your body again.”* [4]

Similarly, CBT encourages self-efficacy, positive expectations, and reductions in symptom focusing and catastrophizing which are likely to impact how patients report their symptoms, even if the intervention has no beneficial effect. [5] Exercise therapy and CBT were also already recommended for ME/CFS by healthcare institutions when the largest trials (PACE, FINE, FITNET, GETSET) were being conducted. Therefore, in trials of CBT and exercise therapy, it is likely that participants’ reporting of the outcome was influenced by knowledge of the intervention received. In such cases, the Cochrane handbook (second edition, 2019) indicates that the risk of bias is high. [6]

In contrast, this review rated most trials of CBT and exercise therapy as medium risk of bias, even if subjective outcome measures were used. The review also does not explain this important limitation in the abstract. The abstract only mentions heterogeneity, imprecision, inconsistency, uncertain generalizability as methodological limitations.

We would like to clarify that risk of bias assessments should not be influenced by practical limitations of study design. The AHRQ report ‘Methods Guide for Effectiveness and Comparative Effectiveness Review’ states that *“the inability to blind outcome assessors does not obviate the risk of bias from [...] lack of blinding.”*[7] Similarly, the Cochrane handbook states: *“The potential for bias cannot be ignored even if the outcome assessor cannot be blinded.”* [6]

**Recommendation:** The risk of bias assessment in non-blinded trials for subjective outcome measures should be changed from medium to high. Because the bulk of evidence in this review is affected by this risk of bias, it is recommended to mention and explain this important limitation in the abstract.

## References

- [1] Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, Als-Nielsen B, Balk EM, Gluud C, Gluud LL, Ioannidis JP. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technology Assessment*. 2012;16(35):1-82.
- [2] Hróbjartsson A, Emanuelsson F, Skou Thomsen AS, Hilden J, Brorson S. Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. *International journal of epidemiology*. 2014 Aug 1;43(4):1272-83.
- [3] Bavinton J, Darbishire L, White PD. PACE manual for therapists. Graded Exercise Therapy (GET) for CFS/ME. Final trial version: version 7 (MREC Version 2). PACE Trial Management Group; 2004. Available from: <https://media.org/images/8/89/PACE-get-therapist-manual.pdf>
- [4] Powell P. FINE trial patient booklet version 9. Royal Liverpool University Hospital; 2005. Available from: <https://web.archive.org/web/20140811161130/http://www.fine-trial.net/downloads/Patient%20PR%20Manual%20ver9%20Apr05.pdf>
- [5] Surawy C, Hackmann A, Hawton K, Sharpe M. Chronic fatigue syndrome: a cognitive approach. *Behaviour research and therapy*. 1995 Jun 1;33(5):535-44.
- [6] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd Edition. Chichester (UK): John Wiley & Sons, 2019.
- [7] Agency for Healthcare Research and Quality (AHRQ). *Methods Guide for Effectiveness and Comparative Effectiveness Reviews*. 2014. [https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/ceer-methods-guide\\_overview.pdf](https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/ceer-methods-guide_overview.pdf)

## Objective outcomes not reported

Objective outcomes are less prone to bias than patient-reported outcome measures. One of the largest studies to date on bias in randomized trials, the BRANDO project, gave the following recommendation: *“Our results suggest that, as far as possible, clinical and policy decisions should not be based on trials in which blinding is not feasible and outcome measures are subjectively assessed. Therefore, trials in which blinding is not feasible should focus as far as possible on objectively measured outcomes.”*[1]

Objective outcomes such as actigraphy, employment figures, and various fitness tests were used in numerous trials of exercise therapy and CBT. Unfortunately, this review included only data on school attendance and the 6-minute walk test. It is unclear why data on employment is not reported given that it was explicitly mentioned as a main outcome measure in the trial protocol. [2] Similarly, it is unclear why the SF-36 questionnaire was used as an indicator of physical functioning while actigraphy, cardiopulmonary exercise testing, and various other fitness tests were not.

The following objective outcome measures were used in the included trials on exercise therapy and CBT but not reported in this review:

Trial	Objective outcome measurements	Publication
Fulcher 1997	Various physiological assessments were made during a treadmill walking test including: <ul style="list-style-type: none"> <li>• Peak oxygen consumption with exercise (ml/kg/min)</li> <li>• Maximum ventilation (l/min)</li> <li>• Maximum heart rate (beats/min)</li> </ul>	Fulcher KY, White PD. Randomised controlled trial of graded exercise in patients with the chronic fatigue syndrome. <i>Bmj</i> . 1997 Jun 7;314(7095):1647.

	<ul style="list-style-type: none"> <li>• Percentage of predicted maximum heart rate</li> <li>• Recovery heart rate three minutes after test (beats/min)</li> <li>• Test duration (min)</li> <li>• Submaximal blood lactate (mmol/l)</li> <li>• Post-test blood lactate (mmol/l)</li> <li>• Maximal quadriceps voluntary contraction (with twitch interpolation)</li> </ul>	
Wearden 1998	Physiological assessment included measurement of height, weight, body fat, grip strengths and functional work capacity. The last was determined using a Bosch ERG 551 electronically braked cycle ergometer.	Wearden AJ, Morriss RK, Mullis R, Strickland PL, Pearson DJ, Appleby L, Campbell IT, Morris JA. Randomised, double-blind, placebo-controlled treatment trial of fluoxetine and graded exercise for chronic fatigue syndrome. The British Journal of Psychiatry. 1998 Jun 1;172(6):485-90.
Wallman 2004	<p>Participants were assessed on the on the Aerobic Power Index test where various measurements were recorded including:</p> <ul style="list-style-type: none"> <li>• Resting heart rate (bpm)</li> <li>• Resting systolic BP (mmHg)</li> <li>• Resting diastolic BP (mmHg)</li> <li>• Oxygen uptake (mL·kg<sup>-1</sup>·min<sup>-1</sup>)</li> <li>• Respiratory exchange ratio</li> <li>• Net blood lactate production (mmol/L)</li> <li>• Proportion of patients in which the target heart rate was reached.</li> <li>• Power output (W/kg)</li> <li>• Rate of perceived exertion / power</li> <li>• Activity level (kJ/week)</li> </ul> <p>Cognitive function was tested using a computerised version of the modified Stroop Colour Word test</p>	Wallman KE, Morton AR, Goodman C, Grove R, Guilfoyle AM. Randomised controlled trial of graded exercise in chronic fatigue syndrome. Medical Journal of Australia. 2004 May;180(9):444-8.
Moss-Morris 2005	<p>Participants underwent incremental exercise testing to determine maximum aerobic capacity (VO<sub>2</sub> peak) on a motorized treadmill. Physiological assessments include:</p> <ul style="list-style-type: none"> <li>• Maximum heart rate achieved</li> <li>• Percentage of predicted maximum heart rate</li> <li>• VO<sub>2</sub> peak (ml/kg/min)</li> </ul>	Moss-Morris R, Sharon C, Tobin R, Baldi JC. A randomized controlled graded exercise trial for chronic fatigue syndrome: outcomes and mechanisms of change. Journal of health psychology. 2005 Mar;10(2):245-59.
Jason 2007	Employment status	Jason LA, Torres-Harding S, Friedberg F, Corradi K, Njoku MG, Donalek J, Reynolds N, Brown M, Weitner BB, Rademaker A, Papernik M. Non-pharmacologic interventions for CFS: A randomized trial. Journal of Clinical

		Psychology in Medical Settings. 2007 Dec 1;14(4):275-96.
Wearden 2010 (FINE)	<p>Capacity to exercise was assessed using a timed step test.</p> <p>The mediation analysis published three years after the main trial publication reports: <i>“There were no between group differences in any of the step test measures at 20 or 70 weeks.”</i></p> <p>The full data of the step test have made publicly available due to a Freedom of Information Request by Kathryn Dickenson at:  <a href="https://www.whatdotheyknow.com/request/cfs_fine_what_story_does_the_obj_2#incoming-1026066">https://www.whatdotheyknow.com/request/cfs_fine_what_story_does_the_obj_2#incoming-1026066</a></p>	Wearden AJ, Emsley R. Mediators of the effects on fatigue of pragmatic rehabilitation for chronic fatigue syndrome. Journal of consulting and clinical psychology. 2013 Oct;81(5):831.
White 2011 (PACE)	<p>Data on service use was reported for the following medical services:</p> <ul style="list-style-type: none"> <li>• Primary care</li> <li>• Other doctor</li> <li>• Health professional</li> <li>• Inpatient</li> <li>• Accident and emergency</li> <li>• Medication</li> <li>• Complementary healthcare</li> <li>• Other health/social services</li> <li>• Informal care</li> <li>• Total health costs</li> </ul> <p>The following data were also reported in the economic analysis:</p> <ul style="list-style-type: none"> <li>• Lost employment</li> <li>• Income benefits</li> <li>• Illness/disability benefits</li> <li>• Payments from income protection schemes or private pensions</li> </ul>	McCrone P, Sharpe M, Chalder T, Knapp M, Johnson AL, Goldsmith KA, White PD. Adaptive pacing, cognitive behaviour therapy, graded exercise, and specialist medical care for chronic fatigue syndrome: a cost-effectiveness analysis. PLoS One. 2012 Aug 1;7(8):e40808.
White 2011 (PACE)	<p>Fitness and perceived exertion were measured using a step test. The data was not reported in full. The authors reported included a graph in the meditation analysis that showed that there was no significant difference between exercise therapy, specialist medical care, adaptive pacing therapy, or CBT.</p>	Chalder T, Goldsmith KA, White PD, Sharpe M, Pickles AR. Rehabilitative therapies for chronic fatigue syndrome: a secondary mediation analysis of the PACE trial. The Lancet Psychiatry. 2015 Feb 1;2(2):141-52.
Sharpe 1996	<p>The authors reported the percentage that improved in work status in both groups.</p>	Sharpe M, Hawton K, Simkin S, Surawy C, Hackmann A, Klimes I, Peto T, Warrell D, Seagroatt V. Cognitive behaviour therapy for the chronic fatigue syndrome: a randomised

		controlled trial. <i>Bmj.</i> 1996 Jan 6;312(7022):22-6.
Janse 2018	Physical activity assessed with actigraphy (mean waken score). The data was published in the supplementary material.	Janse A, Worm-Smeitink M, Bleijenberg G, Donders R, Knoop H. Efficacy of web-based cognitive-behavioural therapy for chronic fatigue syndrome: randomised controlled trial. <i>The British Journal of Psychiatry.</i> 2018 Feb;212(2):112-8.
Knoop 2008	Physical activity was assessed with actigraphy. The data was reported in Wiborg et al. 2010.	Wiborg JF, Knoop H, Stulemeijer M, Prins JB, Bleijenberg G. How does cognitive behaviour therapy reduce fatigue in patients with chronic fatigue syndrome? The role of physical activity.
O' Dowd 2006	<p>Participants performed an incremental shuttle walk test. The number of shuttles walked and the walking speed were reported as outcome measures.</p> <p>Cognitive function was assessed with the short-form neurocognitive battery. Outcomes include mood (alertness, hedonic tone, anxiety) recall (total words recalled, correct words), simple reaction time (reaction time, trials completed), and repeated digits detection (reaction time, hit rate).</p> <p>A health economic questionnaires assessed personal expenses, medication use, private treatments, informal help, and employment details.</p>	O'dowd H, Gladwell P, Rogers CA, Hollinghurst S, Gregory A. Cognitive behavioural therapy in chronic fatigue syndrome: a randomised controlled trial of an outpatient group programme. <i>HEALTH TECHNOLOGY ASSESSMENT- SOUTHAMPTON-</i> . 2006 Oct 1;10(37).
Stubhaug 2008	Cardiorespiratory fitness was assessed by the Åstrand-Ryhming test (indirect test of maximal oxygen uptake (VO <sub>2</sub> max)), performed on an ergometer bicycle.	Stubhaug B, Lie SA, Ursin H, Eriksen HR. Cognitive-behavioural therapy v. mirtazapine for chronic fatigue and neurasthenia: randomised placebo-controlled trial. <i>The British Journal of Psychiatry.</i> 2008 Mar;192(3):217-23.
Stulemeijer 2005	Physical activity assessed with actigraphy. The data was reported by Wiborg et al. 2010.	Wiborg JF, Knoop H, Stulemeijer M, Prins JB, Bleijenberg G. How does cognitive behaviour therapy reduce fatigue in patients with chronic fatigue syndrome? The role of physical activity.

**Recommendation:** Objective outcomes are less prone to bias when patients and therapists are not blinded to treatment allocation. They can therefore provide more reliable information than patient-reported outcomes in trials where neither patients nor therapists can be blinded. It is recommended to include all

the objective outcome measurements that were used in trials of exercise therapy and CBT in this review and to mention their results in the abstract.

## References

[1] Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, Als-Nielsen B, Balk EM, Gluud C, Gluud LL, Ioannidis JP. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technology Assessment*. 2012;16(35):1-82.

[2] Roger Chou, Marian McDonagh, Jessica Griffin, Sara Grusing. Diagnosis and treatment of myalgic encephalomyelitis/chronic fatigue syndrome: a systematic evidence review. PROSPERO 2019 CRD42019142805 Available from: [https://www.crd.york.ac.uk/prospERO/display\\_record.php?ID=CRD42019142805](https://www.crd.york.ac.uk/prospERO/display_record.php?ID=CRD42019142805)

## Reports on harms not included

The AHRQ report [1] states that: *“relying solely on published RCTs [Randomized Controlled Trials] to evaluate harms in CERs [Comparative Effectiveness Reviews] is problematic. First, most RCTs lack prespecified hypotheses for harms. Rather, hypotheses are usually designed to evaluate beneficial effects, with assessment of harms a secondary consideration. As such, the quality and quantity of harms reporting in clinical trials is frequently inadequate.”* The report advises to *“gather evidence on harms from a broad range of sources, including observational studies, particularly when clinical trials are lacking.”*

Observational studies refer to a broad range of study designs, including case reports and uncontrolled series of patients receiving an intervention. According to the AHRQ report *“all can yield useful information as long as their specific limitations are understood.”* Similarly, AMSTAR-II emphasizes that *“the failure to include non-randomised studies in a review of adverse outcomes of treatment may be a critical flaw.”* [2]

Unfortunately, this review includes only data on harms from randomized controlled trials. The abstract states that there is limited evidence that exercise was *“not associated with increased risk of serious adverse events or worsening of symptoms.”* The data on harms for exercise therapy, however, came from only two studies: PACE and GETSET. The GETSET trial did not assess the safety of graded exercise therapy (GET) but a self-help guide to exercise. Therefore its findings may not be generalizable to full courses of GET in clinical practice. As noted by Kindlon in 2017, data on harms in the PACE trial was not reported in accordance with the pre-specified protocol. [3] GETSET and PACE also did not provide objective evidence of adherence to GET.

This review ignores the multiple patient surveys where participants indicate that their health deteriorated after trying GET. This finding has been reported consistently for more than 20 years by ME/CFS patients. Surveys were conducted by patient organizations in various countries, including the United States, Norway, the Netherlands, Australia, and the United Kingdom. These were summarized and reviewed by Geraghty and colleagues in 2019. [4]

**Recommendation:** Relying solely on randomized controlled trials to evaluate harms can be misleading, especially if there is a large body of observational evidence that points to a different conclusion. It is therefore recommended to include other reports of safety and harm of the interventions assessed in this review, in particular graded exercise therapy.

## References

[1] Agency for Healthcare Research and Quality (AHRQ). Methods Guide for Effectiveness and Comparative Effectiveness Reviews. 2014. [https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/ceer-methods-guide\\_overview.pdf](https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/ceer-methods-guide_overview.pdf)

[2] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *bmj*. 2017 Sep 21;358.

[3] Kindlon T. Do graded activity therapies cause harm in chronic fatigue syndrome?. *Journal of health psychology*. 2017 Aug;22(9):1146-54.

[4] Geraghty K, Hann M, Kurtev S. Myalgic encephalomyelitis/chronic fatigue syndrome patients' reports of symptom changes following cognitive behavioural therapy, graded exercise therapy and pacing treatments: Analysis of a primary survey compared with secondary surveys. *Journal of health psychology*. 2019 Sep;24(10):1318-33.

## Heterogeneity

The abstract states that graded exercise is “*more effective than inactive control therapies (usual care, usual specialist care, or an attention control) in improving fatigue, function, and other outcomes.*” These meta-analyses, however, are all affected by high heterogeneity (I<sup>2</sup>= 85%-95%). The main cause of this heterogeneity seems to be the outlier trial by Powell and colleagues. [1] It reported reductions in fatigue and physical functioning that were more than 4 times as large as reported by more recent, large studies such as FINE, PACE, and GETSET.

For many outcomes, heterogeneity is greatly reduced when the outlier by Powell et al. is eliminated but this also lowers the effect sizes substantially (see overview below). Without the outlier by Powell et al., the differences found are consistently lower than the minimum clinically important difference (MCID). In other words, if heterogeneity is reduced by removing the outlier the point estimates of effect are no longer clinically significant.

Effect of exercise therapy versus inactive control with the outlier trial (Powell 2001) excluded						
Outcome measure	Effect size (mean difference or SMD)	95% confidence interval	MCID	Number of trials	Number of participants	Heterogeneity (I <sup>2</sup> )
Fatigue, post-treatment	SMD -0.32	-0.52 to -0.12	2.3 (SMD of 0.36)*	5	886	46%
Fatigue, follow-up	SMD -0.35	-0.58 to -0.12	2.3 (SMD of 0.36)*	2	477	35%
Physical function, post-treatment	5.89	2.52 to 9.25	10	4	817	0%
Physical function, follow-up	6.37	1.89 to 10.85	10	2	563	0%
Depression, post-treatment	-0.97	-1.71 to -0.23	1.7	3	540	8%
Depression, follow-up	-0.85	-1.61 to -0.08	1.7	2	551	0%
Anxiety, post-treatment	-1.31	-2.12 to -0.51	1.7	2	472	0%
Anxiety, follow-up	-0.38	-1.52 to 0.76	1.7	2	549	49%
Sleep, post-treatment	SMD -0.31	-0.57 to -0.05	/	1	272	/
Sleep, follow-up	SMD -0.26	-0.56 to 0.03	/	2	552	65%



Pain	“A post-hoc analysis from the PACE trial found exercise associated with decreased severity of muscle pain (mean difference -0.42, 95% CI -0.11 to -0.73) and joint pain (mean difference - 0.25, 95% CI -0.70 to - 0.57) at post-intervention follow-up (each measured on a 0 to 4 scale).”
*The conversion of the minimum clinically important difference of 2.3 for fatigue to an SMD of 0.36 was taken from the Cochrane review by Larun et al. 2019.	

**Recommendation:** The meta-analyses for exercise therapy versus no intervention (usual care, usual specialist care, or an attention control) suffers from unacceptably high heterogeneity. This is mostly due to the outlier Powell et al. that reported much larger treatment effects than the other trials. If this outlier is excluded, the point estimates of the mean differences are no longer clinically significant. Therefore, statements that exercise therapy is effective at improving fatigue, function, and other outcomes should be avoided. We recommend including an explanation in the main text that exercise therapy did not lead to a clinically significant effect if the outlier Powell et al. is excluded from the analyses.

## References

[1] Powell P, Bentall RP, Nye FJ, Edwards RH. Randomised controlled trial of patient education to encourage graded exercise in chronic fatigue syndrome. *Bmj*. 2001 Feb 17;322(7283):387.

## Reporting bias

It is unclear why in ‘Appendix F. Risk of Bias for Randomized Controlled Trials’, all trials received the label “Yes” for “Outcomes Pre-specified”. It is recommended to preserve this judgment to trials where a trial protocol or registration was publicly available prior to the start of the trial and where the results were published according to this pre-specified plan.

The FITNET trial (Nijhof et al. 2012) is at high risk of reporting bias because physical performance, as measured with an actometer, was listed in the protocol as an outcome measure but its results were never reported. Similarly, the PACE trial (White et al. 2011) did not report outcome measures in accordance with the trial protocol, and these changes were never clarified or explained. In the SMILE study (Crawley et al. 2018) the authors switched their primary outcome measure as their study transformed from a feasibility study into a full trial.

**Recommendation:** Several studies such as FITNET, PACE, and SMILE should be rated as high risk of reporting bias because the authors have not reported the results in accordance with a pre-specified trial protocol.

## Recovery criteria in the PACE trial

The PACE trial protocol specified a definition of recovery but the authors changed this definition after viewing the results. They clarified that the change was made so that rates were “*more consistent both with the literature, and with our clinical experience.*” [1] The changed definition, however, lowered recovery thresholds so that, for some measurements, it became lower than the threshold used to include patients in the trial. As noted by Wilshire et al. the revised recovery threshold score for physical functioning, for example, became “*so low that it is close to the mean score of patients with osteoarthritis of the hip, rheumatoid arthritis, and Class II congestive heart failure.*” [2]

**Recommendation:** We recommend using the pre-specified definition of recovery in the PACE trial. The revised definition was formulated post-hoc and suffers from multiple inconsistencies.

## References

[1] Sharpe M, Goldsmith K, Chalder T. The PACE trial of treatments for chronic fatigue syndrome: a response to WILSHIRE et al. *BMC psychology*. 2019 Dec;7(1):1-5.

[2] Wilshire C, Kindlon T, Matthees A, McGrath S. Can patients with chronic fatigue syndrome really recover after graded exercise or cognitive behavioural therapy? A critical commentary and preliminary re-analysis of the PACE trial. *Fatigue: Biomedicine, Health & Behavior*. 2017 Jan 2;5(1):43-56.

## **Problems with the Chalder Fatigue Scale**

Cochrane's risk of bias tool (second version) asks authors to consider bias in measurement of the outcome, namely "*Whether the method of measuring the outcome is appropriate.*" [1] Similarly, the methods guide for effectiveness and comparative effectiveness reviews by the AHRQ encourages reviewers to evaluate if outcomes were assessed using valid and reliable measures. [2] Unfortunately, in this review, little attention is paid to the validity and reliability of outcome measures.

One of the most used outcome measures in randomized trials of ME/CFS is the Chalder Fatigue Questionnaire (CFQ) which suffers from ceiling effects and interpretation problems when tracking symptoms over time.

### Ceiling effects

Ceiling effects have especially been noted with the bimodal scoring system (0-11) of the CFQ. [3,4] In the trial by Powell et al., for example, patients had a fatigue score of 10.28 out of 11 points at baseline. In the FINE Trial, patients had a score of 10.45 out of 11 points at baseline. An increase in fatigue might not be recorded in these trials as most participants already had a score close to the maximum of the scale.

If a worsening of fatigue is equally likely in the exercise- and passive control group, ceiling effects might not have favored one over the other. But this assumption is rather unlikely as a worsening of symptoms following (physical) exertion is one of the characteristics of ME/CFS and in multiple surveys, ME/CFS patients report to have worsened following GET. More generally, participating in an exercise intervention has been shown to increase the relative risk of non-serious adverse events. [5] Therefore, it seems reasonable to assume that more ME/CFS patients in the exercise than in the control group could have experienced an increase in fatigue after scoring (close to) the maximum on the CFQ. This would have distorted the results and caused a false impression of improvement.

I would like to spell out this argument more clearly as it could easily be overlooked or misinterpreted. To clarify we could use an imaginary exercise trial where all participants have a fatigue score of 9 out of 11 at the start of the trial. In the passive control group, half of the participants' fatigue scores increase by 2 points while in the other half it decreases by 2 points. The average of the control group at the end of the trial would still be 9 out of 11. In the exercise group, half of the participants' fatigue scores increase by 6 points while for the other half it decreases by 6 points. Their average is not 9 but 7 out of 11 because an increase of 6 points could not be fully recorded on the scale. Something similar might have happened in exercise trials for ME/CFS where patients scored close to the maximum of the CFQ at the start of the trial.

### Interpretation problems

The CFQ also has problems of interpretability as it asks trial participants if they experience fatigue symptoms less than usual compared to when they were last well. [6,7] When questionnaires are completed after treatment ends, patients might be confused and compare themselves to how they were before the trial started, rather than when they were last well. This misinterpretation occurred in a Japanese trial exploring the effects of yoga in ME/CFS. [8] One of the participants recorded very low scores on the

CFQ post-treatment because she was confused by the baseline comparison. The authors note: “*whereas the intent was to compare her current condition to when she last felt well, she had been sick and almost bed-bound for some years so she misunderstood ‘than usual’ as ‘than the days sick in bed’ because it had become a regular part of life for her.*” In a trial on CBT in multiple sclerosis, patients in both the intervention and control group reported having less fatigue on the CFQ than healthy controls at the end of the trial. [9] A plausible explanation is that patients wanted to indicate that they had less fatigue since the start of the trial rather than compared to when they were last well. These interpretation problems question the validity of the CFQ in measuring improvements over time.

**Recommendation:** The ceiling effects and interpretation problems of the CFQ should be mentioned as a limitation in this review.

## References

- [1] Agency for Healthcare Research and Quality (AHRQ). Methods Guide for Effectiveness and Comparative Effectiveness Reviews. 2014. [https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/cer-methods-guide\\_overview.pdf](https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/cer-methods-guide_overview.pdf)
- [2] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions. 2nd Edition. Chichester (UK): John Wiley & Sons, 2019.
- [3] Stouten B. Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. BMC Health Serv Res. 2005 May 13;5:37.
- [4] Morriss RK, Wearden AJ, Mullis R. Exploring the validity of the Chalder Fatigue scale in chronic fatigue syndrome. J Psychosom Res. 1998 Nov;45(5):411–7.
- [5] Niemeijer A, Lund H, Stafne SN, Ipsen T, Goldschmidt CL, Jørgensen CT, et al. Adverse events of exercise therapy in randomised controlled trials: a systematic review and meta-analysis. Br J Sports Med. 2019 Sep 28;pii: bjsports-2018-100461.
- [6] S4ME: Submission to the public review on common data elements for ME/CFS: Problems with the Chalder Fatigue Questionnaire [Internet]. Science for ME. [cited 2019 Nov 27]. Available from: <https://www.s4me.info/threads/s4me-submission-to-the-public-review-on-common-data-elements-for-me-cfs-problems-with-the-chalder-fatigue-questionnaire.2065/>
- [7] Vink M, Vink-Niese A. Graded exercise therapy for myalgic encephalomyelitis/chronic fatigue syndrome is not effective and unsafe. Re-analysis of a Cochrane review. Health Psychol Open. 2018 Dec;5(2):2055102918805187.
- [8] Takakura S, Oka T, Sudo N. Changes in circulating microRNA after recumbent isometric yoga practice by patients with myalgic encephalomyelitis/chronic fatigue syndrome: an explorative pilot study. Biopsychosoc Med. 2019 Dec 2;13(1):29.
- [9] van Kessel K, Moss-Morris R, Willoughby E, Chalder T, Johnson MH, Robinson E. A randomized controlled trial of cognitive behavior therapy for multiple sclerosis fatigue. Psychosom Med. 2008 Feb;70(2):205–13.

## Publication bias

AMSTAR-II lists “*Assessment of presence and likely impact of publication*” as a critical domain that can affect the validity of a review and its conclusion. It is considered a key issue that authors have “*shown an awareness of the likely impact of PB [publication bias] in their interpretation and discussion of the results and performed a sensitivity analyses [sic] to determine how many missing ‘null’ studies would be needed to invalidate the results they obtained.*” [1] In the Cochrane handbook it is stated that: “*failure to consider the potential impact of non-reporting biases on the results of the review can lead to the uptake of ineffective and harmful interventions in clinical practice.*” [2]

Unfortunately, this review did not evaluate publication bias because no meta-analysis included 10 trials or more. It could, however, mention that the review by Marques and colleagues, which combined all 'behavioral interventions with a graded physical activity component' for ME/CFS, found evidence of publication bias. [3]

**Recommendation:** The review by Marques and colleagues found evidence of publication bias when all behavioral interventions for ME/CFS with a graded physical activity component were combined. This finding could be mentioned in the section 'limitations'.

## References

[1] Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, Moher D, Tugwell P, Welch V, Kristjansson E, Henry DA. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *bmj*. 2017 Sep 21;358.

[2] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd Edition. Chichester (UK): John Wiley & Sons, 2019.

[3] Marques MM, De Gucht V, Gouveia MJ, Leal I, Maes S. Differential effects of behavioral interventions with a graded physical activity component in patients suffering from chronic fatigue (syndrome): an updated systematic review and meta-analysis. *Clinical psychology review*. 2015 Aug 1;40:123-37.

## Inclusion and exclusion of studies

Regarding Key Question 1: 'In patients undergoing evaluation for possible ME/CFS, what is the frequency of non-ME/CFS conditions?' One relevant study is not mentioned in the list of included or excluded studies, namely:

Johnston SC, Staines DR, Marshall-Gradisnik SM. Epidemiological characteristics of chronic fatigue syndrome/myalgic encephalomyelitis in Australian patients. *Clin Epidemiol*. 2016 May 17;8:97-107.

The authors performed a clinical examination of patients enrolled in an Australian ME/CFS research database. Recruitment was based on self-identification in response to an advertisement in ME/CFS community support networks across Australia, as well as a general advertisement on local radio and social media. Patients had to report a diagnosis of ME/CFS by their primary physician. According to the authors *"there were 14.58% reporting chronic fatigue but did not meet criteria for CFS/ME and 23.18% were considered noncases due to exclusionary conditions."*

The study by Stubhaug et al. selected patients on the basis of ICD-10 criteria for neurasthenia. These differ from ME/CFS criteria. Even if most patients in the study by Stubhaug et al. fulfilled the latter, participants might have differed from a representative sample of ME/CFS patients as they were first selected with the ICD-10 neurasthenia criteria. The previous AHRQ report on the management of ME/CFS excluded this study.

Stubhaug B, Lie SA, Ursin H, et al. Cognitive-behavioural therapy v. mirtazapine for chronic fatigue and neurasthenia: randomised placebo-controlled trial. *Br J Psychiatry*. 2008;192(3):217-23. PMID: 18310583. Exclusion code: 5

The study by Chan et al. 2013 was included but selected patients without clinical confirmation of their diagnosis as ME/CFS criteria require. Participants were not diagnosed with CFS but "as having CFS-like illness."

Chan JSM, Ho RTH, Wang CW, et al. Effects of qigong exercise on fatigue, anxiety, and depressive symptoms of patients with chronic fatigue syndrome-like illness: a randomized controlled trial. Evid Based Complement Alternat Med. 2013

**Recommendation:** We advise to include the study by Johnston et al. and exclude the study by Stubhaug et al. and the study by Chan et al.