Critical Review

HEALTH PSYCHOLOGY OPEN

Health Psychology Open
July-December 2018: 1–12
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2055102918805187
journals.sagepub.com/home/hpo
SAGE

Graded exercise therapy for myalgic encephalomyelitis/chronic fatigue syndrome is not effective and unsafe. Re-analysis of a Cochrane review

Mark Vink[1] and Alexandra Vink-Niese[2]

Abstract

The analysis of the 2017 Cochrane review reveals flaws, which means that contrary to its findings, there is no evidence that graded exercise therapy is effective. Because of the failure to report harms adequately in the trials covered by the review, it cannot be said that graded exercise therapy is safe. The analysis of the objective outcomes in the trials provides sufficient evidence to conclude that graded exercise therapy is an ineffective treatment for myalgic encephalomyelitis/chronic fatigue syndrome.

Keywords

chronic fatigue syndrome, Cochrane review, graded exercise therapy, myalgic encephalomyelitis

Introduction

A recent Cochrane review of graded exercise therapy (GET) for chronic fatigue syndrome (CFS) concluded that GET is effective and safe (Larun et al., 2017). The review was to determine the effects of exercise therapy for patients with CFS as compared with any other intervention or control. It included eight randomised controlled studies and reported data from 1510 participants: Fulcher and White (1997) (66), Wearden et al. (1998) (136), Powell et al. (2001) (148), Wallman et al. (2004) (61), Moss-Morris et al. (2005) (49), Jason et al. (2007) (114), Wearden et al. (2010b) (296) and White et al. (2011) (640) (number of patients in each study between parentheses).

Larun et al. (2017) state that CFS is sometimes called myalgic encephalomyelitis (ME), but in the rest of their document, the term CFS is used. The same has been done here to avoid any confusion.

The review has been subject to criticism, for example, Kindlon (2015) and Courtney (2016). Here, we analyse the review and find seven general issues. We also identify a number of problems revealed in the original studies, including a failure to report harms. Our analysis shows the review's conclusion that GET is effective and safe is not supported by the evidence. When the objective outcomes of the trials are considered, it is possible to state that GET is ineffective for CFS.

Seven general issues

1. There are questions about conflicts of interest.

The review itself and seven of the eight studies (Jason et al., 2007, the exception) were conducted by researchers with an allegiance to a particular model of CFS and to two interventions, cognitive behavioural therapy (CBT) and GET.

In studies examining more than one treatment approach, the treatment favoured by the researchers tends to outperform other treatments (Luborsky et al., 1999, 2002; Munder et al., 2012). Several factors may contribute to this effect, but one is likely to be the manner in which the non-favoured 'comparison' treatment is conceptualised and implemented.

[1]Soerabaja Research Center, The Netherlands
[2]Independent Researcher, Germany

Corresponding author:
Mark Vink, Family and Insurance Physician, Soerabaja Research Center, 1096 HH Amsterdam, The Netherlands.
Email: markvink.md@outlook.com

Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (http://www.creativecommons.org/licenses/by-nc/4.0/) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access pages (https://us.sagepub.com/en-us/nam/open-access-at-sage).

A researcher's enthusiasm for a particular treatment can also lead his/her to overinterpret their findings or overlook limitations (Wilshire, 2017).

Selection of the editorial group from the Cochrane Common Mental Disorders Group suggests a bias to a particular view of CFS as ME has been classified as a neurological disease by the World Health Organization (WHO) since 1969 with CFS as an equivalent. The four authors of the review are all known proponents of the biopsychosocial model, which views GET as an effective treatment for CFS.

Before they conducted their research, Fulcher (and White), Powell, Wearden, Wallman, Moss-Morris and White are all known to have favoured the approach to the illness being tested. It is notable that the one study conducted by a researcher without an allegiance to the model concluded that none of the four treatment strategies was superior to another treatment strategy in all areas (Jason et al., 2007).

2. A study was excluded, which contradicted the main findings.

Núñez et al. (2011) was excluded from the Cochrane review because exercise therapy was a minor part of the intervention and it did not measure outcomes viewed as primary outcomes in the review (Larun et al., 2017).

The trial compared multidisciplinary treatment combining CBT, GET and pharmacological treatment with usual treatment, with 1-year follow-up after the end of treatment. It concluded that at 12 months, the interventions did not improve health-related quality of life scores and led to worse physical function and bodily pain scores. Nuñez found that the combination of CBT and GET is ineffective and not evidence-based and may in fact be harmful.

3. Criteria used in the trials were too broad.

As inclusion criteria, five studies (Fulcher and White, 1997; Powell et al., 2001; Wearden et al., 1998, 2010b; White et al., 2011) used the Oxford criteria. Three studies (Moss-Morris et al., 2005; Jason et al., 2007; Wallman et al., 2004) used the Centres for Disease Control and Prevention (CDC) criteria, also known as the Fukuda Criteria (Fukuda et al., 1994).

The only requirement for the Oxford criteria (Sharpe et al., 1991) is 6 months or more of unexplained fatigue. It was created as an alternative, less strict, operational definition which is essentially chronic fatigue in the absence of neurological signs with psychiatric symptoms as common associated features (David, 1991).

The Oxford criteria are untenable because they inappropriately select healthy subjects with mild fatigue and chronic idiopathic fatigue and mislabel them as CFS (Baraniuk, 2017). The American National Institutes of Health (NIH) concluded in 2014 that the Oxford criteria are flawed and include people with other conditions, confounding the ability to interpret the science (Green et al., 2014a). Continuing to use the Oxford definition may impair progress and cause harm (Green et al., 2014a, 2014b). The Agency for Healthcare Research and Quality (AHRQ) stated that using the Oxford case definition results in a high risk of including patients who may have an alternate fatiguing illness or whose illness resolves spontaneously with time (Smith et al., 2016). Both the NIH and AHRQ recommend that the Oxford definition should be retired.

The Fukuda criteria are the most commonly used criteria for CFS. Patients need to have 6 months or more of unexplained chronic fatigue and have a minimum of four out of eight criteria. However, ME is characterised by an abnormally delayed muscle recovery after trivial exertion (Ramsay, 1988), commonly referred to as post-exertional malaise (PEM). PEM is now accepted as a core symptom (IOM, 2015) of CFS but is only optional and not compulsory for diagnosis in Fukuda as it is one of the eight additional criteria (Fukuda et al., 1994). A total of 15 per cent of people labelled by these criteria as having CFS were in fact healthy people (Friedberg et al., 2000).

Moreover, around 40 per cent of participants in seven of the eight studies (Wallman et al., 2004, the exception with 12%) suffered from co-morbid psychiatric disorders. The presence of a medical or psychiatric condition that may explain the chronic fatigue state excludes the classification as CFS in research studies because overlapping pathophysiology may confound findings specific to CFS (Reeves et al., 2003).

The use of the Oxford and the Fukuda criteria in all eight of the studies means that they may have included patients who did not have CFS but who were susceptible to the interventions.

4. Entry score requirements were not sufficiently strict.

In five studies (Fulcher and White, 1997; Jason et al., 2007; Moss-Morris et al., 2005; Powell et al., 2001; Wearden et al., 1998), high-functioning patients may have been included.

Three studies (Fulcher and White, 1997; Jason et al., 2007; Moss-Morris et al., 2005) did not have entry score requirements. In two further studies, entry requirements were such that relatively high-functioning participants were included: Wearden et al. (1998) had a physical functioning score of up to 83.3, and in Powell et al. (2001), patients could be selected with an almost normal physical functioning score of up to 24 (included) out of 30.

Further questions about inclusion are raised in Moss-Morris et al. (2005), where 77.6 per cent of patients were well enough to be in work, and in both Fulcher and White (1997) and Moss-Morris et al. (2005), where participants had normal VO_{2max} scores.

It is likely that in five of the studies in the review patients were therefore included who may not have had CFS.

5. The review used subjective fatigue measured by questionnaires as the primary outcome.

All eight trials in the review were by definition unblinded, yet the review used two subjective primary outcomes, fatigue and adverse outcomes.

Patient self-report is an unreliable measure (Wechsler et al., 2011). Lack of patient blinding combined with self-reporting of outcomes leads to pronounced bias as patients become prone to outside influences leading to the erroneous inference of efficacy in its absence, thus making subjectively assessed outcomes unreliable (Hróbjartsson et al., 2014; Lilienfeld et al., 2014). Low correlation between objective and subjective activity measurements (Scheeres et al., 2009) is not confined to the chronically ill but is also present in the healthy population (Van den Berg-Emons et al., 2011). For patients with CFS, there is a particular problem with subjective outcomes as they may feel better able to cope with daily activities because they have reduced their expectations of what they should achieve, rather than because they have made any recovery as a result of the intervention (Whiting et al., 2001).

The review itself acknowledges this problem and states that all studies were at risk of performance bias, as they were unblinded (Larun et al., 2017). The only way to correct for this problem in unblinded trials is by using well-designed control groups and objective primary outcomes (Edwards, 2017; Lilienfeld et al., 2014). All trials in the review, apart from Powell et al. (2001), used objective outcomes, so it would have been possible for the Cochrane review to have used them.

Any conclusion that the intervention was effective must be seen as unreliable because of the use of subjective outcomes in unblinded studies.

6. The Chalder Fatigue Scale is flawed.

Seven of the eight trials (Jason et al. (2007), the exception) used the Chalder Fatigue Scale to measure fatigue.

Four flaws have been identified with the use of the Scale.

First, it does not provide a comprehensive reflection of fatigue-related severity, symptomology, or functional disability in CFS (Haywood et al., 2011), as it was developed by mental health professionals, and many questions are geared towards depression and not CFS (Chalder et al., 1993).

Second, the ceiling effect means that a maximum score at baseline cannot increase even if there is deterioration during the trial. As a consequence, for example, if a participant deteriorated during the trial on eight items and improved on three, the score should reflect a deterioration of five points. However, if they had scored the maximum at baseline, then since eight scores cannot get worse and three scores have improved, the Chalder Fatigue Scale would classify the participant, who had deteriorated by five points, as improved by three points.

Analysis of the use of the Chalder Fatigue Scale in CFS patients who were well enough to attend an outpatient clinic found high rates of maximal scoring (Morriss et al., 1998). This issue was a particular concern in Wearden et al. (2010b). A review of the mean and standard deviation data for that trial calculated that between 65 and 82 per cent in the pragmatic rehabilitation group (PR) must have recorded the maximum score at baseline (Stouten, 2010). Since the patients were more severely afflicted and unable to attend outpatient clinics, there was a high likelihood of maximum scores at baseline

Third, the Scale has been found to be unreliable in distinguishing between healthy controls and fatigue. In a trial of CBT for patients with multiple sclerosis (MS), it was found that after treatment, fatigued MS patients had less fatigue than healthy controls (Van Kessel et al., 2008).

Fourth, few items on the Chalder Fatigue Scale appear clearly related to fatigue and there is a focus on change in fatigue, rather than intensity (Wilshire et al., 2018a).

The use of the flawed Chalder Fatigue Scale to measure subjective fatigue and hence the primary outcome further casts doubt on the trial's and review's conclusions.

7. There are concerns about dropouts.

The percentage of dropouts differed substantially between the trials.

Participants who do not respond to treatment or are negatively affected by it are more likely to drop out or be lost to follow up (Lilienfeld et al., 2014). In Wearden et al. (1998), 37.3 per cent dropped out in the two exercise groups combined but only 21.7 per cent in the two control groups combined. In Powell et al. (2001), 18 per cent dropped out from the exercise groups compared to only 5.9 per cent from the control group. In Moss-Morris et al. (2005), this was 12 and 0 per cent, respectively.

Only 6 per cent dropped out in the GET group of White et al. (2011); however, according to the supplement to the secondary mediation paper, there were missing step test data for 34 per cent (GET) (Chalder et al., 2015), which may have inflated any improvement in the GET group on that test. These dropouts add further doubts about the reliability of the review's findings.

Analysis of the trials

Fulcher and White (1997)

The trial concluded that graded aerobic exercise is an effective treatment for CFS. Patients reported an improvement on the primary outcome (self-rated clinical global impression change score) represented by a score of 1 or 2.

However, if a categorisation into 1–3 (all scores representing an improvement) and 4–7 (the rest) had been used, then there would have been no significant difference between the two treatment groups.

Because of their separate effects on fatigue, Fulcher excluded patients with a current psychiatric disorder or symptomatic insomnia, apart from simple co-morbid phobias, yet 30.3 per cent (20/66) in the trial were on full dose antidepressants. Symptomatic insomnia (sleep reversal and/or unrefreshing sleep) is a common and important symptom of CFS, so excluding such patients would seem to have excluded patients with a common symptom of CFS.

Participants in the exercise group had sessions of 5–15 minutes, increasing to a maximum of 30 minutes, at least 5 days a week. Such a workload would exclude most patients with CFS.

Doubts about the diagnosis are further cast by the VO_{2max} scores. The average VO_{2max} for healthy sedentary controls of a similar age as in the trial, 37, is 30–36 (De Becker et al., 2000). The American Medical Association (AMA) (Gunnar and Occhiarella, 2000) categorises impairment as follows (VO_{2max}; mL/kg/min): no impairment, >25; mild impairment, 20–25; moderate impairment, 15–20; and severe impairment, <15. Scores for CFS patients in another study (VanNess et al., 2003) were 22.1 (mild impairment), 17.2 (moderate impairment), and 12.3 (severe impairment). In Fulcher and White (1997), the VO_{2max} score at baseline in the GET group was 31.8, in other words unimpaired according to the AMA categorisation and well above even mildly impaired CFS patients in VanNess et al. (2003).

The exclusion of patients with a common symptom of CFS, the inclusion of patients on full dose antidepressants, the high level of fitness of the participants and the relatively onerous workload completed by those in the exercise group raise very serious concerns about whether this was in fact a trial for patients with CFS.

Wearden et al. (1998)

Three problems are apparent with this trial.

First, it used too broad criteria to select patients. Not only did it use the Oxford criteria, but it set a physical functioning score of <83.3 to designate caseness and 46 per cent of participants had a current psychiatric diagnosis. Subjects randomised to graded exercise were instructed to carry out aerobic activity for 20 minutes, at least three times a week, a level of activity beyond most patients with CFS.

Second, the exercise and placebo group and its control group, the exercise control and placebo group, are difficult to compare with each other because of the big difference (30.7%) in VO_{2max} scores at baseline, 19.9 and 26.0, respectively.

Third, since more participants dropped out from and fewer complied fully with the GET groups, it would seem impossible to draw any safe conclusions. In total, 37.3 per cent dropped out of the two GET groups combined and only a small percentage (34.3%) of participants complied fully with GET. In contrast, only 21.7 per cent dropped out in the two exercise placebo groups and 78.3 per cent complied fully with exercise placebo. The difference in these two rates is particularly a concern since patients who dropped out were significantly more likely to have changed or given up their occupation as a result of their illness.

Even with these apparent biases towards the exercise groups, by 26 weeks the exercise capacity in the GET group improved by only 4.3 per cent more than in its control group. Increases of 20 per cent have been observed in patients with stable chronic heart failure after only 3 weeks of training (Meyer et al., 1997). Furthermore, the mean fatigue scores at the end of treatment in the two GET groups were 29.9 and 28.0 so that patients were still ill enough to meet the entry criteria for the trial (cut-off of 4 or more to designate caseness; 0–42).

Powell et al. (2001)

Similar problems afflict Powell et al. (2001). The entry physical functioning score was <25, a score of 25 deemed similar to normal daily functioning for the UK general population. However, the main outcome measure and predetermined criterion for clinically important improvement was a score of 25 or more or an increase of >10 on the SF-36 physical functioning subscale (range: 10–30) 1 year after randomisation. In other words, a patient could enter the trial with a score of 24, improve to 25 over the course of the trial, and this minimal improvement would be deemed clinically important.

The trial employed non-equivalent controls which favoured the interventions. A minimum number of sessions were stipulated for the treatment groups: 3 hours face-to-face and 1 hour telephone contact for the minimum intervention group, 3 hours face-to-face and 4.5 hours telephone contact for the telephone intervention group and 10 hours face-to-face and 1.5 hour telephone contact for the maximum intervention group. There was no such specification for the control group. It is possible those in the control group had zero hours of contact and it was therefore no more than a waitlist group. This was recognised in the article where it is acknowledged that one of the limitations was the lack of a placebo control group that received equivalent therapist time and attention.

The hospital anxiety and depression scale scores suggest that instead of treating the CFS, the interventions addressed the patients' depression. At baseline, patients had high depression (9.3, 9.0 and 9.0) and anxiety scores (10.6, 10.0 and 10.2) for the minimum intervention, telephone intervention and maximum intervention groups, respectively. One year after randomisation, the depression scores improved dramatically to 4.2, 4.6 and 4.2, and the anxiety scores less so to 7.1, 6.5 and 7.7. In the control group,

however, the scores remained roughly the same: 10.4 and 10.1 for depression and 11.2 and 10.1 for anxiety, at baseline and 1 year after randomisation, respectively.

Once again, there was a problem with selection criteria, 33.8 per cent (50/148) of the participants were working at the start of the trial and dropouts. The dropout rate was 16.2 per cent (6/37) for the minimum intervention, 18 per cent (7/39) for the telephone intervention and 18.4 per cent (7/38) for the maximum intervention. In contrast, the control group dropout rate was much lower at 5.9 per cent (2/34).

These flaws make any conclusion that the interventions were effective impossible.

Wallman et al. (2004)

Despite its claim to show graded exercise improves functional ability and minimises deconditioning, which can result in more symptoms, and that these improvements may be associated with the abandonment of avoidance behaviours, it would appear that the trial was not of graded exercise but of a form of pacing.

In graded exercise therapy for CFS, planned physical activity and not symptoms dictate what participants do (Bavinton et al., 2004). Participants in this trial, though, were instructed to exercise every second day, unless they had a relapse. If this occurred, or if symptoms became worse, the next exercise session was shortened or cancelled. Subsequent exercise sessions were reduced to a length the patient felt manageable.

Since the exercise treatment allowed patients to decrease as well as increase how much exercise they did, depending on their symptoms, the intervention being evaluated was in fact not GET but a form of pacing (Goudsmit et al., 2012).

Moss-Morris et al. (2005)

This trial had a biased sample: participants were from a private clinic and had contacted the university, so were self-selected. They were, then, invested in the trial and more likely to believe in the possible effectiveness of the interventions. Furthermore, patients were high functioning: only 22.4 per cent were unemployed and unable to work due to disability, indicating 77.6 per cent were employed and therefore only mildly affected. As many as 30 and 42 per cent of the sample, though, were possibly suffering from depression and anxiety, respectively.

The groups were poorly matched: there was a difference in mean ages (36.7 in the GET group, 45.5 in the SMC control), illness duration (2.67 years GET, 5.0 years SMC) and SF-36 physical functioning scores (53.10, GET; 45.65 SMC) at baseline.

Although it was concluded that graded exercise appears to be an effective treatment for CFS, patients showed no objective improvement. In fact, there was a significant physical deterioration after 'effective' GET as seen in the large drop in VO_{2max} of 15 per cent. Since there is an inverse relation between fatigue and activity (Rongen-van Dartel et al., 2014), a decrease in fatigue should have been matched by an objective increase in physical functioning.

The study acknowledges that the intervention (GET) did not have a significant effect on fitness. Any conclusions, then, about the effectiveness of the treatment cannot be sustained.

Jason et al. (2007)

The review says this trial showed little or no difference in fatigue between anaerobic exercise and treatment as usual at follow-up. It states that the better physical functioning scores for relaxation therapy were because the relaxation group (RELAX) had a higher mean score at baseline than the anaerobic activity group (ACT), 53.77 and 39.17, respectively. However, the objective 6-minute walk test results at baseline were similar: 1335 (ACT) versus 1317 (RELAX; higher scores indicating better outcome). The physical functioning scores did improve more for the RELAX group: an increase of 12.1 per cent from baseline for RELAX, compared to 1.4 per cent for ACT.

In other tests, the RELAX group also performed better. Improvement in the 6-minute walk test was 3.2 per cent (ACT) and 8.4 per cent (RELAX), and in quality of life scores 3.5 per cent (ACT) and 12.3 per cent (RELAX). The RELAX group performed better for difference in symptom outcome, improvement for muscle pain, unrefreshing sleep and PEM.

The trial would seem to show, contrary to the conclusion in the review, that participants in the relaxation group improved more than those in the exercise group.

Wearden et al. (2010b)

There were once again problems with entry requirements for this trial: Criteria included 70 (out of 100) or lower on the SF-36 physical functioning scale, but one patient was admitted to the trial with a score of 75; 20.3 per cent (60/296) of the participants suffered from anxiety and 17.9 per cent (53/296) from depression.

The trial was not properly controlled. The PR group on average received 11.63 sessions (9.63 with the therapist + 2 general practitioner (GP) consultations), the supportive listening group (SL) had 12.5 (9.5 + 3 GP consultations) and the treatment as usual (GPTAU) control group had only three GP consultations.

The entry criteria and outcomes were switched after the trial had started. Dates for the trial are given as from 21 June 2004 to 25 July 2008 in the protocol registered with the International Standard Randomised Controlled Trials Number (ISRCTN). The selection criteria were changed in February 2005, from the Fukuda to the even wider Oxford criteria. No reason was given (Wearden, 2001).

According to the FINE trial protocol (Wearden et al., 2006), primary outcomes were to be self-reported physical functioning and fatigue at 1 year. Yet in the 2010 paper, the measures were taken instead at the end of treatment, 20 weeks and 70 weeks from recruitment.

The step test was an objective secondary outcome measure in the protocol but was omitted from the 2010 paper. Not publishing results jeopardises the validity of a study (Heneghan et al., 2017). The results were published 3 years later (Wearden and Emsley, 2013), and there were no differences between pragmatic rehabilitation and GP treatment. Three Dutch studies that also did not publish their step test results on analysis many years later by proponents of the biopsychosocial model also showed no objective differences between treatment and control (Wiborg et al., 2010).

The fatigue scores were changed from bimodal (0–11) to Likert (0–33) in a Rapid Response in the BMJ (Wearden et al., 2010a) and in Wearden and Emsley (2013). Re-scored there was now a clinically modest, but statistically significant effect of PR compared with GPTAU at both outcome points. However, altering measures in this way after the trial to find a small effect suggests a form of p-hacking.

The entry criteria, outcome switching and null objective improvement in this trial mean that it is unsafe to claim any effect for the interventions.

White et al. (2011)

This trial was the largest, 640 patients in a review that included 1510. A number of similar flaws have been identified.

Patients were included with depression or anxiety: 47 per cent of participants had a co-morbid depression or anxiety disorder and only 20 per cent (640/3158) of participants screened were selected for the study.

The control group did not have the same contact hours: the CBT group received on average 16 sessions, the GET group 17 sessions (both including three sessions of SMC), yet the SMC group only had five. This imbalance creates serious biases towards finding a positive effect for the intervention, regardless of whether it is effective or not (Lilienfeld et al., 2014).

A null effect at long-term follow-up was spun as positive. Outcomes with SMC alone or adaptive pacing therapy (APT) improved from the 1-year outcome and were similar to CBT and GET at long-term follow-up, but it was claimed that the data should be interpreted in the context of additional therapies having being given after the 1-year trial final assessment (Sharpe et al., 2015). However, the long-term follow-up (Supplementary Appendix) shows the majority of participants did not have any additional CBT (76%) or GET (83%) after the trial. It also shows that patients in all four groups who did not receive additional treatment subsequent to trial completion exhibited lower fatigue and higher physical functioning scores relative to those of patients who received additional treatment (Vink, 2016).

Baseline figures were used for one objective test, an actometer, a reliable measure of activity to assess improvement objectively (Scheeres et al., 2009), but were not recorded at the end of the trial. The reason given was that it would be too great a burden (Vink, 2016) for patients, even though they had consented to use it; they had completed moderately effective treatment (White et al. (2011) and 22 per cent of those in the CBT and GET groups had recovered (White et al., 2013).

An extensive number of endpoint changes were made (Sharpe et al., 2015; Vink, 2016; White et al., 2011; Wilshire et al., 2018b). The timing of these changes – several months after trial completion – was highly problematic (Wilshire et al., 2018b). As a result, there was an overlap in entry and recovery criteria: 13.3 per cent of participants were already recovered according to one (12.8%) or two (0.5%) of the recovery criteria at trial entry (Vink, 2017a).

These changes affected both the physical functioning scores and the fatigue scores. The minimum physical functioning score required to qualify as recovered was reduced from 85 to 60 (White et al., 2011). The maximum score for trial entry was increased from 60 to 65 (0–100; higher scores indicating better functioning), although according to the PACE trial's recovery article, a score of 65 or less represents 'abnormal levels of physical function' (White et al., 2013) and severe disability according to the literature (Stulemeijer et al., 2005). Participants with a score of 60–65 (inclusive) were thus considered ill enough to participate and to have an abnormal level of physical functioning, yet were also recovered and severely disabled. Three participants (0.45%) saw their physical functioning score go down from 65 to 60, reflecting deterioration, and three others (0.45%) had unchanged physical functioning scores, but all (0.9%) were still classed as recovered, according to the physical functioning recovery criterion (Vink, 2017b).

Something similar happened to the fatigue scores. When PACE was registered with the ISRCTN on 22 May 2003, participants needed a Chalder Fatigue Questionnaire (CFQ) score of 4 or more to be classed as ill enough to take part (White, 2003). The CFQ entry criterion was changed to six or more before the trial started and then during an unblinded trial switched from bimodal to Likert, 18 or more to qualify. To be classed as recovered, a bimodal score of $\leqslant$3 out of 11, which represented a screening threshold for abnormal fatigue, was changed to a Likert score of 18 or less (0–33) (White et al., 2013). Consequently, with a Likert score of 18, one was simultaneously classed as disabled and recovered.

These endpoint changes increased recovery rates of CBT and GET fourfold. Had the PACE trial stuck to the protocol defined endpoints, then there would have been no statistically significant difference in recovery rates between the four treatment groups (Wilshire et al., 2018b).

The net improvement of the quality of life scores (EQ-5D) after GET at 52 weeks over the adaptive pacing control group was 1 per cent (13.5% (0.07/0.52)–12.5% (0.06/0.48)). A study by Olesen et al. (2016) of 20,220 adult patients found a mean quality of life score of 0.84 for the total population and 0.93 for people without a chronic condition. Yet the quality of life at 52 weeks in the GET group (0.59) (McCrone et al., 2012) was similar to the score (0.60) for people with five or more chronic health conditions and still worse than in cerebral thrombosis (0.62), rheumatoid arthritis and angina (0.65), acute myocardial infarction (AMI) (0.66) (Olesen et al., 2016), MS (0.67), lung cancer (0.69), stroke (0.71) or ischaemic heart disease (0.72) (higher scores indicating a better quality of life) (Falk Hvidberg et al., 2015).

Also, there was no statistically significant difference in the improvement in CFS symptom count between GET and SMC ($p = 0.0916$) or APT ($p = 0.23$) at 52 weeks.

These flaws in both the review and the trial render unsafe any conclusion that GET is effective.

Evidence on harms

According to the review, only two studies reported on safety or adverse reactions and in the larger (White et al., 2011) there are questions about the definition used.

The review acknowledges that limited information makes it difficult to draw firm conclusions about the safety of exercise therapy yet claims no evidence to suggest exercise therapy may worsen outcomes. This conclusion relies on the fact that White et al. (2011) reported only two serious adverse reactions (SARs) possibly related to treatment, and Wearden et al. (2010b) reported no SARs due to therapy.

However, the claims that the interventions in White et al. (2011) are safe are based on an unrealistic definition of harms. Adverse events were considered serious when they involved death, hospital admission, increased severe and persistent disability, self harm, were life-threatening or required an intervention to prevent one of these (White et al., 2011). These are not harms normally complained of by patients.

Furthermore, reports in both Moss-Morris et al. (2005) and Wearden et al. (1998) raise concerns. Moss-Morris et al. (2005) found that for 40 per cent of participants exercise led to worsening of symptoms. 12 per cent dropped out from the GET group yet 0 per cent from the control group.

In Wearden et al. (1998), only 34.3 per cent of participants complied fully with GET compared to 78.3 per cent who complied fully with the control group treatment, pacing. 37.3 per cent dropped out of the two graded exercise treatment groups combined, while only 21.7 per cent dropped out of the two control groups combined. Patients who dropped out were significantly more likely to have changed or given up their occupation as a result of their illness (Wearden et al., 1998). A treatment cannot be considered safe if patients do not actually adhere to it (Kindlon, 2017).

Caution about the safety of GET is particularly necessary given the consistent finding in patient surveys that the intervention has caused worsened health in a large percentage of patients (25% ME group, 2004; Action for ME, 2011; Bjørkum et al., 2009; Bringsli et al., 2014; De Kimpe et al., 2016; De Veer and Francke, 2008; Eyssens, 2017; Geraghty et al., 2017; Kindlon, 2011, 2017; ME Association, 2010, 2015). For example, in a survey by Action for ME (2011), 60.2 per cent of people with ME reported that GET had made their condition worse, 44.1 per cent reported that it had made it much worse or very much worse. The ME Association (2015) found that in 74 per cent, GET had made their condition worse. In 82 per cent of patients who were housebound or bedridden, their health had been negatively affected by GET and some patients were not severely affected before trying GET (25% ME group, 2004).

It has also been shown that patients have a low tolerance for physical activity or mild exercise, which can provoke symptoms (Snell et al., 2013; VanNess et al., 2010) and the experiences of harm from GET by CFS patients are well supported by the scientific literature (VanNess et al., 2018).

Furthermore, GET protocols indirectly discourage research participants from reporting harm. A key feature of GET is pushing beyond limits and participants are told to interpret symptom flares experienced during a study as a normal exercise response and reconditioning (Bavinton et al., 2004). The GET manual may instruct patients to view negative experiences as unhelpful (White et al., 2007) or they may be told that 'Activity or exercise cannot harm you' (Powell and the FINE trial research team, 2005).

The failure of most studies to report on safety or adverse reactions and the unrealistic definition used by White et al. (2011) mean that it is not possible to conclude GET is safe.

Review of the objective outcomes

Analysis shows the review's conclusion that GET is effective and safe is not justified by the evidence. It is worth attempting to see whether anything can be properly inferred from the trials. Objective outcomes reveal GET to be ineffective.

It is worth noting that the GET model is at odds with the physiological findings (Helmfrid, 2016). Physical deconditioning does not seem to be a perpetuating factor in CFS (Bazelmans et al., 2001) and CFS patients without a comorbid psychiatric disorder do not have an exercise phobia (Gallagher et al., 2005).

Powell et al. (2001) did not use objective outcomes. The seven other trials that did use objective outcomes, only Fulcher and White (1997) (6.9%) and Wearden et al. (1998) (4.3%), showed a minimal improvement in the intervention group compared to the control. However, there are reasons

to consider this minimal benefit with caution. Both trials used the broad Oxford criteria, had a high percentage of participants with a psychiatric disorder and failed to exclude very high-functioning participants. Fulcher and White (1997) had no entry score requirements so that participants were selected with a normal VO_{2max} score and no impairment according to the AMA, and in Wearden et al. (1998), participants could have a physical functioning score of up to 83.3.

A large number of patients were excluded from Fulcher and White (1997) (60.5%), which used a walking test to measure improvement. It is possible that patients concentrated on the exercise and slightly increased walking at the expense of other areas of their life. In Wearden et al. (1998), a higher percentage of participants dropped out of exercise (37.3%) compared to the control groups (21.7%).

Even then, the difference in improvement in the exercise groups was only 4–7 per cent. Increases of 20 per cent have been observed in patients with stable chronic heart failure after only 3 weeks of exercising (Meyer et al., 1997). A major criterion for defining CFS is a reduction in physical capacity of at least 50 per cent compared to pre-illness levels (De Becker et al., 2000; Fukuda et al., 1994; Holmes et al., 1988) so an improvement of 4–7 per cent would still leave patients considerably worse off than before the illness.

In Wallman et al. (2004), although the oxygen uptake values (VO_{2max}) were 9.6 per cent higher after the intervention in the exercise group, the difference in final values for the groups was not significant and the activity levels did not differ between the groups.

Moss-Morris et al. (2005) found no statistically significant group effect for the physiological variables, including the maximum heart rate achieved, the percentage of the predicted maximum heart rate and the VO_{2max}. The intervention (GET) did not have a significant effect on fitness; in fact, there was a large drop in VO_{2max} of 15 per cent.

In the 6-minute walk test in Jason et al. (2007), the ACT group improved by 3.2 per cent, yet patients in the relaxation control group improved by 5.2 per cent more, an improvement matched in the quality of life scores (3.5%, ACT; 12.3%, RELAX), symptom outcome improvement for muscle pain, unrefreshing sleep and PEM.

When the scores for the only objective outcome used in Wearden et al. (2010b), the step test, were published in 2013, there were no differences between the pragmatic rehabilitation and GP treatment as usual groups on any of the step test measures at 20 or 70 weeks.

An even larger number of patients were excluded from White et al. (2011) (79.7%). Their step test did not show any objective improvements. This is matched by the net improvement of the quality of life scores after GET over APT of only 1 per cent. The number of patients who were unable to work and who were receiving benefits increased and the number of patients receiving income protection in the GET group actually doubled. In addition, there was no statistically significant difference in the improvement in CFS symptom count between GET and SMC ($p = 0.0916$) or APT ($p = 0.23$) at 52 weeks.

There was a slight benefit shown in the 6-minute walk test after exercise for 24 weeks. Patients with stable chronic heart failure improved their 6-minute walk test results by 65 per cent after only 3 weeks of exercising (Meyer et al., 1997). Also, since the actometers were used at the beginning of the trial but not after 'effective' treatments, as with Fulcher and White (1997), it is not known whether patients concentrated on the exercise and slightly increased walking at the expense of other areas of their life. This doubt is all the greater since despite exercising five times a week with a target of 30 minutes a time for 24 weeks, the fitness levels of those in the GET group did not improve. This suggests that there was an underlying physical problem preventing this.

Even after the slight improvement measured on the walk test, patients would still be ill enough to be put on the waiting list for a lung transplant. No one in the trial achieved actual recovery, where symptoms are eliminated and patients return to pre-morbid levels of functioning (Kennedy, 2002), which is the general public's understanding of the meaning of recovery (Vink, 2017a). The PACE trial protocol defined improvement as an increase of 50 per cent. According to the 6-minute walk test results, the only objective individual results that were released, this benchmark was matched by 6.3 per cent in the GET group, but also 5 per cent in the SMC group, implying an effect of GET of only 1.3 per cent (2/160) as participants in all treatment groups also received SMC (Vink, 2017a).

The trial omitted from the review, Núñez et al. (2011), found that at 12 months the intervention did not improve health-related quality of life scores and led to worse physical function and bodily pain scores. They also concluded that the combination of CBT and GET is ineffective and not evidence-based and may in fact be harmful.

Stordeur et al., (2008) analysed the effectiveness of CBT and GET in the Belgian CFS knowledge centres. They found no objective improvements after CBT and GET (VO_{2max}) and fewer people were able to work and more people were receiving illness benefits.

Two American government agencies, the CDC (2017) and the AHRQ (Smith et al., 2016), have recently removed (CDC) and downgraded (AHRQ) their recommendations for CBT and GET, because there is insufficient evidence that GET is effective. The Dutch Health Council (2018) has removed its recommendation for GET in March 2018 for the same reason.

Conclusion

The GET trials reviewed here are inherently biased: use of exercise may attract only the mildly affected and may deter

the more disabled patients from participating. These trials go back for more than two decades, at the cost of considerable money and involving large numbers of patients. The flaws in the review and the trials, as discussed above, all created a bias in favour of the exercise intervention. Despite these flaws, they have found no significant evidence of objective improvement. The analysis of the objective outcomes in the trials provides sufficient evidence to conclude GET is an ineffective treatment for ME/CFS.

Most studies failed to report on safety or adverse reactions, and White et al. (2011), one of the two studies that did, used an unrealistic definition. ME is characterised by an abnormally delayed muscle recovery after trivial exertion (Ramsay, 1988), commonly referred to as PEM. Exercise physiology studies reveal abnormalities in patients' responses to exertion. Yet PEM is not a requirement for diagnosis according to the Oxford criteria, used by five of the studies, and only an optional criterion according to the Fukuda criteria, used by the other three studies in the review. If patients do not suffer from PEM, it is likely that they will have no problems exercising. This might be an important reason why outside of clinical trials things are different with many patients, who do suffer from PEM, reporting deterioration with GET. Given these considerations, one cannot conclude that GET is safe.

Acknowledgements

The authors would like to thank M.V.'s parents for typing out his speech memos. The authors would also like to thank the reviewers for their help in improving the article.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

References

25% ME Group (2004) Severely affected ME (myalgic encephalomyelitis) Analysis Report on Questionnaire Issued January 2004. *Troon*, 1 March. Available at: https://huisartsvink.files.wordpress.com/2018/09/severely-affected-me.pdf (accessed 5 September 2018).

Action for ME (2011) GET, GAT and exercise on prescription survey results. *Interaction* 75: 11–13.

Baraniuk J (2017) Chronic fatigue syndrome prevalence is grossly overestimated using Oxford criteria compared to Centers for Disease Control (Fukuda) criteria in a U.S. population study. *Fatigue: Biomedicine, Health & Behavior* 5(4): 215–230.

Bavinton J, Dyer N and White PD (2004) PACE manual for participants: Graded exercise therapy. Available at: https://huisartsvink.files.wordpress.com/2018/09/6-get-participant-manual.pdf (accessed 22 September 2018).

Bazelmans E, Bleijenberg G, Van Der Meer JW, et al. (2001) Is physical deconditioning a perpetuating factor in chronic fatigue syndrome? A controlled study on maximal exercise performance and relations with fatigue, impairment and physical activity. *Psychological Medicine* 31(1): 107–114.

Bjørkum T, Wang CE and Waterloo K (2009) Patients' experience with treatment of chronic fatigue syndrome. *Tidsskr Nor Laegeforen* 129(12): 1214–1216.

Bringsli GJ, Gilje A and Getz Wold BK (2014) The Norwegian ME Association National Survey (Abridged English Version). Oslo: Norwegian ME Association. Available at: https://huisartsvink.files.wordpress.com/2018/08/bringsli-2014-me-nat-norwegian-survey-abr-eng-ver-2.pdf (accessed 3 August 2018).

Centers for Disease Control and Prevention (CDC) (2017) Treatment update ME/CFS. *CDC*, 30 May. Available at: https://www.cdc.gov/me-cfs/treatment/index.html (accessed 9 August 2018).

Chalder T, Berelowitz G, Pawlikowska T, et al. (1993) Development of a fatigue scale. *Journal of Psychosomatic Research* 37(2): 147–153.

Chalder T, Goldsmith KA, White PD, et al. (2015) Rehabilitative therapies for chronic fatigue syndrome: A secondary mediation analysis of the PACE trial. *Lancet Psychiatry* 2: 141–152.

Courtney R (2016) Comments on Larun L, Brurberg KG, Odgaard-Jensen J, et al. (2017) Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews* 4: CD003200. 119–133.

David AS (1991) Post viral syndrome and psychiatry. *British Medical Bulletin* 47: 966–988.

De Becker P, McGregor N, De Meirleir K, et al. (2000) Exercise capacity in chronic fatigue syndrome. *Archives of Internal Medicine* 160: 3270–3277.

De Kimpe A, Crijnen B, Kuijper J, et al. (2016) Care for ME survey of ME patients about their experiences with health care in the Netherlands. Available at: https://huisartsvink.files.wordpress.com/2018/07/de-kimpe-2016-rapport-2.pdf (accessed 3 August 2018).

De Veer AJE and Francke AL (2008) Zorg voor ME/CVS-patiënten: Ervaringen van de achterban van patiëntenorganisatiesmet de gezondheidszorg. Utrecht: NIVEL. Available at: https://www.nivel.nl/sites/default/files/bestanden/Rapport-draagvlakmeting-CVS-ME-2008.pdf (accessed 3 August 2018).

Dutch Health Council (2018) Aan: De Voorzitter van de Tweede Kamer der Staten-Generaal Nr. 2018, Den Haag, 19 maart. Gezondheidsraad Nr 2018/07. Available at: https://www.gezondheidsraad.nl/sites/default/files/grpublication/kernadvies_me_cvs_1.pdf (accessed 3 August 2018).

Edwards J (2017) PACE team response shows a disregard for the principles of science. *Journal of Health Psychology* 22: 1155–1158.

Eyssens S (2017) Post-Exertional Malaise (PEM) and Graded Exercise Therapy (GET) in myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). Available at: https://

huisartsvink.files.wordpress.com/2018/08/eyssens-pem-get-primer-december-2017.pdf (accessed 3 August 2018).

Falk Hvidberg M, Brinth LS, Olesen AV, et al. (2015) The health-related quality of life for patients with myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *PLoS ONE* 10: e0132421.

Friedberg F, Dechene L, McKenzie MJ, et al. (2000) Symptom patterns in long-duration chronic fatigue syndrome. *Journal of Psychosomatic Research* 48(1): 59–68.

Fukuda K, Straus SE, Hickie I, et al. (1994) The chronic fatigue syndrome: A comprehensive approach to its definition and study. *Annals of Internal Medicine* 121: 953–959.

Fulcher KY and White PD (1997) Randomised controlled trial of graded exercise in patients with chronic fatigue syndrome. *BMJ* 314(7095): 1647–1652.

Gallagher AM, Coldrick AR, Hedge B, et al. (2005) Is the chronic fatigue syndrome an exercise phobia? A case control study. *Journal of Psychosomatic Research* 58(4): 367–373.

Geraghty K, Hann M and Kurtev S (2017) 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*. Epub ahead of print 29 August. DOI: 10.1177/1359105317726152.

Goudsmit EM, Nijs J, Jason LA, et al. (2012) Pacing as a strategy to improve energy management in myalgic encephalomyelitis/chronic fatigue syndrome: A consensus document. *Disability and Rehabilitation* 34(13): 1140–1147.

Green CR, Cowan P, Elk R, et al. (2014a) Advancing the research on myalgic encephalomyelitis/ chronic fatigue syndrome (Draft Executive Summary). National Institutes of Health-Pathways to Prevention Workshop. Available at: https://huisartsvink.files.wordpress.com/2018/08/green-draftreport-odp-mecfs-1.pdf (accessed 3 August 2018).

Green CR, Cowan P, Elk R, et al. (2014b) Advancing the research on myalgic encephalomyelitis/ chronic fatigue syndrome, executive summary (Final Report). National Institutes of Health-Pathways to Prevention Workshop. Available at: https://huisartsvink.files.wordpress.com///green-odp-final-report-pp-mecfs.pdf2018082 (accessed 3 August 2018).

Gunnar AB and Occhiarella L (2000) *Guide to the Evaluation of Permanent Impairment* (3rd edn). Chicago, IL: American Medical Association.

Haywood KL, Staniszewska S and Chapman S (2011) Quality and acceptability of patient-reported outcome measures used in chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): A systematic review. *Quality of Life Research* 21(1): 35–52.

Helmfrid S (2016) Studier av kognitiv beteendeterapi och gradvis ökad träning vid ME/CFS är missvisande. *Socialmedicinsk tidskrift* 93: 433–444. English version: Studies on cognitive behavioral therapy and graded exercise therapy for ME/CFS are misleading. Available at: https://huisartsvink.files.wordpress.com/2018/08/helmfrid_smt2016_englishversion.pdf (accessed 16 August 2018).

Heneghan C, Goldacre B and Mahtani KR (2017) Why clinical trial outcomes fail to translate into benefits for patients. *Trials* 18: 122.

Holmes GP, Kaplan JE, Gantz NM, et al. (1988) Chronic fatigue syndrome: A working case definition. *Annals of Internal Medicine* 108: 387–389.

Hróbjartsson A, Emanuelsson F, Thomsen AS, et al. (2014) 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* 43(4): 127–283.

IOM (2015). *Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Redefining an Illness* (Institute of Medicine; Committee on the Diagnostic Criteria for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome; Board on the Health of Select Populations). Washington, DC: National Academies Press.

Jason L, Torres-Harding S, Friedberg F, et al. (2007) Non-pharmacologic interventions for CFS: A randomized trial. *Journal of Clinical Psychology in Medical Settings* 14: 275–296.

Kennedy S (2002) Full remission: A return to normal functioning. *Journal of Psychiatry & Neuroscience* 27: 233–234.

Kindlon T (2011) Reporting of harms associated with graded exercise therapy and cognitive behavioural therapy in myalgic encephalomyelitis/chronic fatigue syndrome. *Bulletin of the IACFS ME* 19(2): 59–111.

Kindlon T (2015) Comments on Larun L, Brurberg KG, Odgaard-Jensen J, et al. (2017) Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews* 4: CD003200. 113–118.

Kindlon T (2017) Do graded activity therapies cause harm in chronic fatigue syndrome? *Journal of Health Psychology* 22(9): 1146–1154.

Larun L, Brurberg KG, Odgaard-Jensen J, et al. (2017) Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews* 4: CD003200.

Lilienfeld SO, Ritschel LA, Lynn SJ, et al. (2014) Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. *Perspectives on Psychological Science* 9: 355–387.

Luborsky L, Diguer L, Seligman DA, et al. (1999) The researcher's own therapy allegiances: A 'wild card' in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice* 6(1): 95–106.

Luborsky L, Rosenthal R, Diguer L, et al. (2002) The dodo bird verdict is alive and well – Mostly. *Clinical Psychology: Science and Practice* 9(1): 2–12.

McCrone P, Sharpe M, Chalder T, et al. (2012) Adaptive pacing, cognitive behaviour therapy, graded exercise, and specialist medical care for chronic fatigue syndrome: A cost-effectiveness analysis. *PLoS ONE* 7: e40808.

ME Association (2010) Managing my M.E.: What people with ME/CFS and their carers want from the UK's health and social services. *ME Association*, 27 May. Available at: http://www.meassociation.org.uk/wp-content/uploads/2010/09/2010-survey-report-lo-res10.pdf (accessed 2 August 2018).

ME Association (2015) ME/CFS illness management survey results: No decisions about me without me (Patient Survey). *ME Association*, 29 May. Available at: http://www.meassociation.org.uk/wp-content/uploads/2015-ME-Association-Illness-Management-Report-No-decisions-about-me-without-me-30.05.15.pdf (accessed 2 August 2018).

Meyer K, Schwaibolda M, Westbrook S, et al. (1997) Effects of exercise training and activity restriction on 6-minute walking test performance in patients with chronic heart failure. *American Heart Journal* 133(4): 447–453.

Morriss RK, Wearden AJ and Mullis R (1998) Exploring the validity of the Chalder Fatigue Scale in chronic fatigue syndrome. *Journal of Psychosomatic Research* 45(5): 411–417.

Moss-Morris R, Sharon C, Tobin R, et al. (2005) A randomized controlled graded exercise trial for chronic fatigue syndrome: Outcomes and mechanisms of change. *Journal of Health Psychology* 10(2): 245–259.

Munder T, Flückiger C, Gerger H, et al. (2012) Is the allegiance effect an epiphenomenon of true efficacy differences between treatments? A meta-analysis. *Journal of Counseling Psychology* 59(4): 631–637.

Núñez M, Fernández-Solà J, Nuñez E, et al. (2011) Health-related quality of life in patients with chronic fatigue syndrome: Group cognitive behavioural therapy and graded exercise versus usual treatment. A randomised controlled trial with 1 year of follow-up. *Clinical Rheumatology* 30: 381–389.

Olesen AV, Oddershede L and Petersen KD (2016) Health-related quality of life in Denmark on a relative scale: Mini-catalogue of mean EQ-5D-3L index scores for 17 common chronic conditions. *Nordic Journal of Health Economics* 4(2): 44–56.

Powell P and the FINE Trial Research Team (2005) FINE Trial Patient Booklet (Version 9), 29 April. Available at: https://huisartsvink.files.wordpress.com/2018/08/fine-trial-patient-pr-manual-ver9-apr05.pdf (accessed 4 August 2018).

Powell P, Bentall RP, Nye FJ, et al. (2001) Randomised controlled trial of patient education to encourage graded exercise in chronic fatigue syndrome. *BMJ* 322(7283): 387–390.

Ramsay MA (1988) *Myalgic Encephalomyelitis and Post Viral Fatigue States: The Saga of Royal Free Disease* (2nd edn). London, UK: Gower Medical Publishing.

Reeves WC, Lloyd A, Vernon SD, et al. (2003) Identification of ambiguities in the 1994 chronic fatigue research case definition and recommendations for resolution. *BMC Health Services Research* 3: 25.

Rongen-van Dartel SA, Repping-Wuts H, van Hoogmoed D, et al. (2014) Relationship between objectively assessed physical activity and fatigue in patients with rheumatoid arthritis: Inverse correlation of activity and fatigue. *Arthritis Care & Research (Hoboken)* 66(6): 852–860.

Scheeres K, Knoop H, van der Meer J, et al. (2009) Clinical assessment of the physical activity pattern of chronic fatigue syndrome patients: A validation of three methods. *Health and Quality of Life Outcomes* 7: 29.

Sharpe MC, Archard LC, Banatvala JE, et al. (1991) A report – Chronic fatigue syndrome: Guidelines for research. *Journal of the Royal Society of Medicine* 84(2): 118–121.

Sharpe MC, Goldsmith KA and Johnson AL (2015) Rehabilitative treatments for chronic fatigue syndrome: Long-term follow-up from the PACE trial. *The Lancet* 2: 1067–1074.

Smith MEB, Nelson HD, Haney E, et al. (2016) Diagnosis and treatment of myalgic encephalomyelitis/chronic fatigue syndrome. Evidence Reports/Technology Assessments No. 219. (Prepared by the Pacific Northwest Evidence-based Practice Center under Contract No. 290-2012-00014-I. AHRQ Publication No. 15-E001-EF). Rockville, MD: Agency for Healthcare Research and Quality, July. Available at: https://huisartsvink.files.wordpress.com/2018/08/ahrq-smith-et-al-2016-chronic-fatigue_research.pdf (accessed 4 August 2018).

Snell C, Stevens S, Davenport T, et al. (2013) Discriminative validity of metabolic and workload measurements for identifying people with chronic fatigue syndrome. *Physical Therapy* 93(11): 1484–1492.

Stordeur S, Thiry N and Eyssen M (2008) Chronisch Vermoeidheidssyndroom: Diagnose, behandeling en zorgorganisatie [Fatigue Syndrome: Diagnosis, treatment and organisation of care]. Technical report 88A (in Dutch). Brussels: Belgian Healthcare Knowledge Center (KCE). Available at: https://kce.fgov.be/sites/default/files/page_documents/d20081027358.pdf (accessed 4 August 2018).

Stouten B (2010) Fatigue Scale. *BMJ Rapid Response*, 12 May. Available at: https://www.bmj.com/rapid-response/2011/11/02/fatigue-scale (accessed 2 August 2018).

Stulemeijer M, De Jong LW, Fiselier TJ, et al. (2005) Cognitive behaviour therapy for adolescents with chronic fatigue syndrome: Randomised controlled trial. *BMJ* 330(7481): 14.

Van den Berg-Emons RJ, Ortye L, Buffart AA, et al. (2011) Validation of the Physical Activity Scale for individuals with physical disabilities. *Archives of Physical Medicine and Rehabilitation* 92: 923–928.

Van Kessel K, Moss-Morris R, Willoughby L, et al. (2008) A randomized controlled trial of cognitive behavior therapy for multiple sclerosis fatigue. *Psychosomatic Medicine* 70: 205–213.

VanNess JM, Snell CR, Davenport T, et al. (2018) Opposition to Graded Exercise Therapy (GET) for ME/CFS. *Workwell Foundation*, 1 May. Available at: https://huisartsvink.files.wordpress.com/2018/08/vanness-et-al-2018-mecfs-get-letter-to-health-care-providers-v4-30-2.pdf (accessed 8 August 2018).

VanNess JM, Snell CR, Strayer DR, et al. (2003) Subclassifying chronic fatigue syndrome through exercise testing. *Medicine & Science in Sports & Exercise* 35(6): 908–913.

VanNess JM, Stevens SR, Bateman L, et al. (2010) Post-exertional malaise in women with chronic fatigue syndrome. *Journal of Women's Health (Larchmt)* 19(2): 239–244.

Vink M (2016) The PACE trial invalidates the use of cognitive behavioral and graded exercise therapy in myalgic encephalomyelitis/chronic fatigue syndrome: A review. *Journal of Neurology and Neurobiology* 2(3).

Vink M (2017a) Assessment of individual PACE trial data: In myalgic encephalomyelitis/ chronic fatigue syndrome, cognitive behavioral and graded exercise therapy are ineffective, do not lead to actual recovery and negative outcomes may be higher than reported. *Journal of Neurology and Neurobiology* 3(1).

Vink M (2017b) PACE trial authors continue to ignore their own null effect. *Journal of Health Psychology* 22: 1134–1140.

Wallman KE, Morton AR, Goodman C, et al. (2004) Randomised controlled trial of graded exercise in chronic fatigue syndrome. *Medical Journal of Australia* 180(9): 444–448.

Wearden A (2001) A randomised controlled trial of nurse facilitated self-help treatment for patients in primary care with chronic fatigue syndrome. *ISRCTN Registry*, 18 May. Available at: https://huisartsvink.files.wordpress.com/2018/08/wearden-isrctn-2001-archive_org.pdf (accessed 17 August 2018).

Wearden AJ and Emsley R (2013) Mediators of the effects on fatigue of pragmatic rehabilitation for chronic fatigue syndrome. *Journal of Consulting and Clinical Psychology* 81: 831–838.

Wearden AJ, Dowrick C, Chew-Graham C, et al. (2010a) Fatigue Scale. *BMJ Rapid Response*, 27 May. Available at: http://www.bmj.com/rapid-response/2011/11/02/fatigue-scale-0 (accessed 17 August 2018).

Wearden AJ, Dowrick C, Chew-Graham C, et al. (2010b) Fatigue Intervention by Nurses Evaluation (FINE) Trial Writing Group and the FINE Trial Group. Nurse led, home based self help treatment for patients in primary care with chronic fatigue syndrome: Randomised controlled trial. *BMJ* 340: c1777.

Wearden AJ, Morriss RK, Mullis R, et al. (1998) Randomised, double-blind, placebo-controlled treatment trial of fluoxetine and graded exercise for chronic fatigue syndrome. *British Journal of Psychiatry* 172: 485–490.

Wearden AJ, Riste L, Dowrick C, et al. (2006) Fatigue Intervention by Nurses Evaluation–the FINE Trial. A randomised controlled trial of nurse led self-help treatment for patients in primary care with chronic fatigue syndrome: Study protocol [ISRCTN74156610]. *BMC Medicine* 4: 9.

Wechsler ME, Kelley JM, Boyd IO, et al. (2011) Active albuterol or placebo, sham acupuncture, or no intervention in asthma. *New England Journal of Medicine* 365(2): 119–126.

White PD (2003) RCT of CBT, graded exercise, and pacing versus usual medical care for the chronic fatigue syndrome [ISRCTN5 4285094], 22 May. Available at: https://huisartsvink.files.wordpress.com/2018/09/pace-isrctn54285094.pdf (accessed 4 September 2018).

White PD, Goldsmith K, Johnson AL, et al. (2013) Recovery from chronic fatigue syndrome after treatments given in the PACE trial. *Psychological Medicine* 43: 2227–2235.

White PD, Goldsmith KA, Johnson AL, et al. (2011) Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): A randomised trial. *The Lancet* 377: 823–836.

White PD, Sharpe MC, Chalder T, et al. (2007) Protocol for the PACE trial: A randomised controlled trial of adaptive pacing, cognitive behaviour therapy, and graded exercise as supplements to standardised specialist medical care versus standardised specialist medical care alone for patients with the chronic fatigue syndrome/myalgic encephalomyelitis or encephalopathy. *BMC Neurology* 7(1): 6.

Whiting P, Bagnall AM, Sowden AJ, et al. (2001) Interventions for the treatment and management of chronic fatigue syndrome: A systematic review. *JAMA* 286: 1360–1368.

Wiborg JF, Knoop H, Stulemeijer M, et al. (2010) How does cognitive behaviour therapy reduce fatigue in patients with chronic fatigue syndrome? The role of physical activity. *Psychological Medicine* 40: 1281–1287.

Wilshire C (2017) The problem of bias in behavioural intervention studies: Lessons from the PACE trial. *Journal of Health Psychology* 22(9): 1128–1133.

Wilshire CE and McPhee G; The Science for MECFQ (2018a) Submission to the public review on common data elements for ME/CFS: Problems with the Chalder Fatigue Questionnaire. *Science for ME*, 22 January. Available at: https://huisartsvink.files.wordpress.com/2018/08/wilshire-mcphee-cfq-cde-critique-for-s4me-final.pdf (accessed 17 August 2018).

Wilshire CE, Kindlon T, Courtney R, et al. (2018b) Rethinking the treatment of chronic fatigue syndrome-a reanalysis and evaluation of findings from a recent major trial of graded exercise and CBT. *BMC Psychology* 6(1): 6.

World Health Organization (WHO) (2010) *International Statistical Classification of Diseases and Health Related Problems (the) ICD-10*, vol. 1: Tabular List (2nd edn). Geneva: WHO (10 revision Volume 3 Alphabetical index Fifth edition).