

Graded exercise therapy does not restore the ability to work in ME/CFS – Rethinking of a Cochrane review

Mark Vink^{a,*} and Friso Vink-Niese^b

^a*Family and Insurance Physician, Amsterdam, The Netherlands*

^b*Independent Researcher, Germany*

Received 12 November 2019

Accepted 14 January 2020

Abstract.

BACKGROUND: Cochrane recently amended its exercise review for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) in response to an official complaint.

OBJECTIVE: To determine if the amended review has addressed the concerns raised about the previous review and if exercise is an effective treatment that restores the ability to work in ME/CFS.

METHOD: The authors reviewed the amended Cochrane exercise review and the eight trials in it by paying particular interest to the objective outcomes. We also summarised the recently published review of work rehabilitation and medical retirement for ME/CFS.

RESULTS: The Cochrane review concluded that graded exercise therapy (GET) improves fatigue at the end of treatment compared to no-treatment. However, the review did not consider the unreliability of subjective outcomes in non-blinded trials, the objective outcomes which showed that GET is not effective, or the serious flaws of the studies included in the review. These flaws included badly matched control groups, relying on an unreliable fatigue instrument as primary outcome, outcome switching, p-hacking, ignoring evidence of harms, etc. The review did also not take into account that GET does not restore the ability to work.

CONCLUSION: GET not only fails to objectively improve function significantly or to restore the ability to work, but it is also detrimental to the health of $\geq 50\%$ of patients, according to a multitude of patient surveys. Consequently, it should not be recommended.

Keywords: Bias, occupational health, patient safety, return to work, work rehabilitation

1. Introduction

The Cochrane exercise review by Larun et al. for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) was published in 2015 [1]. This

triggered responses from a number of people, most notably Kindlon and Courtney [2, 3]. In their replies, the reviewers did not address the concerns that were raised. Consequently, Courtney lodged a formal complaint [4]. He discussed a number of problems with the review which led him to determine that the authors came to an incorrect conclusion. An independent review of the Cochrane exercise review was published [5] in October 2018, in which the authors confirmed

*Address for correspondence: Mark Vink, Family and Insurance Physician, Amsterdam, The Netherlands. E-mail: markvink.md@outlook.com.
ORCID: <http://orcid.org/0000-0003-1079-5647>

the concerns raised by Kindlon, Courtney and others, namely, that the studies in the Cochrane review exhibited serious problems that extended to the Cochrane review itself.

According to that independent review, the Cochrane review not only ignored those problems but also ignored the objective outcomes which showed that exercise therapy does not lead to significant objective improvements in patients with ME/CFS. In October 2019, in response to the complaint by Courtney, Larun et al. published an amended Cochrane exercise review [6]. In this article, we will summarise and review this, to determine if the reviewers adequately addressed the concerns of Kindlon, Courtney and the aforementioned independent review. This determination is vital as the Cochrane review is an important document for patients and doctors worldwide.

We also intend to summarise the recently published review of work rehabilitation and medical retirement for ME/CFS [7]. We will focus on the most important things in both papers - with regards to occupational therapists, occupational health doctors and others involved in rehabilitation of ME/CFS patients - to answer the following questions:

- Should ME/CFS patients be treated with GET?
- Is GET safe to use in ME/CFS?
- Does GET lead to clinically significant objective improvements?
- Does GET restore the ability to work in ME/CFS?

2. The amended Cochrane exercise review

According to Larun et al., “Chronic fatigue syndrome (CFS) or myalgic encephalomyelitis (ME) is a serious disorder characterised by persistent post-exertional fatigue and substantial symptoms related to cognitive, immune and autonomous dysfunction.” “The review included eight randomised controlled trials (RCTs) with data from 1518 participants (adults) with a primary diagnosis of CFS, from all diagnostic criteria, who were able to participate in exercise therapy.” The reviewers concluded that: “Most studies had a low risk of selection bias. Exercise therapy probably has a positive effect on fatigue in adults with CFS and they may have moderately better physical functioning [at the end of treatment] compared to usual care or passive therapies. The evidence regarding adverse effects is uncertain. All studies

were conducted with outpatients diagnosed with 1994 criteria of the Centers for Disease Control and Prevention or the Oxford criteria, or both. Patients diagnosed using other criteria may experience different effects” [6].

It is important to note that *persistent post-exertional fatigue* is *not* the main characteristic of the disease. ME/CFS is most commonly characterised by post-exertional malaise (PEM). The prestigious American Institute of Medicine (IOM, now the National Academy of Medicine), defined PEM as an exacerbation of some or all of an individual’s ME/CFS symptoms that occurs after physical or cognitive exertion and leads to a reduction in functional ability. The IOM concluded that ME/CFS is a systemic exertion intolerance disease [8]. Modern diagnostic criteria of ME/CFS, like the Canadian consensus criteria from 2003 [9], or the international consensus criteria from 2011 [10], require post-exertional malaise to be present in order to be diagnosed with the illness.

2.1. Selection criteria

Despite concerns identified by Kindlon and Courtney [2, 3], the Cochrane review relied on studies known for using poor selection criteria. Three studies in the review used the 1994 CDC criteria, often referred to as the Fukuda criteria (Jason et al., 2007; Moss-Morris et al. 2005; Wallman et al., 2004 [11–13], and the five other studies used the Oxford criteria (Fulcher and White, 1997; Powell et al., 2001; Wearden et al., 1998; Wearden et al., 2010; White et al., 2011 [14–18]). The only criterion according to the Oxford criteria, is six months or more of chronic disabling fatigue [19]. The Fukuda criteria also require four or more of eight specified other symptoms [20]. PEM, the main characteristics of the disease, is not a requirement according to the Oxford criteria and only an optional requirement according to the Fukuda criteria. In a study by Friedberg et al. [21], 15% of people labelled by the Fukuda criteria as having ME/CFS, were in fact healthy people. Baraniuk [22] found that the Oxford criteria inappropriately select healthy subjects with mild or chronic idiopathic fatigue and mislabel them as ME/CFS. A report commissioned by the American National Institute of Health (NIH), concluded in 2014 that the Oxford criteria are flawed and lead to the inclusion of people with other conditions, confounding the ability to interpret the science [23]. That report stated that: “Continuing to use the Oxford definition may impair progress and cause harm” [23,

24]. The Agency for Healthcare Research and Quality (AHRQ) stated in 2016 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” [25]. Both the NIH and AHRQ recommend that the Oxford definition should be retired.

The use of the Oxford and the Fukuda criteria by the eight studies of the review, means that it is likely that these studies included patients who did not have ME/CFS. Because of these challenges with the selection criteria, the conclusions of the Cochrane review are questionable.

2.2. Problems with the controls

According to the Cochrane review [6], there was a problem with matching in Jason et al. (2007) [11], because the relaxation group (RELAX) had a higher mean physical functioning score at baseline than the anaerobic activity group (ACT), 53.77 versus 39.17. However, the objective 6-minute walk test results at baseline were very similar: 1335 (ACT) versus 1317 (RELAX; higher scores indicating better outcome).

But in five of the other eight studies, the review ignored the fact that the control groups were not evenly matched. For instance, in Wearden et al. (1998) [16], there was a difference in age between the exercise and placebo drug group (40.4) and the exercise placebo and placebo drug group (37.6); in illness duration (34.5 and 22.0 months) and fitness according to the oxygen uptake during cardiopulmonary exercise testing (CPET; 19.9 versus 26.0 for the treatment and control group respectively).

In Fulcher and White (1997) [14], there was a difference in physical fatigue on the visual analogue scale with scores of 161 (GET) and 177 (flexibility) and oxygen consumption scores during CPET (31.8 versus 28.2 flexibility); the Chalder fatigue scores were 28.9 and 30.5 at baseline and the SF 36 general health scores were 41 and 33 for exercise and flexibility respectively. This indicates that the participants in the flexibility control group had worse fitness and physical health than participants in the treatment group. Also, the oxygen consumption scores in the exercise group indicate that participants in this group had normal fitness at baseline.

In Moss-Morris et al. (2005) [12], there were big differences in mean age (36.7 exercise versus 45.48 control); illness duration (2.67 years exercise versus 5.0 years control) and physical functioning scores (53.1 exercise versus 45.65 control).

In Wearden et al. (2010) [17], there was a significant difference in comorbidities: 44.2% (treatment) and 33.7% (no-treatment) had no comorbidities and 33.0% (treatment) and 43.0% (no-treatment), had two or more comorbidities. In White et al. (2011) [18], there was a difference in quality of life between the adaptive pacing (APT) control group (0.48) and GET (0.52). This suggests that the APT group was more disabled.

The control groups in a number of trials were in reality no-treatment control groups. No-treatment was labelled standard medical care in Moss-Morris et al. (2005) [12], general practitioner treatment as usual in Wearden et al. (2010) [17], specialist or standardised medical care (SMC) in Powell et al. (2001) [15] and White et al. (2011) [18] and exercise placebo in Wearden et al. (1998) [16]. These studies had not organised regular comparison treatment where patients would receive the same number of sessions, the same care and attention etcetera, from practitioners who were as positive about their control treatment as the practitioners would be in the treatment group. Assignment to ‘no-treatment’ may strengthen participants’ beliefs that they will not improve, thereby reducing the chance of spontaneous improvement [26]. White et al. (2011) [18] documented therapeutic alliance and adherence to manual. Figures were provided for both questions for APT, CBT and GET yet no figures were provided for SMC, confirming that SMC was a no-treatment control group.

In conclusion, in five of the eight trials, there was a significant difference in level of disability between participants in the treatment groups and those in the control groups. Also, instead of receiving a control treatment, in most trials they received no-treatment. In view of these problems, one cannot safely conclude that GET is an effective treatment.

2.3. GET definition challenges

There were additional concerns about the definitions of GET and pacing in Wallman et al. (2004) [13]. The researchers stated that they studied GET and pacing, which they described in the following manner. “Subjects choose walking, cycling or swimming. Subjects 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 that the subject felt was manageable. This form of exercise, which allows for flexibility in exercise routines, is known as pac-

Table 1
Potential forms of bias in the trials

Study	Crit	% of screened patients selected	Blinded study	Researchers alliance to the treatment	Objective primary/secondary outcomes used	Protocol published/published before the start of the trial	Changes to protocol after the start of the trial	All outcomes reported	ItT analysis performed	Participant newsletter or manual used to promote treatment as effective
Fulcher and White (1997) [14]	Ox	39.5% (66/167)	No	Yes	No/yes	No	—	?	Yes	No
Jason et al. (2007) [11]	FuC	100% (114/114)	No	No	No/yes	No	—	?	No	No
Moss-Morris et al. (2005) [12]	FuC	96.1% (49/51)	No	Yes	No/yes	No	—	?	Yes	No
Powell et al. (2001) [15]	Ox	47.4% (148/312)	No	Yes	No	No	—	?	Yes	No
Wallman et al. (2004) [13]	FuC	82.9% (68/82)	No	Yes	No	No	—	?	No	No
Wearden et al. (1998) [16]	Ox	59.9% (136/227)	No	Yes	No	No	—	?	Yes	No
Wearden et al. (2010) [17, 28]	Ox	65.9% (296/449)	No	Yes	No/yes	Yes/no	Yes	No	Yes	Yes
White et al. (2011) [18, 29]	Ox	20.3% (641/3158)	No	Yes	No/yes	Yes/no	Yes	No	Yes	Yes

Crit: criteria; FuC: Fukuda criteria; ItT: intention to treat; Ox: Oxford criteria.

ing.” In graded exercise therapy for ME/CFS, planned physical activity and not symptoms dictate what participants do [27]. Wallman et al. was therefore a study of pacing and not one of graded exercise and pacing.

2.4. Bias

Larun et al. [6] report that selection bias in the studies they included is low. Selection bias is the bias introduced by how participants are selected and who researchers include or exclude from their studies [26]. It is also introduced when a large percentage of patients who are eligible for a study are not included in it. In this manner, a study could select participants who the researchers think are the most likely to benefit from their intervention. This was a particular problem in the PACE trial by White et al. (2011) [18] and Fulcher and White (1997) [14], which both used the very wide and not well defined Oxford criteria; only 20.3% (640/3158) and 39.5% (66/167) of the screened participants, respectively, were selected as can be seen in Table 1. This table also shows a number of other forms of bias in the studies.

Three studies (Fulcher and White, 1997; Jason et al., 2007; Moss-Morris et al., 2005) [11, 12, 14] did not have entry score requirements. In Wear-

den et al. (1998) [16] entry requirements were such that relatively high-functioning participants could be included as they used a physical functioning entry score of up to 83.3, whereas a score of 84 is the mean score for the UK working age population according to White et al. (2011) [18] (0–100, high score meaning better physical functioning). In Powell et al. (2001) [15], a physical functioning entry score of up to 24 (included, which corresponds to a score of 70 on the scale of 0–100 [28]) out of 30 was used. This is barely below the score of 25 (equivalent to 75) or more - an improvement on just one of the 20 questions/items - which was used to deem the study’s treatment successful. Only three studies (Wearden et al., 1998; Wearden et al., 2010; White et al., 2011) [16–18] had entry requirements for fatigue even though fatigue was a primary outcome of most studies included in the Cochrane review.

Further questions about selection bias are raised in Moss-Morris et al. (2005) [12], where 77.6% of patients were well enough to be in work. But also in both Fulcher and White (1997) [14] and Moss-Morris et al. (2005) [12], where participants already had normal objective mean physical functioning/fitness scores at trial entry as was found by the reanalysis of the Cochrane exercise review [5]. Moreover,

Moss-Morris et al. (2005) [12] had a biased sample: participants were from a private clinic and had contacted the university, so were self-selected just like the participants in Jason et al. (2007) [11]. These participants were thus invested in the trial and more likely to believe in the possible effectiveness of the intervention. But at the same time, contrary to most ME/CFS patients, they were also less likely to have problems with exercise, because if they had, it is unlikely that they would have volunteered for an exercise study.

2.4.1. Selective reporting (reporting bias)

2.4.1.1. *p-hacking and endpoint changes.* According to the Cochrane review [6], “Wearden 2010 and White 2011 referenced published protocols. We checked these against the published results, and found that reporting was adequate and that the risk of bias was low.”

Wearden et al. (2010) is more commonly known as the FINE trial (FINE: **F**atigue **I**ntervention by **N**urses **E**valuation). The entry criteria of this trial were changed from the Fukuda to the even wider Oxford criteria, eight months into a four-year long trial. No reason was given [31]. According to the FINE trial protocol [28], primary outcomes were to be self-reported physical functioning and fatigue at one year. Yet in the 2010 paper, the outcomes at 20 weeks (end of treatment) were added and were suddenly the most important ones. The reason for this might be that the fatigue outcome at 70 weeks showed a null effect.

Further, the FINE trial’s fatigue scores were changed from bimodal (0–11) to Likert (0–33) after the 2010 paper was published, in a Rapid Response in the BMJ [34] and in their economic analysis [33]. This change was made despite the fact that two of the authors (including the Principal Investigator) concluded in a paper, devoted to analysing the use of the Chalder Fatigue scale in ME/CFS [35], that near-maximal scoring on six physical fatigue scale items from the total of 14 items (five if the 11 item scale is used) supports using the two-point bimodal, rather than the four-point Likert scoring. Once re-scored, there was now a clinically modest, but statistically significant, effect of pragmatic rehabilitation compared with no-treatment at both outcome points. However, altering measures in this way after the trial to find a small effect suggests a form of *p-hacking*. This is a type of bias which “occurs when researchers collect or select data or statistical analyses until non-significant results become significant” [36].

White et al. (2011) [18] is better known as the PACE trial (PACE: **P**acing, **A**ctivity, and **C**ognitive behaviour therapy: a randomised **E**valuation). The Cochrane review [6] notes that some readers claim that the study should be viewed as a *post hoc* study because “The protocol and the statistical analysis plan were not formally published prior to recruitment of participants.” They also note that “The study authors oppose this, and have published a minute from a Trial Steering Committee (TSC) meeting stating that any changes made to the analysis since the original protocol was agreed by TSC and signed on before the analysis commenced.” Because of this, the Cochrane review concluded that there was a low risk of selective reporting in this trial. However, as Evans noted in his article on changing endpoints after the start of a clinical trial, “A fundamental principle in the design of randomized trials involves setting out in advance the endpoints that will be assessed in the trial, as failure to prespecify endpoints can introduce bias into a trial and creates opportunities for manipulation” [37]. Moreover, the PACE trial involved the recruitment of “641 participants” “between March 18, 2005, and Nov 28, 2008” [18] and according to a Freedom of information request [38], the approval by the TSC, as mentioned by the Cochrane reviewers, was given on 4 September 2009. This means that the researchers had ample time to have a look at the results of most or all of the participants before this date. Finally, it was a non-blinded trial and researchers in such trials usually have a good indication of how participants are responding to the treatments under investigation, even if they do not look at the data.

2.4.1.2. *Not publishing and/or ignoring objective outcomes.* The FINE trial’s only objective (secondary) outcome measure (step test) was omitted from the 2010 paper, even though not publishing results jeopardizes the validity of a study [32]. The step test results were published three years later [33] and they showed that there were no objective differences between exercise (pragmatic rehabilitation or PR) and no-treatment.

In the PACE trial, baseline figures were captured for one objective test, the actometer, a reliable measure of activity to assess improvement objectively [39], but were not recorded at the end of the trial. According to the PACE trial researchers, that would have been too great a burden [40] for patients, even though they had consented to use the actometer before and after treatment. On top of this, participants had completed moderately effective treatment according

to the trial itself [18]; around 60% of those in the CBT and GET groups had substantially improved and 22% had actually recovered according to the researchers [41]. Consequently, it would have been more of a burden at the beginning than at the end of the PACE trial.

Also, the PACE trial's step test results were *not published* in the original 2011 paper. The researchers *waited till 2015* to publish a figure labelled fitness, depicting the outcome and they wrote the following about it: "There were no effects on...physical fitness" [42]. In other words, the step test showed that CBT and GET did not lead to objective improvement. The mean fitness scores, however, have still not been published and the investigators *refuse* to release the individual data of the step test.

2.4.1.3. Overlap in entry and recovery criteria. An important problem of the PACE trial is that the investigators made an extensive number of endpoint changes as mentioned before [18, 40, 43, 44]. As a result there was suddenly an overlap in entry and recovery criteria. Consequently, 13.3% of participants *were already recovered*, according to one (12.8%) or two (0.5%) of the recovery criteria, *at trial entry* - before receiving any treatment [45].

These changes affected both the physical function scores (PF) and the fatigue scores. The minimum PF required to qualify as recovered was reduced from 85 to 60 [18], even though a score of 65 or less represents "abnormal levels of physical functioning" according to the PACE trial's own recovery article [41]. The maximum score for trial entry was increased from 60 to 65 (0–100; higher scores indicating better functioning). Participants with a score of 60 to 65 (inclusive) were thus considered ill enough to participate, to have an abnormal level of physical functioning, *and be severely disabled* according to the literature [46]. *Yet at the same time*, with the same physical function scores, they were *also recovered* according to the PACE trial's recovery article [41]. One cannot be severely disabled and recovered at the same time. Moreover, 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 [47]. Clearly, this classification of recovery - even when patients had no change or negative change from baseline, or when they were still severely disabled - could lead to misleading conclusions. For

example, newspapers, doctors and medical guidelines could conclude that patients can talk and exercise their way out of this illness, when in reality, the PACE trial's own data shows that this is not the case and an incorrect conclusion.

A similar misclassification happened to the fatigue scores. When PACE was registered with the ISRCTN on 22 May 2003, participants needed a Chalder Fatigue Scale (ChFS) score of four or more to be classed as ill enough to take part [48]. The ChFS entry criterion was changed to six or more before the trial started and then, during a non-blinded trial, switched from bimodal to Likert, 18 or more to qualify. To be classed as recovered, a bimodal score of ≤ 3 out of 11, which represented a screening threshold for abnormal fatigue, and which equates to a Likert score of 9 or less, was changed to a Likert score of 18 or less (0–33) [41]. Consequently, with a Likert score of 18, one was *simultaneously classed as disabled and recovered*. Yet in a properly conducted trial this should not happen. Nor should participants be selected for a trial who already fulfill one or more of the recovery criteria at trial entry - before receiving any treatment.

These endpoint changes increased recovery rates of CBT and GET 4-fold. 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 [44].

Consequently, the risk of selective reporting in Wearden et al. (2010) [17] and White et al. (2011) [18] was high, not only because some of the outcomes were not reported or only reported many years later, but also because not all outcomes were reported in accordance with their pre-defined protocols [28, 29].

2.4.2. Allegiance bias

The allegiance effect points to significant researcher and clinician bias. Psychotherapeutic treatments and treatments based on presumed psychiatric aetiology have been reported to be especially susceptible to the allegiance effect [49]. Part of the issue of allegiance bias in psychological therapies in particular is that measures are often subjective, and the clinician may unconsciously prod the subject to respond to their favoured therapy or not respond to a therapy they consider ineffective. For example, a therapist's confidence in their therapy of choice is almost certainly perceptible to the patient, even when this is not overtly advertised [50].

The review itself and 7 of the 8 studies (Jason et al., 2007, the exception [11]) were conducted by

researchers with an allegiance to a particular model of ME/CFS and to two interventions, CBT and GET, who wanted to prove their own theories. Generally, in studies examining more than one treatment approach, the treatment favoured by the researchers tends to outperform other treatments [51–53]. 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. Often, when a treatment is used as a comparison condition, it may be implemented in a weaker form than when it is used clinically. Usually investigators may not believe in the effectiveness of the control condition [54]. Consequently, treatments might not be presented to participants as equally likely to lead to improvement. This is especially important when the primary outcomes are self-report measures which can be strongly influenced by patients' expectations. Finally, a researcher's enthusiasm for a particular treatment can also lead them to overinterpret their findings or overlook limitations [54].

Before they conducted their research, White (principal investigator of the PACE trial by White et al. (2011) [18] and co-author in Fulcher and White (1997) [14]), Moss-Morris, Wearden, Powell, Chalder and Sharpe (co-authors of the PACE trial) were all known to have favoured the approach to the illness being tested. These investigators are *strong proponents* of the 'unhelpful cognitions' theory of ME/CFS, which they and other colleagues, had originated and/or actively promoted. If their trials had failed to show significant improvement and recovery through GET, this would have undermined the very theories of reversibility to which the investigators have dedicated their careers. Consequently, the risk of latent bias was palpable from the outset [55]. It is notable that Jason et al. (2007) [11], the only 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."

2.5. Primary outcomes of the review

The primary outcomes of the review are fatigue and adverse effects. Larun et al. concluded that after 12 to 26 weeks, exercise therapy probably reduces fatigue by 3.4 points on the Chalder fatigue scale (0 to 33) compared to no-treatment (usual care and waiting-list) or relaxation and flexibility. They also concluded that the effect on fatigue after 52 to 70 weeks, is uncertain [6].

Larun et al. also conducted a sensitivity analysis to assess the impact, effect or influence of key assumptions such as different definitions of outcomes, protocol deviations and missing data, on the overall conclusions of a study [56]. Their sensitivity analysis showed that there was considerable heterogeneity at the end of treatment which was caused by the deviating results in Powell et al. (2001). Exclusion of that study from their analysis removed the heterogeneity but it also reduced the treatment effect or standardised mean difference (SMD) of exercise therapy according to Larun et al., from 0.66 to 0.44, which equates to the aforementioned 3.4 points and 2.2 points on the Chalder fatigue scale, respectively [6].

Chalder Fatigue scores from the two studies (Powell et al. (2001) [15] and Wearden et al. (2010) [17]) that used the 11-point bimodal scale were re-scored by the reviewers on the 33-point Likert scale. However, as discussed earlier, when this was done by Wearden et al. (2010) [17], a non-statistically significant difference was transformed into a clinically modest, but statistically significant effect [33, 34]. It is a matter of concern, that changing one official way of scoring an instrument to the other official way can lead to a different outcome. Consequently, the Chalder fatigue scale should not be re-scored. It also raises questions about its suitability as a primary outcome in the Cochrane review.

To help quantify the minimal important difference (MID) for fatigue, the Cochrane reviewers used a study by Goligher et al. from 2008 [57] among people with systemic lupus erythematosus (SLE). Goligher et al. reported a threshold of around 2.3 points for a minimal clinically important difference (MCID) change on the 33-point Chalder Fatigue Scale, an effect size that corresponds to an SMD of about 0.36. But other studies reported a higher MCID than that chosen by Cochrane. For example, a study by Pouchot - one of the authors of Goligher et al. - which was also from 2008, found that the MCID for fatigue in rheumatoid arthritis (RA) was 3.3. (Most persons with RA complain of fatigue). Pouchot was also part of the Ad Hoc Committee on SLE Response Criteria for Fatigue [59]. This committee concluded that the MCID should be a decrease of fatigue of 15%. Pouchot and Liang, two of the authors of Goligher et al., Pouchot et al. and the Ad Hoc Committee, were also part of the most recent study by Pettersson et al. [60] that was set up to determine the MCID for seven measures of fatigue in SLE. Pettersson et al. concluded in 2015 that the standardized MCID was 0.54, rep-

resenting an improvement in fatigue of 4.4 on the 33-point Chalder Fatigue Scale. Finally, a study by Ridsdale et al. [61], into the efficacy of exercise therapy for people who presented with chronic fatigue in primary care, concluded that the mean difference between exercise and control group on the Chalder Fatigue Scale should be 4 points.

The Cochrane review analysed the outcome of the Chalder fatigue questionnaire, which was used by seven studies (Jason et al. (2007) [11] the exception as it used the fatigue severity scale or FSS). However, contrary to the other studies, Fulcher and White (1997) [14] actually used two different fatigue scales. Unfortunately, the outcome on the VAS scale, which showed less improvement than the Chalder fatigue scale, was not analysed by the Cochrane review. After 12 weeks of treatment (end of treatment), the Chalder fatigue score had improved by 18.7% (5.4/28.9) compared to the control group, yet the total fatigue score on the visual analogue scale (VAS) had only improved by 6.4% (20/312). The VAS also highlighted the fact that the two groups in Fulcher and White (1997) were not evenly matched for physical fatigue with scores of 161 (GET) and 177 (flexibility control group). The scores for mental fatigue 151 (GET) and 148 (flexibility) were evenly matched. The 10% difference in physical fatigue scores is matched by a 11.3% (3.6/31.8) difference in oxygen consumption between the two groups (31.8, GET and 28.2, flexibility).

There are also other problems with choosing fatigue, measured by the Chalder fatigue scale or questionnaire, at the end of treatment as the primary outcome. First of all, the protocol of Wearden et al. (2010) [17], better known as the FINE trial, noted that "Assessment at week 70 is required because short-term assessments of outcome in a chronic health condition such as CFS/ME can be misleading" [28]. Secondly, there are many issues with the Chalder fatigue scale (ChFS). The reanal-

yses of the Cochrane GET and CBT reviews listed a total of 10 different problems with it [5, 62]. The following five are some of the most important ones:

- The ChFS does not provide a comprehensive reflection of fatigue related severity, symptomology or functional disability in ME/CFS;
- The ChFS suffers from the ceiling effect so that a maximum score at baseline cannot increase even if there is deterioration during the trial;
- Few items on the ChFS clearly relate to fatigue;
- The ChFS is unable to distinguish between ME/CFS and primary depression;
- Its scoring has limited evidence of test-retest reliability.

An important problem of the Chalder fatigue scale is the above mentioned ceiling effect. This means that a maximum score at baseline cannot increase, even if there is deterioration during the trial. The released individual participant data [63, 64] of both Wearden et al. (2010) [17] and White et al. (2011) [18], the only two trials of the eight that released these data, show that a large percentage of participants in both trials had the maximum Chalder fatigue scores at baseline as can be seen in Table 2. If these patients with maximum scores at baseline, for example, had improved on three of the 11 items, and deteriorated on eight, then they should have been classified as deteriorated by five points (eight minus three). But as the eight scores could not get worse, these patients would actually be classified as having improved by three points.

Table 2 also includes the figures and percentages for some other scores, because for example with a Likert score of 30, patients had the maximum score on at least eight of the 11 questions. This means that even if they did deteriorate on those eight questions, their scores could not reflect that.

Table 2
Percentage of participants with (near) maximum Chalder fatigue scores at baseline

Study	Bimodal 11	Bimodal 10	Bimodal 9	Total bimodal 9 – 11	Likert 33	Likert 32	Likert 31	Likert 30	Total Likert 30 – 33
Wearden et al. (2010) [17, 63] (n = 196)	147 (75%)	19 (9.7%)	14 (7.1%)	180 (91.8%)	57 (29.1%)	17 (8.7%)	16 (8.2%)	18 (9.2%)	108 (55.1%)
White et al. (2011) [18, 64] (n = 640)	417 (65.2%)	106 (16.6%)	57 (8.9%)	580 (90.6%)	93 (14.5%)	53 (8.3%)	58 (9.1%)	78 (12.2%)	282 (44.1%)

Number of participants (%); bimodal scoring: 0, 0, 1, 1 (0–11); Likert scoring: 0, 1, 2, 3 (0–33); Wearden et al. (2010) [17] released individual data for the treatment group and the supportive listening group which they used for another study [65]. Data for the no-treatment control group was not released.

Table 3
Fatigue scores at the end of treatment compared to entry requirement

Study	Baseline	End of treatment	Fatigue score entry requirement
Wearden et al. (1998) [16]	33.7	28.7	4 or more
Wearden et al. (2010) [17]	10.49	8.39	4 or more
White et al. (2011) [18]	28.2	21.7	18 or more

Another problem of the Cochrane conclusion based on the fatigue scores is highlighted in Table 3. It shows that at the end of treatment, in the three studies that specified a minimum fatigue entry requirement, patients were still ill enough to re-enter that study and receive the same treatment again. If a treatment would be truly effective, this would not be the case.

A minimum fatigue entry score was not specified in the other five studies - Fulcher and White (1997), Jason et al. (2007), Moss-Morris et al. (2005), Powell et al. (2001) and Wallman et al. (2004) [11–15]. Why they did not specify this is unclear, especially as fatigue was a primary outcome in most of them.

A 2003 study by Tench et al. [66] on the efficacy of exercise therapy for SLE further highlighted the problems of the Chalder fatigue scale and the fact that it can lead to false-positive outcomes. This study, which included the principal investigator of White et al. (2011) [18], compared the efficacy of 12 weeks of graded exercise therapy with two other interventions (relaxation and no intervention). It used a number of subjective outcomes but also objective ones including VO_2 peak. This study is of importance because it used three different measures for fatigue: the fatigue severity score (FSS), the Chalder fatigue scale and the visual analogue scale (VAS). The mean age in the trial was 39, which is similar to the trials in the Cochrane review, and the groups were evenly matched. The Chalder fatigue scores after GET improved by 3 points compared to relaxation and by 4 compared to the no-treatment control group. This represents an improvement on the Chalder fatigue scale of 13.6% (3/22) and 18.2% (4/22) compared to relaxation and no-treatment, respectively. If Tench et al. had been part of the current Cochrane review, then the reviewers would have concluded that this study provided evidence of the efficacy of exercise therapy for SLE at the end of treatment. However, GET did not lead to improvement on the other two fatigue scales, nor did it lead to objective improvement. Also, improvements in Chalder fatigue scores were not maintained at (three months) follow-up.

In conclusion, one cannot safely conclude that a treatment is effective if, after treatment, patients are still ill enough to re-enter the same trial to receive the same treatment again. Also, there are many problems in using the Chalder fatigue scale in ME/CFS studies. In particular, the ceiling effect, in combination with a high percentage of patients with the maximum score at baseline, make this scale unreliable and therefore unsuitable for use in ME/CFS studies.

2.6. Physical functioning

Larun et al. concluded that exercise therapy may moderately improve physical functioning after 12 to 24 weeks (low evidence) and the effect after 52 to 70 weeks is uncertain (very low evidence) compared to no-treatment (usual care and waiting-list) or relaxation and flexibility. They also stated that, when compared to adaptive pacing, “the available evidence suggests that exercise therapy may slightly improve physical functioning” after 24 and 52 weeks (low evidence for both) [6]. The reviewers noted that “Jason 2007 observed better results among participants in the relaxation group. The latter results were distorted by very large baseline differences in physical functioning between the exercise and relaxation groups (39/100 versus 54/100), and we therefore decided not to include these results in the meta-analysis” [6]. However, as discussed earlier, the objective physical functioning scores of the 6MWT in Jason et al. (2007) [11] are almost identical for both groups - 1335.27 (exercise) versus 1317.78 (relaxation). Consequently, their results were not distorted because there were no objective baseline differences in physical functioning between these two groups and Jason et al. should not have been excluded from the meta-analysis.

The reviewers came to their conclusions about the improvement of physical functioning based on a threshold for minimal important differences (MID) on the physical functioning subscale of the SF-36 of around 7 points based on studies in people with rheumatoid arthritis or chronic heart disease [6]. Minimally important clinical differences (MCIDs) should be estimated by an evaluation study similar in its pre-conditions to an investigative study that determines effects. Ideally this should be done in the same disease or a disease which is similar [67]. Just like the Cochrane reviewers, our review team could not find a study estimating the MCID in ME/CFS, or similar illnesses like Multiple Sclerosis or fibromyalgia, for fatigue. However, such a study to determine the MCID for physical functioning for ME/CFS was done

by Brigden et al. [68], albeit in children. They concluded in 2018 that the MCID was 10. In adults such a study was not done in ME/CFS but in fibromyalgia, which is more similar to ME/CFS than rheumatoid arthritis or chronic heart disease. It was done by Kaleth et al. ($n = 187$) [69] and they concluded that the MCID for the SF-36 physical functioning in fibromyalgia (FM) is 10 and that this corresponds with a MCID for the six-minute walk test (6MWT) in patients with FM of 167 meters.

It makes a difference if a review uses a score of 7, as the reviewers did, or a score of 10. However, on an individual level, this does not make a difference because the SF36-PF consists of 20 questions and each question counts for five points (0 to 100). So to improve by 7, a respondent would need to improve on two questions, which equates to 10 points. The only study in the review to release individual participant data for the SF36-PF and the 6MWT is White et al. (2011) [18, 64]. Table 4 shows the number and percentage of participants in each of the groups of that study who improved by 10 or more points on the SF36-PF. The data also shows that in each group only a small percentage of them improved by 167 m or more on the 6MWT. Moreover, GET (8.0%) was not more effective than no-treatment (8.3%) and CBT (3.9%) was the least effective of the four. It is also known that there is an inverse relationship between fatigue and physical functioning/activity [70]. This suggests that the small subjective improvement of fatigue after GET over no-treatment, was simply an artefact. At the same time, the analysis of the individual PACE trial data [64], shows that in 20% of cases, patients *improved subjectively* even though they *deteriorated objectively*, as can be seen in Table 4. This table also shows that in a considerable number of cases, for those whose subjective physical functioning scores had improved by 10 or more, the 6MWT

scores were missing. This means that we actually do not know if these patients improved or deteriorated according to objective measures.

In conclusion, an *increase* in subjective physical functioning with an *objective decrease* in physical functioning in a considerable number of cases, highlights the unreliability of subjective outcomes.

2.7. Quality of life

According to the Cochrane review [6], quality of life was only measured in Jason et al. (2007) [11], “which observed an MD [mean difference] of 9.00. The estimate is biased in favour of the control arm because of baseline differences between groups.” By this, the reviewers meant the baseline difference in *subjective* physical functioning scores between anaerobic activity therapy (ACT) and the relaxation control group (RELAX). However, as discussed earlier, the *objective* physical functioning scores (6-minute walk test) of the groups, *were almost identical*. Consequently, there were no baseline differences in physical functioning. Improvement in quality of life scores was 3.5% (ACT) and 12.3% (RELAX) and in the 6-minute walk test this was 3.2% (ACT) and 8.4% (RELAX). The quality of life scores, on the Quality of Life Scale (QOLS), at the twelve-month follow-up were 63.0 (ACT) and 72.0 (RELAX) (QOLS range: 16 to 112; higher scores indicating higher quality-of-life). These QOLS scores were still worse than for people with fibromyalgia (70), COPD, psoriasis and urinary incontinence (82), rheumatoid arthritis (83), systemic lupus erythematosus (84), osteoarthritis (87) and young adults with juvenile rheumatoid arthritis (92) [71].

Contrary to Cochrane’s findings, quality of life was not only measured in Jason et al. (2007) [11], but also in the PACE trial by White et al. (2011), who pub-

Table 4
Improvement in subjective and objective physical functioning in the PACE trial according to the released individual participant data

Outcome measure	CBT ($n = 161$)	APT ($n = 159$)	SMC ($n = 160$)	GET ($n = 160$)	Total
SF-36 PF improvement ≥ 10	103 (64.0%)	71 (44.7%)	84 (52.5%)	112 (70%)	370 (57.8%)
Combined improvement of SF-36 PF $\geq 10 + 6MWT \geq 167$ m	4 (2.5%)	4 (2.5%)	7 (4.5%)	9 (5.5%)	24 (6.5%)
SF-36 improved ≥ 10 yet 6MWT worsened	21 (20.4%)	11 (15.5%)	14 (16.7%)	16 (14.3%)	62 (16.8%)
Missing 6MWT data of those whose SF-36 PF improved ≥ 10	16 (15.5%)	16 (22.5%)	17 (20.2%)	25 (22.3%)	74 (20%)
SF-36 improved ≥ 10 yet 6MWT worsened for those with 6MWT data	24.1% (21/87)	20% (11/55)	20% (14/70)	16.7% (16/96)	20.1% (62/308)

6MWT: Six minute walk test; PACE trial: White et al. (2011) [18]; SF-36 PF: Short Form 36 Physical Functioning Health Survey Questionnaire; White et al. released individual participant data [64].

Table 5
Quality of life scores and QALYs in the PACE trial

Time point	APT	GET	SMC (no-treatment)
Missing data (%)	11/159 (6.9%)	17/160 (10.6%)	9/160 (5.6%)
Baseline score	0.48	0.52	0.50
24 weeks (EOT)	0.54	0.60	0.52
52 weeks follow-up	0.54	0.59	0.53
QALYs accrued	0.53	0.57	0.52
Net improvement after GET from control at 24 wks (%)	0.02 (3.8%)	–	0.06 (11.5%)
Net improvement after GET from control at 52 wks (%)	0.01 (1.9%)	–	0.04 (8.4%)
Net improvement in QALYs accrued	0.05 (10.4%)	0.05 (9.6%)	0.02 (4%)

EOT: end of treatment; FU: follow-up; QALYs: according to White et al. (2011) [18], quality adjusted life years were generated from the EQ-5D health-related quality of life questionnaire; wks: weeks. Quality of life scores from the PACE trial [18] were published in their cost effectiveness analysis [72].

lished this in their cost effectiveness analysis [72]. This analysis was published one year after their original publication [18] and it was one of the references to studies included in the Cochrane review [6].

The net improvement of the quality of life scores (EQ-5D) after GET at 52 weeks over the adaptive pacing control group in the PACE trial, was 1.9%. A study by Olesen et al. [73] ($n=20,220$) 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) [72] was similar to the score (0.60) for people with five or more chronic health conditions and worse than in cerebral thrombosis (0.62), rheumatoid arthritis and angina (0.65), acute myocardial infarction (0.66) [73], MS (0.67), lung cancer (0.69), stroke (0.71) or ischemic heart disease (0.72) (linear scale ranging from -0.624 to 1.000 where negative values are conditions considered worse than death; higher scores indicating a better quality of life) [74].

Finally, the net improvement in QALYs (quality-adjusted life-year) - which refers to gains in health, combining a time dimension and an adjustment for quality of life [75] - was marginally better after pacing than after GET. In conclusion, one cannot safely conclude that GET is an effective treatment in view of the lack of significant improvement of quality of life scores after exercise treatment in Jason et al. (2007) and White et al. (2011). In fact, even after treatment, the quality of life scores remain lower than in many other disabling diseases.

2.8. Objective outcomes

The objective outcomes were not discussed by the Cochrane review, even though they were used by seven of the eight studies. Powell et al. (2001), the study that was removed by the Cochrane review after

their sensitivity analysis, was the only study not to use them. As can be seen in Table 6, in most studies, exercise therapy did not lead to objective improvement. In Fulcher and White (1997) [14], there was a small improvement at the end of treatment, yet the control group was badly matched and patients in the exercise group had normal fitness at baseline. Moreover, participants in the exercise group had sessions of five to fifteen minutes, increasing to a maximum of thirty minutes, at least five days a week. Such a workload would exclude most patients with ME/CFS. In Wearden et al. (1998) [16], the groups were badly matched as well and in Wallman et al. (2004) [13], pacing was labelled as GET and pacing, as discussed earlier.

In conclusion, the objective outcomes used by almost all the studies in the Cochrane review show that GET does not lead to clinically significant objective improvement.

2.9. Chronic Fatigue Syndrome symptom count

This was measured by White et al. (2011) [18]. Figures for end of treatment were not released, but at 52 weeks there was no statistically significant difference in the improvement in Chronic Fatigue Syndrome symptom count between GET and SMC ($p=0.0916$) or GET and APT ($p=0.23$).

2.10. Did patients comply with graded exercise therapy?

The only trial that answered this question is Wearden et al. (1998) [16]. In this study, 67 patients were randomised to graded exercise and fluoxetine or to graded exercise and drug placebo. Of these 67 patients, only 34.3% (23/67) complied fully with graded exercise. In the exercise placebo groups,

Table 6
Objective improvements after exercise therapy compared to control

Study	Rx	Control	Objective outcome	Baseline	Obj improvement compared to control (EOT)	Obj improvement compared to control (FU)
Fulcher and White (1997) [14] ($n=66$);	GET	Flex + relax	VO ₂ peak	31.8	7.5% (2.4/31.8; 12 wks)	—
Jason et al. (2007) [11] ($n=114$)	ACT	Relax	6MWT	1335.27	—	Relax improved 5.1% (68.42/1335.27) more than ACT
Moss–Morris et al. (2005) [12] ($n=49$)	GET	NT	VO ₂ peak	31.99	Both groups deteriorated; by 14.9% (4.78/31.9) exercise and 16.8% (5.22/31.0) NT; difference not statistically significant $p=0.39$)	—
Powell et al. (2001) [15] ($n=148$)	GET	NT	—	—	—	—
Wallman et al. (2004) [13] ($n=61$)	Pacing	Relax + flex	O ₂ uptake	15.6	9% improvement + flexibility worsened by 8%	—
Wearden et al. (1998) [16] ($n=136$)	Exercise + placebo drug	Exercise control (NT) + placebo drug	O ₂ uptake (exercise test)	19.9	10% (26 wks)	—
Wearden et al. (2010 and 2013) [17, 33] ($n=296$)	GIA	NT	ST	Not published	“No between group differences” ($p=0.819$; 20 wks)	“No between group differences” ($p=0.832$; 70 wks)
White et al. (2011) [18] ($n=640$)	GET	APT and NT	6MWT; Step test; Actometer	312 6MWT; Step test: figures not published; using Actometer at FU cancelled	—	6MWT: 41.0 APT; 35.3 NT; Step test: GET had no effect on fitness; exact mean scores not published [38]

6MWT: six-minute walk test; APT: adaptive pacing treatment; EOT: end of treatment; FU: follow-up; GIA: gradually increasing activity; NT: no-treatment; O₂: oxygen; Rx: treatment; wks: weeks. In Wallman et al. (2004) [13], pacing was labelled as GET and pacing (see problems with the controls). Also, they used a submaximal cycle test to measure oxygen uptake.

patients “were not offered any specific advice on how much exercise they should be taking, but were told to do what they could when they felt capable and to rest when they felt they needed to.” 78.3% (54 of the 69 patients) in the two control groups, who were treated with exercise placebo and fluoxetine or exercise placebo and drug placebo, complied fully with exercise placebo.

In conclusion, only one trial measured if patients complied with exercise therapy. They found that only one out of three participants treated with exercise therapy, actually complied with it. Consequently, one cannot safely conclude that exercise therapy is effective

in that study. As far as the other studies are concerned, if one does not know if patients actually adhered to the prescribed exercise regime, then one cannot conclude if that treatment is safe or effective.

2.11. A study was excluded that contradicted the main findings

Núñez et al. (2011) [76], was excluded from the Cochrane review, because, according to the reviewers, exercise therapy was a minor part of the intervention and it did not measure outcomes viewed as primary outcomes in the review [6].

The trial compared multidisciplinary treatment combining CBT, GET and pharmacological treatment with usual treatment. It found that at twelve months follow-up, the interventions did not improve health-related quality of life scores, and led to worse physical function and bodily pain scores. Núñez et al. concluded that “the results of our study tend to support the somewhat controversial findings of Twisk and Maes that the combination of CBT and GET is ineffective and not evidence-based and may in fact be harmful” [76].

3. Work rehabilitation and medical retirement in ME/CFS

3.1. Presenting to an occupational health physician

According to a recent review of work rehabilitation and medical retirement in ME/CFS [7], most patients who present to their occupational health physician with chronic fatigue (CF) do not suffer from ME/CFS. The single most important factor in discriminating ME/CFS from idiopathic CF or psychiatrically explained CF, is postexertional malaise (PEM), the main characteristic of ME/CFS. PEM is also an important prognostic indicator of poorer outcome at follow-up [77].

Routine testing does not reveal any abnormalities in most patients with ME/CFS. Consequently, large parts of the medical profession view ME/CFS as a psychological disorder. However, increasingly advanced tests have become available over the last 5–10 years, and as noted by Komaroff in a recent overview [78], more and more studies are documenting underlying biological abnormalities, involving many organ systems, in patients with ME/CFS. *These abnormalities* include metabolic changes, lactic acid production irregularities, immunological abnormalities in lymphocytes - especially in T cells and poorly functioning natural killer cells - and significant elevation of many blood cytokines, especially in the first three years of illness, which are correlated with the severity of the illness. These studies have also shown widespread neuro-inflammation of the brain and cognitive impairments, not explained by concomitant psychiatric disorders. Multiple cardiopulmonary exercise test (CPET) studies have demonstrated an impairment in the cellular energy production in patients with ME/CFS. This energy production impairment is much more prominent dur-

ing a second exercise test repeated 24 hours after the first [78]. A study by Melamed et al. that used invasive CPET - so that arterial blood samples could be taken repeatedly during the test - concluded that “exertional intolerance is caused solely by poor systemic oxygen extraction.” Abnormal peripheral oxygen extraction “is also the exercise hallmark of the mitochondrial myopathies” for example [79].

3.2. ME/CFS problems interfering with work

ME/CFS is far from a rare disease [80]. In the Netherlands, for example, it is more common than MS [81]. Diagnosing can be difficult and up to 90% of patients remain undiagnosed [82]. This occurs because, on the one hand, most doctors do not know much about the disease and on the other hand because there is no diagnostic test. Treatment is based on symptom management. The most important problems interfering with work, according to research by TNO [81], an independent Dutch research institute, are muscle pain, severe and disabling chronic fatigue, cognitive dysfunction - concentration or short-term memory impairments, difficulty with reading or information processing. The other really important problem interfering with work is PEM [83]. To be considered disabled in the US, patients have to be able “to do sustained work activities in an ordinary work setting on a regular and continuing basis...A “regular and continuing basis” means [being able to work] 8 hours a day, for 5 days a week” [84]. PEM keeps them from being able to do that.

Other symptoms which occur in more than 80% of cases, according to a large nationwide population-based cohort study by Castro-Marrero et al. [85], are muscle weakness, dizziness, generalized chronic pain and/or joint pains, hypersensitivity to noise and/or light, new onset headaches or migraines, and episodes of postural orthostatic hypotension (POTS). Characteristic of the disease is that symptoms and impairments increase following exertion. But they can fluctuate in nature and severity throughout the course of the disease and they can differ from patient to patient.

ME/CFS can interfere with the ability to show up at work every day and/or stay all day long and with work-related physical functions like walking, sitting, standing, lifting, carrying, pushing, pulling, reaching, and handling. It can also interfere with cognitive functions including the ability to remember, understand, and carry out simple instructions, the ability to use appropriate judgment, and the ability to respond

appropriately to supervision, co-workers, and usual work situations, including changes in a routine work setting [86].

3.3. Prognostic factors

The most important prognostic factor is how the illness is managed in its initial stages, according to Dr. Melvin Ramsay [87]. He was the infectious disease specialist who was involved in the management of the almost 300 patients, mainly doctors and nurses, who fell ill during the outbreak in the Royal Free Hospital in London in 1955. He also noted that most patients will try to go back to work in the initial stages when they are improving. With many other illnesses, that does not pose a problem but with ME/CFS it does. As documented by Dr Ramsay, “those patients who are given a period of enforced rest from the onset have the best prognosis” [87].

Other factors associated with a worse outcome are illness duration and severity, older age, or having a comorbid psychiatric disorder when patients fall ill [7].

3.4. Work rehabilitation

According to the report by the IOM, “ME/CFS symptoms often are so debilitating” that “35 to 69 percent” are unable to work as found by a review of 15 studies [8]. For the minority who have improved enough and whose own work, or adapted work, is physically light enough to consider a return to work, work rehabilitation will usually need to start with a dramatically reduced workload and number of hours [88, 89]. For these patients, an individualised return to work plan should be developed, taking the symptoms and specifics of the disease and the way it is affecting the individual employee into account. In particular, care should be taken to match the subject’s capabilities to the proposed employment duties with the need for flexibility to adjust time of day to start work, the ability to work from home, etc. Work that is likely to place sustained high pressure on the employee like strenuous physical work, long working hours, work requiring sustained high levels of attention and concentration or rapidly changing shift patterns, is inadvisable or at the very least, requires careful monitoring until it is clear that the employee is able to sustain this level of work. Definite deadlines in anticipating recovery and future employability should not be set to avoid causing relapses [9].

3.5. Disability Discrimination Act

In the UK, most employees with ME/CFS used to fall under the Disability Discrimination Act 1995 [89], which was replaced by the Equality Act 2010 in England, Scotland and Wales but not in Northern Ireland [90]. The Equality Act covers the provisions in the Disability Discrimination Act. It also offers additional protection from indirect discrimination, discrimination arising from disability and discrimination on the basis of association or perception. Disability is defined as “a physical or mental impairment” which “has a substantial and long-term adverse effect on a person’s ability to carry out normal day-to-day activities” [90].

According to this Act, employers are required to make ‘reasonable adjustments’ to the workplace and to working practices, so that a disabled or chronically ill employee is not at a disadvantage, compared to healthy and able-bodied employees, and is able to work despite his or her disability. Workplace adjustments that fall under this disability act could include: changing work and/or location of work, limiting working hours, reducing workload, working from home, and limiting or reducing physical tasks [88, 89, 91]. Most other western countries will likely have a similar Act in place to protect disabled workers.

3.6. Important factors enabling a return to work

According to a report by NIVEL, the Netherlands Institute for Health Services Research [92], there are a number of important things which enabled ME/CFS patients to (return to) work. For 92%, the most important thing was support in finding the right balance between work and spare time. The second most important thing (84%), was support and cooperation from the employer to enable patients to continue to work. Other important things were the following:

- supplying information about ME/CFS to colleagues and superiors (62%);
- changing tasks (61%);
- reducing the number of hours they had to work (61%);
- more rest periods during working times (60%);
- the availability of a special rest place at work (45%);
- working from home (52%);
- individual support and coaching in general (51%);

- and by an occupational health physician in particular (44%);
- adjustments to working conditions (furniture, physical aids) (38%);
- and a regulation or provision for commuting to work (36%).

3.7. Medical retirement

Many patients with ME/CFS are too disabled to work and as a consequence are receiving long term disability benefits. In cases when incapacity is prolonged, work rehabilitation is impossible or unsuccessful, and prognosis appears to be poor, then medical retirement is often the only option. The occupational health physician/doctor may then be asked to advise on this. Qualifying criteria of a company pension scheme inevitably vary, although permanent inability to undertake normal duties for reasons of ill health is a common requirement in the UK [88, 89, 91]. In contrast, patients in the Netherlands can be granted full or partial temporary medical retirement, without the need to prove permanent inability [93]. Nyland et al. did a long-term follow-up study of seven years in young adults ($n=111$) who had developed ME/CFS after glandular fever / mononucleosis and had been ill for a mean of 4.7 years at the start of the study. Their study showed that in younger patients, who have a much better prognosis than older patients in general and even more so after glandular fever, “long-term compensations to secure the socioeconomic position does not inhibit return to work, but may be essential contributors to...becoming employed” later on [94].

4. Discussion

At the beginning of October 2019, Cochrane published a long-awaited amendment to its review of exercise therapy for ME/CFS [6] following a formal complaint to its Editor in Chief [4]. In this article we analysed the amended version. We also summarised the recently published review of work rehabilitation and medical retirement in ME/CFS [7] as they analysed a large number of studies, which reported on work outcomes in trials of CBT and GET, including those in this Cochrane review. Unfortunately, the published amendment does not address the main flaws of the Cochrane review and the studies in it, continues to overestimate the evidence for exercise therapy in ME/CFS and downplays the flaws of those studies.

The main conclusions of the Cochrane reviewers were the following:

- 1 “Exercise therapy probably has a positive effect on fatigue [at the end of treatment] in adults with CFS compared to usual care or passive therapies”;
- 2 “The evidence regarding adverse effects is uncertain”;
- 3 “All studies were conducted with outpatients diagnosed with 1994 criteria of the Centers for Disease Control and Prevention or the Oxford criteria, or both. Patients diagnosed using other criteria may experience different effects”;
- 4 Limited evidence makes it difficult to draw conclusions about the effectiveness of exercise therapy compared to adaptive pacing or other interventions.

However, the main characteristic of ME/CFS is an exacerbation of some or all of an individual’s ME/CFS symptoms that occurs after physical or cognitive exertion and leads to a reduction in functional ability - also known as postexertional malaise (PEM) [8]. This core symptom is not required according to the Oxford criteria [19] and only optional according to the Fukuda criteria [20]. “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,” according to the American Agency for Healthcare Research and Quality (AHRQ) [25]. Both AHRQ and the American National Institute of Health (NIH) [23, 24], recommended that the flawed Oxford criteria should be retired.

4.1. Bias and flaws

Unfortunately, serious issues with the selection criteria were not the only flaws of the studies in the review and the review itself. Another important issue is the fact that all studies in the review were non-blinded by definition and they were using subjective primary outcomes. The reviewers acknowledge that relying on subjective primary outcomes in such trials increases the risk of performance and detection bias but they do not think that is a problem because “many groups representing the interests of those with CFS are opposed to exercise therapy, and this may in contrast reduce the outcome estimate” [6]. However, if that would be the case, then the majority of patients would not take part in those studies and only those who are only very mildly affected and do not have a

major problem with exercising would take part. Consequently, the self-selection of participants would reduce the generalisability of the results dramatically and increase instead of reduce the bias. Moreover, the following figures from the PACE trial, one of the eight trials in the Cochrane review, contradict the statement from the reviewers. At baseline and before treatment was started, the PACE trial researchers asked the participants if treatment is logical and 84% (APT), 71% (CBT), 84% (GET) and 49% (SMC) answered yes to that question [18].

The four Cochrane reviewers, just like the investigators of seven of the eight trials in their review, are supporters of the biopsychosocial model. This model is based on the assumption that there is no underlying illness in ME/CFS. Instead, patients are deemed to have false illness or dysfunctional beliefs that exercise is bad for them, and patients subsequently develop fear of exercise and become deconditioned. According to this model, deconditioning is the reason for the symptoms of ME/CFS. However, *it is illogical* to then determine if exercise therapy is effective *by using subjective outcomes in patients who don't know how to interpret their symptoms correctly*. The only way to adequately check if fitness has improved and patients are not deconditioned anymore, is by using objective fitness outcomes. Yet, none of the eight studies in the Cochrane review used objective outcomes as primary outcomes. The Cochrane review itself, left out an analysis of the objective outcomes which were used as secondary outcomes by seven of the eight studies in the review. A simple, cheap, reliable and easy objective outcome to use is the six-minute walk test [69], which was used by two of the eight studies. There was no clinically significant objective improvement according to the six-minute walk test in the PACE trial (White et al. (2011) [18]) and in Jason et al. (2007) [11], where patients objectively improved more with relaxation than with exercise.

A systematic review by Hróbjartsson et al. [95], concluded that there is pronounced bias due to lack of patient blinding in clinical trials with patient-reported outcomes and that non-blinded patients exaggerated the effect size by an average of 0.56 standard deviation. According to the Cochrane review, at the end of treatment when compared to no-treatment, the effect size was 0.66. They also concluded that one of the studies in the review did not pass their sensitivity analysis and after excluding that study, the effect size dropped to 0.44. This is less than the effect size of relying on subjective

outcomes in non-blinded studies as found by Hróbjartsson et al.

The BRANDO project (Bias in Randomised and Observational studies) [96], which amongst others included Stanford professor Ioannidis, concluded in 2012 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” because lack of blinding is “associated with an average 13% exaggeration of intervention effects”. “Therefore, trials in which blinding is not feasible should focus as far as possible on objectively measured outcomes.” The Cochrane review and the studies in it, failed to do this.

4.2. *Methods that help to show that treatment is effective, even when it is not*

Cuijpers and Cristea [97] concluded that there are several methods available to help researchers show that their therapy is effective, even when it is not. According to them, these methods “include a strong allegiance towards the therapy, anything that increases expectations and hope in participants, making use of the weak spots of randomised trials (risk of bias), small sample sizes and waiting list control groups” or no-treatment groups. Many of these methods were seen in the studies in the Larun review.

First, researchers of seven of the eight trials in the review had a strong allegiance towards the therapy. According to a systematic review by Dragioti et al., “experimenter’s allegiance effect inflates the reported effect sizes in randomized controlled trials in psychotherapy by 30 %” [50]. Jason et al. (2007) [11], the one study where this was not the case, “found few differential results among the [four] non-pharmacologic interventions.”

Second, during the PACE trial - the largest CBT and GET trial ever conducted - patients were sent a newsletter which stressed that in the new NICE guidelines, “recommended therapies include Cognitive Behavioural Therapy, Graded Exercise Therapy and Activity Management” [98].

Further, a key feature of GET is pushing beyond limits. In the PACE trial’s GET manual for participants, participants are told to interpret symptom flares as “a normal part of CFS/ME recovery” and not as a worsening of the disease [27]. In the FINE trial patient booklet, patients are told that “Activity or exercise cannot harm you” and that “medical research evidence shows: no underlying serious disease.” “You will have conquered CFS by your own effort and you

will be back in control of your body again.” That booklet also states that “you cannot relapse because you now know how to combat it” [99].

Third, five of the studies in the review used no-treatment control groups, which were labelled treatment as usual care, specialist medical care, exercise placebo etcetera. The Cochrane review itself relied on exercise compared to the no-treatment control group at the end of treatment to label exercise therapy moderately effective for fatigue in ME/CFS.

Fourth, the reviewers introduced further bias into their review in two other ways. First of all by publishing it in the Cochrane Common Mental Disorders Group and giving the wrong impression about ME, which has been classified as a neurological disease since 1969 by the World Health Organisation with CFS as an equivalent [100]. Secondly, according to the acknowledgement section of their review, the Cochrane reviewers asked two proponents of the biopsychosocial model, who inappropriately view ME/CFS as a behavioural problem and CBT and GET as effective treatments, for advice on an exercise review. This introduced further bias into their review as one of them is not only a psychiatrist, but also a co-author of Fulcher and White (1997) [14] and the principal investigator of White et al. (2011) [18], two of the studies in the review. This is akin to asking students for advice on how to mark their own exam papers. If the reviewers thought it was necessary to ask for advice for their exercise review, then it would seem more logical and appropriate to ask for advice from exercise physiologists. For example, Professor Keller, or Professors Davenport, VanNess and Snell who have published many exercise physiological papers on ME/CFS [101–106] yet had no involvement with the studies in the review. Maybe the reason why they were not asked for advice is that in May 2018, VanNess et al. concluded that “graded exercise aimed at training the aerobic energy system, not only fails to improve function, but is detrimental to the health of [ME/CFS] patients and should not be recommended” [106].

4.3. *Problems with the primary outcome measure*

Fatigue was used as the main primary outcome of the review - the other one was safety of exercise therapy - and seven of the eight studies used the Chalder fatigue scale. As discussed earlier, there are many problems with this instrument. One of the problems is the ceiling effect whereby scores of patients who have

the maximum score at baseline - or have the maximum score for individual items - cannot get worse (on these items) if they deteriorate. The consequence of this can be for example that if patients deteriorate on eight items and improve on three, that their scores reflect an improvement by three even though in reality they have deteriorated by five. The magnitude of this problem was highlighted by the individual participant data of the FINE trial and PACE trial [63, 64], as can be seen in Table 2. This shows for example, that 75% of participants in the FINE trial - Wearden et al. (2010) [17] - for whom individual data was released, had the maximum Chalder fatigue score at baseline.

Wallman et al. [107] analysed the reliability of the outcome variables in their own study (Wallman et al. (2004) [13]), which was part of the Cochrane review. They concluded that the mental and physical Chalder fatigue scores at baseline “were of questionable reliability in both groups” and of moderate reliability after treatment. Yet the “post-intervention scores for peak oxygen uptake...and peak power (W·kg⁻¹) were all similar to baseline values (i.e. highly reliable)”.

The Cochrane review analysed the Chalder fatigue scale (ChFS) outcome in Fulcher and White (1997) [14]. They did not analyse the outcome of the second fatigue questionnaire - visual analogue scale (VAS) - used by the same trial. Improvements in fatigue compared to the control group were the following: 18.6% (ChFS) yet only 6.4 % (VAS). The VAS also highlighted the fact that the two groups were not evenly matched for physical fatigue. A study into the efficacy of exercise therapy for SLE, in patients without active disease, highlighted the problems of relying on the Chalder fatigue scale [66]. At the end of treatment, the ChFS scores after GET had improved by 3 points compared to relaxation and by 4 compared to the no-treatment control group which is similar to the treatment effect on fatigue of 3.4 after exercise in the ME/CFS studies according to the Cochrane exercise review. However, GET did not lead to improvement on the other two fatigue scales - FSS and VAS - nor did it lead to objective improvement in the SLE study. Also, at three months follow-up the improvement on the ChFS had disappeared.

It is also unclear why the review relied on an outcome at the end of treatment, as a systematic review by Whiting et al. concluded in 2001 that because of the relapsing nature of ME/CFS, “follow-up should continue for at least an additional 6 to 12 months after the intervention period has ended, to confirm that any improvement observed was due to the intervention itself and not just to a naturally occurring fluctua-

tion in the course of the illness” [108]. And a study that was part of the Cochrane review - Wearden et al. (2010) [17] - noted in their protocol that “short-term assessments of outcome in a chronic health condition such as CFS/ME can be misleading” [28]. Therefore, their primary outcome was at 70 weeks and not at the end of treatment and the Cochrane review should have done the same thing.

Whiting et al. also concluded that subjective outcomes may be unreliable because “persons 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. A more objective measure of the effect of any intervention would be whether participants have increased their working hours, returned to work...or increased their physical activities” [108]. Moreover, unlike self-report methods, objective measures are less susceptible to subjectivity and tend to yield more accurate results regarding fitness and reconditioning [109], not only in ill patients but also in the healthy population [110]. The unreliability of subjective outcomes is highlighted by the following examples from three trials that were part of the Cochrane review:

1. In Jason et al. (2007) [11], there was a *substantial difference in subjective* physical functioning scores at baseline between exercise and control group, yet *objectively there wasn't* (6MWT);
2. In Moss-Morris et al. (2005) [12], after GET, physical functioning *subjectively improved* by 30%, yet *objectively deteriorated* by 15% (CPET);
3. In the PACE trial by White et al. [18], physical functioning *improved subjectively* yet *deteriorated objectively* (6MWT) in a considerable number of participants, as can be seen in Table 4.

4.4. Core set of outcome measures in exercise trials

A Cochrane review into the efficacy of exercise therapy for multiple sclerosis (MS) - a disease with many similarities with ME/CFS - concluded in 2005 that there is an urgent need for a consensus on a core set of measurements of outcome to be used in exercise trials. These outcome measures should be reliable and valid and reflect activities of daily living and quality of life domains. In addition, these studies should be methodologically sound and also use objective outcomes [111]. A systemic review by Latimer-Cheung et al. [112] into the efficacy of exercise therapy for

MS implemented this urgent need for a core set in 2013 by using indicators of physical capacity that included aerobic capacity, measured via CPET, most commonly defined as maximal oxygen consumption (VO₂ max), and anaerobic threshold as these fitness outcomes are relevant to mobility, performance of activities of daily living, fatigue, and quality of life among individuals with MS according to the authors. This review did not focus on one outcome (fatigue) as was done by Larun et al., even though fatigue is often an important problem in MS.

Walking impairment is one of the most common and life-altering features in MS, just like it is in ME/CFS. Walking impairment is most commonly assessed with performance measures such as the 6-minute walk test. This test was used by Jason et al. (2007) [11] and White et al. (2011) [18], and just like the other objective outcome measures used by the trials in the review, showed that graded exercise therapy does not lead to clinically significant objective improvement. This was found by the reanalysis of the Cochrane exercise review [5] and can also be seen in Table 6.

4.5. Safety

The Cochrane review also concluded that “The evidence regarding adverse effects is uncertain” because only two of the eight studies reported on the safety of the intervention. Consequently, the review provided no evidence that exercise therapy is actually safe. Safety of patients should always come first. If a study cannot guarantee safety, then that treatment should not be recommended. This is particularly so when patient surveys over the last 20 years have repeatedly shown that exercise therapy is harmful in at least 50% of cases, as was found by Kindlon [113] and Geraghty et al. [114] who pooled surveys in 2011 and 2017, respectively. Moreover, in the UK, the National Institute for Health and Clinical Excellence (NICE) is reviewing their ME/CFS guideline. A survey amongst ME/CFS patients ($n = 2,274$) carried out for the NICE review process by the Oxford Brookes University [115], dated the 27th of February 2019, found the following. 98.5% of the patients who took part in the survey experienced post-exertional malaise, the core symptom of the disease.

Worsening of symptoms after treatment was reported by:

- 81.1% (GET);
- 85.9% (GET combined with CBT);

- 58.3% (CBT combined with GET).

After treatment, the percentage of severely affected patients increased from:

- 12.9% to 35.3% (GET);
- 13.2% to 41.9% (GET combined with CBT);
- 12.6% to 26.6% (CBT combined with GET).

Black and McCully [116] concluded in 2005 “that CFS patients may develop exercise intolerance...after 4–10 days. The inability to sustain target activity levels, associated with pronounced worsening of symptomology, suggests the subjects with CFS had reached their activity limit.” A recent study by Lien et al. [117] concluded again that exercise deteriorates physical performance and increases lactate in patients with ME/CFS, whereas in the healthy population the exact opposite happens.

Also, as found by the PACE trial [18], exercise does not improve ME/CFS symptom count. The net improvement in QALYs (quality-adjusted life-year), in the PACE trial was marginally better after pacing than after GET and quality of life was still worse than in lung cancer, acute myocardial infarction, MS and other debilitating illnesses. And in Jason et al. (2007) [11], quality of life improved more after relaxation than after exercise therapy.

Finally, a study can only report that a treatment is safe or effective, if one knows if the participants have actually adhered to the treatment. Only one trial - Wearden et al. (1998) [16] - answered this question. 34.3% complied fully with exercise treatment versus 78.3% with no-treatment. As concluded by Lilienfeld et al., 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 [118]. In other words, two out of every three patients in the exercise group in Wearden et al. (1998) did not comply with the exercise regimen. Most likely this happened because they had no benefits from the treatment or were adversely affected by it. In the other seven trials, it is unclear if participants actually adhered to the treatment or not. In view of this, one cannot safely conclude that a treatment, in this case exercise treatment, is effective.

4.6. ME/CFS and prognosis

ME/CFS is a debilitating multisystem disease [8]. Most cases tend to start as an unremarkable viral infection. However, instead of recovering, patients begin to experience profound muscular (and cognitive) fatigue - for example heavy legs - following

activities which were previously completed without difficulty. Also typical is an abnormally prolonged delay in the restoration of muscle (and brain) power [119]. Prior to developing ME/CFS, most patients were sporty, healthy, and active [120]. There is no diagnostic test, therefore diagnostic criteria are used to diagnose ME/CFS. Illness severity can differ from patient to patient, but also throughout the course of the disease. Most people believe that fatigue is the main characteristic. However, as concluded by the Institute of Medicine in 2015, the main characteristic is postexertional malaise (PEM), which is an increase in symptoms after physical or mental exertion and further loss of functioning [8]. This core symptom distinguishes ME/CFS from psychiatric fatigue and from idiopathic chronic fatigue. A progression of ME/CFS is seen in 10 to 20% of cases according to Peterson et al. [121] and 13 to 26% according to a systematic review by Cairns and Hotopf [122]. Overall, according to the same systematic review, only 5% [122] will recover and the prognosis for severely affected patients - those who are homebound or bedridden - is even worse [123]. Early management of the illness appeared the most important determinant of severity [7, 87].

Illness severity, cognitive problems, a comorbid psychiatric disorder [7, 122], or having comorbid fibromyalgia [124], at baseline, are associated with a poor outcome. Psychosocial factors, smoking, personality or attitude show little relationship to recovery. Spontaneous recovery is rare and only occurred in patients with an illness duration of less than 1 1/2 years [125]. The recently published review of work rehabilitation and medical retirement found that recovery and substantial improvement are uncommon in patients if they have been ill for longer than 2 to 3 years [7]. This confirmed the conclusion by the Inspectorate Work and Pay of the Dutch Ministry of Work and Social Affairs [93], that if patients have been on long-term sick leave for two years or more and treatment with CBT did not make a difference, then the prognosis for a return to work is poor.

4.7. Work rehabilitation and medical retirement

People with ME/CFS are often unable to engage in economically productive work and typically request sick leave as a response to their health crisis [126]. Between 27% and 65% are reported not to be working due to ME/CFS according to a systematic review by Cairns and Hotopf [122]. The American CDC - Centers for Disease Control and Prevention - reported

that as many as 75% are not working due to ME/CFS [127]. Men, people in older age groups and those who have been ill for longer are more likely to have ceased employment due to their illness [7]. According to research by TNO, an independent Dutch research institute [81], only 7% of patients ($n=924$) had never been on long-term sick leave. Those who had been able to go back to work after long-term sick leave were working less hours, more often did less physically demanding work, were doing sedentary work behind a computer and were less often involved in management. Also, only 1/4 of patients who worked, were able to work more than 24 hours a week according to Nivel, another independent Dutch research institute [92].

4.8. Does GET restore the ability to work and relieve the economic burden on patients and society?

ME/CFS puts a heavy economic burden on patients, their partners, families and society [82]. CBT and GET have been recommended by guidelines as effective treatments for the last two decades and an influential systematic review concluded in 2005 that medical retirement should be postponed until patients have been treated with these two treatments [122]. In the Netherlands, many patients are still being forced to be treated with CBT and GET, because the chairman of the Dutch insurance physician association [128] does not agree with the conclusion from the Dutch Health Council that ME/CFS is a debilitating multisystem disease [129] and that patients should not be forced to be treated with CBT and GET. In April 2019, he told insurance physicians in an interview in a Dutch medical journal [128], that they should question patient's recovery behaviour if they do not want to be treated with CBT and GET. He also stated that an unwillingness to be treated, should have consequences for their disability benefits/medical pension. In the Netherlands, the more than 700 insurance physicians of the UWV (Uitvoeringsinstituut Werknemersverzekeringen or Employee Insurance Agency) [93] decide if employees will be granted (temporary) full or partial medical pensions.

A recently published review of work rehabilitation and medical retirement [7] also looked at the question of whether CBT and GET restore the ability to work in ME/CFS. Amongst the studies reviewed was the PACE trial ($n=641$) [18], the largest CBT and GET trial ever conducted. The efficacy of these treatment has also been assessed in real life outside of

clinical trials, in the Belgium CFS knowledge centres ($n=655$) and the NHS CFS clinics ($n=952$) [130, 131]. These evaluations, just like the PACE trial itself, showed that *more patients were unable to work and more were receiving illness benefits after being treated with CBT and GET than before treatment*. Consequently, CBT and GET do not restore the ability to work and in fact actually *increase* the economic burden on patients and society. These evaluations of the efficacy of CBT and GET in real life, by proponents of the biopsychosocial model themselves, also make it clear that questioning patients' recovery behaviour to force patients to be treated with CBT and GET, does not benefit patients or their employer. Nor does it reduce the economic burden on society.

4.9. Implications for research

According to the Cochrane reviewers, the implications for research are that "Further randomised controlled trials are needed to clarify the most effective type, intensity and duration of exercise therapy" [6]. Yet the therapy's claimed mechanism of action cannot be reconciled with what's known about the disease pathology in ME/CFS. Also, a therapy that only leads to a very small subjective improvement in fatigue if we were to ignore all the serious flaws of the studies; that does not lead to objective improvement, does not improve quality of life or symptom count, does not restore the ability to work and according to patients surveys, is harmful in more than 50% [113, 114], or more than 80% of cases according to the most recent survey by Britain's Oxford Brookes University [115], should not be used or recommended. Nor should it be investigated further. Twenty years of recommending this treatment have confirmed what patients have been saying for a long time. In a time that medicine should be patient centred, it is now important to listen to patients instead of continuing to ignore them and find treatments that restore their ability to work.

5. Conclusion

The recently amended Cochrane exercise review for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) concluded that graded exercise therapy (GET) improves fatigue at the end of treatment compared to no-treatment. Larun et al. also concluded that there is no evidence that GET is safe. However, the review continues to ignore the unrelia-

bility of subjective outcomes in non-blinded studies and fails to address other key flaws of the studies in the review. *These flaws included:*

- 1) using criteria that also select people who do not have the disease;
- 2) not excluding patients with a psychiatric or self-limiting illness;
- 3) badly matched control groups;
- 4) relying on an unreliable fatigue instrument as primary outcome;
- 5) not using objective outcomes and/or ignoring them;
- 6) outcome switching;
- 7) p-hacking;
- 8) ignoring evidence of multi-system biological pathologies that can not be explained by their psychological treatment rationale;
- 9) ignoring evidence of harms.

Analysis of the objective outcomes shows that GET does not lead to clinically significant objective improvement. It also does not lead to improvement of CFS symptoms count or quality of life measurements, which remains lower for those with ME/CFS, than in many other debilitating illnesses.

Only 5% of patients recover. Many patients are unable to work and those who can work, often need a reduction in hours and/or reduction of physical intensity. *Unfortunately, GET doesn't restore the ability to work.* Instead, more patients are unable to work and more are reliant on illness benefits after being treated with GET than before treatment with it.

Finally, to use the words of three leading exercise physiologists in the field of ME/CFS, “graded exercise [therapy]...not only fails to improve function, but is detrimental to the health of [ME/CFS] patients and should not be recommended” [106].

Acknowledgments

The authors would like to thank M.V.'s parents for typing out his speech memos and Kasia for her help in improving the article.

Conflict of interest

The authors declare no conflicts of interest.

Funding

This research received no external funding.

References

- [1] Larun L, Brurberg KG, Odgaard-Jensen J, Price JR. Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews* 2015;(2):CD003200.
- [2] Kindlon T. Comments on Larun L, Brurberg KG, Odgaard-Jensen J, et al. Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews Issue 4. 2015 Art. No.: CD003200.* (p.113-118)
- [3] Courtney R. Comments on Larun L, Brurberg KG, Odgaard-Jensen J, et al. Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews Issue 4. 2016 Art. No.: CD003200.* (p.119 to 133)
- [4] Tuller D, Trial By Error: Cochrane's Report on Courtney's Complaint. 12 MARCH 2019 <http://www.virology.ws/2019/03/12/trial-by-error-cochrane-report-on-courtneys-complaint/> (accessed 12 November 2019)
- [5] Vink M, Vink-Niese F. Graded exercise therapy for myalgic encephalomyelitis/chronic fatigue syndrome is not effective and unsafe. Re-analysis of a Cochrane review. *Health Psychology Open* July-December 2018:1–12
- [6] Larun L, Brurberg KG, Odgaard-Jensen J, Price JR. Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews* 2019, Issue 10. Art. No.: CD003200.
- [7] Vink M, Vink-Niese F. Work Rehabilitation and Medical Retirement for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Patients. A Review and Appraisal of Diagnostic Strategies. *Diagnostics (Basel)*. 2019;9(4).
- [8] Institute of Medicine (IOM); Committee on the Diagnostic Criteria for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome; Board on the Health of Select Populations. *Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Redefining an Illness*. Washington, DC: National Academies Press, 2015.
- [9] Carruthers BM, Jain AK, De Meirleir KL, Peterson DL, Klimas NG, Lerner AM. Myalgic encephalomyelitis/chronic fatigue syndrome: clinical working case definition, diagnostic and treatment protocols. *Journal of chronic fatigue syndrome*. 2003;11(1):7–115.
- [10] Carruthers BM, van de Sande MI, De Meirleir KL, Klimas NG, Broderick G, Mitchell T, Staines D, Powles AC, Speight N, Vallings R, Bateman L, Baumgarten-Austrheim B, Bell DS, Carlo-Stella N, Chia J, Darragh A, Jo D, Lewis D, Light AR, Marshall-Gradisnik S, Mena I, Mikovits JA, Miwa K, Murovska M, Pall ML, Stevens S. Myalgic encephalomyelitis: international consensus criteria. *Journal of Internal Medicine*. 2011;270(4):327–38.
- [11] Jason LA, Torres-Harding S, Friedberg F, Corradi K, Njoku MG, Donalek J, Reynolds N, Brown M, Weiner BB, Rademaker A, Papernik M. Non-pharmacologic interventions for CFS: A randomized trial. *J Clin Psychol Med Settings*. 2007;14(4):275–296.
- [12] Moss-Morris R, Sharon C, Tobin R, Baldi JC. A randomized controlled graded exercise trial for chronic fatigue syndrome: outcomes and mechanisms of change. *J Health Psychol*. 2005;10(2):245–59.
- [13] Wallman KE, Morton AR, Goodman C, Grove R, Guilfoyle AM. Randomised controlled trial of graded exercise in chronic fatigue syndrome. *Med J Aust*. 2004;180(9):444–8.
- [14] Fulcher KY, White PD. Randomised controlled trial of graded exercise in patients with chronic fatigue syndrome. *BMJ*. 1997;314(7095):1647–52.

- [15] Powell P, Bentall RP, Nye FJ, Edwards RH. Randomised controlled trial of patient education to encourage graded exercise in chronic fatigue syndrome. *BMJ*. 2001;322(7283):387-90.
- [16] 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. *Br J Psychiatry*. 1998;172:485-90. Erratum in: *Br J Psychiatry*. 1998;173:89.
- [17] Wearden AJ, Dowrick C, Chew-Graham C, Bentall RP, Morriss RK, Peters S, Riste L, Richardson G, Lovell K, Dunn G; 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*. 2010;340:c1777.
- [18] White PD, Goldsmith KA, Johnson AL, Potts L, Walwyn R, DeCesare JC, Baber HL, Burgess M, Clark LV, Cox DL, Bavinton J, Angus BJ, Murphy G, Murphy M, O'Dowd H, Wilks D, McCrone P, Chalder T, Sharpe M; PACE trial management group. 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*. 2011;377:823-836.
- [19] Sharpe MC, Archard LC, Banatvala JE, Borysiewicz LK, Clare AW, David A, Edwards RH, Hawton KE, Lambert HP, Lane RJ. A report—chronic fatigue syndrome: guidelines for research. *J R Soc Med*. 1991;84(2):118-21.
- [20] Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. The chronic fatigue syndrome: a comprehensive approach to its definition and study. *Ann Intern Med*. 1994;121(12):953-9.
- [21] Friedberg F, Dechene L, McKenzie MJ 2nd, Fontanetta R. Symptom patterns in long-duration chronic fatigue syndrome. *J Psychosom Res*. 2000;48(1):59-68.
- [22] Baraniuk, J. 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*. 2017;5(4).
- [23] Green CR, Cowan P, Elk R, O'Neil KM, Rasmussen AL. Draft executive summary: national institutes of health- Pathways to Prevention Workshop: Advancing the Research on Myalgic Encephalomyelitis/ Chronic Fatigue Syndrome. 2014 <https://huisartsvink.files.wordpress.com/2018/08/green-draftreport-odp-mecfs-1.pdf> (accessed 12 November 2019)
- [24] Green CR, Cowan P, Elk R, O'Neil KM, Rasmussen AL. Final Report: National Institutes of Health-Pathways to Prevention Workshop: Advancing the Research on Myalgic Encephalomyelitis/ Chronic Fatigue Syndrome, Executive summary. 2014 <https://huisartsvink.files.wordpress.com/2018/08/green-odp-finalreport-p2p-mecfs.pdf> (accessed 12 November 2019)
- [25] Smith MEB, Nelson HD, Haney E, Pappas M, Daeges M, Wasson N, McDonagh M. 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; December 2014. Addendum July 2016. Available at: https://huisartsvink.files.wordpress.com/2018/08/ahrq-smith-et-al-2016-chronic-fatigue_research.pdf (accessed 12 November 2019)
- [26] Mohr DC, Spring B, Freedland KE, Beckner V, Arean P, Hollon SD, Ockene J, Kaplan R. The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychother Psychosom*. 2009;78(5):275-84.
- [27] Bavinton J, Dyer N, White PD. PACE Manual for participants: Graded exercise therapy. 2004 <http://www.wolfson.qmul.ac.uk/images/pdfs/6.get-participant-manual.pdf> (accessed 12 November 2019)
- [28] Wearden AJ, Riste L, Dowrick C, Chew-Graham C, Bentall RP, Morriss RK, Peters S, Dunn G, Richardson G, Lovell K, Powell P. 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 Med*. 2006;4:9.
- [29] White PD, Sharpe MC, Chalder T, DeCesare JC, Walwyn R. PACE trial group. 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 Neurol*. 2007;7:6.
- [30] Ware JE, Kosinski M, Dewey JE. How to Score Version 2 of the SF-36® Health Survey. Lincoln, RI: Quality Metric Incorporated, 2000.
- [31] Wearden A. A randomised controlled trial of nurse facilitated self-help treatment for patients in primary care with chronic fatigue syndrome ISRCTN registry 18/05/2001 https://huisartsvink.files.wordpress.com/2018/08/wearden-isrctn-2001-archive_org.pdf (accessed 12 November 2019)
- [32] Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients *Trials*. 2017;18(1):122.
- [33] Wearden AJ, Emsley R. Mediators of the effects on fatigue of pragmatic rehabilitation for chronic fatigue syndrome. *J Consult Clin Psychol*. 2013;81(5):831-8.
- [34] Wearden AJ, Dowrick C, Chew-Graham C, Bentall RP, Morriss RK, Peters S, Riste L, Richardson G, Lovell K, Dunn G. Fatigue scale BMJ rapid response 27 May 2010 <http://www.bmj.com/rapid-response/2011/11/02/fatigue-scale-0> (accessed 12 November 2019)
- [35] Morriss RK, Wearden AJ, Mullis R. Exploring the validity of the Chalder Fatigue scale in chronic fatigue syndrome. *J Psychosom Res*. 1998;45(5):411-7.
- [36] Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD. The Extent and Consequences of P-Hacking in Science. *PLoS Biol*. 2015;13(3):e1002106.
- [37] Evans S. When and how can endpoints be changed after initiation of a randomized clinical trial? *PLoS Clinical Trials*. 2007;2:e18.
- [38] Kemp P. Freedom of Information request to Medical Research Council. When were the MRC informed that the PACE Trial Primary Outcome Measures of 'Positive Outcome' and 'Recovery' were eliminated from the Research Protocol September/October 2016 <https://huisartsvink.files.wordpress.com/2019/11/kemp-freedom-of-information-request-to-medical-research-council.pdf> (accessed 12 November 2019)

- [39] Scheeres K, Knoop H, Meer vd, Bleijenberg G. Clinical assessment of the physical activity pattern of chronic fatigue syndrome patients: a validation of three methods. *Health Qual Life Outcomes*. 2009;7:29.
- [40] Vink M. The PACE Trial Invalidates the Use of Cognitive Behavioral and Graded Exercise Therapy in Myalgic Encephalomyelitis/ Chronic Fatigue Syndrome: A Review. *J Neurol Neurobiol*. 2016;2(3).
- [41] White PD, Goldsmith K, Johnson AL, Chalder T, Sharpe M. Recovery from chronic fatigue syndrome after treatments given in the PACE trial. *Psychological Medicine*. 2013;43:2227-2235.
- [42] Chalder T, Goldsmith KA, White PD, Sharpe M, Pickles AR. Rehabilitative therapies for chronic fatigue syndrome: a secondary mediation analysis of the PACE trial. *Lancet Psychiatry*. 2015;2(2):141-52.
- [43] Sharpe, M, Goldsmith, KA, Johnson, AL. Rehabilitative treatments for chronic fatigue syndrome: Long-term follow-up from the PACE trial. *The Lancet*. 2015;2:1067-1074.
- [44] Wilshire CE, Kindlon T, Courtney R, Matthees A, Tuller D, Geraghty K, Levin B. Rethinking the treatment of chronic fatigue syndrome-a reanalysis and evaluation of findings from a recent major trial of graded exercise and CBT. *BMC Psychol*. 2018;6(1):6.
- [45] Vink M. 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. *J Neurol Neurobiol*. 2017;3(1).
- [46] Stulemeijer M, de Jong LW, Fiselier TJ, Hoogveld SW, Bleijenberg G. Cognitive behaviour therapy for adolescents with chronic fatigue syndrome: Randomised controlled trial. *BMJ*. 2005;330(7481):14, Erratum in *Science* 9:355 *BMJ*. 2005;330(7495):820.
- [47] Vink M. PACE trial authors continue to ignore their own null effect. *Journal of Health Psychology*. 2017; 1-7.
- [48] White PD. RCT of CBT, graded exercise, and pacing versus usual medical care for the chronic fatigue syndrome ISRCTN54285094. 22 May 2003 <https://huisartsvink.files.wordpress.com/2018/09/pace-isrctn54285094.pdf> (accessed 12 November 2019)
- [49] Westen D, Novotny CM, Thompson-Brenner H. The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. *Psychol Bull*. 2004;130(4):631-63.
- [50] Dragioti E, Dimoliatis I, Fountoulakis KN, Evangelou E. A systematic appraisal of allegiance effect in randomized controlled trials of psychotherapy. *Ann Gen Psychiatry*. 2015;14:25.
- [51] Luborsky L, Diguier L, Seligman DA, Rosenthal R, Krause ED, Johnson S, Halperin G, Bishop M, Berman JS, Schweizer E. The researcher's own therapy allegiances: A 'wild card' in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice*. 1999;6(1):95-106.
- [52] Luborsky L, Rosenthal R, Diguier L, Andrusyna TP, Berman JS, Levitt JT, Seligman DA, Krause ED. The dodo bird verdict is alive and well – Mostly. *Clinical Psychology: Science and Practice*. 2002;9(1):2-12.
- [53] Munder T, Flückiger C, Gerger H, Wampold BE, Barth J. Is the allegiance effect an epiphenomenon of true efficacy differences between treatments? A metaanalysis. *Journal of Counseling Psychology J Couns Psychol*. 2012;59(4):631-7.
- [54] Wilshire C. The problem of bias in behavioural intervention studies: Lessons from the PACE trial. *J Health Psychol*. 2017;22(9):1128-1133.
- [55] Lubet S. Investigator bias and the PACE trial *J Health Psychol*. 2017;22(9):1123-1127.
- [56] Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, Thabane M, Giangregorio L, Dennis B, Kosa D, Borg Debono V, Dillenburg R, Fruci V, Bawor M, Lee J, Wells G, Goldsmith CH. A tutorial on sensitivity analyses in clinical trials the what, why, when and how *BMC Med Res Methodol*. 2013;13:92
- [57] Goligher EC, Pouchot J, Brant R, Kherani RB, Avina-Zubieta JA, Lacaillle D, Lehman AJ, Ensworth S, Kopec J, Esdaile JM, Liang MH. Minimal clinically important difference for 7 measures of fatigue in patients with systemic lupus erythematosus. *J Rheumatol*. 2008;35(4):635-42.
- [58] Pouchot J, Kherani RB, Brant R, Lacaillle D, Lehman AJ, Ensworth S, Kopec J, Esdaile JM, Liang MH. Determination of the minimal clinically important difference for seven fatigue measures in rheumatoid arthritis. *J Clin Epidemiol*. 2008;61(7):705-13.
- [59] Ad Hoc Committee on Systemic Lupus Erythematosus Response Criteria for Fatigue. Measurement of fatigue in systemic lupus erythematosus: a systematic review. *Arthritis Rheum*. 2007;57(8):1348-57.
- [60] Pettersson S, Lundberg IE, Liang MH, Pouchot J, Henriks-son EW. Determination of the minimal clinically important difference for seven measures of fatigue in Swedish patients with systemic lupus erythematosus. *Scand J Rheumatol*. 2015;44(3):206-10.
- [61] Ridsdale L, Hurley M, King M, McCrone P, Donaldson N. The effect of counselling, graded exercise and usual care for people with chronic fatigue in primary care: a randomized trial. *Psychol Med*. 2012;42(10):2217-24.
- [62] Vink M, Vink-Niese A. Cognitive behavioural therapy for myalgic encephalomyelitis/chronic fatigue syndrome is not effective. Re-analysis of a Cochrane review. *Health Psychol Open*. 2019;6(1):2055102919840614
- [63] FINE trial released individual participant data <https://huisartsvink.files.wordpress.com/2019/11/fine-trial-released-individual-participant-data.xls> (accessed 12 November 2019)
- [64] FOIA request to QMUL (2016) FOIA request 2014/F73 Dataset file available at: <https://sites.google.com/site/pacefoir/pace-ipd.foia-qmul-2014-f73.xlsx> Readme file available at: <https://sites.google.com/site/pacefoir/pace-ipd-readme.txt> (accessed 12 November 2019)
- [65] Goldsmith LP, Dunn G, Bentall RP, Lewis SW, Wearden AJ. Therapist Effects and the Impact of Early Therapeutic Alliance on Symptomatic Outcome in Chronic Fatigue Syndrome *PLoS One*. 2015;10(12):e0144623.
- [66] Tench CM, McCarthy J, McCurdie I, White PD, D'Cruz DP. Fatigue in systemic lupus erythematosus: a randomized controlled trial of exercise. *Rheumatology (Oxford)*. 2003;42(9):1050-4.
- [67] Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. *J Clin Epidemiol*. 2017;82:128-136.
- [68] Brigden A, Parslow RM, Gaunt D, Collin SM, Jones A, Crawley E. Defining the minimally clinically important difference of the SF-36 physical function subscale for paediatric CFS/ME: triangulation using three different methods. *Health Qual Life Outcomes*. 2018;16(1):202.

- [69] Kaleth AS, Slaven JE, Ang DC. Determining the Minimal Clinically Important Difference for 6-Minute Walk Distance in Fibromyalgia. *Am J Phys Med Rehabil.* 2016;95(10):738-45.
- [70] Rongen-van Dartel SA, Repping-Wuts H, van Hoogmoed D, Knoop H, Bleijenberg G, van Riel PL, Franssen J. Relationship between objectively assessed physical activity and fatigue in patients with rheumatoid arthritis: inverse correlation of activity and fatigue. *Arthritis Care Res (Hoboken).* 2014;66(6):852-60.
- [71] Burckhardt CS, Anderson KL. The Quality of Life Scale (QOLS): Reliability, validity, and utilization. *Health Qual Life Outcomes.* 2003;1:60.
- [72] 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;7:e40808.
- [73] Olesen AV, Oddershede L, Petersen KD. Health-related quality of life in Denmark on a relative scale: Minicatalogue of mean EQ-5D-3L index scores for 17 common chronic conditions. *Nordic Journal of Health Economics.* 2016;4(2):44-56.
- [74] Falk Hvidberg M, Brinth LS, Olesen AV, Petersen KD, Ehlers L. The Health-Related Quality of Life for Patients with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *PLoS One.* 2015;10(7):e0132421.
- [75] Neumann PJ, Cohen JT. QALYs in 2018-Advantages and Concerns. *JAMA.* 2018;319(24):2473-2474.
- [76] Núñez M, Fernández-Solà J, Núñez E, Fernández-Huerta JM, Godás-Sieso T, Gomez-Gil E. 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. *Clin Rheumatol.* 2011;30(3):381-9.
- [77] Jason LJ, Taylor RR. Applying cluster analysis to define a typology of chronic fatigue syndrome in a medically-evaluated, random community sample. *Psychology & Health* Volume 17, 2002 - Issue 3 Pages 323-337 | Published online: 27 Oct 2010
- [78] Komaroff AL. Advances in Understanding the Pathophysiology of Chronic Fatigue Syndrome. *JAMA.* 2019.
- [79] Melamed KH, Santos M, Oliveira RKF, Urbina MF, Felsenstein D, Opatowsky AR, Waxman AB, Systrom DM. Unexplained exertional intolerance associated with impaired systemic oxygen extraction. *Eur J Appl Physiol.* 2019;119(10):2375-2389.
- [80] NICE Guidelines ME/CFS Criteria Chronic Fatigue Syndrome/Myalgic Encephalomyelitis (or Encephalopathy): Diagnosis and Management Clinical Guideline [CG53] Published date: August 2007. <https://www.nice.org.uk/guidance/cg53/resources/chronic-fatigue-syndromemyalgic-encephalomyelitis-or-encephalopathy-diagnosis-and-management-pdf-975505810885> (accessed 12 November 2019).
- [81] Blatter BM, van den Berg R, van Putten DJ. Werk, uitval en reïntegratie bij patiënten met ME/CVS TBV 13, nr. 7 (juli 2005) <https://huisartsvink.files.wordpress.com/2019/05/blatter-werk-uitval2005-werk.pdf> (accessed 12 November 2019).
- [82] Valdez AR, Hancock EE, Adebayo S, Kiernicki DJ, Proskauer D, Attewell JR, Bateman L, DeMaria A. Jr, Lapp CW, Rowe PC, Proskauer C. Estimating Prevalence, Demographics, and Costs of ME/CFS Using Large Scale Medical Claims Data and Machine Learning. *Front Pediatr.* 2019;6:412.
- [83] Stevens S, Snell C, Stevens J, Keller B, VanNess JM. Cardiopulmonary Exercise Test Methodology for Assessing Exertion Intolerance in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *Front Pediatr.* 2018;6:242.
- [84] Social Security Administration, Disability Insurance SSR 96-8p: Policy interpretation ruling titles ii and xvi: assessing residual functional capacity in initial claims effective/publication date: 07/02/96 https://www.ssa.gov/OP_Home/rulings/di/01/SSR96-08-di-01.html (accessed 28 February 2020)
- [85] Castro-Marrero J, Faro M, Aliste L, Sáez-Francàs N, Calvo N, Martínez-Martínez A, de Sevilla TF, Alegre J. Comorbidity in Chronic Fatigue Syndrome/Myalgic Encephalomyelitis: A Nationwide Population-Based Cohort Study. *Psychosomatics.* 2017;58(5):533-543.
- [86] Taylor RR, Kielhofner GW. An Occupational Therapy Approach to Persons with Chronic Fatigue Syndrome: Part Two, Assessment and Intervention, *Occupational Therapy In Health Care.* 2003;17:2, 63-87.
- [87] Ramsay M. Myalgic Encephalomyelitis: A Baffling Syndrome With a Tragic Aftermath. *ME Association* 1986, UK.
- [88] Mounstephen A, Sharpe M. Chronic fatigue syndrome and occupational health. *Occup Med (Lond).* 1997;47(4):217-27.
- [89] Action for ME, M.E. and work, 2013 <https://huisartsvink.files.wordpress.com/2019/05/actionforme-returnwro rk.pdf> (accessed 12 November 2019)
- [90] Equality Act 2010, UK Public General Acts 2010, Table of contents <http://www.legislation.gov.uk/ukpga/2010/15/contents>
- [91] NHS Plus Evidence based guideline project. Occupational Aspects of the Management of Chronic Fatigue Syndrome: a National Guideline Workplace management of chronic fatigue syndrome. October 2006 https://huisartsvink.files.wordpress.com/2019/05/nhs-occupational-aspects-cfs_full_guideline.pdf (accessed 12 November 2019)
- [92] De Veer AJE, Francke AL. Zorg voor ME/CVS-patiënten: Ervaringen van de achterban van patiënten organisaties met de gezondheidszorg. Utrecht: NIVEL. 2008 Available at: <https://www.nivel.nl/sites/default/files/bestanden/Rapport-draagvlakmeting-CVS-ME-2008.pdf> (accessed 12 November 2019)
- [93] Inspectorate Work and Pay, now the Inspectorate SZW of the Dutch Ministry of Work and Social Affairs Het Chronisch vermoeidheidssyndroom De beoordeling door verzekeringsartsen. Programma Inkomenszekerheid Nummer R 10/08, november 2010 ISSN 1383-8733 ISBN 978-90-5079-239-4 https://huisartsvink.files.wordpress.com/2019/05/inspectie-2010_cvs-verzekeringsartsen.pdf (accessed 12 November 2019)
- [94] Nyland M, Naess H, Birkeland JS, Nyland H. Longitudinal follow-up of employment status in patients with chronic fatigue syndrome after mononucleosis. *BMJ Open.* 2014;4(11):e005798.
- [95] 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. *Int J Epidemiol.* 2014;43(4):1272-83.

- [96] Savović J, Jones H, Altman D, Harris R, Jüni P, Pildal J, Als-Nielsen B, Balk E, Gluud C, Gluud L, Ioannidis J, Schulz K, Beynon R, Welton N, Wood L, Moher D, Deeks J, Sterne J. Influence of reported study design characteristics on intervention effect estimates from randomised controlled trials: combined analysis of meta-epidemiological studies. *Health Technol Assess.* 2012;16(35):1-82. Review.
- [97] Cuijpers P, Cristea IA. How to prove that your therapy is effective, even when it is not: a guideline. *Epidemiol Psychiatr Sci.* 2016;25(5):428-435.
- [98] PACE trial participants newsletter. December 2008 Issue 3 <https://huisartsvink.files.wordpress.com/2019/11/participantsnewsletter3.pdf> (accessed 12 November 2019)
- [99] FINE trial patient booklet Version 9, 29/04/05 <https://huisartsvink.files.wordpress.com/2019/11/fine-trial-patient-pr-manual-ver9-apr05.pdf> (accessed 12 November 2019)
- [100] WHO (World Health Organization). International Statistical Classification of Diseases and Health Related Problems (the) ICD-10, Volume 1: Tabular List, 2nd ed.; WHO: Geneva, Switzerland, 10th revision Volume 3 Alphabetical index Fifth edition 2010.
- [101] Davenport TE, Stevens SR, VanNess JM, Stevens J, Snell CR. Checking our blind spots: current status of research evidence summaries in ME/CFS. *Br J Sports Med.* 2019;53(19):1198.
- [102] Keller BA, Pryor JL, Giloteaux L. Inability of myalgic encephalomyelitis/chronic fatigue syndrome patients to reproduce VO₂-peak indicates functional impairment. *J Transl Med.* 2014;12:104.
- [103] Snell CR, Stevens SR, Davenport TE, Van Ness JM. Discriminative validity of metabolic and workload measurements for identifying people with chronic fatigue syndrome. *Phys Ther.* 2013;93(11):1484-92.
- [104] VanNess JM, Snell CR, Strayer DR, Dempsey L 4th, Stevens SR. Subclassifying chronic fatigue syndrome through exercise testing. *Med Sci Sports Exerc.* 2003;35(6):908-13.
- [105] VanNess JM, Stevens SR, Bateman L, Stiles TL, Snell CR. Postexertional malaise in women with chronic fatigue syndrome. *J Womens Health (Larchmt).* 2010;19(2):239-44.
- [106] VanNess JM, Snell C, Davenport T. Opposition to Graded Exercise Therapy (GET) for ME/CFS. Workwell Foundation. 1 May 2018 <https://huisartsvink.files.wordpress.com/2018/08/vanness-et-al-2018-mecfs-get-letter-to-health-care-providers-v4-30-2.pdf> (accessed 12 November 2019)
- [107] Wallman KE, Morton AR, Goodman C, Grove R. Reliability of physiological, psychological and cognitive variables in chronic fatigue syndrome and the role of graded exercise. *J Sports Sci Med.* 2005;4(4):463-71.
- [108] Whiting P, Bagnall AM, Sowden AJ, Cornell JE, Mulrow CD, Ramirez G. Interventions for the Treatment and Management of Chronic Fatigue Syndrome - A Systematic Review. *JAMA.* 2001;286(11):1360-8. Review. Erratum in: *JAMA* 2002;287(11):1401.
- [109] Rogers R, Reybrouck T, Weymans M, Dumoulin M, Van der Hauwaert L, Gewillig M. Reliability of subjective estimates of exercise capacity after total repair of tetralogy of Fallot. *Acta Paediatr.* 1994;83(8):866-9.
- [110] Scheeres K, Knoop H, Meer vd, Bleijenberg G. Clinical assessment of the physical activity pattern of chronic fatigue syndrome patients: a validation of three methods. *Health Qual Life Outcomes.* 2009;7:29
- [111] Rietberg MB, Brooks D, Uitdehaag BMJ, Kwakkel G. Exercise therapy for multiple sclerosis. *Cochrane Database of Systematic Reviews* 2005, Issue1. Art. No.: CD003980.
- [112] Latimer-Cheung AE, Pilutti LA, Hicks AL, Martin Ginis KA, Fenuta AM, MacKibbin KA, Motl RW. Effects of exercise training on fitness, mobility, fatigue, and health-related quality of life among adults with multiple sclerosis: a systematic review to inform guideline development. *Arch Phys Med Rehabil.* 2013;94(9):1800-1828.e3
- [113] Kindlon T. Reporting of harms associated with graded exercise therapy and cognitive behavioural therapy in myalgic encephalomyelitis/chronic fatigue syndrome. *Bull IACFS ME.* 2011 19(2):59-111. <https://huisartsvink.files.wordpress.com/2018/08/kindlon-reporting-of-harms-associated-with-get-and-cbt-in-me-cfs.pdf> (accessed 12 November 2019)
- [114] 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 *J Health Psychol.* 2017 Aug 1
- [115] Oxford Clinical Allied Technology and Trials Services Unit (OxCATTS), Oxford Brookes University. Evaluation of a survey exploring the experiences of adults and children with ME/CFS who have participated in CBT and GET interventional programmes FINAL REPORT. 27th February 2019 <https://huisartsvink.files.wordpress.com/2019/11/nice-patient-survey-outcomes-cbt-and-get-oxford-brookes-full-report-03.04.19.pdf> (accessed 10 November 2019)
- [116] Black CD, McCully KK. Time course of exercise induced alterations in daily activity in chronic fatigue syndrome. *Dyn Med.* 2005;4:10.
- [117] Lien K, Johansen B, Veierød MB, Haslestad AS, Bøhn SK, Melsom MN, Kardel KR, Iversen PO. Abnormal blood lactate accumulation during repeated exercise testing in myalgic encephalomyelitis/chronic fatigue syndrome. *Physiol Rep.* 2019;7(11):e14138.
- [118] Lilienfeld SO, Ritschel LA, Lynn SJ, Cautin RL, Latzman RD. Why Ineffective Psychotherapies Appear to Work: A Taxonomy of Causes of Spurious Therapeutic Effectiveness. *Perspectives on Psychological Science.* 2014;9(4):355-387.
- [119] Howes S, Goudsmit, E.M. Progressive Myalgic Encephalomyelitis (ME) or A New Disease? A Case Report *Phys Med Rehabil Int.* 2015;2(6):1052.
- [120] Bested AC, Marshall LM. Review of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: an evidence-based approach to diagnosis and management by clinicians. *Rev Environ Health.* 2015;30(4):223-49.
- [121] Peterson PK, Schenck CH, Sherman R. Chronic fatigue syndrome in Minnesota. *Minn Med.* 1991;74:21-26.
- [122] Cairns R, Hotopf MA. systematic review describing the prognosis of chronic fatigue syndrome. *Occup Med (Lond).* 2005;55(1):20-31.
- [123] Pheby D, Saffron L. Risk factors for severe ME/CFS. *Biol Med.* 2009;1:50-74.
- [124] Ciccone DS, Chandler HK, Natelson BH. Illness trajectories in the chronic fatigue syndrome: A longitudinal study of improvers versus non-improvers. *J Nervs Ment Dis.* 2010;198:486-493.
- [125] Van der Werf SP, De Vree B, van der Meer JWM, Bleijenberg G. Natural course and predicting self-reported

- improvement in patients with a relatively short illness duration. *J Psychosom Res.* 2003;53:749–753.
- [126] Fennell PA. The Four Progressive Stages of the CFS Experience. *Journal of Chronic Fatigue Syndrome.* 1995;1:3-4, 69-79.
- [127] Unger ER, Lin JS, Tian H, Natelson BH, Lange G, Vu D, Blate M, Klimas NG, Balbin EG, Bateman L, Allen A, Lapp CW, Springs W, Kogelnik AM, Phan CC, Danver J, Podell RN, Fitzpatrick T, Peterson DL, Gottschalk CG, Rajeevan MS; MCAM Study Group. Multi-Site Clinical Assessment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (MCAM): Design and Implementation of a Prospective/Retrospective Rolling Cohort Study. *Am J Epidemiol.* 2017;185(8):617-626.
- [128] Maassen H. Interview verzekeringarts Rob Kok: ‘Wij mogen patiënten aanspreken op hun herstelgedrag’ *MEDISCH CONTACT* 16-17 | 18 APRIL 2019 <https://www.medischcontact.nl/nieuws/laatste-nieuws/artikel/verzekeringarts-rob-kok-wij-mogen-patienten-aanspreken-op-hun-herstelgedrag.htm> (accessed 12 November 2019)
- [129] Dutch Health Council (2018) Aan: de Voorzitter van de Tweede Kamer der Staten-Generaal Nr. 2018, Den Haag 19 maart 2018 Gezondheidsraad | Nr. 2018/07 https://www.gezondheidsraad.nl/sites/default/files/grpublication/kernadvies_me_cvs.1.pdf (accessed 12 November 2019)
- [130] Stordeur S, Thiry N and Eyssen M. 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). 2008 Available at: https://kce.fgov.be/sites/default/files/page_documents/d20081027358.pdf (accessed 12 November 2019)
- [131] Collin SM, Crawley E, May MT, Sterne JA, Hollingworth W, UK CFS/ME National Outcomes Database. The impact of CFS/ME on employment and productivity in the UK: a cross-sectional study based on the CFS/ME national outcomes database. *BMC Health Serv Res.* 2011;11:217.