

Confirmed: The Conclusion by NICE that CBT is not an Effective Treatment for ME/CFS; Re-Analysis of a Systematic Review

Mark Vink^{1*}; Alexandra Vink-Niese²

¹Independent Researcher, 1096 HZ Amsterdam, The Netherlands. <https://orcid.org/0000-0003-1079-5647>

²Independent Researcher, 30159 Hannover, Germany. <https://orcid.org/0000-0001-8018-1419>

Corresponding Author: Mark Vink

Independent Researcher, 1096 HZ Amsterdam, The Netherlands.

Email: markvink.md@outlook.com

Article Information

Received: Sep 03, 2024

Accepted: Oct 17, 2024

Published: Oct 24, 2024

SciBase Neurology - scibasejournals.org

Vink M et al. © All rights are reserved

Citation: Vink M, Vink-Niese A. Confirmed: The Conclusion by NICE that CBT is not an Effective Treatment for ME/CFS; Re-Analysis of a Systematic Review. SciBase Neurol. 2024; 2(3): 1022.

Abstract

In this article, we analyzed the systematic review by Kuut et al. into the efficacy of cognitive behavioral therapy (CBT) for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS), a disease that predominantly affects women, and the eight trials in it. We found many issues with the studies in the review, but also with the review itself. For example, the systematic review by Kuut et al. included a researcher who was involved in seven of the eight studies in their review, and another one who was involved in five of them. Moreover, at least one of them was involved in every study in the review. On top of that, the three professors who were involved in the systematic review, have all built their career on the CB model and the reversibility of ME/CFS through CBT and GET and two of the systematic reviewers have a potential financial conflict of interest. Yet they failed to inform the readers about these conflicts of interest. Conducting a review in this manner and not informing the readers, undermines the credibility of a systematic review and its conclusion.

Regarding outcome differences between treatment and control group, it's highly likely that the combination of non-blinded trials, subjective outcomes and poorly chosen control groups, alone or together with response shift bias and/or patients filling in questionnaires in a manner to please the investigators, allegiance bias, small study effect bias and other forms of bias, produced the appearance of positive effects, despite the lack of any substantial benefit to the patients, leading to the erroneous inference of efficacy in its absence. That CBT is not an effective treatment is highlighted by the fact that patients remained severely disabled after treatment with it. The absence of objective improvement as shown by the actometer, employment status and objective cognitive measures, confirms the inefficacy of CBT for ME/CFS. The systematic review did not report on safety but research by the Oxford Brookes University shows that CBT, which contains an element of graded exercise therapy, is harmful for many patients. Finally, our reanalysis highlights the fact that researchers should not mark their own homework.

Keywords: CBT; CFS; Chronic Fatigue Syndrome; Cognitive Behavior Therapy; Myalgic Encephalomyelitis; NICE.

Introduction

Myalgic encephalomyelitis (ME), a name which goes back to the 1950s, is also known as chronic fatigue syndrome (CFS). In this day and age it's often referred to as ME/CFS. ME is characterized by an abnormally delayed muscle recovery after trivial exertion [1], which over time has evolved into post exertional malaise (PEM) [2]. Common other symptoms are muscle fatigue, which manifests itself as muscle weakness/heavy legs, myalgia, cognitive disturbances, headaches/migraines, reversal of sleep

rhythm and hypersensitivity to light and / or sound [3] leading to (severe) functional impairment [4]. Up to 25% of patients are severely or very severely affected and are homebound or bed-bound and dependent on help and care from others for even the most basic things [5]. Yet, for many years, treatment for this disease has been based on the biopsychosocial model, also known as the cognitive behavioral model (CBmodel), which is based on the assumption that there is no underlying illness, but that after a viral illness, which has been resolved, patients attribute their

symptoms to disease and they have become deconditioned as a consequence of avoidance of activity and exercise thereby producing a vicious circle which leads to further deconditioning [6,7]. Cognitive behavior therapy (CBT), which contains an element of graded exercise therapy (GET), together with GET, were designed to tackle factors assumed to maintain and perpetuate ME/CFS symptoms and associated disability as well as deconditioning and lead to recovery. According to Surawy et al. [6], who formalized that model in 1995, one of the limitations of their model is the absence of objective proof for it. A number of studies [8,9] showed that patients do not possess the behavioral characteristics targeted by the treatments of the CBmodel. Geraghty et al. published a detailed review of this model. They concluded “that the model lacks high-quality evidential support, conflicts with accounts given by most patients and fails to account for accumulating biological evidence of pathological and physiological abnormalities found in patients. There is little scientific credibility in the claim that psycho-behavioural therapies are a primary treatment for this illness” (p. 1 [10]). Thoma et al. came to a similar conclusion in 2023 in an article entitled “Why the Psychosomatic View on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Is Inconsistent with Current Evidence and Harmful to Patients” (p. 1 [11]).

The prestigious American Institute of Medicine (IOM, now the National Academy of Medicine) confirmed that in 2015 when it concluded that ME/CFS is a “complex, multisystem, and often devastating disorder” and that it is “is a medical — not a psychiatric or psychological — illness” for which there is no effective treatment (p. 1 [2]).

The National Institute for Health and Care Excellence (NICE) confirmed the conclusion by the IOM, that ME/CFS is a complex multisystem chronic medical condition for which there is no effective treatment in its updated guidelines on the diagnosis and management of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) in October 2021 [12]. It specifically stated that CBT and GET do not lead to improvement or recovery, but that CBT might be offered as a supportive therapy. A systematic review by Dutch proponents of the CBmodel however, has recently concluded “that CBT for ME/CFS can lead to significant reductions of fatigue, functional impairment, and physical limitations. There is no indication patients meeting different case definitions or reporting additional symptoms benefit less from CBT. Our findings do not support recent guidelines in which evidence from studies not mandating PEM was downgraded” (p. 1 [13]).

In this article we will review the evidence presented in the systematic review by Kuut et al. [13] to assess whether or not the aforementioned conclusion of this systematic review is justified by the data contained within the primary studies included in the review. A study by Smakowski et al. into the effectiveness of GET, which included professor Chalder, one of the world’s leading CBT and GET proponents for ME/CFS, concluded that “self-report measures...can be problematic because they are dependent on a patient’s perception of their own illness” (p. 7 [14]). Consequently, in our analysis, we will concentrate on the objective outcome measures to establish if improvements in self-report (fatigue, functional impairment, and physical limitations) translate to observable improvement in objective tests (physical ability, fitness, etc.) as there is an inverse relationship between fatigue and physical activity [15]. We also concentrated on work related outcomes, because “significant reductions of fatigue, functional impairment, and physical limitations” should

lead to a significant improvement in work status and a significant reduction in illness and benefits status. Moreover, according to, for example, one of the studies in the review (Tummers et al.), part of their treatment was “to make a plan for work resumption” (p. 6 [16]). Finally, according to Stevelink et al., which also included professor Chalder, “given the evidence that meaningful occupation is important for well-being and psychosocial needs, work-related outcomes should be targeted in CFS treatment” (p. 6 [17]).

Our analysis shows the review’s conclusion, that CBT is effective, irrespective of the case definitions, is not supported by the evidence. When the subjective outcomes of the trials are considered, it is possible to state that patients remain severely disabled after treatment with CBT. Additionally, when the objective outcomes are considered, it is possible to state that CBT is ineffective for ME/CFS.

Requirements for evidence from studies for ME/CFS to be scientifically valid

Lack of patient blinding combined with self-reporting of outcomes

All trials in the review were non-blinded by definition, yet the review only used a subjective primary outcome, fatigue severity (CIS) and two subjective secondary outcomes: functional impairment (SIP8) and physical functioning (SF-36). Even though it is well known that patient self-report is an unreliable measure [18].

As noted by Demetriou et al., “one of the main advantages of the self-report questionnaire is that it can be administered to a large sample of people quickly without much effort or financial cost” (p. 1 [19]). Moreover, questionnaires are quick and easy to administer and score [20]. Yet the lack of patient blinding in combination with self-reporting of outcomes leads to pronounced bias as patients become prone to outside influences leading to the erroneous inference of efficacy in its absence, thus making subjectively assessed outcomes unreliable [21,22]. Also, according to, for example, the BRANDO project (Bias in Randomised and Observational studies) [23], which, amongst others, included Stanford Professor Ioannidis, it is important 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 the 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” (p. 23, 45 [23]). In non-blinded studies, self-report measures are highly vulnerable to response bias, the size of which is not trivial. No such inflation was observed when objective outcome measures were used [21]. According to a systematic review by Whiting et al., there is a particular problem with subjective outcomes for patients with CFS, as they “may feel better able to cope with daily activities, because they have reduced their expectations of what they should achieve, rather than because they have made any recovery as a result of the intervention” (p. 1366 [24]). Whiting et al. continued by stating that “a more objective measure of the effect of any intervention would be whether participants have increased their working hours, returned to work or school, or increased their physical activities” (p. 1366 [24]).

Moreover, Vercoulen et al., which included one of the researchers of the systematic review (Bleijenberg), concluded

that “one has to be very careful with using self-report questionnaires as measures for actual activity level: none of the self-report questionnaires had strong correlations with the Actometer. Thus, self-report questionnaires are no perfect parallel tests for the Actometer” (p. 670 [25]). But also that “subjective instruments do not measure actual behaviour. Responses on these instruments appear to be an expression of the patients’ views about activity and may be biased by cognitions concerning illness and disability. In healthy subjects such cognitions do not exist and therefore their responses were not biased by these cognitions” (p. 670 [25]). Van der Werf et al., which also included Bleijenberg, concluded, “that self-report measures of activity and behavioural data often correlate poorly” (p. 378 [26]).

Healey et al. concluded that “caution should be taken when using self-reported PA [Physical Activity] measures. It is also important to note the wider limitations of all self-report measures i.e. potential for social desirability bias, recall bias, over and underestimation of activities/ misclassification of activities. Therefore, where possible, the use of objective measures of PA (e.g. accelerometry) should be considered. There is greater evidence of their validity and reliability and they can objectively capture all dimensions of PA” (p. 15 [27]). Consequently, “self-reported physical activity tends to overestimate the level of physical activity compared to the objective method” (p. 568 [28]).

Quinlan et al. who “investigated the associations between... self reported...and objectively measured physical activity in middle-aged adults” concluded that “in adults, self-report measures of physical activity tend to have low correlations with objective measures. Accurately measuring physical activity using self-reported tools may be difficult as individuals cannot accurately estimate the amount and type of physical activity completed in the surveyed time, or precisely report the intensity of physical activity” (p. 1, 2 [29]).

Additionally, the reanalysis of the amended Cochrane exercise review [30] found that objective outcomes from two CBT trials for ME/CFS confirmed the unreliability of the subjective outcomes in non-blinded studies, as shown by the following examples:

In Jason et al. [31], there was a substantial difference in the subjective physical functioning scores at baseline between the exercise and control groups, yet, objectively, there was not (six-minute walk test or 6MWT);

In the PACE trial, the released individual participant data showed that 20% of participants whose physical functioning improved subjectively had deteriorated objectively (6MWT) [32–34].

The German Institute for Quality and Efficiency in Healthcare (Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, IQWiG) methods guidance handbook [35] states the following about this problem. The value of patient-reported outcomes (PROs) in non-blinded studies is limited because of the subjective nature of them. (In German, “Da Angaben zu PROs aufgrund ihrer Natur subjektiv sind, sind offene, d. h. nicht verblindete Studien in diesem Bereich nur von eingeschränkter Validität” (p. 61). The same handbook also states that non-blinded studies should, as far as possible, rely on objective endpoints, because subjective ones can be influenced by the person collecting them. (In German, “Falls eine verblindete Zielgrößenerhebung nicht möglich ist, sollte ein möglichst objektiver Endpunkt gewählt werden, der in seiner Ausprägung und in der Stringenz

der Erfassung so wenig wie möglich durch diejenige Person, die den Endpunkt (unverblindet) erhebt, beeinflusst werden kann” p. 171). Finally, Lilienfeld et al. [22] concluded that non-blinded trials, and CBT studies are non-blinded by definition, should not rely on subjective primary outcomes, but use either objective primary outcomes alone, or combined with subjective ones, as a methodological safeguard against erroneous inference of efficacy in its absence.

Control groups

A key principle of a randomized controlled trial (RCT) to ensure a fair comparison is that groups should be similar with respect to all factors that might affect the outcome, besides the intervention, including the number of treatment sessions, to ensure a fair comparison [36] but, also, to make sure that an RCT is ‘internally valid’, which refers to the extent that the outcome for a trial can be attributed to the experimental treatment and not to any alternative explanation, such as the natural course of the target problem [37]. Yet, in a waitlist control group, patients do not get any treatment from (a doctor from) that study.

Additionally, “being allocated to a waiting list control group may evoke feelings of being denied support and disappointment” (p. 9 [38]) because participants expect to get some form of treatment in return for taking part in a trial “and that participants can decide to wait to attempt to change until receiving the support sought” (p. 9 [38]) resulting in participants putting less effort into change.

Also, in a no treatment, usual care, or waiting list control group (WLC), participants must attend several assessments without any direct benefit for themselves. These patients will be disappointed that they have been denied treatment benefits they anticipated from participation in a study. Assignment to those sort of control groups may strengthen participants’ beliefs that they will not improve, thereby reducing the chance of spontaneous improvement. “Direct evidence of this phenomenon was found in one exploratory trial [38], which showed that participants who rated themselves as ready to change their alcohol consumption, and who were allocated to a waiting list group, waited to reduce their drinking”. As a matter of fact, “28% of participants had markedly higher consumption at follow-up...after having joined the study hoping to reduce it” because of being allocated to a waiting list group. “Being made to wait may invite negative research participation effects” (p. 9, 10 [38]) artificially inflating the treatment effect of the treatment under investigation. Moreover, a meta analysis by Furukawa et al. entitled “waiting list may be a nocebo condition”, concluded that waiting list control groups do not control “for regression towards the mean and the natural course of the disease but instead it may introduce negative psychological expectation of ‘waiting for the desired active treatment” (p. 189 [39]). Also, using waitlist, usual care or no-treatment control conditions does not adequately correct for the placebo effect, regression to the mean and other forms of biases and confounding factors [37].

Researchers often assume that with waiting list control groups and other no-treatment control designs, the absence of treatment equates with the absence of an effect. Yet participants randomized to these designs may improve less than would be expected compared to participants not enrolled in a trial which may threaten the internal validity of a trial. Consequently, subjective baseline–follow-up differences cannot be

assumed to be the natural history of what would have occurred in the absence of patients enrolling in the study [37].

Moreover, Janse et al. [40], one of the studies in the systematic review, who used a waiting list control group, noted that “the use of a waiting-list control [group] does not control for non-specific therapy factors and limits the external validity” (p. 116). Therefore, using waitlist, usual care or no-treatment control conditions can lead to the overestimation of the effectiveness of a treatment [37].

Study size

A review of homeopathy studies by the Australian National Health and Medical Research Council (NHMRC), used a minimum number of 150 participants in randomised controlled trials (evenly distributed across the therapy and control group) because, according to the NHMRC, the results may be distorted in studies with a smaller number of participants [41]. Consequently, “trials with limited sample sizes are more likely to report larger beneficial effects than large trials” (p. 1 [42]). This is also known as the small study effect. We not only mention this review by the NHMRC because of the distortion of results if the number of participants in a study is small, but also because it was an integral part of the recent advice of the EASAC Homeopathy Working Group—The European Academies Science Advisory Council—to the EU on homeopathy [43]. Furthermore, one of its members (Van der Meer) is one of the world’s leading CBT proponents for ME/CFS and Van der Meer, Bleijenberg and Knoop are the three leading ME/CFS experts in the Netherlands. Also, the three of them are/were the leaders of the Dutch Expertise Centre for Chronic Fatigue in the Netherlands which has been promoting CBT as an effective treatment for ME/CFS for many years. Consequently, one should be careful with interpreting therapeutic responses in RCTs with less than the aforementioned minimum number of participants and those studies were excluded from the recent advice of the EASAC Homeopathy Working Group.

Response-shift bias and allegiance bias

As noted by Howard, when “using self-report instruments, researchers assume that a subject’s understanding of the standard of measurement for the dimension being assessed will not change from one testing to the next (pretest to posttest). If the standard of measurement were to change, the ratings would reflect this shift in understanding in addition to any actual changes in the subject. Consequently, comparisons of the ratings would not accurately reflect change due to treatment and would be invalid” (p. 93, 94 [44]). This “instrumentation related source of contamination is known as response-shift bias” (p. 93 [44]). This is even more of a problem when the therapy used, in this case different forms of CBT for ME/CFS, aims to modify participants’ beliefs and perception of their symptoms [45]. According to Lilienfeld et al., one of the things that can “help to eliminate response-shift biases as explanations for apparent improvement” is not relying “exclusively on self-report ratings” (p. 372 [22]).

One of the other problems with CBT and psychotherapy studies is allegiance of the researchers with the treatment. “Allegiance in psychotherapy represents the therapist’s personal belief both in the superiority and the efficacy of a particular treatment” (p. 1 [46]). A systematic review by Dragioti et al. concluded that “experimenter’s allegiance influences the effect sizes of psychotherapy RCTs and can be considered non-financial conflict of interest introducing a form of optimism bias,

especially since blinding is problematic in this kind of research” (p. 1 [46]). This is of particular interest because two of the researchers of Kuut et al. (Bleijenberg and Knoop) have based their career on the CBmodel and the efficacy of CBT for ME/CFS.

Methods

Design

A reanalysis was conducted of the systematic review by Kuut et al. and the studies in it.

Study selection criteria

Studies were only eligible if they were part of that systematic review.

Analysis of the results

We examined the characteristics of the studies in the review by paying particular attention to the size of the treatment groups and the sort of control group that was used. We also checked each study for dropouts because not only can that give valuable information about the acceptability of a treatment to participants, but at the same time, a treatment can only be deemed to be safe and effective if patients actually adhered to it.

We examined the clinical relevance of the subjective effects of CBT, by using the scoring by the systematic review on the CIS-fatigue, no longer severely fatigued, i.e. scoring <35. However, the review classified someone who scored <700 on the SIP8 as no longer functionally impaired, and someone scoring >70 on the SF-36 subscale physical functioning (SF-36 PF), as no longer severely impaired in physical functioning. In other words, participants were deemed to be recovered on those scales, even though with scores of 700 or more (SIP8) or 70 or less they were deemed to be severely functionally impaired and severely impaired in physical functioning (SF-36 PF), respectively. Yet, no impairments and severely impaired do not border on each other, instead, there is a spectrum of improvement in between. This is highlighted by the following from a study by two of the systematic reviewers (Knoop and Bleijenberg) which noted that the mean CIS-fatigue score in healthy adults (mean age of 37) is 17.3, the mean SF-36 PF score in healthy adults without a chronic condition, is 93.1 and the mean SIP8 total score of healthy women is 65.5 [47]. Consequently, we changed, no longer being functionally impaired, and no longer being impaired in physical functioning into no longer being severely functionally impaired (scoring <700 on the SIP8) and no longer being severely impaired in physical functioning (scoring >70 on the SF-36 PF).

We assessed the number of published studies that had a physical functioning entry requirement. Only studies that had a physical functioning entry requirement were used to assess this because, as concluded by Janse et al., “the fact that our study did not select on the level of physical functioning will make it more difficult to find an effect of iCBT on physical functioning” (p. 116 [40]).

To investigate the overall clinical relevance of the effect of CBT, we used objective outcomes. Not only to correct for potential biases and confounding factors caused by using subjective outcomes in non-blinded studies, or caused by using a waiting list or treatment as usual control group, but also, and more importantly, because according to the CBmodel, symptoms in ME/CFS are caused by deconditioning. Consequently, by definition,

symptoms can only be substantially less and there can only be a relevant improvement for patients, if there is a substantial improvement in physical conditioning.

According to a Dutch multidisciplinary ME/CFS guideline from 2013 which included one of the authors of the systematic review (Bleijenberg), “it is not unusual for interventions that have been proven to be effective in a research setting to perform less well in (clinical) practice. An important question is therefore whether information is available about the effectiveness of CBT for CFS and graded exercise therapy outside the confines of clinical trials” (in Dutch: “Het is niet ongebruikelijk dat interventies die in een onderzoekssetting effectief zijn gebleken, in de (klinische) praktijk minder goed presteren. Belangrijke vraag is dan ook of er informatie voorhanden is over de werkzaamheid van CGT voor CVS en graded exercise therapie in de praktijk” (p. 46 [48]). We will therefore also analyze evaluation studies that have been conducted to answer this question by paying particular attention to the objective outcomes because as noted before, there can only be a relevant improvement for patients, if there is a substantial improvement of their deconditioning (fitness).

Table 1: Characteristics of the studies in the systematic review.

Study	n	Participants	CBT	Control	Duration of treatment
Prins et al. (2001) [49]	278	Adults	Individual face-to-face CBT (n=93) Guided support (n=94)	Natural course (n=91)	8 months
Stulemeijer et al. (2005) [50]	71	Adolescents	Individual face-to-face CBT (n=36)	Waiting list (n=35)	5 months
Knoop et al. (2008) [51]	171	Adults	Guided self-instructions with email contact with a therapist (n=85)	Waiting list (n=86)	At least 16 weeks
Nijhof et al. (2012) [52]	135	Adolescents	Internet-based (n=68)	Care as usual (n=67)	6 months
Tummers et al. (2012) [16]	123	Adults	Guided self-instructions with email contact with a therapist (n=62)	Waiting list (n=61)	At least 20 weeks
Van der Schaaf et al. (2015) [53, 54]	Unpublished	Non-peer reviewed study	Unpublished study	Waiting list	6 months
Wiborg et al. (2015) [55]	204	Adults	Group face-to-face (n=136)	Waiting list (n=68)	6 months
Janse et al. (2018) [40]	240	Adults	Internet-based (n=160)	Waiting list (n=80)	6 months

NC: natural course; WL: waiting list.

As can be seen in table 2, in three studies, the dropout rate from the CBT group was substantially higher than in the control group (15% versus 5%, 19% versus 12% and 19% versus 6%), and in one study, it was much higher than in the control group (41% versus 23%). A systematic review by Whiting et al. entitled “Interventions for the Treatment and Management of Chronic Fatigue Syndrome A Systematic Review” noted that “where dropout rates are higher in the intervention group than in the control group it may be the case that there is something about the intervention that trial participants find unacceptable. It may be the method or frequency of administration, or adverse effects arising from the intervention” (p. 1367 [24]).

Finally, as concluded by psychology professor Lilienfeld in an article entitled “Why Ineffective Psychotherapies Appear to Work: A Taxonomy of Causes of Spurious Therapeutic Effectiveness”, “in contrast to clients who remain in treatment, those who drop out of treatment tend to be lower functioning” (p. 373 [22]). But also that “clients who drop out of therapy are not a random subsample of all clients. Research demonstrates that clients who are not improving are especially likely to leave psychotherapy. As a result, therapists may conclude erroneously that their treatments are effective merely because their remaining clients are those that have improved” (p. 367 [22]). He

The systematic review by Kuut et al.

Analysis of the review and the studies in it: The systematic review by Kuut et al. included eight randomized controlled studies (RCTs) of CBT for ME/CFS and “the total sample consisted of 1298 patients” [13], yet one of them is a non-published non-peer reviewed study, as can be seen in table 1. Only 640 participants (93 + 36 + 85 + 68 + 62 + 136 + 160) received some form of CBT in the remaining seven studies. All seven published studies relied on subjective primary outcomes. In two of the seven published studies, participants received individual face-to-face CBT, in one study they received group CBT, in two studies guided self-instructions with email contact with a therapist and in two others, Internet-based CBT. In two studies the participants were adolescents, in the other ones they were adults. One study used a natural course control group and one a care as usual one. The other five studies used a waiting list control group. Three of the seven published studies had less than 75 participants in the treatment group and would therefore not have been included in the earlier mentioned review of homeopathy studies.

also noted that with high levels of dropout “clients who remain in these treatments...are generally faring better than when they began, but they are unrepresentative of the clients who initially enrolled. The clients who dropped out may not have been helped or may have even been harmed by the intervention” (p. 367 [22]).

Only four of the seven published studies used one or more objective outcome measures and only one of these studies published the actometer results in (the supplement of) their original publication. Prins et al. waited nine years and Stulemeijer et al. five years before they published those results in an article by Wiborg et al. [56], which involved authors of the original two studies but also two of the authors of the systematic review (Bleijenberg and Knoop). Bleijenberg was also involved in the FITNET study by Nijhof et al. who did not publish these results at all. Instead, they stated the following in a response to comments by Kindlon and Crawford: “actual physical activity as measured by actigraphy is not likely to be the mediator of reduction in fatigue” (p. 562 [57]). Thereby acknowledging the null effect on the actometer in an indirect way. The only study that used the actometer and published its findings in supplement two of their article, was the study by Janse et al. who found an improvement, but the researchers themselves noted

in the article itself that “this might be an accidental finding, taking the amount of missing data into account” (p. 116 [40]). Moreover, as noted earlier, Janse et al. also concluded that “the fact that our study did not select on the level of physical functioning will make it more difficult to find an effect of iCBT on physical functioning” (p. 115 [40]). Or to put it differently, there was no physical functioning entry score requirement, nor was there a maximum physical functioning score. The consequence of this was that the mean physical functioning score at baseline was 62.4 with a standard deviation of 21.1 [40]. This means that patients were included in the study with a physical functioning score of 80 or more. These scores suggest that patients were already very high functioning, which suggests that their diagnosis of ME/CFS might have been incorrect. Moreover, according to supplement two of the study by Janse et al., only 16% (13/80)

adhered to the treatment in the protocol-driven feedback and only 19% (15/80) in the feedback on demand treatment group [40]. Consequently, 81% and 84%, respectively, did not adhere to treatment in the two treatment groups. According to an article by Baryakova et al. on overcoming barriers to patient adherence, “experiencing adverse effects or anxiety about potential adverse effects is a major deterrent to patient adherence” (p. 392 [58]). An improvement can only be down to a treatment if patients actually adhered to it. Or as Janse et al. themselves noted, the improvement on the actometer, “might be an accidental finding, taking the amount of missing data into account” (p. 116 [40]). Janse et al. also decided to deviate from the original study protocol by not determining quality of life scores. This might suggest that treatment did not have an effect on the quality of life.

Table 2: Results from the objective outcome measures.

Study	CBT	Objective outcome used	Objective improvement	Drop outs
Prins et al. (2001) [49]	Individual face-to-face CBT and guided support	Yes (work status, actometer and objective neuropsychological tests)	No [49,56,59]	41% (CBT, 38/93); 23% (NC, 21/91)
Stulemeijer et al. (2005) [50]	Individual face-to-face CBT	Yes (actometer)	No [56]	19% (CBT, 7/36); 6% (WL, 2/35)
Knoop et al. (2008) [51]	Guided self-instructions (GSI)	No objective outcomes were used, despite setting goals, such as returning to work.	Not used	15% (GSI, 13/85); 5% (WL, 4/86)
Nijhof et al. (2012) [52]	Internet-based (IB)	Yes (actometer)	No [57]	6% (IB, 4/68); 6% (CAU, 4/67)
Tummers et al. (2012) [16]	Guided self-instructions (GSI)	Not used	Not used	11% (GSI, 7/62); 6% (WL, 5/61)
Van der Schaaf et al. (2015) [53,54]	Unpublished study	BOLD signal as measured with fMRI; Cerebral tissue properties as measured with MRI, DTI and MR-spectroscopy.	Unpublished study	Unpublished study
Wiborg et al. (2015) [55]	Group CBT	No	No	19% (Group CBT, 26/136); 12% (WL, 8/68)
Janse et al. (2018) [40]	Internet-based (IB)	Yes (actometer)	Improvement, but “this might be an accidental finding, taking the amount of missing data into account” (p. 116 [40]).	6% (IB, 10/160); 5% (WL, 4/80); but high amount of missing data for the objective outcome.

According to the Dutch trial register, Van der Schaaf et al. [53] was a randomised control trial with a no treatment, waiting list control group. The study used the following 4 primary outcomes.

1) “Blood Oxygenation Level Dependent (BOLD) signal as measured with functional Magnetic Resonance Imaging (fMRI)

2) Cerebral tissue properties as measured with Magnetic Resonance Imaging (MRI), Diffusion Tensor Imaging (DTI) and MR-spectroscopy.

3) Behavioural performance on computerized tasks

4) Fatigue severity: Checklist individual strength (CIS) (p. 1 [53]). This study has not been published so far and consequently is a non-peer reviewed study. The fact that it has not been published despite having an anticipated start date of the 1st of January 2014 might suggest that the objective primary outcomes did not support the hypothesis of the study and contradicted the efficacy of CBT.

The objective neuropsychological tests (two reaction time tests and a symbol digit modalities task) from Prins et al. were published six years later in 2007 by authors of the original ar-

ticle, which included two of the authors of the systematic review (Bleijenberg and Knoop). The results were available for 83.8 per cent (233/278) of participants equally divided over the three groups (78 CBT; 79 support group; 76 no treatment) and showed that CBT did not lead to objective improvement [59].

According to Tummers et al., “activity patterns are usually assessed with an actometer, a small device worn around the ankle, and activity levels are assessed over a period of 12 days. However, as this was an implementation study, actometers were not available because of the high costs involved” (p. 2207 [16]). Tummers et al. is the only study in the review that acknowledged the importance of the actometer. However, according to Peters et al., “implementation research is the scientific inquiry into questions concerning implementation—the act of carrying an intention into effect, which in health research can be policies, programmes, or individual practices” (p. 1 [60]). Additionally, “implementation outcome variables describe the intentional actions to deliver services. These implementation outcome variables—acceptability, adoption, appropriateness, feasibility, fidelity, implementation cost, coverage, and sustainability—can all serve as indicators of the success of implementation” (p. 2 [60]). An implementation study “contrasts with typical randomised controlled trials that look at the efficacy of an

intervention in an “ideal” or controlled setting and with highly selected patients and standardised clinical outcomes, usually of a short term nature” (p. 3 [60]) whereas an implementation study doesn’t do that. The study by Tummers et al. “evaluated the effectiveness of guided self-instruction for CFS implemented in an MHC [a community-based mental health centre], delivered by nurses.” AND “one hundred and twenty-three patients were randomly assigned to either guided self-instruction (n = 62) or a waiting list (n=61)” (p. 2205 [16]). Consequently, it was not an implementation study, but simply a randomized controlled trial into the efficacy of an intervention, just like it said in the title of their article (“Implementing a minimal intervention for chronic fatigue syndrome in a mental health centre: a randomized controlled trial”). Consequently, the study by Tummers et al. should have used the actometer as an objective outcome measure. The same applies to the other studies, which did not

use the actometer as an objective measure of activity.

As noted earlier, the ultimate test to see if a treatment is effective or not, is the evaluation of its efficacy in real life, outside the confines of a RCT. A number of evaluation studies of the use of CBT in real life have been conducted [61-65] and as can be seen in table 3, two of those studies did not use objective outcome measures. The three others that did, showed that CBT has a negative instead of a positive effect on employment and disability/illness benefit status. Moreover, the Belgium evaluation study showed that CBT did not lead to an improvement in fitness [62].

Consequently, one cannot safely conclude that CBT is effective in view of the above named problems, the negative effect on employment status and disability benefits, and the lack of objective improvement of fitness.

Table 3: Objective outcome measures from evaluation studies.

Study	Treatment	N	Work and disability/illness benefit status	Physical capacity/fitness
Koolhaas et al. (2008) [61]	Evaluation of CBT in The Netherlands	100	41% were employed before and 31% after CBT; patients who worked, worked 5 h less after CBT	Not used
Stordeur et al. (2008) [62]	CBT and GET evaluation in Belgian CFS clinics	655	Employment status decreased from 18.3% to 14.9%; percentage of incapacitated persons increased from 54% to 57%	“Physical capacity (maximal or sub-maximal) did not change between start and end of the treatment” (p. 80)
Collin and Crawley (2017) [63]	Evaluation of CBT and GET in 11 English CFS clinics	952	After therapy: 47.2% unchanged working status; 18.0% worked again or longer; 30.0% stopped working or worked less because of CFS	Not used
Collin et al. (2018) [64]	Evaluation of the presence/absence of 5 symptoms (muscle pain, joint pain, headache, sore throat, and painful lymph nodes) which can occur in addition to the 3 symptoms (post-exertional malaise, cognitive dysfunction, and disturbed/unrefreshing sleep) that are present for almost all patients in 12 specialist CFS/ME services (11 UK, 1 NL)	918 (UK) and 1392 (Dutch)	Not used	Not used
Adamson et al. (2020) [65]	Outcomes from a specialist clinic in the UK	995	Not used	Not used

Scores after CBT in comparison to healthy people

Kuut et al. investigated the clinical relevance of the effects of CBT by using a CIS-fatigue score of <35 (no longer severely fatigued), physical functioning score of >70 (“no longer impaired in physical functioning”) and a SIP8 score of <700 (“no longer functionally impaired”) (p. 5 [13]). Most studies in the review had a fatigue severity entry requirement of 35 or more on the Checklist Individual Strength (CIS) fatigue severity sub-scale. Many of the included studies did not have an entry requirement on the Sickness Impact Profile–8 (SIP8) but in Knoop et al. [51] and Wiborg et al. [55], the entry requirement was a total score of more than 700. According to the same study by Knoop et al., “the CIS sub-scale ‘fatigue severity’ was used to measure the level of fatigue over the past 2 weeks. Scores ranged from 8 (no fatigue) to 56 (severely fatigued). The weighted total score on eight sub-scales of the SIP8 (SIP8 total score) was used to assess functional disability in all domains of functioning. Physical disabilities were measured with the physical functioning sub-scale of the 36-item Short Form Health Survey (SF–36). Scores ranged from 0 (maximum physical limitations) to 100 (ability

to do vigorous activity)” (p. 340 [51]). Additionally, in the Qure study, which investigated the efficacy of CBT and doxycycline for Q fever fatigue syndrome and which included two researchers of this systematic review (Bleijenberg and Knoop), “significant disabilities in daily functioning” was defined by a “score ≥450 on the Sickness Impact Profile [SIP8]” (p. 999 [66]).

Kuut et al. reported that the mean scores after treatments were 34.49 (fatigue severity), 920.98 (functional impairment) and 73.42 (physical functioning) [13]. Table 4 puts those into perspective by comparing them to severe impairment and scores of healthy people.

Table 4: Scores after CBT in comparison to healthy people.

Outcome	After CBT	Severe impairment	Scores of healthy people
CIS fatigue severity	34.49	35 or more	17.3
Functional impairment (SIP8 total score)	920.98	More than 700 or ≥450 in the Qure study	65.5
SF–36 physical functioning	73.42	score of 70 or below	93.1

According to a non-randomised study without a control group by Knoop et al. entitled, “Is a full recovery possible after cognitive behavioural therapy for chronic fatigue syndrome?”, which also included Bleijenberg, “healthy adults with a mean age of 37.1 have a mean score on the CIS-fatigue of 17.3”, “healthy adults without a chronic condition...[have]...a mean [physical functioning] score of 93.1” and “the mean SIP8 total score of healthy...women is 65.5” (p. 172 [47]).

Consequently, as can be seen in table 4, patients were still severely disabled after treatment labeled as effective by the systematic review.

Table 5 highlights the fact that only two of the seven published studies had a physical functioning entry requirement. Studies that did not have a physical functioning entry requirement, showed the following. In Janse et al. [40], the baseline physical functioning score was 62.4 with a standard deviation of 21.1 so that many patients already had a physical functioning score of 80 or more before receiving any treatment. In Wiborg et al. [55], this was 55.4 with a standard deviation of 18.8 so that many patients already had a physical function score of 74 or more. In Nijhof et al. [52], the physical function score at baseline was 60.7 with a standard deviation of 14.5 so that many patients already had a physical function score of 75 or more at baseline. These three studies without a physical functioning entry requirement, therefore artificially inflated the mean physical functioning score after treatment of the studies combined, which, according to the systematic review, was 73.42. Moreover, in Nijhof et al. in only 28% of cases was the onset of the disease in the treatment group after an infection, even though ME/CFS is a post-infectious disease. Consequently, that puts the diagnosis of ME/CFS in doubt in 72% of cases.

In for example Tummers et al., severely disabled, was operationalized as a SF 36 physical and/or social functioning score of 70 or below [16] and a CIS fatigue score of 35 or higher, indicated severe fatigue. In the two studies that did have a physical functioning entry requirement, the scores after treatment show that patients remained severely disabled.

Table 5: Physical functioning entry requirements and scores after CBT.

Study	Physical functioning after therapy	Entry requirement?
Prins et al. (2001) [49]	Not an outcome measure	No
Stulemeijer et al. (2005) [50]	69.4	65 or less
Knoop et al. (2008) [51]	65.9	No
Nijhof et al. (2012) [52]	88.5	No
Tummers et al. (2012) [16]	65.4	70 or less
Van der Schaaf et al. (2015) [53,54]	Non-published study	Non-published study
Wiborg et al. (2015) [55]	74.4	No
Janse et al. (2018) [40]	73.3	No

Physical functioning scores “range from 0 (maximum limitations) to 100 (no limitations)” (p. 375 [42]).

Work related outcome

In the study by Tummers et al., guided self-instruction consisted of a booklet and in chapter 9 patients were invited “to make a plan for work resumption. This plan contains the date when a patient will resume work, and how the patients will increase the number of hours worked.” And in chapter 13, “pa-

tients attain the goals as formulated in chapter 1 step by step, including resumption of work” (p. 2207 [16]). The intervention in Knoop et al. “consisted of a self-instruction booklet containing information about chronic fatigue syndrome and weekly assignments. The programme took at least 16 weeks, but often more if patients formulated long-term goals such as returning to work” (p. 340 [51]).

Why Tummers et al. and Knoop et al. did not assess work status before and after treatment, if resuming work is an important goal of the intervention, is unclear. Unless that’s because the study by Prins et al. already showed that CBT doesn’t lead to an improvement in work and employment status [49].

Conflict of interest and involvement in the studies

The systematic review was done by a group of Dutch researchers which included psychology professors Bleijenberg and Knoop. Both researchers have successfully built a career on the CBmodel and the claimed efficacy of CBT for ME/CFS. They have both become professors as a consequence of that and together with a professor of internal medicine (Van der Meer), they are the three leading experts in the Netherlands. Moreover, together with four mental health professors from the UK, they are seen as the leading worldwide experts and for many years, CBT has been promoted by guidelines all over the world based on their research. These investigators are deeply committed to the ‘unhelpful cognitions’ theory of ME/CFS, which they and other colleagues had originated and/or continue to actively promote. If their systematic review would have failed to show that CBT is an effective treatment for ME/CFS, then that would have undermined their careers.

As can be seen in table 6, Bleijenberg was involved in 7 of the 8 RCTs that were analyzed in the systematic review and Knoop was involved in 5 of them. In the world of academia, this is known as marking your own homework. However, this is not listed in the competing interest section of their article. The only competing interest of both researchers, according to the article itself, is the following: “HK [Knoop] and GB [Bleijenberg] receive royalties for a published manual of CBT for ME/CFS” (p. 8 [13]). On top of that, seven of the eight studies in this systematic review were done by the same Institute from the Netherlands (the Expert Centre for Chronic Fatigue, in Dutch, het Nederlands Kenniscentrum Chronische Vermoeidheid also known as the NCKV) and the systematic review was conducted by the leader of that Institute (Knoop). The only study in this systematic review which was conducted by researchers from a different Institute, was the FITNET study by Nijhof et al., yet Bleijenberg from the NCKV, was one of the researchers of that study as well. If the non-published study would be excluded, then Bleijenberg was involved in all seven studies of the systematic review and Knoop in four of them. Moreover, in four of these seven studies, both of them were involved as researchers and at least one of them was involved in all studies. Additionally, there is also a potential financial conflict of interest as the NCKV [67] is the main center in the Netherlands which earns money from treating ME/CFS patients with CBT. As noted, one of the systematic reviewers (psychologist Knoop) is the leader of that Institute and one of the other systematic reviewers (Kuut) is a psychologist in that center. A negative outcome of the systematic review might have meant that Dutch healthcare insurance companies would stop reimbursing CBT for ME/CFS.

Finally, the three professors who were involved in the systematic review (Bleijenberg, Moss-Morris and Knoop), have all

built their career on the CBmodel and the reversibility of ME/CFS through CBT and GET. Why none of this is mentioned as a conflict of interest, and why it wasn't brought to the attention of the readers of the article, is unclear even though the risk of latent bias was palpable from the outset.

Table 6: Involvement of the researchers of the systematic review as researchers in the studies in the review.

Study	Bleijenberg	Knoop	Other authors of the systematic review
Prins et al. (2001) [49]	Yes	No	No
Stulemeijer et al. (2005) [50]	Yes	No	No
Knoop et al. (2008) [51]	Yes	Yes	No
Nijhof et al. (2012) [52]	Yes	No	No
Tummers et al. (2012) [16]	Yes	Yes	No
Van der Schaaf et al. (2015) [53,54]	No	Yes	No
Wiborg et al. (2015) [55]	Yes	Yes	No
Janse et al. (2018) [40]	Yes	Yes	No
Total number of studies involved	7 of the 8	5 of the 8	0 of the 8

Other authors of the systematic review: Kuut, Buffart, Braamse, Csorba, Nieuwkerk, Moss-Morris, Müller.

Safety

Generally, systematic reviews focus on the efficacy or effectiveness of therapeutic interventions, however, a balanced assessment of interventions also requires the analysis of harms. According to a systematic review by Ernst and Pittler who investigated the reporting of safety aspects of therapeutic interventions by systematic reviews and meta-analyses, "information on safety is equally important for making informed, evidence based decisions on the value of a given treatment" (p. 546 [68]). Yet most systematic reviews "provide little information on the safety aspects of therapeutic interventions" (p. 546 [68]). This "may be explained, in part, because safety is not the primary aim of most reviews" (p. 133 [69]). The PRISMA harms checklist which specifically measures the reporting of harms, was published in 2016 to improve it [69]. Safety was not reported on by Kuut et al. [13]. Why they did not use the PRISMA harms checklist is unclear.

Discussion

Issues with the review and the studies in it

In this article, we analyzed the systematic review by Kuut et al. and the eight trials in it. In our analysis we found many issues with the studies in the review, but also with the review itself. For example, one of the eight included studies was a non-published, non-peer reviewed study which is not available on the Internet, and as a consequence this study should have been excluded from the review. All studies were by definition non-blinded, yet just like this systematic review, they relied on usually one subjective primary outcome. The combination of non-blinding and subjective outcomes is known to exaggerate the treatment effect of a treatment under investigation. This problem was made bigger by using badly designed control groups yet the systematic review didn't flag these things up as problematic. Of note is that the non-published study used two subjective and two objective primary outcomes. The study started in January 2014 with 60 participants in the treatment and control group each. Why the study has not been published so far, is unclear, although it might well be that the objective primary

outcomes showed that CBT did not lead to objective improvement and the researchers did not want to publish a null effect. It might also be that medical journals did not want to publish a study with negative results although in the last five years or more, more and more medical journals do want to publish studies with negative results, to counter publication bias.

Other systematic reviews

According to Kuut et al., "CBT leads to a significant and clinically relevant reduction of fatigue and functional impairment, and improvement of physical functioning in ME/CFS patients. This is in accordance with findings of one systematic review (Ingman et al., 2022) and of seven previous meta-analyses" (p. 5 [13]). They do not mention the systematic review by Ahmed et al. entitled "Assessment of the scientific rigour of randomized controlled trials on the effectiveness of cognitive behavioural therapy and graded exercise therapy for patients with myalgic encephalomyelitis/chronic fatigue syndrome: A systematic review" (p. 1 [70]). Ahmed et al. concluded that "in order to securely demonstrate the efficacy of CBT/GET within a non-blinded design, researchers need to show that self-reported improvements are supported by objectively measurable outcomes" (p. 13 [70]). Moreover, the final conclusion of Ahmed et al. was that "the findings of this systematic review do not support the claim that CBT and GET are effective treatments for ME/CFS patients" (p. 13 [70]).

Moreover, the systematic review by Ingman et al. noted among the limitations of their own review "that most studies reported subjective, self-report measures, which may have increased the risk of observer or detection bias" (p. 10 [71]) but also that "subjective and objective measures do not necessarily correlate" (p. 11 [71]). Ingman et al. go on to conclude that "results suggest some support for the positive effects of CBT and GET at short-term to medium-term follow-up although this requires further investigation given the inconsistent findings of previous reviews" (p. 1 [71]). Their conclusion about "the inconsistent findings of previous reviews" shows that it is incorrect when Kuut et al. state that more reviews have provided evidence for the efficacy of CBT and GET. Of note is that one of the researchers of Ingman et al. is Professor Chalder, one of the four leading proponents of CBT from the UK, and one of the principal investigators of the PACE trial, the largest CBT and GET trial for ME/CFS so far. Ingman et al. also concluded that "a final limitation is that treatment in most trials involved at least some face-to-face sessions, thereby excluding less mobile or housebound participants, therefore findings may not be generalizable to severe CFS" (p. 11 [70]). This is of particular interest, because Kuut et al. concluded that "patients with less severe functional impairment benefitting more [from treatment with CBT] as compared to patients with severe functional impairment" (p. 5 [13]). The (very) severely disabled, those who are home or bedbound – 25 per cent according to most estimates [5] – are unable to attend outpatient clinics and take part in those studies. Consequently, those patients with severe functional impairment as mentioned by Kuut et al., are not patients with severe ME/CFS, as noted by Ingman et al., but are in fact moderately affected ME/CFS patients, who are able to attend outpatient clinics and take part in those studies, yet have more functional impairment than the mildly affected patients in the same studies. Consequently, findings from RCTs are not generalisable to the wider CFS population and the CBT trials reviewed here are inherently biased as a consequence of that. On top of that, the FINE trial [72], the sister trial of the PACE trial, examined the

efficacy of CBT and GET for the more severely affected. They found that both treatments do not lead to objective improvements in the severely affected.

Consequently, what Kuut et al. mean is that the only patients who benefit subjectively from treatment with CBT according to their review, are the mildly affected patients. But even that is incorrect as can be seen in Table 3. The mean fatigue score of 34.5 borders on the score of severely fatigued (35 or more). The mean physical functioning score of 73.4 is less than the change on one question (one gets five points per question on this item with 20 questions) away from being severely disabled, operationalized as scoring 70 or less by for example, Tummers et al. [16]. As noted earlier, only two of the seven published studies used an entry requirement for physical functioning. One of the studies that didn't is the study by Janse et al. [40] and many patients in the study already had a physical functioning score of 80 at baseline. This not only puts the diagnosis of ME/CFS in doubt, but it also artificially inflates the physical functioning score after treatment, because these patients are hardly ill, and consequently do not have a problem exercising. Another study, in which this was a major issue, is the study by Nijhof et al. [52]. In this study, in only 28% of participants in the treatment group, their ME/CFS followed an infection, even though ME/CFS is a post-infectious disease. This is of particular importance because of the very high physical functioning score (88.5) after treatment in comparison to the other studies which will have artificially inflated the mean physical functioning score of the systematic review. This suggests that this score is a problematic outlier representing poor sampling [73]. Hawkins described an outlier as an observation that "deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism" (p. 1 [74]).

In the two studies that did have an entry requirement, the physical functioning score after treatment was less than 70. As noted by Janse et al., "the fact that our study did not select on the level of physical functioning will make it more difficult to find an effect of iCBT on physical functioning" (p. 116 [40]). Why the systematic review did not take this into account is unclear, even more so because two of the systematic reviewers (Bleijenberg and Knoop) were part of that study. Why the systematic review didn't remove the outlier score from Nijhof et al. is also unclear.

After treatment deemed to be effective by the systematic review, the mean SIP8 score of 921 is far away from the score from no longer being functionally impaired (less than 700 or less than 450 in a study by the same researchers for Q fever fatigue syndrome [66]. All three scores (fatigue, physical functioning and SIP8) are a long way away from healthy individuals of the same age as highlighted by table 4, even though the basis of the CBmodel as an explanation for the symptoms in ME/CFS, is that there is no underlying disease and that symptoms are a result of avoidance of exercise and deconditioning. CBT, but also GET, are designed to reverse this and by doing so lead to recovery. Yet after treatment deemed to be effective by Kuut et al., patients remain severely disabled. Moreover, the aforementioned systematic review by Ahmed et al. concluded that "the methodological quality of the 18 included studies was found to be relatively low, as bias was prominently found, affecting the main outcome measures of the studies (fatigue, physical functioning and functional impairment/status)" (p. 1 [70]).

Objective outcomes

Only four studies in the review used an objective outcome (actometer) yet three of these studies did not publish those results in their initial publications. Two studies published them five and, respectively, nine years later and the results showed that CBT does not lead to objective improvement in adults and adolescents, respectively. The third study did not publish their objective outcome (actometer) at all, but acknowledged later in a comment that Internet-based CBT does not lead to objective improvement. The fourth study concluded that the improvement in their study according to the actometer, was artificially caused by a high number of missing data on this outcome.

One study also used two other objective outcomes (an objective outcome of cognitive functioning and employment status). Both these objective outcomes showed no objective improvement either. Consequently, the objective outcome measures confirm the inefficacy of CBT for ME/CFS. Of note is that three of the four studies that used objective outcomes, used selective reporting to avoid publishing a null effect.

Vercoulen et al., which included one of the authors of the systematic review (Bleijenberg), concluded that "the actual level of physical activity was related to fatigue severity", and that "fatigue severity was related to the Actometer" (p. 671 [25]). But also that "the CIS [fatigue severity] also is a self-report questionnaire requiring a general subjective interpretation. Thus, responses on these instruments are susceptible to the same biases" (p. 671 [25]). In view of that, it is unclear why the systematic review by Kuut et al. ignored the actometer results but also why they ignored the objective outcomes in general. Even more so because according to Kuut et al., "in all included trials, a specific CBT protocol was used. All CBT protocols for ME/CFS are based on the cognitive-behavioral model of fatigue assuming that cognitive-behavioral factors perpetuate fatigue and associated disability. They generally focus on similar perpetuating factors. Further, all CBT protocols for ME/CFS contain graded exposure to activity, a central element of the intervention" (p. 6 [13]). In most studies, the treatment lasted six months or more and if there is no underlying physical disease, then one would expect a substantial improvement in fitness and activity after such a time frame. Yet the objective outcome measures showed that a graded exposure to activity doesn't lead to objective improvement. By definition of the CBmodel, if deconditioning, i.e. the level of fitness of patients did not change, then symptoms cannot have changed either. Consequently, the subjective improvement after CBT is an artifact caused by all the problems with the design of the studies and the different forms of bias. It also shows that CBT is not the right treatment for ME/CFS.

High risk of bias

Kuut et al. acknowledge that a limitation of their review is that "none of the included studies were rated as having a low risk of bias. All included studies used patient-reported outcomes, namely subjectively experienced symptoms and functional impairments." They then downplayed the relevance of this by stating the following. "All case definitions of ME/CFS rely on reports of patients of subjectively experienced symptoms. Therefore the efficacy of interventions aimed at symptoms of ME/CFS can only be determined with patient-reported outcome measures" (p. 6, 8 [13]). Even though they also acknowledge that "in the Cochrane risk of bias tool studies are penalized if the outcome assessor (the patient) was aware of the intervention received. However, this limitation is inherent to the evaluation of behavioral/psychotherapeutic interventions using a subjective outcome measure" (p. 8 [13]). Yet this

is puzzling for a number of reasons. First of all, the basis of the CBmodel as an explanation for ME/CFS is that patients interpret their symptoms incorrectly. It is illogical and unscientific to then rely on patients interpreting their symptoms by using subjective outcomes because that would imply that patients are interpreting their symptoms wrongly but at the same time they are also interpreting the same symptoms rightly. Moreover, CBT studies are non-blinded by definition, and relying on subjective outcomes in those studies is very prone to all sorts of different forms of bias. Consequently, one could get the impression of efficacy, even though patients have not benefited from the treatment. An additional problem of the CBT studies is that patients are instructed to interpret their symptoms differently. A subjective improvement could then simply be caused by answering questionnaires in a different way. This is also known as response shift bias, and the only way to correct for this is by using objective outcomes. Moreover, the GETSET study by White, one of the world's leading CBT and GET proponents for ME/CFS, noted the following about using subjective outcomes: "All outcomes were self-rated, which might lead to bias by expectation." But also that "objective outcomes, such as actigraphy,...might have tested the validity of our self-rated measures of physical activity" (p. 10 [75]). Finally, if Kuut et al. want to rely on patient reported outcomes, then why do they ignore that patients have been saying for a long time that CBT is ineffective?

Badly designed control groups

Six of the eight studies used a waiting list control group, even though the basic requirement of a properly conducted randomized controlled study is that the only difference between the treatment and control group is the treatment. Interestingly enough, this problem was documented by one of the studies in the meta analysis as long ago as in 2008 (Knoop et al., which included two of the authors of the meta analysis, Knoop and Bleijenberg) when they noted that "as we did not use a control condition [they used a waiting list control group] we cannot be sure that the specific elements in the minimal intervention condition were responsible for the reduction of fatigue and disabilities" (p. 341 [51]).

Moreover, a meta analysis by Furukawa et al. entitled "waiting list may be a placebo condition", concluded that "the effect size estimates for CBT [in the treatment of depression] were substantively different, depending on the control condition" (p. 181 [39]). Also, that "in individual trials of psychotherapy, the use of WL [a waiting list control group] as control should be more carefully deliberated, as it "doesn't control "for regression towards the mean and the natural course of the disease but instead it may introduce negative psychological expectation of 'waiting for the desired active treatment'" (p. 189 [39]) which artificially inflates the treatment effect. According to the same meta analysis by Furukawa et al., "the apparent existence of small study effects is another major threat" (p. 189 [39]) which affected three of the seven studies, or four of the eight studies, when the non-published, non-peer reviewed study is included, which all had less than 75 participants in the treatment group.

Work rehabilitation

The study by Prins et al. noted that "the final goal of CBT for CFS included work rehabilitation for patients who used to be active in a job" (p. 846 [49]). They tried to downplay the null effect on employment by stating that "only 33% had a job at baseline, whereas 76% had been employed before the onset of CFS. For the unemployed patients, securing employment within

the limited period of treatment and follow-up would be difficult" (p. 846 [49]). Yet treatment finished after eight months and follow-up was after 14 months. Consequently, participants had at least six months to find work but if the improvement on the subjective outcomes had been a real improvement, then patients were already improving in the last few months of treatment and therefore would have had most likely two, three or four extra months during the last months of treatment to start and look for a job. In reality, they would then have had six to ten months to look for a job. One would assume that many people aged 36.2, the mean age in Prins et al., should be able to find a job in that sort of timeframe if the "clinically significant improvement...in fatigue severity" which occurred in 35% of cases, in Karnofsky performance status (49%), and self-rated improvement (50%)" (p. 841 [49]) as reported by Prins et al., were real improvements and not artifacts.

Important other limitations

Tummers et al. noted a few important limitations of their study, which might also apply to the other studies in the systematic review. During the study, the authors found out that 10% of patients (12/123) in their study were misdiagnosed and therefore, "were wrongfully included in this study" yet they "were not excluded from data analyses" (p. 2213 [16]) even though patients who do not have the disease under investigation, should be excluded from a study and its data analyses. Most other studies did not record misdiagnosing.

The second limitation of their study, as noted by Tummers et al. themselves, was that the "assessment of the physical activity patterns of patients was not based on actometer scores, a valid and reliable method to determine the activity pattern" (p. 2213 [16]). Moreover, "treatment adherence...was not assessed" (p. 2213 [16]). This was not assessed by most of the other studies in the systematic review with the notable exception of Janse et al.. According to supplement number two of this study, only 16% (13/80) adhered to the treatment in the protocol-driven feedback and only 19% (15/80) in the feedback on demand treatment group [40]. Consequently, 81% and 84%, did not adhere to treatment. Yet one cannot conclude that a treatment is safe and effective if one does not know if patients have actually adhered to the treatment or not, or if the large majority of them has not adhered to the treatment as was the case in Janse et al..

Safety

An important part of a systematic review is also reporting about the safety of a treatment. Kuut et al. however, did not report about the safety of CBT yet it is unclear why they didn't use the PRISMA harms checklist to do so. Although, as noted by the aforementioned systematic review by Ahmed et al., "not reporting adverse events is typical for this field as psychotherapy trials generally report infrequently on adverse outcomes" (p. 13 [70]).

According to a systematic review by Ernst and Pittler, the reporting of safety aspects of therapeutic interventions by systematic reviews and meta-analyses, is poor. But also that randomized clinical trials typically assess only a small number of patients which limits the chances of detecting adverse events. Therefore, they recommend that the "assessment of safety has to go far beyond randomized clinical trials" (p. 546 [68]). One of those ways is the reporting by patients of adverse events due

to a particular treatment in real life. Patients have been saying since as early as 1990 that CBT and GET are harmful [76]. Why Kuut et al. ignored this, especially as they emphasize the importance of patient reported outcome measures, and their systematic review relies on them, is unclear.

The British National Institute for Health and Clinical Excellence (NICE) published its updated ME/CFS guidelines in October 2021 [12]. As part of that review process, it commissioned the Oxford Brookes University to carry out a survey amongst ME/CFS patients (n = 2274) on the safety of CBT and GET. The Oxford Brookes University published its report in February 2019 [77]. In it, they concluded 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 ME/CFS symptoms after treatment with CBT, which contains an element of GET, was reported by 58.3% and the percentage of severely affected patients increased from 12.6% to 26.6%. Or to put it differently, 14% of patients were made homebound or bedridden by CBT.

Finally, in an interview on Dutch TV [78], pediatrician Terheggen-Lagro and medical psychologist Oostrom from the Amsterdam UMC (University Medical Centre), a leading hospital in the Netherlands, stated in January 2024, that they started to treat children with long Covid with CBT as designed for ME/CFS, which contains an element of graded exercise therapy (GET), and GET. However, they noted that instead of helping patients recover, it made them worse, just like in ME/CFS. Consequently, they stopped using these treatments.

NICE updated ME/CFS guideline

NICE published its updated ME/CFS guideline in October 2021 [12], in which it concluded that CBT does not lead to improvement or recovery and that it should only be used as an adjunctive treatment if patients need it. Two of the researchers of the systematic review have built their career on the CBmodel and the efficacy of CBT for ME/CFS and one of them (Knoop) has already co-written two articles to criticize the new NICE guidelines and its processes. One of those is entitled “Anomalies in the review process and interpretation of the evidence in the NICE guideline for chronic fatigue syndrome and myalgic encephalomyelitis” and it was published in a journal in which one of the co-authors is a paid associate editor according to the competing interests section of the article [79]. The other one entitled “New NICE guideline on chronic fatigue syndrome: more ideology than science?” was published in *The Lancet* [80]. Our analysis of this article found that it was based on rhetoric and ideology, instead of evidence based science [81]. This all gives the impression that the systematic review was the third article in a row, to counter the conclusion by NICE because the authors of those three articles cannot accept the pragmatic shift away from the opinion and eminence based biopsychosocial and cognitive behavioral model. This seems to be a typical example of the following noted by Ioannidis. “Investigators working in any field are likely to resist accepting that the whole field in which they have spent their careers is a “null field.”” (p. 0700 [82]).

Additionally, the systematic review by Kuut et al. included studies from 2001 till 2018 and all those studies were already known to NICE and included in their extensive analysis of the CBT literature. This showed that all the CBT trials included in the review process to update the NICE ME/CFS guideline, were

of low or very low quality [12,83]. Our analysis of the systematic review by Kuut et al. and all the studies in it, highlighted the extensive number of problems with those studies and confirmed the conclusion by NICE that none of those studies were of higher quality.

Non-reported conflict of interest

A potentially important limitation of the systematic review is that one of its authors (Bleijenberg) was involved in seven of the eight RCTs that were analyzed and another one (Knoop) was involved in five of them. After excluding the non-published non-peer reviewed study, then one of the authors (Bleijenberg) was involved in all of the seven studies, and the other one (Knoop) in four of the seven. Or to put it differently, at least one of them was involved in all seven studies and both of them were investigators in four of the seven studies. In the world of academia, this is known as marking your own homework which might have influenced the review and its outcome [84]. Also, two of the systematic reviewers might have financial interests in the outcome of their review and three professors (Bleijenberg, Moss-Morris and Knoop) who were involved in the systematic review, have all built their career on the CBmodel and the reversibility of ME/CFS through CBT and GET. As noted by Dana and Loewenstein [85], when individuals have a stake in reaching a particular conclusion, they weigh arguments in a biased fashion that favors a specific conclusion. Moreover, as noted by Groopman [86], scientists often ignore what they don't want to see and seek confirmation of what they believe. Consequently, the three professors had a preexisting professional stake in the outcome of their systematic review and their impartiality might reasonably be questioned. If the systematic review had failed to show that CBT/GET is an effective treatment for ME/CFS, then that would have undermined the CBmodel and the theories of reversibility to which three of the systematic reviewers have dedicated their careers. In conclusion, some of the systematic reviewers have an interest that could be affected substantially by the outcome of the systematic review, whether financially or otherwise, yet, none of these conflicts of interests are listed in the competing interest section of their article, nor is it mentioned in the abstract, the discussion and the conclusion. As a consequence, most readers will not be aware of these.

According to the publication ethics section of the journal (*Psychological Medicine*) that the systematic review was published in, “competing interests are situations that could be perceived to exert an undue influence on the presentation, review or publication of a piece of work” (p. 1 [84]). Moreover, according to, for example, a leading medical journal (the *British Medical Journal*, usually referred to as the *BMJ*), their journal “should know about any competing interests that authors may have, and that if we publish the article readers should know about them too” (p. 1 [87]) yet, the researchers did not inform their readers.

A systematic review by Dragioti et al. not only concluded that “experimenter's allegiance influences the effect sizes of psychotherapy RCTs and can be considered non-financial conflict of interest introducing a form of optimism bias, especially since blinding is problematic in this kind of research” [46]. But also that “reported effect sizes were found to be larger by almost 30% when the allegiant therapist had participated in the respective RCT compared to studies in which he was not included in the authorship list” (p. 1 [46]). This is of particular importance because Bleijenberg and Knoop have built their career on the CBmodel and the efficacy of CBT for ME/CFS and Bleijenberg

participated in seven of the eight included RCTs in the systematic review and Knoop in five.

Finally, in March 2009, “the World Association of Medical Editors revised its policy on competing interests to emphasize nonfinancial competing interests, including academic commitments, personal relationships, political or religious beliefs and institutional affiliations” (p. 11 [88]). As a consequence, the Canadian Medical Association Journal changed its policy accordingly, and in an editorial it emphasized that “full disclosure [of competing interest] is particularly important for authors of commentaries, editorials and *review articles* [italic by us]. Because such articles often offer explicit guidance, readers expect a stronger guarantee of integrity” (p. 12 [88]). Why the researchers of the systematic review ignored all that, is unclear.

Checklist of recommendations on how to design and conduct treatment studies for ME/CFS

There are no effective treatments for post infectious diseases like ME/CFS and long Covid because the medical profession has been psychologising those diseases for decades. If we had taken them seriously, as we should have, then by now we would have had effective pharmacological treatments for the estimated 400 million people with long Covid [89] and the medical profession would not be clutching at straws on how to treat those patients. The same applies to the estimated 17 to 24 million people with ME/CFS [90]. As noted by Rohrhofer et al., “given the substantial health and socioeconomic burden associated with ME/CFS, urgent attention and research efforts are needed to define causative treatment approaches” (p. 1 [91]).

Treatment studies to define and find causative treatment approaches need to adhere to the highest standards of clinical research to be able to find effective pharmacological treatments for post infectious diseases. Requirements for those studies, which will allow for reliable and valid conclusions, are:

- The research population needs to be homogenous;
- There need to be at least 75 participants in the treatment group and a similar number in the control group unless it's a phase one or feasibility study [41,42];
- The Oxford criteria [92], that do not require the main characteristic of the disease (PEM) for diagnosis, or the Fukuda criteria [93], according to which the main characteristic is optional, should not be used anymore;
- Studies should use the International Consensus Criteria [94];
- Primary outcomes should be objective (i.e. the steptest, six minute walk test, actometer, hours worked, as discussed earlier) [17,22,24-27,35];
- One additional subjective primary outcome (quality of life scores) could be used [95];
- Fatigue should not be used as a primary outcome because a subjective “improvement” of a few points on a fatigue scale, is not relevant to patients, even more so because many ME/CFS patients do not suffer from fatigue;
- Studies of psychological treatments and gradually increasing activity should not be performed or financed anymore, to stop wasting time, money and resources and to prevent harming more patients;

- Control groups should be properly matched and designed as discussed earlier;
- Studies should not use a wait list, no treatment or usual care, or other control groups where patients do not receive a control treatment with the same frequency, duration, expectation of improvement, et cetera from the study as people in the treatment group [36-40];
- Patients and/or carers should be involved in the design and conduct of these studies;
- Definitions of recovery should be based on objective outcomes;
- Non-blinded studies that deem treatments to be effective, based on subjective outcomes, should be rejected for publication;
- Misdiagnosed patients should be removed from a study and its analysis;
- Future trials are particularly needed for individuals with severe and very severe ME/CFS because they are devoid of any medical care, despite being severely ill;
- Studies that resort to selective reporting of objective outcomes should be rejected by journals;
- Studies that make extensive outpatient changes, use a post hoc definition of recovery, have an overlap in entry and recovery criteria, label the severely ill as recovered, et cetera, should also be rejected by journals.

Conclusion

In summary, the systematic review by Kuut et al. included a researcher who was involved in seven of the eight studies in their review, and another one who was involved in five of them. Moreover, at least one of them was involved in every study. On top of that, the three professors who were involved in the systematic review, have all built their career on the CBmodel and the reversibility of ME/CFS through CBT and GET and two of the systematic reviewers have a potential financial conflict of interest. Yet they failed to inform the readers about these conflicts of interest. Conducting a review in this manner and not informing the readers, undermines the credibility of a systematic review and its conclusion. Regarding outcome differences between treatment and control group, it's highly likely that the combination of non-blinded trials, subjective outcomes and poorly chosen control groups, alone or together with response shift bias and/or patients filling in questionnaires in a manner to please the investigators, allegiance bias, small study effect bias and other forms of bias, produced the appearance of positive effects, despite the lack of any substantial benefit to the patients, leading to the erroneous inference of efficacy in its absence. That CBT is not an effective treatment is highlighted by the fact that patients remain severely disabled after treatment with it. The absence of objective improvement as shown by the actometer, employment and disability benefit status, and objective cognitive measures, confirms the inefficacy of CBT for ME/CFS. Kuut et al. did not report about the safety of CBT. The Oxford Brookes University did, and their research shows that CBT, which contains an element of graded exercise therapy, is harmful for many patients. Finally, our reanalysis highlights the fact that researchers should not mark their own homework.

Declarations

Author contributions: Conceptualization, M.V.; methodology, M.V. and A.V.-N.; validation, M.V. and A.V.-N.; writing—original draft preparation, M.V.; writing—review and editing, M.V. and A.V.-N.; and supervision, M.V. and A.V.-N. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional review board statement: Not applicable.

Informed consent statement: Not applicable.

Data availability statement: Not applicable.

Acknowledgments: Not applicable.

Conflicts of interest: The authors declare that they have no conflict of interest.

References

1. Ramsay MA. Myalgic Encephalomyelitis and Post Viral Fatigue States: The Saga of Royal Free Disease, 2nd ed, Gower Medical Publishing: London, UK, 1988.
2. Committee on the Diagnostic Criteria for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome; Board on the Health of Select Populations; Institute of Medicine. Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Redefining an Illness. Washington (DC): National Academies Press (US); 2015. PMID: 25695122. <https://pubmed.ncbi.nlm.nih.gov/25695122/> (accessed on 30 October 2024).
3. Dowsett EG, Ramsay AM, McCartney RA, Bell EJ. Myalgic encephalomyelitis—a persistent enteroviral infection? *Postgrad Med J.* 1990; 66(777): 526–30. doi: 10.1136/pgmj.66.777.526.
4. Reeves WC, Lloyd A, Vernon SD, Klimas N, Jason LA, Bleijenberg G, et al. International Chronic Fatigue Syndrome Study Group. Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. *BMC Health Serv Res.* 2003; 3(1): 25. doi: 10.1186/1472-6963-3-25.
5. Pendergrast, T, Brown, A, Sunnquist, M, tke, R, Newton, J.L, Strand, E.B, Jason, L.A. Housebound versus nonhousebound patients with myalgic encephalomyelitis and chronic fatigue syndrome. *Chronic Illn.* 2016; 12: 292–307.
6. Surawy, C, Hackmann, A, Hawton, K, Sharpe, M. Chronic Fatigue Syndrome: A cognitive approach. *Behav. Res. Ther.* 1995; 33: 535–544.
7. Vercoulen JH, Swanink CM, Galama JM, Fennis JF, Jongen PJ, Hommes OR, et al. The persistence of fatigue in chronic fatigue syndrome and multiple sclerosis: development of a model. *J Psychosom Res.* 1998; 45(6): 507–17. doi: 10.1016/s0022-3999(98)00023-3.
8. Song S, Jason LA. A population-based study of Chronic Fatigue Syndrome (CFS) experienced in differing patient groups: An effort to replicate Vercoulen et al.'s model of CFS. *J. Ment. Health.* 2005; 14: 277–289.
9. Sunnquist M, Jason LA. A reexamination of the cognitive behavioral model of Chronic Fatigue Syndrome. *J. Clin. Psychol.* 2018; 74: 1234–1245.
10. Geraghty, K, Jason, L.A, Sunnquist, M, Tuller, D, Blease, C, Adeniji, C. The 'cognitive behavioural model' of Chronic Fatigue Syndrome: Critique of a flawed model. *Health Psychol. Open.* 2019; 6: 2055102919838907.
11. Thoma, M, Froehlich, L, Hatteso, D.B.R, Quante, S, Jason, L.A, Scheibenbogen, C. Why the Psychosomatic View on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Is Inconsistent with Current Evidence and Harmful to Patients. *Medicina* 2024; 60: 83. <https://doi.org/10.3390/medicina60010083>
12. NICE. MyalgicEncephalomyelitis (or Encephalopathy)/Chronic Fatigue Syndrome: Diagnosis and Management. NICEGuideline [NG206]. 29 ober 2021. Available online: <https://www.nice.org.uk/guidance/ng206> (accessed on 30 October 2024).
13. Kuut TA, Buffart LM, Braamse AMJ, Csorba I, Bleijenberg G, Nieuwkerk P, et al. Does the effect of cognitive behavior therapy for chronic fatigue syndrome (ME/CFS) vary by patient characteristics? A systematic review and individual patient data meta-analysis. *Psychol Med.* 2023. 6: 1–10. doi: 10.1017/S0033291723003148.
14. Smakowski, A, Adamson, J, Turner, T, Chalder, T. Graded exercise therapy for patients with chronic fatigue syndrome in secondary care—A benchking study. *Disabil Rehabil.* 2022; 44: 5878–5886.
15. Rongen-van Dartel SA, Repping-Wuts H, van Hoogmoed D, Knoop H, Bleijenberg G, van Riel PL, Fransen 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. doi: 10.1002/acr.22251.
16. Tummers M, Knoop H, van Dam A, Bleijenberg G. Implementing a minimal intervention for chronic fatigue syndrome in a mental health centre: a randomized controlled trial. *Psychol Med.* 2012; 42(10): 2205–15. doi: 10.1017/S0033291712000232. Epub 2012. 21.
17. Stevelink SAM, k KM, Fear NT, Hotopf M, Chalder T. Chronic fatigue syndrome and occupational status: a retrospective longitudinal study. *Occup Med (Lond).* 2022; 72(3): 177–183. doi: 10.1093/occmed/kqab170.
18. Wechsler ME, Kelley JM, Boyd IO, et al. Active albuterol or placebo, sham acupuncture, or no intervention in asthma. *New England Journal of Medicine.* 2011; 365(2): 119–126.
19. Demetriou C, Ozer B.U., Essau C.A. Self-Report Questionnaires, The Encyclopedia of Clinical Psychology. 23 uary 2015. <https://doi.org/10.1002/9781118625392.wbecp507>
20. Healey EL, Allen KD, Bennell K, Bowden JL, Quicke JG, Smith R. Self-Report Measures of Physical Activity. *Arthritis Care Res (Hoboken).* 2020; 72 Suppl 10(Suppl 10): 717–730. doi: 10.1002/acr.24211.
21. Hróbjartsson A, Emanuelsson F, Thomsen AS, et al. Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. *International Journal of Epidemiology.* 2014; 43(4): 1272–1283.
22. Lilienfeld SO, Ritschel LA, Lynn SJ, et al. Why ineffective psychotherapies appear to work: A taxonomy of causes of spurious therapeutic effectiveness. *Perspectives on Psychological Science.* 2014; 9: 355–387
23. Savović J, Jones H, Altman D, Harris R, Jūni 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. doi: 10.3310/hta16350.
24. Whiting P, Bagnall AM, Sowden AJ, Cornell JE, Mulrow CD, Ramírez G. Interventions for the treatment and management of

- chronic fatigue syndrome: a systematic review. *JAMA*. 2001 Sep 19; 286(11): 1360-8. doi: 10.1001/jama.286.11.1360. Erratum in: *JAMA* 2002; 287(11): 1401.
25. Vercoulen JH, Bazelmans E, Swanink CM, Fennis JF, Galama JM, Jongen PJ, Hommes O, van der Meer JW, Bleijenberg G. Physical activity in chronic fatigue syndrome: assessment and its role in fatigue. *J Psychiatr Res* 1997; 31(6): 661±73.
 26. van der Werf SP, Prins JB, Vercoulen JH, van der Meer JW, Bleijenberg G. Identifying physical activity patterns in chronic fatigue syndrome using actigraphic assessment. *J Psychosom Res*. 2000; 49(5): 373-9. doi: 10.1016/s0022-3999(00)00197-5.
 27. Healey EL, Allen KD, Bennell K, Bowden JL, Quicke JG, Smith R. Self-Report Measures of Physical Activity. *Arthritis Care Res (Hoboken)*. 2020; 72 Suppl 10(Suppl 10): 717-730. doi: 10.1002/acr.24211.
 28. Luo J, Lee RYW. Opposing patterns in self-reported and measured physical activity levels in middle-aged adults. *Eur J Ageing*. 2021; 19(3): 567-573. doi: 10.1007/s10433-021-00657-z.
 29. Quinlan C, Rattray B, Pryor D, Northey JM, Anstey KJ, Butterworth P, Cherbuin N. The accuracy of self-reported physical activity questionnaires varies with sex and body mass index. *PLoS One*. 2021 11; 16(8): e0256008. doi: 10.1371/journal.pone.0256008.
 30. Vink M, Vink-Niese F. Graded exercise therapy does not restore the ability to work in ME/CFS - Rethinking of a Cochrane review. *Work*. 2020; 66(2): 283-308. doi: 10.3233/WOR-203174.
 31. Jason, L, Torres-Harding, S, Friedberg, F, Corradi, K, Njoku, M.G, Donalek, J, Reynolds, N, Brown, M, Weitner, B.B, Rademaker, A, et al. Nonpharmacologic interventions for CFS: A randomized trial. *J. Clin. Psychol. Med. Settings*. 2007; 172: 485–490.
 32. FOIA Request to QMUL. Dataset File. FOIA Request 2014/F73a. 2016. Available online: https://huisartsvink.files.wordpress.com/2022/03/pace-ipd_foia-qmul-2014-f731.pdf (accessed on 30 October 2024).
 33. White, P.D, Goldsmith, K.A, Johnson, A.L, Potts, L, Walwyn, R, esare, J.C, Baber, H.L, Burgess, M, Clark, L.V, Cox, D.L, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): A randomised trial. *Lancet*. 2011; 377: 823–836.
 34. FOIA Request to QMUL.A Selection of Individual Participant Data from the PACE Trial for Chronic Fatigue Syndrome. Readme File. FOIA Request 2014/F73b. 2016. Available online: <https://huisartsvink.files.wordpress.com/2022/03/pace-ipd-readme.pdf> (accessed on 30 October 2024).
 35. IQWiG Methods Guidance Handbook (Allgemeine Methoden Version 6. 1 vom 24.01.2022). Available online: <https://www.iqwig.de/methoden/allgemeine-methoden-v6-1.pdf> (accessed on 30 October 2024).
 36. An, M.W, Duong, Q, Le-Rademacher, J, Mandrekar, S.J. Principles of Good Clinical Trial Design. *J. Thorac. Oncol*. 2020; 15: 1277–1280.
 37. Mohr, D.C, Spring, B, Freedland, K.E, Beckner, V, Areal, P, Holton, S.D, Ockene, J, Kaplan, R. The Selection and Design of Control Conditions for Randomized Controlled Trials of Psychological Interventions. *Psychother. Psychosom*. 2009; 78: 275–284.
 38. Gunnarsson KU, McCambridge J, Bendtsen M. Reactions to being allocated to a waiting list control group in a digital alcohol intervention trial. *Patient Educ Couns*. 2023; 107: 107572. doi: 10.1016/j.pec.2022.11.014. Epub 2022 18.
 39. Furukawa TA, Noma H, Caldwell DM, Honyashiki M, Shinohara K, Imai H, Chen P, Hunot V, Churchill R. Waiting list be a nocebo condition in psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatr Scand*. 2014; 130(3): 181-92. doi: 10.1111/acps.12275. Epub 2014 4.
 40. se, A, Worm-Smeitink, M, Bleijenberg, G, Donders, R, Knoop, H. Efficacy of web-based cognitive-behavioural therapy for chronic fatigue syndrome: Randomised controlled trial. *Br. J. Psychiatry*. 2018; 212: 112–118.
 41. National Health and Medical Research Council. NHMRC Information Paper: Evidence on the Effectiveness of Homeopathy for Treating Health Conditions; National Health and Medical Research Council: Canberra, Australia, 2015.
 42. Zhang Z, Xu X, Ni H. Small studies overestimate the effect sizes in critical care meta-analyses: a meta-epidemiological study. *Crit Care*. 2013; 17(1): R2. doi: 10.1186/cc11919.
 43. Fears, R, Griffin, G, Larham, D, Ter Meulen, V, Van Der Meer, J.W.M. Assessing and regulating homeopathic products. *J. Intern. Med*. 2017; 282: 563–565.
 44. Howard GS. Response-Shift Bias: A Problem in Evaluating Interventions with Pre/Post Self-Reports. *Evaluation Review*. 1980; 4(1): 93-106. <https://doi.org/10.1177/0193841X8000400105>
 45. Ghatineh, S, Vink, M. FITNET's Internet-Based Cognitive Behavioural Therapy Is Ineffective and Impede Natural Recovery in Adolescents with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. A review. *Behav. Sci*. 2017; 7: 52.
 46. 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. doi: 10.1186/s12991-015-0063-1.
 47. Knoop H, Bleijenberg G, Gielissen MF, van der Meer JW, White PD. Is a full recovery possible after cognitive behavioural therapy for chronic fatigue syndrome? *Psychother Psychosom*. 2007; 76(3): 171-6. doi: 10.1159/000099844.
 48. CBO. Richtlijn Diagnose, behandeling, begeleiding en beoordeling van patiënten met het chronisch vermoeidheidssyndroom (CVS). <https://huisartsvink.wordpress.com/wp-content/uploads/2024/11/cbo-richtlijn-chronisch-vermoeidheidssyndroom-cvs.pdf> (accessed 30 October 2024)
 49. Prins JB, Bleijenberg G, Bazelmans E, Elving LD, de Boo TM, Sevens JL, van der Wilt GJ, Spinhoven P, van der Meer JW. Cognitive behaviour therapy for chronic fatigue syndrome: a multicentre randomised controlled trial. *Lancet*. 2001; 357(9259): 841-7. doi: 10.1016/S0140-6736(00)04198-2.
 50. 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 1; 330(7481): 14. doi: 10.1136/bmj.38301.587106.63. Epub 2004 7. Erratum in: *BMJ*. 2005 9; 330(7495): 820.
 51. Knoop H, van der Meer JW, Bleijenberg G. Guided self-instructions for people with chronic fatigue syndrome: randomised controlled trial. *Br J Psychiatry*. 2008; 193(4): 340-1. doi: 10.1192/bjp.bp.108.051292.
 52. Nijhof SL, Bleijenberg G, Uiterwaal CS, Kimpfen JL, van de Putte EM. Effectiveness of internet-based cognitive behavioural treatment for adolescents with chronic fatigue syndrome (FITNET): a randomised controlled trial. *Lancet*. 2012; 379(9824): 1412-8. doi: 10.1016/S0140-6736(12)60025-7. Epub 2012 3.
 53. Van der Schaaf, M. Neural correlates of chronic fatigue syndrome. Dutch Trial Register Published: 09-12-2013. <https://>

- www.onderzoekmetmensen.nl/en/node/20349/pdf (accessed on 30 October 2024).
54. van Der Schaaf ME, Schmits IC, Roerink M, Geurts DE, Toni I, Roelofs K, De Lange FP, Nater UM, van der Meer JW, Knoop H. Investigating neural mechanisms of change of cognitive behavioural therapy for chronic fatigue syndrome: a randomized controlled trial. *BMC Psychiatry*. 2015; 15: 144. doi: 10.1186/s12888-015-0515-9.
 55. Wiborg JF, van Bussel J, van Dijk A, Bleijenberg G, Knoop H. Randomised controlled trial of cognitive behaviour therapy delivered in groups of patients with chronic fatigue syndrome. *Psychother Psychosom*. 2015; 84(6): 368-76. doi: 10.1159/000438867. Epub 2015 Sep 25.
 56. Wiborg JF, Knoop H, Stulemeijer M, Prins JB, Bleijenberg G. How does cognitive behaviour therapy reduce fatigue in patients with chronic fatigue syndrome? The role of physical activity. *Psychol Med*. 2010; 40(8): 1281-7. doi: 10.1017/S0033291709992212. Epub 2010 5.
 57. Nijhof SL, Bleijenberg G, Uiterwaal CSPM, Kimpen JLL, van de Putte EM. Reply to Kindlon and Crawford. *The Lancet* Vol 380 ust 11, 2012. <https://huisartsvink.wordpress.com/wp-content/uploads/2024/01/nijhof-fitnet-august-11-2012-reply-to-kindlon-and-crawford-actometer.pdf> (accessed on 30 October 2024).
 58. Baryakova TH, Pogostin BH, Langer R, McHugh KJ. Overcoming barriers to patient adherence: the case for developing inative drug delivery systems. *Nat Rev Drug Discov*. 2023; 22(5): 387-409. doi: 10.1038/s41573-023-00670-0. Epub 2023 27. PMID: 36973491; PMCID: PMC10041531.
 59. Knoop H, Prins JB, Stulemeijer M, van der Meer JW, Bleijenberg G. The effect of cognitive behaviour therapy for chronic fatigue syndrome on self-reported cognitive impairments and neuropsychological test performance. *J Neurol Neurosurg Psychiatry*. 2007; 78(4): 434-6. doi: 10.1136/jnnp.2006.100974.
 60. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. *BMJ*. 2013; 20: 347: f6753. doi: 10.1136/bmj.f6753.
 61. Koolhaas, M.P, de Boorder, H, van Hoof, E. Cognitive Behavioural Therapy in Chronic Fatigue Syndrome ME/CFS) from the Perspective of the Patient. 2008. Available online: <https://huisartsvink.files.wordpress.com/2021/01/koolhaas-csmrt2008.pdf> (accessed on 30 October 2024).
 62. Stordeur, S, Thiry, N, Eyssen, M. Chronic Fatigue Syndrome: Diagnosis, Treatment and Care Organisation [Fatigue Syndrome: Diagnosis, Treatment and Organisation of Care]; KCE Technical Report 88A; Belgian Healthcare Knowledge Center: Brussels, Belgium, 2008. Available online: <https://kce.fgov.be/sites/default/files/2021-11/d20081027358.pdf> (accessed on 30 October 2024).
 63. Collin, S.M, Crawley, E. Specialist treatment of chronic fatigue syndrome/ME: A cohort study among adult patients in England. *BMCHealth Serv. Res*. 2017; 17: 488.
 64. Collin SM, Heron J, Nikolaus S, Knoop H, Crawley E. Chronic fatigue syndrome (CFS/ME) symptom-based phenotypes and 1-year treatment outcomes in two clinical cohorts of adult patients in the UK and The Netherlands. *J Psychosom Res*. 2018; 104: 29-34. doi: 10.1016/j.jpsychores.2017.11.007. Epub 2017 8.
 65. Adamson J, Ali S, Santhouse A, Wessely S, Chalder T. Cognitive behavioural therapy for chronic fatigue and chronic fatigue syndrome: outcomes from a specialist clinic in the UK. *J R Soc Med*. 2020; 113(10): 394-402. doi: 10.1177/0141076820951545. Epub 2020 Sep 15.
 66. Keijmel SP, Delsing CE, Bleijenberg G, van der Meer JWM, Donders RT, Leclercq M, Kampschreur LM, van den Berg M, Sprong T, Nabuurs-Franssen MH, Knoop H, Bleeker-Rovers CP. Effectiveness of Long-term Doxycycline Treatment and Cognitive-Behavioral Therapy on Fatigue Severity in Patients with Q Fever Fatigue Syndrome (Qure Study): A Randomized Controlled Trial. *Clin Infect Dis*. 2017; 64(8): 998-1005. doi: 10.1093/cid/cix013.
 67. NKCv: Over NKCv en Team. <https://huisartsvink.wordpress.com/wp-content/uploads/2024/07/nkcv-over-nkcv-en-team.pdf> (accessed 30 October 2024).
 68. Ernst E, Pittler MH. Assessment of therapeutic safety in systematic reviews: literature review. *BMJ*. 2001; 323(7312): 546. doi: 10.1136/bmj.323.7312.546.
 69. Li L, Xu C, Deng K, Zhou X, Liu Z, Busse JW, Ren Y, Zou K, Sun X. The reporting of safety among drug systematic reviews was poor before the implementation of the PRISMA harms checklist. *J Clin Epidemiol*. 2019; 105: 125-135. doi: 10.1016/j.jclinepi.2018.09.014. Epub 2018 Sep 29.
 70. Ahmed SA, Mewes JC, Vrijhoef H. Assessment of the scientific rigour of randomized controlled trials on the effectiveness of cognitive behavioural therapy and graded exercise therapy for patients with myalgic encephalomyelitis/chronic fatigue syndrome: A systematic review. *J Health Psychol*. 2020; 25(2): 240-255. doi: 10.1177/1359105319847261. Epub 2019 10.
 71. Ingman T, Smakowski A, Goldsmith K, Chalder T. A systematic literature review of randomized controlled trials evaluating prognosis following treatment for adults with chronic fatigue syndrome. *Psychol Med*. 2022; 52(14): 2917-2929. doi: 10.1017/S0033291722002471. Epub 2022 Sep 5.
 72. 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 priy care with chronic fatigue syndrome: randomised controlled trial. *BMJ*. 2010 23; 340: c1777. doi: 10.1136/bmj.c1777.
 73. Osborne, Jason W. and Overbay, Amy. The power of outliers (and why researchers should ALWAYS check for them)," *Practical Assessment, Research, and Evaluation*. 2019;0 9: 6. DOI: <https://doi.org/10.7275/qf69-7k43>
 74. Hawkins, D.M. Identification of outliers. London: Chapman and Hall. 1980.
 75. Clark LV, Pesola F, Thomas JM, Vergara-Williamson M, Beynon M, White PD. Guided graded exercise self-help plus specialist medical care versus specialist medical care alone for chronic fatigue syndrome (GETSET): a pragmatic randomised controlled trial. *Lancet*. 2017; 22: 390(10092): 363-373. doi: 10.1016/S0140-6736(16)32589-2. Epub 2017 22.
 76. 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: 59-111. Available online: <https://huisartsvink.wordpress.com/wp-content/uploads/2018/08/kindlon-reporting-of-harms-associated-with-get-and-cbt-in-me-cfs.pdf> (accessed on 30 October 2024).
 77. Oxford Clinical Allied Technology and Trials Services Unit (Ox-CATTS), 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. 27 ruary 2019. Available online: <https://huisartsvink.files.wordpress.com/2019/11/nice-patient-survey->

- outcomes-cbt-and-get-oxford-brookes-full-report-03.04.19.pdf (accessed on 30 October 2024).
78. Terheggen-Lagro, S, Oostrom, K. Exercise makes long Covid patients sicker just like ME/CFS patients. NPO, Kassa, Afl. 3, Za 20 Jan, 19:14. https://npo.nl/start/serie/kassa/seizoen-20_1/kassa_701/afspelen (accessed on 30 October 2024).
 79. White P, Abbey S, Angus B, Ball HA, Buchwald DS, Burness C, et al. Anomalies in the review process and interpretation of the evidence in the NICE guideline for chronic fatigue syndrome and myalgic encephalomyelitis. *J Neurol Neurosurg Psychiatry*. 2023; 94(12): 1056-1063. doi: 10.1136/jnnp-2022-330463. Epub 2023 10.
 80. Flottorp SA, Brurberg KG, Fink P, Knoop H, Wyller VBB. New NICE guideline on chronic fatigue syndrome: more ideology than science? *Lancet*. 2022; 12; 399(10325): 611-613. doi: 10.1016/S0140-6736(22)00183-0.
 81. Vink, M, Vink-Niese, A. The Updated NICE Guidance Exposed the Serious Flaws in CBT and Graded Exercise Therapy Trials for ME/CFS. *Healthcare* 2022; 10: 898. <https://doi.org/10.3390/healthcare10050898>
 82. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005; 2(8): e124. doi: 10.1371/journal.pmed.0020124. Epub 2005 30. Erratum in: *PLoS Med*. 2022; 25; 19(8): e1004085.
 83. National Institute for Health and Care Excellence. Myalgic Encephalomyelitis (or Encephalopathy)/Chronic Fatigue Syndrome: Diagnosis and Management [G] Evidence Reviews for the Non-Pharmacological Management of ME/CFS. NICE Guideline NG206 Evidence Reviews Underpinning Recommendations and Research Recommendations in the NICE Guideline Final October 2021. Available online: <https://www.nice.org.uk/guidance/ng206/evidence/g-nonpharmacological-management-of-mecfs-pdf-9265183028> (accessed on 30 October 2024).
 84. Psychological Medicine Publishing ethics, Cambridge University Press, 2024. <https://www.cambridge.org/core/journals/psychological-medicine/information/journal-policies/publishing-ethics#> (accessed on 30 October 2024).
 85. Dana J, Loewenstein G. A social science perspective on gifts to physicians from industry. *JAMA*. 2003 9; 290(2): 252-5. doi: 10.1001/jama.290.2.252. PMID: 12851281.
 86. Groopman J. *How Dors Think*. Boston, MA: Houghton Mifflin Company. ch 2007
 87. The BMJ. Competing interest policy. <https://www.bmj.com/about-bmj/resources/authors/forms-policies-and-checklists/declaration-competing-interests> (accessed on 30 October 2024).
 88. Stanbrook MB, Flegel K, MacDonald N, Attaran A, Eggertson L, Sibbald B, Kelsall D, Fletcher J, Ramsay J, Kale R, Hébert PC. Competing interests of authors: We have revised our policy. *CMAJ*. 2009 7; 181(1-2): 11-2. doi: 10.1503/cmaj.091042.
 89. Al-Aly Z, Davis H, McCorkell L, Soares L, Wulf-Hanson S, Iwasaki A, Topol EJ. Long COVID science, research and policy. *Nat Med*. 2024; 30(8): 2148-2164. doi: 10.1038/s41591-024-03173-6. Epub 2024 Aug 9. PMID: 39122965.
 90. Lim EJ, Ahn YC, g ES, Lee SW, Lee SH, Son CG. Systematic review and meta-analysis of the prevalence of chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME). *J Transl Med*. 2020 24; 18(1): 100. doi: 10.1186/s12967-020-02269-0.
 91. Rohrhofer, J, Hauser L, Lettenmaier L, Lutz L, Koidl L, Gentile SA, et al. Immunological Patient Stratification in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. *J. Clin. Med*. 2024, 13, 275. <https://doi.org/10.3390/jcm13010275>
 92. Sharpe MC, Archard LC, Banatvala JE, Borysiewicz LK, Clare AW, David A, et al. A report--chronic fatigue syndrome: guidelines for research. *J R Soc Med*. 1991; 84(2): 118-21. doi: 10.1177/014107689108400224.
 93. Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Kooff A. The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. *Ann Intern Med*. 1994; 15: 121(12): 953-9. doi: 10.7326/0003-4819-121-12-199412150-00009.
 94. Carruthers BM, van de Sande MI, De Meirleir KL, Klimas NG, Broderick G, Mitchell T, et al. Myalgic encephalomyelitis: International Consensus Criteria. *J Intern Med*. 2011; 270(4): 327-38. doi: 10.1111/j.1365-2796.2011.02428.x. Epub 2011 22. Erratum in: *J Intern Med*. 2017; 282(4): 353.
 95. Haywood KL, Staniszevska S, Chapman S. Quality and acceptability of patient-reported outcome measures used in chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. *Qual Life Res*. 2012; 21(1): 35-52. doi: 10.1007/s11136-011-9921-8. Epub 2011 18.