

Are cognitive behavioral therapy and a group physical and mental health rehabilitation programme effective treatments for long COVID? Rethinking of a systematic review

Mark Vink^{1*}; Friso Vink-Niese²

¹Family and Insurance Physician, 1096 HZ Amsterdam, The Netherlands.

²Independent Researcher, Germany.

Corresponding Author: Mark Vink

Family and Insurance Physician, 1096 HZ Amsterdam,
The Netherlands.

Email: markvink.md@outlook.com

Article Information

Received: Jan 10, 2025

Accepted: Mar 03, 2025

Published: Mar 26, 2025

SciBase Neurology - scibasejournals.org

Vink M et al. © All rights are reserved

Citation: Vink M, Vink-Niese F. Are cognitive behavioral therapy and a group physical and mental health rehabilitation programme effective treatments for long COVID? Rethinking of a systematic review. SciBase Neurol. 2025; 3(1): 1026.

Abstract

In this article, we analyzed the systematic review by Zeraatkar et al. which concluded that Cognitive Behavioral Therapy (CBT) and a group physical and mental health rehabilitation programme, are effective treatments for long COVID. Our analysis of the review highlights the problems with the 2 studies that were used for this claim but also with the systematic review itself. These problems included relying on subjective outcomes in non-blinded studies with poorly chosen control groups, selection, volunteer and self referral bias, response shift and allegiance bias, small study effect bias, selective reporting of the objective outcomes and selecting patients who didn't have Post-Exertional Malaise (PEM) but then claiming that exercise treatment is safe for long COVID patients with PEM. Moreover, the CBT study and the systematic review ignored the fact that CBT did not lead to objective improvement. The group physical and mental health rehabilitation programme was labelled effective based on its primary outcome even though the threshold of minimal important difference was not reached at any of the three outcome points and the scores of its primary outcome (quality of life) remained lower than in diseases like cerebral thrombosis, acute myocardial infarction, MS, lung cancer and stroke. Also, the study selected obese and older patients aged 56, with pre-existing health issues, who had been hospitalised for a severe COVID-19 infection and more than a third of them had been admitted to IC/HDU. Yet the average adult long COVID patient has a normal BMI, is much younger, used to be fit and well and developed long COVID after a mild infection with COVID-19. Consequently, one can not generalise the findings from that study to the average long COVID patient. It's unclear why the systematic review ignored all of that but also ignored the biases created by selective reporting and the widespread deviations from the intended interventions in both studies. Even though these two forms of bias are an important part of the risk of bias assessment according to the systematic review itself.

In conclusion, our analysis does not lend any support for the claim that CBT or a group physical and mental health rehabilitation programme are safe and effective treatments for long COVID patients who suffer from PEM.

Keywords: Chronic fatigue syndrome (CFS); COVID-19; Long COVID; ME/CFS; Post-COVID-19 condition; Post-infectious disease; Rehabilitation; SARS-CoV-2.

Introduction

More than seven hundred million people worldwide have fallen ill during the COVID-19 pandemic with major consequences for patients, countries, their health care system and its economies [1-4]. Most patients recover from a COVID-19 infection, but according to a conservative estimate, at least 400 million people [5] globally develop post-COVID-19 condition

(most commonly known as long COVID). The WHO [6] defined Post-COVID-19 condition as the continuation or development of new symptoms, for which there is no other explanation, 3 months after the initial SARS-CoV-2 infection, with symptoms lasting for at least 2 months. Symptoms might include exercise intolerance, fatigue, myalgia, impaired cognitive function and a

whole range of other symptoms. Effective treatments are currently lacking but according to a recent systematic review by Zeraatkar et al., “it is reasonable to offer CBT and mental and physical rehabilitation to [long COVID] patients. We emphasise that the effectiveness of CBT and physical rehabilitation for long COVID neither indicates the condition is psychological nor negates a possible somatic cause” (p. 11 [7]). According to them, there is a striking resemblance between long COVID and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). They also conclude that “CBT and graduated physical activity are offered to patients with long COVID and ME/CFS” to help “addressing patients’ unhelpful beliefs about fatigue and activity” (p. 11 [7]). Is it unclear why they then claim that a condition caused by unhelpful beliefs is not psychological.

However, the British National Institute for Health and Care Excellence (NICE) published its updated guideline for ME/CFS in October 2021 and concluded that ME/CFS is a debilitating chronic multisystem disease [8,9]. It also concluded that CBT and gradually increasing activities/exercise, more commonly known as Graded Exercise Therapy (GET), does not lead to improvement or recovery. The conclusion from the systematic reviewers has led to a lot of comments from ME/CFS and long COVID patients on the Internet as they have tried these therapies without effect. Moreover, many state that these therapies have been harmful.

In this article we will analyse the evidence presented by Zeraatkar et al. [7] in regards to the safety and efficacy of CBT and physical rehabilitation for long COVID which the systematic review base on two studies: the ReCOVer study by Kuut et al. [10] and the REGAIN study by McGregor et al. [11]. Therefore, we will pay particular attention to these two studies to see if there is any merit in the conclusion by Zeraatkar et al. or if they should have come to a different conclusion.

The recover study by kuut et al. [10].

In this study by Kuut et al. [10], 57 patients were treated with online CBT and 57 with Care As Usual (CAU) for 17 weeks and the study used one subjective primary outcome (CIS fatigue scale).

A review to advise the European Union (EU) about homeopathy, which included one of the leading CBT proponents for ME/CFS (van der Meer), excluded trials with less than 75 patients in the treatment group because then the chances that improvements were down to chance and not to the treatment, are too high [12]. Consequently, the ReCOVer study would have been excluded.

How to prove that your therapy is effective, even when it is not

Moreover, psychologists Cuijpers and Cristea published an article entitled *How to prove that your therapy is effective, even when it is not: a guideline* [13]. They concluded that there are several methods available to help one show that one’s therapy is effective, even when it is not. Methods that can help include a strong allegiance towards the therapy, and the main investigator (Knoop) has built his career on the efficacy of CBT for the other post infectious disease (ME/CFS). Anything that increases expectations and hope in participants, and the study was informing participants before they started treatment, that CBT was effective for similar diseases and that if they would be allocated to the control group then they could be treated with the effective treatment (CBT) after finishing the study [14]. Other

methods include conducting studies with not enough participants in the therapy group as mentioned before and using badly designed control groups, like a waiting list control group, as used by the ReCOVer study [13]. Because the main principle of a properly conducted and designed study, is that everything in the control group is the same as in the therapy group, including expectations raised about the efficacy of the control treatment, apart from the treatment under investigation [13]. Only then does one know that a change in scores is caused by the treatment and not other, often unknown, variables, confounding factors or combinations of them.

A meta-review of systematic reviews by Fordham et al. concluded that there is “consistent evidence for the general benefit which CBT offers” “for people living with many different mental and physical conditions” (p. 21,28 [15]). One of the conditions included in that review was ME/CFS. We mention this review because it showed the effect of using a non-active control group instead of an active one. Fordham et al. concluded that the effect size was 0.31 if the studies used a non-active comparator control group and 0.09 when an active comparator control group was used [15]. A score between 0.20≤ and <0.50, means that the effect of a treatment is small and below 0.20, that it is ignorable [16,17]. Consequently, using a poorly designed (non-active) control group instead of a properly designed active control group, artificially inflated the results of the ReCOVer study.

One of the other things non-blinded studies can do to create the illusion of efficacy, is relying on subjective outcomes [13]. The ReCOVer study used only one primary outcome, the subjective CIS fatigue scores [10] which might have further artificially inflated the outcome of the study.

Selection bias

Another issue with psychological studies can be selection bias. This, according to Drew, “refers to a distortion in research outcomes caused by the non-random or biased selection of participants for a study, leading to a sample that is not representative of the entire population of interest. This bias can compromise the generalizability and accuracy of research findings, impacting both external and internal validity” (p. 1 [18]). There are many different forms of selection bias for example, sampling bias, self-selection bias, pre-screening of subjects and cherry-picking bias. We mention this for a number of reasons. First, the study assessed 721 people for eligibility by phone call or email off which 265 did not meet the inclusion criteria but the other 305 declined to participate (n=135) or did not respond after referral (n=170). The remaining 151 were assessed by questionnaire and medical records and 37 were excluded, 10 because they declined to participate and one because the study was full. Consequently, up to 43.6% (315/721) of the people that were assessed for eligibility, refused to take part which suggests that volunteer bias might be a problem for the study. This form of bias is created when a disproportionate number of individuals decline to participate because they are fundamentally different from the ones who agreed to participate [18].

Second, 68% of participants in the CBT group self referred to the study. According to Kaźmierczak et al., self selection bias in psychological studies can be a problem because “individuals willing to take part in psychological studies are seeking a therapeutic environment, diagnoses, and/or a meeting with a psychologist” (p. 2 [19]). But also that, “our field may be conducting research on an atypically disordered and motivated group of people leading to biased views of the reality of psychological

effects" (p. 9 [19]). Consequently, participants in the ReCOVER study might not have been representative for the group of long COVID patients in general which means that results might not be generalisable to them.

The minimally important difference

The true treatment effect in medical research is the very thing which a trial is intended to estimate, and which patients and clinicians are interested in. Zeraatkar et al. concluded that "moderate certainty evidence suggests that CBT... probably improve[s] symptoms of long COVID" (p. 1 [7]). They base that on the Minimally Important Difference (MID). This is an important change in the score of a patient-reported outcome measure (either positive or negative) from the patient's or clinician's perspective which would warrant a change in the management of the patient. Zeraatkar et al. were unable to identify MIDs specific to long COVID and we couldn't find them either. Instead, Zeraatkar et al. defined the MID as 0.5 of the Standard Deviation (SD) based on an influential article by Norman et al. about the MID for health-related quality of life scores. Norman et al. suggest that half a SD may be a universal standard [20]. However they also note that "it would be inappropriate for this to be viewed as a fixed benchmark" (p. 590 [20]). Additionally, "the criterion [of half a SD] may be more appropriately thought of as a Minimally Detectable Difference (MDD), not a minimally important difference" (p. 583 [20]). They also noted that "Lydick and Epstein pointed out, [that] expressing minimal changes in terms of statistical quantities is of limited value to clinicians" (p. 589 [20]).

There are also other problems with this concept as noted by for example Beaton. One of those problems being "that the severity of the disorder, recognized by others as an emerging source of variability, should be examined. In addition, it should also be disentangled from regression to the mean" (p. 594 [21]). Beaton notes that "Hays and Woolley warn that 'identifying a single threshold that defines the amount of score difference that is clinically important is potentially misleading' because of the various factors influencing the meaning of a change score" (p. 595 [21]).

Zeraatkar et al. base the moderate certainty evidence on a MID of a difference of only 3 points on the CIS-fatigue questionnaire [7]. The ReCOVER study itself however stated that "a difference of 6 points on the CIS-fatigue score is considered clinically relevant" (p. 5 [10]). The "overall between-group difference of the fatigue severity score was" a "mean difference [of] -8.4" "favoring CBT" (p. 7 [10]). In the supplementary material of the article by Zeraatkar et al. [7], they mention one COPD article by Rebelo et al. [22] in their effort to determine the MID for the CIS-fatigue questionnaire. The MID in COPD is 9.6, yet Zeraatkar et al. do not explain why they mention it but then ignore it instead of using it.

The fatigue score in the CBT group at baseline was 47.8, if we would deduct the MID according to Zeraatkar et al. or the one as mentioned by the ReCOVER study, then that score would be 44.8 or 41.8 respectively. Using the MID for COPD would yield a score of 38.2 yet a score of 35 or more means that participants are severely fatigued [10]. In contrast, the score of healthy people is 17.3 according to one of the researchers of the study (Knoop) [23]. Consequently, patients would still be severely disabled, irrespective of which of the three MIDs would be chosen. Yet for example, if patients with a severe pneumonia would still have severe pneumonia after treatment has finished, then

that would not be classed as a clinically important change. Nor would it be classed as effective and recommended. Interestingly enough, the fatigue score in the study by Rebelo et al. [22] at baseline was 36.9. The above highlights the fact that the more severely patients are affected, the bigger the MID needs to be for a change in scores to be perceived as an important improvement for patients. As noted by Beaton [21], but also by Norman et al. [20], the severity of the disorder should be examined because it is a source of variability so that the universality of defining the MID as 0.5 of the standard deviation might not apply.

Moreover, as noted by Zeraatkar et al., ME/CFS is "a condition with a striking resemblance to long COVID" (p. 11 [7]) and housebound or bed bound ME/CFS patients are too ill and unable to take part in studies [24-27]. It's likely that something similar applies to long COVID patients. Consequently, patients who are labelled as having severe fatigue in studies are in reality patients who are 'only' moderately affected. Therefore, even if any of the aforementioned 3 MIDs would signify real and important improvements and not artefacts caused by the use of a subjective outcome in a non-blinded study with a badly designed control group. Then the findings may not be generalizable to severe long COVID.

The risk of bias tool

Zeraatkar et al. uses the risk of bias 2.0 tool as according to them, it is endorsed by Cochrane. Two of the elements of that tool are bias due to deviations from the intended intervention and selective outcome reporting [7]. 73% of patients (40/55) in the CBT group received care outside of the study for their long COVID during the study. Of those 73% (40/55), 38% (15/40) received more than 17 weeks of treatment by a physical therapist during the study and many of them received physical therapy, two or more times a week. Something like that should never happen in a properly conducted study because then it becomes impossible to know if any improvement is down to the treatment (CBT), the additional therapy, other factors or an unknown combination of those. This is even more of a problem because if the treatment is not controlled by the researchers of a study, then one does not know in what form or way it has been administered. One also doesn't know if patients have received it in the same form, intensity and duration during each session. It also means that 73% deviated from the intended intervention (CBT) because it was not effective and / or they were negatively affected by it. If CBT would have been effective then there would have been no reason or need to seek additional help and treatment.

Moreover, 21% in the CBT group had clinically relevant depressive symptoms and CBT is the most effective treatment for depression. Which leaves 6% of non-depressed patients who might well have suffered from anxiety, for which CBT is an effective treatment, or they needed help coping with their disease.

The aforementioned study by Cuijpers and Cristea also concluded that "if all that [the aforementioned tricks] fails [then] one can always not publish the outcomes" (p. 428 [13]) which is known as selective outcome reporting. This is one of the other forms of bias that should be assessed according to the aforementioned risk of bias tool. According to the protocol of the ReCOVER study, "outcome measures consist of self-reported questionnaires" but "at T0 [baseline] and T1 [end of treatment], data on physical activity level and sleep are also gathered by using an actigraph" (p. 7 [28]). Kuut et al. did not report their objective activity results (actigraphy) which is known as selective

outcome reporting. However, they did mention in a comment that CBT did not lead to objective improvement [29]. There is an inverse relationship between fatigue and activity according to a study which included one of the authors of Kuut et al. (Knoop) [30]. Activity did not improve objectively which means that the change in fatigue is an artefact caused by all the aforementioned methodological problems and/or the different forms of bias and confounding factors of the study. For example, using a subjective outcome in a non-blinded study, using a poorly designed control group together with allegiance, small study, selection and response shift bias.

Kuut et al. also published their long-term follow-up results in a letter in which they concluded that the “favorable outcomes following CBT were maintained” 1 year post-CBT (p. 1078 [31]). They only published the results from the CBT treatment group because patients from the control group crossed over to the treatment group. Consequently, at long term follow-up the non-blinded study also became a non-controlled study, i.e. a study without a control group. So that the results must be interpreted very cautiously and that one cannot come to any causal inference of the efficacy of a treatment [32]. Yet the authors ignored that. They also stated that “all secondary outcomes also favored CBT” (p. 1078 [31]) just like they stated in the original article that “all secondary outcomes were in favor of the CBT group” (p. 7 [7]). Even though they did not publish their objective outcome measure. It also means that in this long-term follow-up letter, they continue to ignore the null effect of that.

In conclusion, the ReCOVer study does not provide evidence that CBT is an effective treatment for long COVID.

The regain study by McGregor et al. [11].

This large study (n=585) concluded that in “adults with post-COVID-19 condition, at least three months after hospital discharge for COVID-19, an eight week, live, online, home based, supervised group rehabilitation programme (REGAIN) was well tolerated and led to sustained improvements in health related quality of life [PROPr score] at three months and one year compared with usual care”. And that “high quality evidence from” their “randomised controlled trial confirmed the clinical benefit and lack of harm” of their treatment (p. 1 [11]).

Participants in the treatment group received a one hour, on-line, one-to-one consultation with a REGAIN practitioner with subsequent weekly practitioner-led live online group exercise sessions under the supervision of a REGAIN practitioner for eight weeks and six live online group psychological support sessions (one hour each) delivered through Zoom. The goal of the REGAIN practitioner exercise sessions was to improve cardiovascular fitness, strength, balance, and fatigue. Yet the study did not use objective outcome measures like Cardiopulmonary Exercise Testing (CPET) or a step test to assess cardiovascular fitness before and after treatment. Why they did not do it in light of their goal to improve cardiovascular fitness, is unclear.

Participants in the usual care group on the other hand received best practice usual care, consisting of a 30 minute, on-line, one-to-one consultation with a trained practitioner. The study specifically stated that “a structured physical activity plan was not provided, and no specific psychological techniques were used” (p. 3 [11]). Consequently, the REGAIN study was a non-blinded study with a poorly designed controlled group that used one subjective primary outcome (PROPr score). Therefore, it was set up in the same way as the aforementioned ReCOVer

study so that any change in PROPr score might simply be down to the way that the study was designed [13].

The REGAIN study was a study of hospitalized patients. According to a systematic review, 51% of adult long COVID patients exhibit PEM and meet ME/CFS criteria. Or to put it differently, COVID-19 triggered their ME/CFS. This cohort of patients is between 20 and 45 years old, were fit and well until they developed long COVID after a (very) mild infection with COVID-19, they had a normal BMI, no pre-existing medical conditions and the majority of them have not been hospitalized. The participants in the REGAIN study on the other hand, were a lot older (mean age of 56.1), and only 12.8% had a healthy weight (BMI 24.9 or lower) according to the supplementary material [11]. No one was underweight (BMI below 18.5) yet 28.2% were overweight (BMI 25 to 29.9) and 59.1% were obese (BMI of 30 to 39.9). The mean BMI in the treatment group was 33. Many of them had pre-existing medical conditions, for example chest or breathing problems (76%) or heart and circulation problems (26%) and 34% of participants had been admitted to the ICU/HDU because of the severity of their COVID-19 infection [11].

Selection bias

The study invited 39697 people by letter and 82 people self-referred. Yet only 1043 expressed interest to take part in the study, 37 of them were ineligible and a further 281 were not contacted but simply excluded without a reason given. 725 were contacted and eligible to take part but 140 of them were not randomised because they did not complete the baseline outcome questionnaire (65 patients) or consent was not received (66 patients) [11]. This all gives the impression that selection bias, which “can be the most important threat to internal validity in intervention research, but is often insufficiently recognized and controlled” (p. 289 [33]) might have been an issue in this study.

Monitoring for post-exertional symptom exacerbation

The researchers noted that the presentation of post-COVID-19 condition and ME/CFS overlap and they “therefore prospectively monitored for Post-Exertional Symptom Exacerbation” [PESE] (p. 4 [11]) PESE, also known as post-exertional malaise or PEM, is the main characteristic of ME/CFS. However, if it would be defined as symptom exacerbation after exertion, then that would be incorrect. PEM or PESE, is an often delayed exacerbation of symptoms after trivial exertion with a loss of function and an abnormally delayed recovery. All those elements need to be present, otherwise patients are not suffering from it [34]. The REGAIN study however, did not investigate how many participants suffered from PEM at baseline nor were these patients excluded from the exercise study as they should have to prevent potentially harming their health with exercise therapy.

Adherence to treatment

According to the study, “adherence to the REGAIN intervention was good” (p. 8 [11]) because “in the intervention group, 141 (47%) participants fully adhered to the programme” (p. 1 [11]). In contrast, 90.2% fully adhered to usual care according to the supplementary data of the REGAIN study [11]. Many researchers use a threshold of ≥80% to distinguish adherent from non-adherent patients based on Haynes’s empirical definition of sufficient adherence [35]. A systematic review by Bullard et al. [36] found that the adherence to physical activity interventions among three chronic conditions (cancer, cardiovascular

disease, and diabetes) was 77%. They also found that adherence rates for clinic-based and home-based activity intervention programs did not differ. Consequently, an adherence rate of 47% is not good but very low. Moreover, the REGAIN study defined full adherence “as attending the initial assessment, plus attending four out of six psychological support sessions, AND five out of eight live exercise sessions” according to its supplement (p. 17 [11]). Therefore, full adherence did not mean that participants had fully adhered to the treatment.

The supplement of the study shows that 78.5% in the treatment group attended the first live session, 31.9% attended all six psychological support sessions yet only 12.4% attended all eight live exercise sessions as can be seen in Table 1. The study doesn't provide figures for how many patients fully attended the first live session, all eight exercise sessions and all six psychological support sessions. It is unclear why they did not provide those that, but it means that the full adherence rate was maximal 12.4%.

Table 1: Full attendance rate to the different elements in the intervention group.

Regain	Attended first live session	Attended all 8 live exercise sessions	Attended all 6 psychological support sessions
REGAIN intervention group	78.5% (234//298)	12.4% (37/298)	31.9% (95/298)

Source: The supplementary data of the REGAIN study [11].

The REGAIN trial concluded that “a structured programme of physical and mental health rehabilitation (REGAIN), delivered in groups online was clinically effective compared with usual care for improving health related quality of life (PROPr) in our primary analysis at three months post-randomisation...Furthermore, the effects of the intervention were also evident at 12 months” (p. 8 [11]). However, as can be seen in table 2, the difference in the health related quality of life scores between the treatment and control group at those time points, just like at six months, was less than the minimal important difference as can be seen in table 2. Therefore, the authors should have concluded that their treatment was clinically not effective.

Table 2: The difference in PROPr scores between groups.

PROPr	3 months	6 months	12 months
Difference between groups*	0.028 (p=0.015)	0.023 (p=0.081)	0.034 (p=0.019)
Minimal important difference**	0.04	0.04	0.04

PROPr: Health related quality of life.

Source: *The supplementary data of the REGAIN study [11].

**The REGAIN study [11].

Is the regain trial population representative for long COVID?

According to the study, “the REGAIN trial population was severely affected by post-COVID-19 condition” (p. 11 [11]). Yet 58.1% of participants in the treatment group thought that they were able to work despite this and the other 41.9% thought they were unable to do so. Consequently, it's unlikely that those 58.1% were severely affected by long COVID as the severely affected are home or bed bound and comprise 25 per cent in the other post infectious disease (ME/CFS) according to most estimates [25]. These patients are unable to attend outpatient clinics and take part in those studies. Consequently, those patients with severe functional impairment in the REGAIN study, are in fact moderately affected long COVID 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 like the REGAIN study, are not generalisable to the wider long COVID population and long COVID trials are inherently biased as a consequence of that.

Missing data

The missing data due to loss to follow-up from the treatment group was 20.5% (237/298) after three months, 24.5% (225/298) after six months and 27.2% (217/298) after 12 months. As noted by Heneghan et al., “the ‘5 and 20 rule’ (i.e., if >20% missing

data, then the study is highly biased; if <5%, then low risk of bias) exists to aid understanding” (p. 3 [37]) of missing data. Consequently, the REGAIN study was highly biased.

Additional bias

The study was biased further by using one subjective outcome in a non-blinded study with a badly designed control group. According to its supplementary material, the study used one objective secondary outcome measure which it did not publish (“Work Status: Time lost from work (paid/unpaid)” (p. 11 [11])). This is a form of selective outcome reporting which biases studies, makes them unreliable and leads to the overestimation of the benefits of an intervention [38-41]. According to Pickett and Roche [41], the general public is the largest stakeholder. Science is primarily paid for with public funds and the public is not impressed by selective outcome reporting because flawed science threatens the public's welfare.

Quality of life scores

The score of the EQ5D-5L, the study's second and secondary quality of life outcome measure, was 0.597 at its primary outcome point at 3 months according to the supplementary data of the study [11]. This score is similar to the score (0.60) for people with five or more chronic health conditions and worse than in cerebral thrombosis (0.62), rheumatoid arthritis and angina (0.65), acute myocardial infarction (0.66) [42], MS (0.67), lung cancer (0.69), stroke (0.71) or ischemic heart disease (0.72) (linear scale ranging from -0.624 to 1.000; higher scores indicating a better quality of life and negative values are conditions considered worse than death [43]). The same applies to the EQ5D-5L scores at 6 months (0.604) and at 12 months (0.615).

In conclusion, the REGAIN study does not provide any evidence that a supervised group rehabilitation programme is a safe and effective treatment for long COVID patients who suffer from PEM.

Exercise and escalation of symptoms (post-exertional malaise)

According to the systematic review, “guidance on the optimal management of patients with long COVID is limited. When guidance has been published, it is largely consensus based, does not base recommendations on rigorous systematic reviews, or provides limited advice on management” (p. 13 [7]). It “prioritises activity management (pacing) over physical activity owing to concerns about post-exertional malaise. This symptom, frequently reported by patients with long COVID and ME/CFS, involves worsening fatigue after physical or mental exertion. The

trial we identified that investigated physical and mental health rehabilitation, however, did not report any instances of post-exertional malaise, despite closely monitoring patients for this symptom. Furthermore, a recent crossover trial [Tryfonos et al.] found tailored exercise rehabilitation can be effective for long COVID without escalation of symptoms. Together, these results suggest that interventions involving supervised, negotiated, and moderate physical activity can be safe for patients with long COVID" (p. 13 [7]).

Wrong definition of post-exertional malaise (PEM)

The systematic review uses three references to wrongly define Post-Exertional Malaise (PEM). First, the article in the New Scientist states that according to campaigners, "encouraging people to raise their exercise levels can trigger post-exertional fatigue" (p. 1 [44]). Whereas the articles by Tuller and the one by us [45,46], don't mention what Post-Exertional Malaise (PEM) is. Yet PEM is not post-exertional fatigue because that is a normal physiological response to exercise. PEM on the other hand, is an abnormal response to trivial mental or physical exertion which is often delayed for up to 48 to 72 hours, with a flare up of symptoms, a loss of function and an abnormally delayed recovery. All these elements need to be present for a diagnosis of PEM [34]. The systematic review then states "that interventions involving supervised, negotiated, and moderate physical activity can be safe for patients with long COVID" (p. 13 [7]). They base it not only on the REGAIN study but also on a crossover trial which according to the systematic reviewers "found [that] tailored exercise rehabilitation can be effective for long COVID without escalation of symptoms" (p. 13 [7]). The REGAIN study did not investigate or document how many patients that were selected for their study actually suffered from PEM. It only states that because of the overlap in presentation between long COVID and ME/CFS, that they "prospectively monitored for post-exertional symptom exacerbation" (p. 4 [11]). But PEM is much more than just post-exertional symptom exacerbation as we just have seen. About safety, the REGAIN trial concluded that they did not observe any instances of PEM "during the trial or follow-up period, indicating that the intervention...was safe and acceptable overall" (p. 11 [11]). Yet as we have seen earlier, the adherence to treatment in the REGAIN study was very low. As concluded by psychology professor Lilienfeld, in contrast to clients who remain in treatment, those who drop out of treatment or do not adhere to it, tend to be lower functioning [47]. Additionally, those patients "are not a random subsample of all patients. Instead, those 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 [47]). Lilienfeld also noted that high levels of dropout or non-adherence to treatment might mean that patients have been harmed by the intervention. The consequence of that is that patients who remain in treatments are generally doing better than when they began, but they are unrepresentative of the clients who were included in the study [47].

The crossover trial by tryfonos et al. [48]

The crossover trial the systematic review uses to conclude that exercise is a safe treatment for a long COVID, is a study by Tryfonos et al. [48] that concluded that non-hospitalised patients with long COVID generally tolerate exercise well. There are however a number of problems with this study and its conclusion. For example, it was a small study with only 31 participants in the treatment group. Yet the basis for their study

seems to be that problems in long COVID are caused by the current recommendation against exercise in long COVID to prevent symptom worsening and the skeletal muscle deconditioning as a consequence of the avoidance of exercise. Healthy controls can exercise as much as they want yet healthy sedentary controls do not exercise. As a consequence, they are deconditioned and their stamina is down to the basal level of their cells and body. Why the study then used healthy controls as a control group instead of healthy sedentary controls is unclear.

According to Tryfonos et al., "it is important to...comprehensively investigate multiple factors in nonhospitalized patients with PEM" (p. 2 [47]). The study states that one of the inclusion criteria was persistent PEM symptoms for three or more months verified by the DePaul symptom questionnaire. The study used three forms of exercise, high intensity interval training of 5x1 minute cycling at 90% maximal workload, moderate intensity continuous training which consisted of 30 minutes continuous cycling at 50% maximal workload and strength training. The latter included three exercises: death lift, push-up and knee extensions using flywheel technology, each with three sets of 10 repetitions and a three minute rest between sets. Patients had to complete three exercise sessions with an approximately 2 to 4 week washout between sessions. The study used a number of objective investigations (six minute walk test, lactate testing and Cardiopulmonary Exercise Testing (CPET) to assess fitness at baseline. CPET was also used to determine the maximal workload. There are a number of references to ME/CFS which is characterised by PEM. It's known from ME/CFS studies that the only way to provide objective evidence for PEM and to distinguish deconditioned individuals from patients with ME/CFS is by repeating CPET on two consecutive days. Why the study then used a second cardiopulmonary exercise test 48 hours after exercise, instead of 24 hours, is unclear.

The study concluded that their "main finding was that participants with PCC (long COVID) generally tolerated all exercise sessions without significant worsening of symptoms or decline in aerobic performance after 48 hours" (p. 13 [48]). However, Tryfonos et al., defined PEM as "persistent fatigue, muscle pain, and cognitive problems that worsen after exertion (referred to as postexertional malaise)" (p. 1 [48]). Moreover, a recent 2-day CPET study for ME/CFS by Keller et al. noted that "the reproducibility of CPET measures is well-established so CPET results are expected to be reproduced within normal variability with confirmation of maximum effort" (p. 23 [49]). But also that "results from CPET-2 [the second test] further substantiated the challenge of ME/CFS to recover normally following CPET1. Despite meeting maximum effort criteria, the total sample of ME/CFS but not CTL [controls], exhibited significant reductions in peak Work (-5.5%), time to peak exercise (-6.6%), ventilatory measures (-4.9% to -7.8%), heart rate (-2.6%), O₂ pulse (-4.0%), and rate-pressure product (-3.4%). In contrast for CTL, only VCO₂ declined significantly by 3% during CPET-2" (p. 23 [49]). One would expect to see something similar in long COVID patients who suffer from PEM. Yet instead, Tryfonos et al. concluded that "nonhospitalized patients with PCC generally tolerated all exercise types without reporting significant symptom exacerbation, performance reductions, or exacerbated inflammation after 48 hours" (p. 15 [48]). This means that the study which wanted to comprehensively investigate PEM, had selected patients who did not suffer from it.

Tryfonos et al. also concluded that "13% had Postural Orthostatic Tachycardia" (POTS) and "62% showed signs of myopa-

thy" (p. 2 [48]). "Given that exercise was generally well tolerated, guidelines cautioning against exercise in similar populations may need to be revised. It seems advisable to cautiously incorporate exercise into rehabilitation protocols and adjust the intensity progressively, considering patients' symptoms and abilities" (p. 15 [48]). Yet as we know from ME/CFS research, if patients do suffer from PEM then progressively increasing the intensity of exercise leads to PEM, flare ups and (severe) relapses. Consequently, advising patients to incorporate exercise and progressively adjust the intensity, potentially puts the health of patients at risk and is therefore contraindicated in ME/CFS and long COVID patients who suffer from PEM.

Moreover, it's also questionable if everybody in the study actually tolerated the exercise program well. The study itself states that "all participants completed [all] three exercise sessions" (p. 3 [48]). Yet according to the flowchart of the study, participants were recruited via advertisements and from the post COVID-19 outpatient clinic and 125 people (long COVID patients and healthy controls) wanted to take part. 41 of them were excluded because they did not meet the inclusion criteria. 39 long COVID patients and 45 healthy controls passed the initial recruitment stage yet five patients declined further participation. One of the remaining 34 patients was excluded after all 34 underwent physiological assessment because of undefined abnormal findings. 2 of them dropped out ("discontinued") from the study and only 26 patients took part in the third round of exercise testing. Consequently, only 68.4% completed all three exercise sessions and 31.6% (12/38) did not do so. This might suggest that the 31.6% suffered from PEM and the 68.4% who completed the study, did not. This also suggests that only participants who did not suffer from PEM tolerated the exercise sessions and you cannot then extrapolate that to all long COVID patients because according to a systematic review only 51% of them suffer from PEM and the other 49% don't [50].

According to Zeraatkar et al., the certainty about the safety of CBT is "very low due to serious risk of bias and very serious imprecision". The first do no harm principle [51] is the main principle of medicine and if you do not know if a treatment is safe or not, then you should not recommend it. Moreover, graded activity, which is part of CBT for post infectious diseases, means an incremental increase in activity, which is contraindicated in patients who suffer from PEM.

Worsening of symptoms after treatment

The British National Institute for Health and Clinical Excellence (NICE) published its updated ME/CFS guidelines in October 2021 [8,9]. NICE commissioned the Oxford Brookes University as part of that review process, 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 [52] in which they concluded that 98.5% of the patients who took part in the survey experienced post-exertional malaise, the core symptom of the disease. Worsening of symptoms after treatment was reported by 58.3% (CBT, which incorporates an element of GET in ME/CFS) and by 81.1% (GET). In addition, the percentage of patients who were bedridden and dependent on help from others due to severe ME/CFS increased from 12.6% to 26.6% after treatment with CBT and from 12.9% to 35.3% after treatment with GET. Or to put it differently, 14% of patients were made homebound or bedridden by CBT and 22.4% by GET [52].

The very high dropout rate of 55%, 73%, and 80% at 6, 9, and 12 months, respectively, in the evaluation study of a 12-month program of GET in a sports medical department of a Dutch hospital as found by a reanalysis [53,54], confirms the findings of the Oxford Brookes University and the unsuitability and harmfulness of GET as a treatment for ME/CFS.

In conclusion, the systematic review and the studies used by it, do not provide any evidence that CBT and exercise therapy for long COVID patients who suffer from PEM, are safe.

Discussion

Zeraatkar et al. recently published a systematic review about different forms of treatment for long COVID in which they concluded that "it is reasonable to offer CBT and mental and physical rehabilitation to [long COVID] patients. We emphasise that the effectiveness of CBT and physical rehabilitation for long COVID neither indicates the condition is psychological nor negates a possible somatic cause" (p. 11 [7]). They also note that there is a striking resemblance between long COVID and ME/CFS and that "CBT and graduated physical activity are offered to patients with long COVID and ME/CFS" to help "addressing patients' unhelpful beliefs about fatigue and activity" (p. 11 [7]).

They base it on one study of CBT (the ReCOVer study) and one study of mental and physical rehabilitation (the REGAIN study). In this article, we have analysed the systematic review and these studies. Both studies were non-blinded studies that used a poorly designed control group and relied on one subjective primary outcome. Setting studies up this way artificially inflates the treatment effect so that it becomes impossible to know if any changes are down to the treatment, the set up of the study, the biases and confounding factors of the studies or an unknown combination of any of these factors. The only way to correct for that is by using objective outcome measures. Both studies however, did not publish their objective outcome measure. The ReCOVer study however, mentioned in a comment that CBT did not lead to objective improvement. The REGAIN study did not publish its objective outcome measure (work status).

The ReCOVer study artificially raised the expectation about the efficacy of their treatment by stating on their recruitment website that the treatment was effective for other diseases and informing patients in the control group that they could cross over to the treatment after the study had finished. By doing so they also informed them that the care as usual 'treatment' in the control group was not effective. Yet this non-blinded study relied on one subjective outcome measure (fatigue). Consequently, the study introduced an extra form of bias into their study by raising the expectation about the efficacy of their treatment, in a study which was already suffering from a number of different forms of bias.

Both studies also suffered from selection bias, as for example 68% of participants in the ReCOVer study self-referred. The REGAIN study invited almost 40,000 patients yet only 1043 expressed interest to take part in the study, 281 of them were not contacted but simply excluded without a reason given. The remaining 725 were contacted and were eligible to take part but 140 of them were not randomised because 131 did not complete the baseline outcome questionnaire or consent was not received. This all gives the impression that both studies were conducting research on an atypically disordered and motivated group of people who thought they would benefit from exercise

and / or CBT. But at the same time they thought that they would not have a problem with that. Consequently, many participants in both studies might not have been representative for the group of long COVID patients in general. This not only leads to biased views of the efficacy of CBT and exercise therapy for long COVID but it also might mean that results are not generalisable to the average long COVID patient.

It's also questionable if the REGAIN study was actually a long COVID study as it selected patients of whom most of them had one or more pre-existing health problems, patients had a mean age of 56, a mean BMI of 33, they all had been hospitalised and 34% of them had been admitted to the IC/HDU. Whereas adult patients who suffer from the post infectious disease long COVID are generally younger people (between 20 and 45), were fit and well before they had a (very) mild infection with COVID-19. Most of them did not require hospitalisation for it. The difference with patients who suffer from a post-infectious disease is highlighted if you compare the patients who were selected for the REGAIN study with the patients who were selected for the ReCOVER study. In the ReCOVER study, 79% of participants in the treatment group were female contrary to 54% in the REGAIN study, they were not obese (BMI of 26.9), were much healthier before contracting COVID-19 (56% had no comorbidities) and they were much younger (mean age was 45.7).

It's therefore much more likely that in the REGAIN study, the severity of the COVID-19 infection had a bad effect on the health of patients and/or patients had organ damage as a consequence of that infection. According to the method section of the study, "participants were adults (26-86 years) who had been discharged from hospital three or more months" (p. 2 [11]). Also, "participants were asked to self-report any substantial lasting effects that they attributed to their hospital admission with covid-19. This was confirmed during an eligibility telephone call with the clinical trial team before study enrolment" (p. 2 [11]). Consequently, participants were not medically examined by a physician from the study nor were any tests or examinations carried out before participants were taking part in the REGAIN study, to exclude organ damage.

The ReCOVER study [10] claimed that 60% of patients had improved because of CBT yet 73% of patients had other treatments during the study for their long COVID and many of them had the other treatment twice a week for more than 17 weeks. This renders it impossible to know if any changes were down to CBT, the new treatment, biases and confounding factors or an unknown combination of those. Moreover, 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 [55]). This "instrumentation related source of contamination is known as response-shift bias" (p. 93 [55]). This is even more of a problem when the therapy used, in this case different forms of CBT for post infectious diseases like ME/CFS and long COVID, aims to modify participants' beliefs and perception of their symptoms [56]. According to Lilienfeld et al., one of the things that eliminates "response-shift biases as explanations for apparent improvement" is not relying "exclusively on self-report ratings" (p. 372 [47]) but use objective measures as well. The REGAIN study

labelled its treatment effective based on its subjective primary outcome (quality of life scores), even though at the primary outcome at three months but also at six and 12 months, the treatment effect was smaller than the minimal important difference. Consequently, the study should have concluded that their treatment was not effective. This was confirmed for example by another quality of life score measure the study used as a secondary outcome measure. According to that outcome measure, patients remained more disabled than in many other diseases.

According to one of the limitations of the systematic review, "it is possible that we missed some problematic trials or misclassified trustworthy trials as problematic" (p. 12 [7]). Yet in this analysis we have seen there is also a third category. That is, problematic trials which have been classified as trustworthy. The systematic reviewers used the Cochrane endorsed risk of bias tool and three of its five domains are, bias due to deviations from the intended intervention, bias due to missing outcome data, and selective outcome reporting. 73% of participants in the treatment group in the ReCOVER study received care outside of the study for their long COVID during the study. 38% of those 73% received more than 17 weeks of treatment by a physical therapist during the study and many of them had treatment twice a week, or more frequently which equated to more than 34 interactions between therapist and patient. In comparison, CBT treatment in the treatment group lasted 18.7 weeks with 11.8 interactions between therapist and patient. The consequence of this is that 73% of cases deviated from the intended intervention but also that it became impossible to know if any improvement was down to CBT, the additional treatment, confounding factors and different forms of biases of the study or an unknown combination of those. Put differently, patients should not be allowed to receive additional non-urgent treatment during a study for the disease under investigation from people outside of the study.

The REGAIN study claimed that adherence to the REGAIN intervention was good and that 47% adhered fully to treatment. Yet the study had defined full adherence in such a way that it included people who did not fully adhere to treatment. Analysis of the data supplied in the supplementary material showed that a maximum of 12.4% of participants adhered fully to the treatment in the REGAIN study. Analysis of adherence in cancer, cardiovascular disease and diabetes shows that a 77% adherence rate is considered to be good. Consequently, the REGAIN study deviated from the intended intervention in at least 87.6% of cases. The fact that so many patients deviated from the intended intervention in both the REGAIN and the ReCOVER study, suggests that both interventions were not effective and / or acceptable to patients. As a matter of fact, patients might have deviated from these treatments because they were harmful.

According to Taylor and Gorman, "Cochrane reviewers perform selective outcome reporting bias assessments very poorly" (p. 6 [38]). Something similar was concluded by Saric et al. who found that "at least 60% of judgments for risk of selective reporting bias of trials in analyzed Cochrane reviews were not in line with the Cochrane Handbook" (p. 53 [57]). The systematic review by Zeraatkar et al. ignored the fact that the ReCOVER and the REGAIN study both resorted to selective reporting of their objective outcome measure. They also ignored the fact that the objective outcome measure from the ReCOVER study showed that CBT did not lead to objective improvement.

If there is missing data in 20% of cases or more, "then the study is highly biased" (p. 3 [37]). The percentage of missing

data in the REGAIN study was 20.5% after three months, 24.5% after six months and 27.2% after 12 months. Why this was also ignored by this systematic review is unclear. Moreover, the study itself noted that “the trend towards the benefit of the REGAIN intervention was consistent for most of the outcome measures” (p. 12 [11]). Which is an indirect way of noting that their treatment was not effective without using those words.

The systematic review also concluded that exercise therapy for long COVID is safe. They based this on the REGAIN study and a study by Tryfonos et al. which concluded that “given that exercise was generally well tolerated, guidelines cautioning against exercise in similar populations may need to be revised” (p. 15 [48]). What they should have concluded however, is that they did not comprehensively investigate PEM as they set out to do as they did not investigate at baseline if patients objectively suffered from it. The REGAIN study did not investigate that either. Also, the objective data from Tryfonos et al. shows that patients who were suffering from PEM, were not included in their study. The evaluation and the analysis of the efficacy and safety of exercise therapy for patients with ME/CFS on the sports medical department in a Dutch hospital, found that in ME/CFS, a disease which is characterised by PEM, exercise at 50 to 60% of maximal workload, was not tolerated by up to 80% of patients [53,54]. The study by Tryfonos et al. included high intensity exercise training which consisted of 5×1 minute cycling at 90% maximum workload. By definition, if patients can do that then they do not suffer from PEM [58,59]. And something similar applies to the REGAIN study. This means that those two studies labelled exercise safe for long COVID patients with PEM, yet patients in these studies did not suffer from PEM. Consequently, both studies do not provide any evidence that (gradually increasing) exercise for long COVID patients with PEM is safe.

Why all this was ignored by the systematic review is unclear. In conclusion, the systematic review and the studies used to claim that CBT are safe and effective, do not provide any evidence to support that claim. Moreover, the systematic review and those studies, ignore all the evidence to the contrary.

Finally, since the publication of the systematic review by Zeraatkar et al., the SIPCOV study by Nerli et al. [59] has been published. Nerli et al. concluded that a brief outpatient rehabilitation based on a cognitive and behavioral approach is a safe and effective treatment for long COVID. They base that on an improvement of 9.2 points on their primary outcome (physical functioning). Yet according to the study itself, a difference of 10 points or more was needed for clinical significance. Why the study ignored that, in a similar manner to the REGAIN study, instead of concluding that their treatment was not effective, is unclear. Other important issues with the study by Nerli et al. are discussed in more detail in Appendix A of this article.

Conclusion

The systematic review by Zeraatkar et al. concluded that CBT and a group physical and mental health rehabilitation programme, are safe and effective treatments for long COVID. Our analysis does not lend any support to use CBT or a group physical and mental health rehabilitation programme for long COVID. Nor does it lend any support for the claim that these treatments are safe for long COVID patients with PEM.

Declarations

Author contributions: Conceptualization, M.V.; methodology, M.V. and F.V.-N.; validation, M.V. and F.V.-N.; writing—original draft preparation, M.V.; writing—review and editing, M.V. and F.V.-N.; supervision, M.V. and F.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.

Conflicts of interest: The authors declare no conflict of interest.

References

1. WHO Health Emergencies Programme, WHO COVID-19 dashboard. 2024.
2. COVID-19 CORONAVIRUS PANDEMIC. Worldometers, the world's leading aggregator of coronavirus data. 2024.
3. Davis HE, Assaf GS, McCorkell L, Wei H, Low RJ, Re'em Y, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. *EClinicalMedicine*. 2021; 38: 101019.
4. Voruz P, Assal F, Péron JA. The economic burden of the post-COVID-19 condition: Underestimated long-term consequences of neuropsychological deficits. *J Glob Health*. 2023; 13: 03019.
5. Al-Aly Z, Davis H, McCorkell L, Soares L, Wulf-Hanson S, Iwasaki A, et al. long COVID science, research and policy. *Nat Med*. 2024; 30: 2148-2164.
6. Soriano JB, Murthy S, Marshall JC, Relan P, Diaz JV. WHO Clinical Case Definition Working Group on Post-COVID-19 Condition. A clinical case definition of post-COVID-19 condition by a Delphi consensus. *Lancet Infect Dis*. 2022; 22: e102-7.
7. Zeraatkar D, Ling M, Kirsh S, Jassal T, Shahab M, Movahed H, et al. Interventions for the management of long COVID (post-covid condition): living systematic review. *BMJ*. 2024; 387: e081318.
8. NICE, Myalgic Encephalomyelitis (or Encephalopathy)/Chronic Fatigue Syndrome: Diagnosis and Management. NICE Guideline [NG206]. 2021.
9. NICE ME/CFS Guideline Outlines Steps for Better Diagnosis and Management. 2021.
10. Kuut TA, Müller F, Csorba I, Braamse A, Aldenkamp A, Appelman B, et al. Efficacy of Cognitive-Behavioral Therapy Targeting Severe Fatigue Following Coronavirus Disease 2019: Results of a Randomized Controlled Trial. *Clin Infect Dis*. 2023; 77: 687-695.
11. McGregor G, Sandhu H, Bruce J, Sheehan B, McWilliams D, Yeung J, et al. Clinical effectiveness of an online supervised group physical and mental health rehabilitation programme for adults with post-covid-19 condition (REGAIN study): multicentre randomised controlled trial. *BMJ*. 2024; 384: e076506.
12. Fears R, Griffin G, Larhammar D, Ter Meulen V, van der Meer JWM. Assessing and regulating homeopathic products. *J Intern Med*. 2017; 282: 563-565.
13. Cuijpers P, Cristea IA. How to prove that your therapy is effective, even when it is not: a guideline. *Epidemiol Psychiatr Sci*. 2016; 25: 428-435.
14. De ReCOVer studie. Moena COVID-19. 27-08-2021. 2024.
15. Fordham B, Sugavanam T, Edwards K, Stallard P, Howard R, das Nair R, et al. The evidence for cognitive behavioural therapy in

- any condition, population or context: a meta-review of systematic reviews and panoramic meta-analysis. *Psychol Med.* 2021; 51: 21-29.
16. Cohen J. Statistical power analysis for the behavioral sciences. 1988; 111: 479.
 17. Juandi D, Kusumah YS, Tamur M, Perbowo KS, Siagian MD, Sulastri R, et al. The Effectiveness of Dynamic Geometry Software Applications in Learning Mathematics: A Meta-Analysis Study. *International Journal of Interactive Mobile Technologies (IJIM).* 2021; 15: 18–37.
 18. Drew C. 16 Selection Bias Examples. *Helpful Professor.* 2023.
 19. Kaźmierczak I, Zajenkowska A, Rogoza R, Jonason PK, Ścigała D. Self-selection biases in psychological studies: Personality and affective disorders are prevalent among participants. *PLoS ONE.* 2023; 18: e0281046.
 20. Norman GR, Sloan JA, Wyrwich KW. Interpretation of changes in health-related quality of life: the remarkable universality of half a standard deviation. *Med Care.* 2003; 41: 582-92.
 21. Beaton DE. Simple as possible? Or too simple? Possible limits to the universality of the one half standard deviation. *Med Care.* 2003; 41: 593-6.
 22. Rebelo P, Oliveira A, Andrade L, Valente C, Marques A. Minimal Clinically Important Differences for Patient-Reported Outcome Measures of Fatigue in Patients With COPD Following Pulmonary Rehabilitation. *Chest.* 2020; 158: 550-561.
 23. 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: 171-6.
 24. Williams LR, Isaacson-Barash C. Three Cases of Severe ME/CFS in Adults. *Healthcare (Basel).* 2021; 9: 215.
 25. Pendergrast T, Brown A, Sunnquist M, Jantke R, Newton JL, Strand EB, et al. Housebound versus non housebound patients with myalgic encephalomyelitis and chronic fatigue syndrome. *Chronic Illn.* 2016; 12: 292–307.
 26. Baxter H, Speight N, Weir W. Life-Threatening Malnutrition in Very Severe ME/CFS. *Healthcare.* 2021; 9: 459.
 27. Dafoe W. Extremely Severe ME/CFS-A Personal Account. *Healthcare (Basel).* 2021; 9: 504.
 28. Kuut TA, Müller F, Aldenkamp A, Assmann-Schuilwerpe E, Braamse A, Geerlings SE, et al. A randomised controlled trial testing the efficacy of Fit after COVID, a cognitive behavioural therapy targeting severe post-infectious fatigue following COVID-19 (ReCOVer): study protocol. *Trials.* 2021; 22: 867.
 29. Kuut TA, Müller F, Nieuwkerk P, Rovers CP, Knoop H. Reply to Crawford and Biere-Rafi et al. *Clin Infect Dis.* 2023; 77: 1075-1077.
 30. Rongen-van Dartel SA, Repping-Wuts H, van Hoogmoed D, Knoop H, Bleijenberg G, van Riel PL, et al. 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: 852-60.
 31. Kuut TA, Müller F, Csorba I, Braamse AMJ, Nieuwkerk P, Rovers CP, et al. Positive Effects of Cognitive-Behavioral Therapy Targeting Severe Fatigue Following COVID-19 Are Sustained Up to 1 Year After Treatment. *Clin Infect Dis.* 2024; 78: 1078-1079.
 32. Butler S, Chalder T, Ron M, Wessely S. Cognitive behaviour therapy in chronic fatigue syndrome. *J Neurol Neurosurg Psychiatry.* 1991; 54: 153-8.
 33. Larzelere RE, Kuhn BR, Johnson B. The intervention selection bias: an underrecognized confound in intervention research. *Psychol Bull.* 2004; 130: 289-303.
 34. Vink M, Vink-Niese F. Using Exercise Therapy for long COVID Without Screening for Post-Exertional Symptom Exacerbation Potentially Increases the Risks for Patients Who Suffer from it: A Reanalysis of Three Systematic Reviews. *Res Inves Sports Med.* 2024; 10: RISM.000748.
 35. Haynes RB, Taylor DW, Sackett DL, Gibson ES, Bernholz CD, Mukherjee J. Can simple clinical measurements detect patient noncompliance? *Hypertension.* 1980; 2: 757–764.
 36. Bullard T, Ji M, An R, Trinh L, Mackenzie M, Mullen SP. A systematic review and meta-analysis of adherence to physical activity interventions among three chronic conditions: cancer, cardiovascular disease, and diabetes. *BMC Public Health.* 2019; 19: 636.
 37. Heneghan C, Goldacre B, Mahtani KR. Why clinical trial outcomes fail to translate into benefits for patients. *Trials.* 2017; 18: 122.
 38. Taylor NJ, Gorman DM. Registration and primary outcome reporting in behavioral health trials. *BMC Med Res Methodol.* 2022; 22: 41.
 39. Kasy Maximilian. Selective publication of findings: Why does it matter, and what should we do about it?, *MetaArXiv xwngs*, Center for Open Science. 2019.
 40. Chan AW, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA.* 2004; 291: 2457-65.
 41. Pickett JT, Roche SP. Questionable, Objectionable or Criminal? Public Opinion on Data Fraud and Selective Reporting in Science. *Sci Eng Ethics.* 2018; 24: 151-171.
 42. Olesen AV, Oddershede L, Petersen KD. Health-related quality of life in Denmark on a relative scale: mini-catalogue of mean EQ-5D-3L index scores for 17 common chronic conditions. *Nordic Journal of Health Economics.* 2016; 4: 44-56.
 43. Falk Hvidberg M, Brinthe LS, Olesen AV, Petersen KD, Ehlers L. The Health-Related Quality of Life for Patients with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). *PLoS One.* 2015; 10: e0132421.
 44. Wilson C. Exercise programme helps people with long COVID, but it's no panacea. 2024. <https://www.newscientist.com/article/2416283exercise-programme-helps-people-with-long-COVID-but-its-nopanacea/> (accessed on 16 December 2024).
 45. Tuller D. Trial By Error: Dutch CBT Study for long COVID Proves that Unblinded Studies with Subjective Outcomes Generate Positive Reports. *Virology blog.* 2023.
 46. Vink M, Vink-Niese A. Could Cognitive Behavioural Therapy Be an Effective Treatment for long COVID and Post COVID-19 Fatigue Syndrome? Lessons from the Qure Study for Q-Fever Fatigue Syndrome. *Healthcare (Basel).* 2020; 8: 552.
 47. Lilienfeld SO, Ritschel LA, Lynn SJ, Cautin RL, Latzman RD. Why Ineffective Psychotherapies Appear to Work: A Taxonomy of Causes of Spurious Therapeutic Effectiveness. *Perspect Psychol Sci.* 2014; 9: 355-87.
 48. Tryfonos A, Pourhamidi K, Jörnåker G, Engvall M, Eriksson L, El-hallos S, et al. Functional Limitations and Exercise Intolerance in Patients With Post-COVID Condition: A Randomized Crossover

- Clinical Trial. JAMA Netw Open. 2024; 7: e244386.
49. Keller B, Receno CN, Franconi CJ, Harenberg S, Stevens J, Mao X, et al. Cardiopulmonary and metabolic responses during a 2-day CPET in myalgic encephalomyelitis/chronic fatigue syndrome: translating reduced oxygen consumption to impairment status to treatment considerations. *J Transl Med.* 2024; 22: 627.
 50. Dehlia A, Guthridge MA. The persistence of myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) after SARS-CoV-2 infection: A systematic review and meta-analysis. *J Infect.* 2024; 89: 106297.
 51. General Medical Council. First Do No Harm: Enhancing Patient Safety Teaching in Undergraduate Medical Education. 2021.
 52. 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. 2019.
 53. Van Berkel S, Brandon T, van Enst GC. Reactivering van patiënten met chronische vermoeidheid middels 'graded exercise therapy' met minimale directe begeleiding (2005–2010). *Sport Geneesk.* 2012; 3: 6–11.
 54. Vink M, Vink-Niese A. The draft updated NICE guidance for ME/CFS highlights the unreliability of subjective outcome measures in non-blinded trials. *J. Health Psychol.* 2022; 27: 9–12.
 55. Howard GS. Response-Shift Bias: A Problem in Evaluating Interventions with Pre/Post Self-Reports. *Evaluation Review.* 1980; 4: 93-106.
 56. Ghatineh S, Vink M. FITNET's Internet-Based Cognitive Behavioural Therapy Is Ineffective and May Impede Natural Recovery in Adolescents with Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. A review. *Behav Sci.* 2017; 7: 52.
 57. Saric F, Barcot O, Puljak L. Risk of bias assessments for selective reporting were inadequate in the majority of Cochrane reviews. *J Clin Epidemiol.* 2019; 112: 53-58.
 58. Charlton BT, Goulding RP, Jaspers RT, Appelman B, van Vugt M, Wüst RCI. Skeletal muscle adaptations and post-exertional malaise in long COVID. *Trends in Endocrinology & Metabolism.* 2024.
 59. Nerli TF, Selvakumar J, Cvejic E, Heier I, Pedersen M, Johnsen TL, et al. Brief Outpatient Rehabilitation Program for Post-COVID-19 Condition: A Randomized Clinical Trial. *JAMA Netw Open.* 2024; 7: e2450744.
 60. Wyller V. Short-time Intervention in Post-Covid Syndrome Syndrome (SIPCOV): A Pragmatic Randomised Controlled Trial (SIPCOV). *ClinicalTrials.gov.* 2025.
 61. Surawy, C, Hackmann, A, Hawton, K, Sharpe, M. Chronic Fatigue Syndrome: A cognitive approach. *Behav Res Ther.* 1995; 33: 535–544.
 62. 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: 507-17.
 63. 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.
 64. Sunnquist M, Jason LA. A reexamination of the cognitive behavioral model of Chronic Fatigue Syndrome. *J Clin Psychol.* 2018; 74: 1234–1245.
 65. Thoma M, Froehlich L, Hattesoehl DBR, Quante S, Jason LA, 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.
 66. Geraghty K, Jason LA, 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.
 67. Institute of Medicine (IOM). Committee on the Diagnostic Criteria for Myalgic Encephalomyelitis/Chronic Fatigue Syndrome; Board on the Health of Select Populations. *Beyond Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: Redefining an Illness*; National Academies Press: Washington, DC, USA. 2015.
 68. Dutch Health Council. To the President of the Lower House of the States-General No. 2018. The Hague 19 March Health Council | No. Available online: https://huisartsvink.files.wordpress.com/2021/01/gezondheidsraad-kernadvies_me_cvs.pdf (accessed on 3 January 2025).
 69. Chalder T. Rehabilitation Based on Cognitive Behavioral Model for Post-COVID-19 Condition. *JAMA Netw Open.* 2024; 7: e2450756.
 70. Jacobsen EL, Bye A, Aass N, Fosså SD, Grotmol KS, Kaasa S, et al. Norwegian reference values for the Short-Form Health Survey 36: development over time. *Qual Life Res.* 2018; 27: 1201-1212.

Appendix A

Appendix A.1. Important issues with the primary outcome of the SIPCOV study [59].

Nerli et al. published the SIPCOV study (Short-time behavioural intervention in post-Covid syndrome) in December 2024. The SF-36 physical functioning score directly at the end of treatment (T1) was its primary outcome measure. Nerli et al. concluded that “the SF-36-PFS improved statistically and clinically significantly in the intervention group” and the score “difference between groups [was] 9.2” (p. 1 [59]). Based on that, the study concluded that a “brief outpatient rehabilitation [with on average 4 outpatient interventions with 2 to 6 weeks in between] based on cognitive and behavioral approach is effective and safe for patients with PCC” (p. 1 [59]). However, before the study started, a difference of 10 points or more was considered to be clinically significant. Why the study ignored that and concluded that their treatment was effective, even though the difference was less than 10 at its primary outcome (9.2 at T1, end of treatment) and 9.0 at T2 (12 months after enrollment), is unclear.

Appendix A.2. Issues with the rehabilitation program

The cognitive and behavioral approach of the study is based on the Cognitive Activation Theory of Stress (CATS) which targets perpetuating factors for patients which, according to the authors might be beneficial because “factors such as loneliness, neuroticism, worrying tendencies, depression, anxiety, and psychological distress, have been associated with persistent post-COVID-19 symptoms and disability. In PIFS [Post Infective Fatigue Syndrome (PIFS) commonly referred to as ME/CFS], subconscious expectancies, worrying tendencies, and associative learning processes have been proposed as perpetuating factors” (p. 2 [59]). The theoretical framework of CATS is similar to the false illness beliefs that forms the basis of the CB-model and CBT for ME/CFS and one of the authors of Nerli et al. (Wyller) is the main proponent of that in Norway. According to the registration of the study at ClinicalTrials.gov, the treatment that will be studied is “Rehabilitation based on CBT principles” (p. 5 [60]). Additionally, in the first stage of the rehabilitation program, patients were seen by a physician who “provided cognitive reassurance that bodily symptoms do not necessarily indicate a disease but rather a disorder that is temporary and amendable” (p. 3 [59]). Evidence for this assumption is not provided and all the evidence to the contrary is ignored. Yet the cognitive behavioural model [61,62] and its assumption that there is no underlying disease in ME/CFS and that CBT is an effective treatment for ME/CFS, has been debunked by a number of studies [63-66], as well as by the Institute of Medicine, the Dutch Health Council and by NICE [8,9,67,68].

Appendix A.3. Issues with subjective outcomes in non-blinded studies

Another problem of the study is the fact that it was a non-blinded study that used one subjective primary outcome (physical functioning). Relying on subjective outcomes in those studies is very prone to all sorts of different forms of bias. On top of that, the study also used a badly designed (non-active) control group. The combination of non-blinding with subjective outcomes and a non-active control group can lead to the aforementioned erroneous inference of efficacy in its absence [13,47] so that one could get the impression of efficacy, even though patients have not benefited from the treatment. The

only way to correct for it, as discussed earlier, is by using one or more objective outcome measures [47].

Moreover, according to their Supplementary Online Content, participants were told during treatment that “progress is not monitored in more or less symptoms, but more or less actions” and “a decrease in symptoms itself is not a goal of the treatment”. Participants were also told that “aiming to foster a belief that patients’ bodies have the potential for recovering and that the disorder is temporary and amendable” (p. 2,3 [59]) without providing any evidence that patients foster the wrong belief. Moreover, these researchers are introducing an extra form of bias (response shift bias) into the study [55] by trying to change the beliefs patients have. The only way then to find out if a treatment in a non-blinded trial has led to recovery or an improvement, is by using one or more objective outcome measures instead of relying on one subjective primary outcome measure. The study acknowledged that “objective measures of physical and social function (eg, steps per day and work attendance) were not included but could have yielded valuable information” (p. 12 [59]). Yet the study didn’t use them because according to the authors, “the diagnosis of PCC is based on patients’ self-report, patient-reported outcomes appear to be most relevant for evaluation of a rehabilitation program” (p. 12 [59]). However, if the basis of the rehabilitation program is that symptoms are caused by wrongful assumptions and beliefs and that patients interpret their symptoms incorrectly as signs of disease. Then 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.

Additionally, professor Chalder, one of the leading proponents of CBT for ME/CFS, concluded in an invited commentary about the study by Nerli et al., that “future studies would benefit from objective behavioral and physiologic assessments” (p. 3 [69]). She also noted that there is a problem with the non-active control group when she stated that “controlling for health care professional time and attention...[is needed to]...provide clarity about active ingredients” (p. 3 [69]).

Appendix A.4. Selection bias.

A total of 473 patients with mild to moderate long COVID were assessed for eligibility (n=364 physician referred; n=109 self-referred) and 9 did not meet the inclusion criteria. 464 were invited to participate yet 130 were not interested and 18 did not return the consent form. A further two were excluded because they were asymptomatic. Why asymptomatic patients could meet the inclusion criteria is unclear. It also means that 32% (148/462) of the 462 invited patients (the 2 asymptomatic ones excluded), did not want to take part. Together with the fact that 109 patients were self-referred, suggests that selection bias played an important role in this study. Consequently, the results might not be generalisable to long COVID patients in general.

Appendix A.5. Definition of recovery.

According to the study, 17% in the treatment group and 20% in the control group already fulfilled the recovery threshold at baseline. That is before receiving any treatment. Patients who are already deemed to be recovered at baseline should be excluded from a properly conducted study. At T1, these percentages were 50% (CATS) and 32% (CAU). Consequently, 18% were

deemed to be recovered at T1 due to CATS. However, “recovery of physical function was defined as a SF-36-PFS score at the population norm (approximately 85) or higher” (p. 4 [59]). The reference used for that is a study by Jacobsen et al. which investigated the Norwegian reference SF-36-PFS scores. It is unclear why a score of approximately 85 or higher was chosen by Nerli et al. for patients with a mean age of 43. Because according to Jacobsen et al., the mean score for Norwegian people aged 40 to 49 is 90.7 for women and 92.8 for men [70]. It is also not clear why recovery was not defined in the protocol, which was posted on the ClinicalTrials.gov website [60] and as a supplement to the article by Nerli et al. [59]. Yet it was not published in a peer reviewed journal before the start of the trial, as is the norm these days.

Additionally, the Norwegian population aged 40 to 49 consists mainly of healthy people but also includes people with chronic diseases which lowers the average physical functioning score. The typical person who falls ill with a post infectious disease on the other hand is someone who was fit, well and active before falling ill. Consequently, their score before falling ill would have been higher than the average score of the population. This also means that the recovery score for long COVID patients is higher than 90.7 for women and 92.8 for men. As a consequence of all that, their definition of recovery was a post-hoc definition which did not reflect recovery.

Appendix A.6. Post-exertional malaise (PEM)

The study used the DePaul SymptomQuestionnaire-2 to diagnose PEM. Nerli et al. [59] concluded that there was “an overall significant decrease in PEM in the intervention group, [and] it does not seem relevant to take PIFS diagnosis or the presence of PEM into account when developing an individualized rehabilitation plan. This notion fits well with another study showing that exercise is well tolerated independent of PEM in patients with PCC” (p. 11 [59]). The study they refer to is the study by Tryfonos et al. [48] which we analysed earlier. Our analysis however showed that that study does not provide any evidence that exercise is well tolerated by long COVID patients who suffer from PEM.

Moreover, Nerli et al. [59] noted themselves that “no direct precautions were taken in regard to PEM; patients were always advised to continue exploring activities. Type of activity was not standardized, but each participant was able to choose the activity and homework based on their values, goals, and interests, as long as the activity was suitable, feasible, and enjoyable. Graded exercise was an opportunity for those who wanted it” (p. 11 [59]).

The study did not document if patients were exploring activities, were exercising, or were using graded exercise or not. Nor did it investigate what graded exercise consisted of if patients used it and at what sort of level, intensity, and frequency it was used. Nerli et al. also did not investigate if patients had increased their level of activity or not and if so by how much, et cetera. Moreover, by not standardising activity or structuring and standardising graded exercise, the researchers had no insight into what activity or graded exercise in their own study actually entailed. This made it even more impossible to make any claims about exercise and its safety and efficacy. The aforementioned research by the Oxford Brookes University [52] (n=2274) on the other hand, highlighted the detrimental effect of gradually increasing exercise/activity in a post-infectious disease (ME/CFS) which is characterised by PEM. Put differently, if gradually increasing activity/exercise has a positive instead of a negative effect on PEM, then patients have been wrongly diagnosed with it and they suffer from deconditioning instead. Consequently, concluding that “it does not seem relevant to take PIFS diagnosis or the presence of PEM into account when developing an individualized rehabilitation plan” is an unsubstantiated and potentially harmful claim for long COVID patients who do suffer from PEM.

Appendix A.7. Missing data.

According to the statistical analysis section of the study, “assuming an SD of 25 in the population under study and a drop-out rate of 20%, a target inclusion of 310 participants would yield a power of 90% ($\alpha=.05$) to detect a small to medium effect size (Cohen $d=0.2-0.5$)” (p. 5 [59]). The study included 314 participants but only “227 patients completed the entire follow-up period, resulting in a total of 35% of individuals having incomplete data” (p. 5 [59]). This is well above the target of 20% which might undermine the statistical power of the study. Additionally, and as mentioned before, if there is missing data in more than 20% of cases, then a study is highly biased [37].

Appendix A.8. Conclusion.

In conclusion, the methodologically challenged study by Nerli et al. does not provide any evidence that a brief outpatient rehabilitation program based on a cognitive and behavioral approach is a safe and effective treatment for long COVID.