Response to CDC Draft Systematic Review Report for Diagnosis and Treatment of ME/CFS

#MEAction Draft Response, August 2, 2021

Note: This is #MEAction's draft public comment response to the CDC draft Systematic Review Report for Diagnosis and Treatment of ME/CFS which can be found here: https://www.regulations.gov/document/CDC-2021-0053-0001

We will continue to revise and update this draft based on further review and feedback as appropriate and feasible. We will submit the final version on the public comment deadline of August 16, 2021 at https://www.regulations.gov/document/CDC-2021-0053-0002.

You can have your name added as a co-signer to #MEAction's response by completing this form by **August 15, 2021**: https://act.meaction.net/page/31886/petition/1

Share feedback on any area for improvement of our draft response here: https://airtable.com/shrJCoP6yodsIfUPJ

#MEAction public recommendation

We find the conclusions of the CDC systematic review on the treatment of ME/CFS to be fundamentally flawed, and we strongly urge that it not be published. The GET and CBT studies included in the systematic review do not reference people with ME/CFS as defined by CDC, and therefore cannot apply to people with ME/CFS as defined by CDC. The review confuses the CDC's public communication about ME/CFS, and most importantly puts people with ME/CFS at risk of harm from inappropriate treatments.

Recommendation Co-signers

The following **[insert number]** members of the myalgic encephalomyelitis (ME) community support #MEAction's criticism of this systematic review as fundamentally flawed and urge the CDC not to publish or proceed with finalizing it in any form:

Full Name	Full Name	Full Name
Full Name	Full Name	Full Name
Full Name	Full Name	Full Name
Full Name	Full Name	Full Name
Full Name	Full Name	Full Name
Full Name	Full Name	Full Name...

Table of Contents

Introduction

Background

This review, conducted by the Pacific Northwest Evidence-based Practice Center of Oregon Health & Science University, follows a 2014 review conducted by the same organization for the Agency for Healthcare Research and Quality (Smith, et al., 2014). Following criticism, AHRQ issued an addendum in 2016 that reanalyzed the evidence and downgraded the original conclusions (NCBI, 2016). The current review suffers from a number of the same flaws as the original 2014 review.

The systematic review's misleading conclusion

The CDC systematic review (Chou, et al., 2020) of graded exercise therapy (GET) and cognitive behavioral therapy (CBT) for myalgic encephalomyelitis/ chronic fatigue syndrome (ME/CFS) included no studies that require post-exertional malaise (PEM),[1] one of three "core symptoms" of ME/CFS, according to CDC (CDC-a). Yet the CDC systematic review concludes that these therapies will likely benefit a population they have not, in fact, examined.

Studies within this review also contain methodological flaws of serious concern, including high risks of bias, and a low strength of evidence. The review authors collapse these concerns into the cursory caveat that "methodological and other limitations (imprecision, inconsistency, uncertain generalizability) precluded strong conclusions" (Chou, et al., 2020, piii). This is insufficient, as the aforementioned flaws should preclude *any* conclusions about GET and CBT for people with ME/CFS.

Publishing this review will lead to further confusion about the appropriate treatment for people diagnosed with ME/CFS as CDC defines it. The review, if published, will put an already vulnerable group of patients at risk of harm.

How this review will harm patients

Healthcare providers, who may have received medical education from CDC on ME/CFS diagnosis, will assume that this systematic review provides useful conclusions on effective treatments that are applicable to their patients with ME/CFS. They will conclude that treatment options for ME/CFS remain limited, that the only thing they can offer patients is CBT and GET,

---

[1] The GETSET trial used the NICE 2007 criteria, which requires either PEM or post-exertional fatigue (PEF), the main difference between the two being that PEF refers to the exacerbation of *fatigue* following exertion, whereas PEM refers to the exacerbation of *symptoms in general* following exertion. NICE has just revised its draft guidelines to require PEM (which they refer to as post-exertional symptom exacerbation, or PESE), saying that the Fukuda criteria "may not accurately represent the ME/CFS population and that people experiencing PEM/PESE are likely to respond differently to treatment than those who do not experience PEM/PESE" (NICE, 2020, sect 3.2.1).

and that those treatments more effectively reduce symptoms than resting or pacing. They will prescribe these treatments, believing they are executing the best, evidence-based practice. While they may not be able to offer "strong conclusions" to their patients that CBT and GET treatments will help, they can at least assure themselves and their patients that trying such treatments won't harm them.

It would be perfectly reasonable for healthcare providers to make such assumptions, but unfortunately they would be incorrect, both about the efficacy of the recommended treatments and about their lack of harm. If published, this review will contribute to the misconception that GET and CBT are appropriate treatments for people with ME/CFS.

GET and CBT are inappropriate ME/CFS treatments

GET and CBT interventions are not appropriate treatments as they have been applied to people with ME/CFS. If a "core symptom" of ME/CFS is post-exertional malaise (PEM)[2] --the worsening of symptoms following physical or cognitive exertion -- then it is an extraordinary claim that graded exercise and talk therapy would be effective treatments for patients. Therefore, the claim requires extraordinary evidence.

The rationale for CBT and GET as treatments for ME/CFS arises from a behavioral/ deconditioning or psychosocial "fear avoidance" model that claims the disease is perpetuated by patients' maladaptive beliefs about being ill. In this model, the patient's avoidance of activity results in deconditioning and therefore a worsening of symptoms. But this runs directly counter to the Institute of Medicine 2015 report on ME/CFS, which stressed that the condition is "a medical — not a psychiatric or psychological — illness" (NCBI, 2015), something the CDC repeats on their website, stating, "ME/CFS is a biological illness not a psychologic disorder" (CDC-b).

The CDC removed recommendations for GET and CBT as treatments for ME/CFS from its website in 2017. In 2020, the UK's National Institute for Health and Care Excellence (NICE) also removed GET from the new draft guidelines on ME/CFS and issued a sweeping warning against "any therapy based on physical activity or exercise as a treatment or cure for ME/CFS" and "any programme based on fixed incremental increases in physical activity or exercise, for example graded exercise therapy" (NICE, 2020, p28). The NICE guidelines allow CBT only as adjunctive psychological support for managing the distress of a chronic illness. Moreover, they explicitly warn that ME/CFS is not perpetuated by "abnormal" illness beliefs or behaviors, and that CBT must not be offered as a treatment or cure for ME/CFS (NICE, 2020, p34). Additional sources of expert opinion that CDC has cited in the past, such as the U.S. ME/CFS Clinician Coalition treatment recommendations, do not include CBT or GET and point to the 2015 IOM report as

---

[2] The CDC requires PEM as a core symptom of ME/CFS (CDC-a), as do all contemporary expert consensus definitions of the disease, which include CCC, ICC, IOM, and NICE 2020 draft guidelines. The final NICE guidelines will not be published until August 18, 2021—two days after the public comment deadline for this systematic review.

directly countering the incorrect understanding that GET and CBT are appropriate treatments for ME/CFS (Clinician Coalition, Treatment Recommendations).

Either these expert conclusions are all biased and are incorrectly constraining the already limited treatment options available for ME/CFS, or this systematic review fails to accurately review the evidence base in the appropriate context and with appropriate methodology. Expert opinion and this systematic review provide contradictory guidance regarding the best approach to treat ME/CFS.

Scope of this public comment

This review is fundamentally flawed, such that there is no clear way that it can be adequately revised within the scope of the CDC's current contract with the systematic review authors, the Pacific Northwest Evidence-based Practice Center (EPC).

Because of the extent of the problems with this systematic review, a line-by-line approach would not be feasible. There are aspects of this review that are deeply troubling, but which it is impossible to explore in sufficient depth at this time. Instead, these public comments are focused on the findings from the review on CBT and GET treatments, which have the potential to do the most harm to people with ME/CFS.

The public comment response below highlights four key arguments for why the findings from this review are fundamentally flawed and should not be published:

1. Applicability of Findings
2. Risk of Bias
3. Exclusion of Harms Evidence
4. Interpretations of Results

Note on terminology

The systematic review includes ten definitions that use the labels "chronic fatigue syndrome (CFS)," "myalgic encephalomyelitis (ME)," "myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS)," or "systemic exertion intolerance disease (SEID)." The review uses the term ME/CFS to refer to all the definitions.

The disease characterized by post-exertional malaise (PEM) and associated with neurological, immunological, autonomic, and energy production impairment is preferentially referred to as ME by #MEAction and by many people living with this disease.

For the purposes of responding to this systematic review, we will mostly use the term ME/CFS, not because this is our preferred term, or the most accurate description of the disease, but in order to clearly articulate the ways in which this review has come to inappropriate conclusions.

As will be discussed in more detail in the "Applicability" section, the CDC uses the IOM report to define and communicate what ME/CFS is to the public. Specifically, the CDC states that the hallmark symptom of PEM is required for an ME/CFS diagnosis and names PEM one of three 'core' symptoms. ME/CFS must therefore not be used in CDC publications to refer to a condition that does not require the presence of PEM.

This systematic review should have used a clearer analytical framework to differentiate studies that examine heterogeneous patient populations.. Another recent systematic review of physiotherapy (i.e., GET) for ME/CFS (Wormgoor and Rodenburg, 2021) utilized just such a framework to clearly communicate their findings across a range of studies that use different case definitions. Wormgoor and Rodenburg used the term "ME" to specify when PEM is a cardinal feature alongside other core symptoms; they used "CFS" when PEM or other core symptoms are optional features; and they used "Chronic Fatigue" (CF) when PEM was not accounted for at all (Wormgoor and Rodenburg, 2021). This CDC systematic review should have used terminology according to a consistent and clear analytical framework such as this. This and related issues are explored further in **Part One: Applicability**.

About #MEAction

The Myalgic Encephalomyelitis Action Network (#MEAction) is a patient advocacy organization working for better recognition, education, and research so that, one day, all people with ME will have access to compassionate and effective care.

Part I: Applicability of Findings

This review fails to adequately analyze and judge the applicability of the research evidence base; this is of particular concern when it comes to the review authors' insufficient handling of the hallmark symptom of ME/CFS: post-exertional malaise (PEM).[3] CDC considers PEM a required symptom of ME/CFS. The systematic review concludes that cognitive behavioral therapy (CBT) and graded exercise therapy (GET) are effective treatments for ME/CFS patients. However, this conclusion reflects a failed analytical framework that does not clearly define the relevant patient population of patients who have PEM.

The CDC commissioned this systematic review of research on ME/CFS treatments to inform development of federal guidelines for the treatment of ME/CFS. The review states that "Many case definitions for ME and CFS have been proposed," and that the "use of multiple case definitions for ME/CFS is an ongoing challenge in the field, as it has resulted in heterogeneous populations in the research literature" (Chou, et al., 2020, p1). Yet it does not acknowledge that the CDC's communication to the public and to healthcare providers about ME/CFS is explicitly based on the Institute of Medicine (IOM)[4] report, which makes clear that PEM is a required symptom for diagnosis (CDC-a). The IOM report concludes "There is sufficient evidence that PEM is a primary feature that helps distinguish ME/CFS from other conditions" (NAM, 2015, p86). The UK's National Institute for Health and Care Excellence (NICE) committee for ME/CFS also aligns with the IOM report and the CDC on this issue in their 2020 draft guidelines: they identified PEM[5] "as an essential symptom that is central to the diagnosis of ME/CFS" (NICE, 2020, p317).

Given the challenge of heterogeneous patient populations arising from differing case definitions in the research literature, it is the responsibility of the review authors to "make clear for which patients and which circumstances the review's conclusions can be used to make clinical… decisions," according to the AHRQ Methods Guide utilized by this review (Atkins, et al., 2020, p2). This review's imprecision about whether CBT and GET are effective treatments for patients with PEM — which is how the CDC has defined ME/CFS — is a fundamental flaw that undermines the validity and applicability of this report. The CDC and the systematic review apply the term ME/CFS in conflicting ways. The lack of clarity regarding which patient populations are referenced in the review is at the heart of the matter.

Of the ME, CFS and ME/CFS case definitions referenced in this review, all but one of the six most recent definitions designate PEM as a required symptom. The sole exception, the UK's NICE ME/CFS treatment guidelines, are now defunct; as of the 2020, they too require PEM.[6]

---

[3] The CDC characterizes post-exertional malaise as "the worsening of symptoms following even minor physical or mental exertion, with symptoms typically worsening 12 to 48 hours after activity and lasting for days or even weeks." (CDC-c)

[4] The Institute of Medicine has since been renamed the National Academy of Medicine.

[5] The committee's preferred term for PEM is post exertional symptom exacerbation (PESE).

[6] A draft of the NICE guidelines update was released in 2020. The final version of the NICE guidelines are to be released two days after this systematic review public comment period has ended (August 20, 2021). So we have referenced the publicly available draft version (NICE, 2020).

Older case definitions are overly inclusive

The older chronic fatigue syndrome definitions used to study GET and CBT treatments have significant issues. Notably, the Oxford criteria requires one symptom only: that of long-term, disabling fatigue of definite onset. The review authors state in the "Limitations" section that using the broadest CFS criteria, the Oxford case definition "classifies more patients with ME/CFS compared with more current case definitions, potentially resulting in misclassification and misleading results" (Chou, et al., 2020, p159). Indeed the review notes that Brurberg, et al. (2014) found that using the Oxford definition would identify 15 times as many patients as the Canadian Consensus Criteria, which requires PEM (Chou, et al., 2020, p1).

The Oxford definition is not the only problematic older case definition; these issues of applicability are also relevant for the CDC's 1994 Fukuda definition of chronic fatigue syndrome. The IOM Report (NAM, 2015, p48) points out that "The Fukuda definition does not require what some consider core symptoms of ME/CFS, such as post-exertional malaise (PEM) and neurocognitive symptoms. The definition has been criticized for being overly inclusive."

The UK's NICE guidelines committee has also articulated concerns about the generalizability of studies using the CDC's 1994 Fukuda definition for the ME/CFS population because it did not include PEM as a required symptom for diagnosis. They concluded that the Fukuda criteria did not accurately represent the ME/CFS population and that people experiencing PEM are likely to respond differently to treatment than those who do not experience PEM. They found that the Fukuda studies were therefore not generalizable to the ME/CFS population and downgraded the strength of the evidence from these studies for "population indirectness" (NICE, 2020, p317).

At a minimum, the review should have prioritized providing clear analysis and conclusions regarding whether CBT and GET are appropriate and effective treatments for patients who have the symptom of PEM, which is now considered central, even definitional to ME/CFS. A systematic review that doesn't explicitly address this question is not only without value, but may subject ME/CFS patients to inappropriate treatments that could harm them. The review's conclusion that CBT and GET are effective treatments for ME/CFS patients but that "uncertain generalizability...precluded strong conclusions" is an abdication of the reviewers' duty to accurately summarize the applicability (i.e., generalizability) of the body of evidence for the patient population in question, which is ME/CFS patients with PEM.

High clinical heterogeneity is not adequately addressed

Taking clinical heterogeneity into account is not an unreasonable expectation for a systematic review of ME/CFS treatments. Another recent systematic review of physiotherapy (i.e., GET) treatments for ME/CFS by Wormgoor and Rodenburg (2021) did just that. The authors created a

specific framework of analysis to differentiate between ME/CFS patient populations in the evidence base according to whether the symptom of PEM was required (ME), optional (CFS) or not required (CF).[7] Wormgoor and Rodenburg analyzed the evidence according to these categories and concluded that "Currently, there is no scientific evidence when it comes to effective physiotherapy for ME patients. Applying treatment that seems effective for CF or CFS patients may have adverse consequences for ME patients and should be avoided" (Wormgoor and Rodenburg, 2021).

Whether the authors of the CDC systematic review agree with the conclusions of the Wormgoor and Rodenburg systematic review, it should be apparent that their analytical framework regarding the referenced patient populations was clear, whereas the CDC systematic review's conclusion inappropriately conflates multiple patient populations. This also demonstrates that the CDC's choice to ignore diagnostic specificity altered the purported finding of "effectiveness" for CBT and GET in ME. By relying on outdated, overly broad definitions of CFS that don't require PEM, the CDC systematic review misleads the reader with conclusions that may be immaterial to this patient population. .

The systematic review's insistence on using ME/CFS to encompass any and all medically unexplained chronic fatigue ultimately makes its analysis unworkable. Instead of the clarity and precision of the language of Wormgoor and Rodenburg, the CDC systematic review unhelpfully offers only repeated statements of uncertainty about the applicability of their findings:

- "the applicability of findings to patients with severe ME/CFS diagnosed using more current, specific case definitions was uncertain." (ii)
- "the applicability of findings to more disabled populations with ME/CFS and those diagnosed using more current ME/CFS criteria was uncertain." (149)
- "The applicability of results to patients with more severe ME/CFS symptoms also remains unclear." (149)
- "The applicability of findings to patients with severe post-exertional fatigue is also uncertain… the applicability of findings to patients diagnosed using more recent, specific ME/CFS case definitions is uncertain." (149-150)
- "Our findings… noted uncertain applicability to patients diagnosed with case definitions other than the Oxford and Fukuda criteria." (151-152)

In place of acknowledging PEM's presence or absence, the reviewers devise a novel, euphemistic category of "more severe" or "more disabled" ME/CFS. The reviewers imply that ME/CFS exists on a spectrum of severity where Oxford's 'medically unexplained chronic fatigue' is "less severe" ME/CFS and all the recent definitions of ME/CFS developed by expert consensus that require PEM as a hallmark symptom is "more severe" ME/CFS (Chou, et al., 2020, p2). There is no evidence to support that this is the case. It is inappropriate to refer to

---

[7] Wormgoor and Rodenburg (2021) used the term "ME" to specify when PEM is a cardinal feature and the other core symptoms are present as well. They used "CFS" when PEM or other core symptoms are optional features. And they used "Chronic Fatigue" (CF) whenever PEM is not accounted for at all.

one condition as a "more severe" version of another, simply to avoid identifying the parameters of both conditions. Severity cannot be a proxy for symptom presentation.

The review justifies its agnostic position about whether PEM is a required symptom of ME/CFS by pointing out that "studies have not been able to determine the accuracy of different ME/CFS case definitions, due to the lack of a reliable and universally accepted reference ('gold') standard" (Chou, et al., 2020, p1). But the lack of a discovery of a biomarker that could be used as a universal reference standard to test the accuracy of the various definitions does not therefore mean it is appropriate to assume that all definitions are the same, or that they are equally valid descriptions of ME/CFS. The review authors acknowledge at multiple points in the review that this is not a credible assumption and that the older CFS definitions utilized by most of the GET and CBT studies likely misclassify patients with ME/CFS, leading to misleading results.

The ME community has been pointing out these flaws since the prior AHRQ systematic review was published in 2014; yet the current systematic review continues to ignore these concerns to reproduce the same problems. The concerns raised by ME advocates in 2014 still apply today:

> *"Most fundamentally, this Evidence Review is grounded in the flawed assumption that eight CFS and ME definitions all represent the same group of patients that are appropriately studied and treated as a single entity or group of closely related entities. Guided by that assumption, this Evidence Review draws conclusions on subgroups, diagnostics, treatments and harms for all CFS and ME patients based on studies done in any of these eight definitions. In doing so, the Evidence Review disregards its own concerns as well as the substantial body of evidence that these definitions do not all represent the same disease and that the ME definitions are associated with distinguishing biological pathologies. It is unscientific, illogical and creates undue risk of harm to lump disparate patients together without regard to substantive differences in their underlying conditions.*
>
> *...The failure to differentiate between patients with the symptom of subjective unexplained fatigue on the one hand, and objective immunological, neurological and metabolic dysfunction on the other, calls into question the entire Evidence Review…"* (Dimmock et al, 2014, pp 1-2).

The review isn't appropriately restricted to reporting results directly applicable to the specific question

In their response to criticisms of the prior AHRQ systematic review, the authors state that they "appreciate the case definitions are very different" but that "there are very few trials and excluding some of these definitions would limit the evidence even further," and that they

believed this inclusive approach could "provide a foundation to determine what interventions may be effective" (Smith, 2014).

A research methodology cannot be justified by its ease or convenience of use alone. If Pacific Northwest Evidence-based Practice Center (EPC) believed CDC had charged them with a project they could not complete as outlined, they might have discussed this challenge with CDC and abandoned the project or changed its goals; they might have analyzed findings by sets of diagnostic criteria and, if there was then not enough data to draw conclusions, reported as much. Instead, the review explicitly states that their methodology was chosen in order to reach a desired conclusion: having enough data to generalize findings about potentially quite dissimilar patient populations.

It should be noted that the Oxford criteria used requires one symptom and one symptom only: that of long-term, severe fatigue. The Canadian Consensus Criteria, preferred by many researchers and clinicians to describe the disease, requires four core symptoms: pain, post-exertional malaise, fatigue, and sleep dysfunction, along with cognitive/neurological, neuroendocrine, autonomic, and/or immune symptoms. It is no wonder that the symptoms of general fatigue, as per Oxford, capture fifteen times as many patients as the Canadian Consensus (Chou, et al., 2020, p1). It is unreasonable to presume that studies on Oxford-diagnosed patients will be generalizable to those patients diagnosed under CCC.

An effective systematic review would have appropriately restricted itself to reporting results from studies directly applicable to the specific question at hand. And that question must be: what are effective treatments for ME/CFS patients as the CDC has described them? It is inappropriate to proceed as though there were any valuable insights to derive from Oxford studies. Treatments for medically unexplained chronic fatigue cannot be extrapolated to ME/CFS patients with PEM.

These studies using the Oxford definition make up approximately half of all the GET and CBT treatment trials reviewed. The decision to include them significantly skews the evidence base. It also threatens the validity, reliability and integrity of this systematic review's meta-analyses, which will be addressed in the **Part Four: Interpretations of Results** section of our response.

Given the dubious applicability of Oxford studies to ME/CFS patients with PEM, it would have been more appropriate to have excluded them entirely from this review. This was in fact done with the 2014 AHRQ review, which then issued a subsequent reanalysis in a 2016 Addendum that examined the effectiveness of GET and CBT when the Oxford studies were removed. AHRQ then downgraded their previous conclusions on the effectiveness of GET and CBT. When Oxford studies were excluded from the analysis, "the Addendum reported that there was insufficient evidence of effectiveness of GET on any outcome," and "the Addendum found insufficient evidence of effectiveness of CBT on function, employment and global improvement and a low strength of evidence of improved fatigue" (Dimmock and Spotila, 2016; NCBI, 2016).

Not only did the 2016 Addendum to the AHRQ systematic review downgrade the findings of effectiveness for CBT and GET, but the authors called for the Oxford case definition to no longer be used in research:

> *"The Oxford case definition is the least specific of the definitions and less generalizable to the broader population of patients with ME/CFS. It could identify individuals who have had 6 months of unexplained fatigue with physical and mental impairment, but no other specific features of ME/CFS such as post-exertional malaise which is considered by many to be a hallmark symptom of the disease. As a result, using the Oxford case definition results in a high risk of including patients who may have an alternate fatiguing illness or whose illness resolves spontaneously with time. In light of this, we recommended in our report that future intervention studies use a single agreed upon case definition, other than the Oxford case definition. If a single definition could not be agreed upon, future research should retire the use of the Oxford case definition. The National Institutes of Health (NIH) panel assembled to review evidence presented at the NIH Pathways to Prevention Workshop agreed with our recommendation, stating that the continued use of the Oxford case definition 'may impair progress and cause harm'" (NCBI, 2016; Green, et al., 2015, p1).*

The 2020 draft review authors fail to mention the 2016 reanalysis, or their previous conclusion that the Oxford definition should no longer be used. The continued use of Oxford studies in systematic reviews for ME/CFS continues to impair progress and cause harm. Stratifying outcome fundings by diagnostic criteria does not effectively solve this problem, which is explored in **Part Four: Interpretations of Results**.

Part II: Risk of Bias

In **Part One** of these public comments, the applicability of the evidence to people with ME/CFS is called into question. Even if the evidence base were applicable to people with ME/CFS, significant risk of bias calls into question the validity and reliability of the review.

In trials included in this systematic review, neither patients nor clinicians were blinded to their intervention group. In addition, these trials relied on subjective primary outcomes. According to Schulz and Grimes (2002), blinding "ensure(s) participants don't change their behaviour as a result of knowing what their group assignment is and do not report their subjective outcome measures differently as a result." Blinding is essential when outcomes are also subjective. According to Fiedorowicz, et al. (2021)*,* un-blinded assessors of subjective binary outcomes may "exaggerate odds ratios by an average of 36%." Lack of blinding in conjunction with subjective measures could invalidate a study's purported conclusions.

Review authors acknowledge the lack of participant and clinician blinding endemic in these trials in the *Discussion*: "It was not possible to blind patients and care providers to the exercise and CBT interventions, potentially resulting in bias...". They also state that the "inability to blind is of particular concern for subjective outcomes such as fatigue and function" (Chou, et al., 2020, p150).

Despite these acknowledgments, most exercise and CBT studies were assigned an overall "medium" risk of bias (RoB). According to the *AHRQ Methods Guide for Effectiveness and Comparative Effectiveness Reviews* (Viswanathan, et al., 2017), the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins, et al., 2021, ch8), and review authors' own "high risk of bias" definition (Chou, et al., 2020, p8), the RoB assignments understate studies' likely "high" risk of bias. Reconsidering these assignments is critical, as they will lead to revised conclusions about the validity and reliability of the studies' results, and the conclusions of the systematic review.

Lack of blinding and delivery of intervention further biases self-reported improvement

Lack of blinding creates participant expectations about an outcome that influence their own perceptions of their illness or behavior (Schultz, et al.,1995; Williams, et al., 2012). Meta-analyses of methodological quality issues in randomized controlled trials (RCTs) show that inadequate blinding creates statistically significant levels of bias, specifically yielding larger estimates of effects, with exaggerated odds ratios (Schultz, et al., 1995; Fiedorowicz, et al., 2021).

But these open-label trials go even farther in creating performance bias, expectation bias, and enthusiasm bias in a way that inflates effect estimates — because the interventions themselves are *explicitly designed* to support positive patient expectations, and to elicit positive responses

using subjective instruments such as questionnaires. Therefore, risk of bias cannot be discussed without underscoring that the design and delivery of CBT for ME/CFS tips the scales in favor of illusory and subjective success.

CBT as delivered to ME patients is distinct from traditional CBT and must be examined and assessed on its own merits (Geraghty, et al., 2019). According to *The CBT for CFS Therapist Manual* by Loades, et al. (2021) from an NHS clinic in Bath, the focus of traditional CBT is on "living the best life you can alongside physical symptoms, rather than necessarily seeking to change the physical symptoms themselves." However, the manual goes on to explain and describe how *CBT for ME* differs: physical symptoms are framed as explicitly arising from patients' thoughts and feelings. The goal of CBT for ME, then, is to change a patient's thoughts and feelings, which should then eliminate physical symptoms.

Thoughts and feelings that cause disease include thinking about physical symptoms, belief in chronicity, "thoughts and beliefs about fatigue," and the belief that activity may cause relapse. Therefore, the CBT framework for ME encourages patients to deny the presence of symptoms, affirm disbelief in chronicity, and believe that activity cannot cause relapse. Changing these beliefs is framed as essential to success, and therefore, to recovery.

Therefore, the purpose of the intervention is to convince the patient of the effectiveness of the intervention -- and that belief in the intervention is necessary to support success, and disbelief in the intervention will lead to failure.

It should be clear that subjective patient responses will be biased towards affirming the success of the intervention, because the intervention is designed to convey that negative thoughts about illness cause illness. This is a source of significant detection bias, since only subjective measures show positive results; importantly, studies show that objective measures such as step counts and back-to-work/school do not reflect these subjective reports of gains in function. For a comprehensive list of objective measures included in the reviewed studies, see Michiel Tacks' public comments (Tack, 2021).

Finally, handbooks for CBT in people with ME explicitly link negative attributes such as perfectionism to symptom presentation, introducing social desirability bias, wherein "self-reporting data can be affected by an external bias caused by social desirability or approval, especially in cases where anonymity and confidentiality cannot be guaranteed at the time of data collection" (Althubaiti, 2016).

Specific trials of CBT and GET included in the review introduce similar sources of bias. In the initial "pragmatic rehabilitation treatment" session of the 2010 FINE trial, nurses were trained to give their patients a detailed explanation of symptoms, and provide them with a manual that does the same. According to the study publication:

> *Our experience is that this initial session, which provides patients with convincing explanations for their symptoms and also indicates how they can be targeted in treatment, is crucial to the success of the intervention* (Wearden, et al., 2006).

In other words, according to study authors, the success of the treatment intervention relies on the influence that nurses can have in shaping patients' perception of their symptoms and illness. In fact, nurses were trained for six months in this CBT protocol, and an interest in psychosomatic medicine was a prerequisite to the training. This same influence is not exerted in either of the trial's two comparator groups.

In another included study, the 2011 PACE trial's Cognitive Behavioural Therapy Manual for Participants taught trial participants that the symptom of fatigue is maintained through resting and reducing activity, focusing on symptoms, and worrying that activity will make symptoms worse. The manual reads:

> *CBT is designed to help you to discover the most useful ways of managing and overcoming your illness. The aim of treatment is to help you to change certain patterns of thinking and behaviour that may be partially responsible for maintaining your CFS/ME (Burgess and Chalder, 2004).*

This CBT intervention, then, was intentionally designed to influence participant knowledge of and expectation about treatment.

The draft systematic review argues that provider preference and enthusiasm might be controlled using a design in which expert clinicians are assigned to different interventions. However, "inactive control therapies" such as usual care, waiting list, or even attention control are unlikely to control for provider preference.The oft-used "control therapy" of the waiting list involves no clinician contact at all, which makes controlling provider preference truly impossible. Additionally, in many of the "usual care" and "usual specialist care" control groups, patients are told that they may visit their regular doctor as needed, so whether or not patients make contact with 'enthusiastic' providers is highly variable.

It is notable that in the FINE trial (Wearden, et al., 2010), improvements in fatigue were observed immediately after treatment, but not one year after treatment. It would be interesting to test the extent to which diminished improvements were a reflection of diminished response bias and provider preference bias.

Primary outcomes are participant-reported outcomes, which calls for a 'high' risk of bias judgment

Reducing bias in open-label trials is possible through a study design that relies on objective, clinically observed outcomes (Viswanathan, et al., 2017, p8). However, fatigue and function as well as other outcomes used throughout exercise and CBT trials were subjective and participant-reported. The *Cochrane* risk of bias tool instructs that when a trial outcome is participant-reported, the assessment of the outcome is likely to be influenced by knowledge of the intervention received (Higgins, et al., 2019, p50). Therefore, risk of bias within the "patient masked?" as well as "outcome assessors masked?" domain should be considered 'high.'

There is empirical evidence showing that bias due to lack of patient blinding, particularly when reliant on participant reported outcomes, tends to lead to exaggerated standardized mean differences (Hrobjartsson, et al., 2014; Nuesch, et al., 2009). In one systematic review, nonblinded patients exaggerated the effect size by an average of 0.56 of a standard deviation (Hrobjartsson, et al., 2014). Effect sizes in many of these CBT and GET trials are small to begin with. Therefore, adjusting for these response biases would likely mean a notable decrease in the statistical and clinical significance of findings.

Overall risk of bias judgment was inaccurately assessed as 'medium' when it should be 'high'

Based on the *revised Cochrane risk of bias tool for randomized trials* (Higgins, et al., 2019, p4), this systematic review should rate exercise and CBT trials as having a 'high' risk of bias. The domains of "Care provider masked?", "Patient masked?" and "Outcome assessors masked?" all have a "high' risk of bias for each primary outcome. *Cochrane* instructs that making a judgment of 'high' risk of bias in even just one domain should lead to an overall judgment of 'high' risk of bias for that outcome or group of outcomes. The tool reads:

> "*Judging a result to be at a particular level of risk of bias for an individual domain implies that the result has an overall risk of bias <u>at least this severe</u>.*" It also reads:"*'Some concerns' in multiple domains may lead the review authors to decide on an overall judgement of 'high' risk of bias for that outcome or group of outcomes (Higgins, et al., 2019, p4).*"

Even if reviewers disagree with #MEAction's judgment of 'high' risk of bias for the three discussed domains, the fact that there are 'Some concerns' in multiple domains should still lead to an overall judgment of 'high' risk of bias for the group of outcomes explored in all exercise and CBT trials.

Direction and magnitude of bias needs to be assessed and factored into meta-analysis conclusions

The *AHRQ Methods Guide* directs reviewers to assess the direction and magnitude of bias when bias is detected. There is no indication that this review does so adequately. This is especially important because estimating direction and magnitude of bias can have significant impacts on the review's conclusions. In meta-analyses, review authors show and report that effect estimates consistently favor treatment groups for several outcomes. Authors use this as support for the conclusion that graded exercise and CBT are effective compared to inactive control therapies. However, it is possible that the consistency noted is in the direction of bias, which has led to a consistent overestimation of true effect sizes in these trials. The treatment effect may no longer be statistically significant once direction and magnitude of bias are taken into account.

Outcomes reporting bias should be incorporated into the risk of bias assessment, and factored into the strength of evidence assessment

The review authors' current risk of bias assessment looks at "outcomes pre-specified," but not "outcomes pre-specified and reported." This is in error, as the *AHRQ Methods Guide* underscores how outcomes reporting has major implications "for both the risk of bias of individual studies and the strength of the body of evidence" (Viswanathan, et al., 2017, p7).

A study looking at the impact of outcome reporting bias in RCTs concluded that:

> *Individuals conducting systematic reviews need to address explicitly the issue of missing outcome data for their review to be considered a reliable source of evidence. Extra care is required during data extraction, reviewers should identify when a trial reports that an outcome was measured but no results were reported or events observed, and contact with trialists should be encouraged (Kirkham, et al., 2010).*

Multiple studies included in this review are prone to reporting bias. For example, the PACE trial did not report primary outcome measures in accordance with the trial protocol (Wilshire, et al., 2017). We urge a review of Michiel Tacks' Public Comments (2021), which identify the FITNET trial and the SMILE trial as two others that concealed or switched outcome measures (respectively) from protocol to publication.

Analyses of PACE's flaws have been publicly available for years. These flaws include (but are not limited to): 1) mid-trial, authors published a newsletter extolling the benefits of the PACE

therapies. At the time, about 200 (or ⅓) of participants were still actively undergoing the trial; 2) At baseline, 13% of trial participants were already "within the normal range" on one or both primary outcome measures. This was not disclosed in the original *Lancet* publication; and 3) changed definitions of "recovery" markedly from trial protocol to reporting, which over-inflated the significance of the recovery rates for treatment groups (Racaniello, 2015; Wilshire, et al., 2017).

PACE and the other trials in this review have other serious methodological limitations that are not assessed within this review's Risk of Bias tool. Nonetheless, they must be used to inform conclusions about the efficacy of these interventions. Some of these other sources of bias are explored in Part Four: Interpretations of Results.

Part III: Exclusion of Harms Evidence

The systematic review's investigation of harms of GET and CBT is insufficient. The review authors used rigid inclusion criteria for finding and selecting data on harms, which does not reflect current best practice for evidence reviews.

The systematic review covers only the inadequate reporting by included RCTs

GET and CBT are widely criticized as treatments for people with ME/CFS. Numerous sources discuss potential harms associated with these therapies in this patient population. Therefore, it is surprising how little focus and attention was placed on addressing the concerns of potential harms in the review.

The Wormgoor and Rodenburg (2021) review, on the other hand, states that "Applying [physiotherapy (e.g., GET)] treatment that seems effective for [chronic fatigue] or CFS patients may have adverse consequences for ME patients and should be avoided"; and the UK's NICE evidence review of non-pharmacological interventions for people with ME/CFS warns that "People with ME/CFS have reported worsening of symptoms with GET and no benefit from CBT" (NICE, 2020).

According to the AHRQ Methods Guide, reviewers should "not assume studies adequately assess harms because methods used to assess and report benefits are appropriate; rather, evaluate how well studies identify and analyze harms" (AHRQ, 2014, p249). However, this review did not assess any evidence for potential harms except those reported by the researchers who conducted the GET and CBT studies. Since "harms were reported poorly in most trials" (Chou, et al., 2020, p159), the only evidence on harms considered for GET was limited to the PACE and GETSET trials, and for CBT, was limited only to PACE. This so skewed the evidence base on harms that the review authors report that there is "limited evidence that exercise and CBT **were not** [emphasis added] associated with increased risk of serious adverse events or worsening of symptoms" (Chou, et al., 2020, pii).

This coverage of harms is not aligned with the patient community's deep, decades-long, well-communicated concerns regarding negative effects of exertion, supported by clinical experts and the array of studies conducted to identify and define PEM. The 2014 IACFS/ME Clinician Primer warns that "exercise worsens symptoms in patients with ME/CFS" (Friedberg, et al., 2014, p17), and the 2017 ME/CFS Pediatric Primer also cautions that "inflexible, graded exercise (GET) is harmful and can lead to worsening of symptoms" (Rowe, et al., 2017, p36). Vink and Vink-Niese's (2018) re-analysis of the Cochrane review concludes, "Because of the failure to report harms adequately in the trials covered by the review, it cannot be said that graded exercise therapy is safe" -- and the same can be said of this review.

On the use of CBT as a treatment for ME/CFS Vink and Vink-Niese (2019) point out, "Patient evidence suggests adverse outcomes in 20 percent of cases. If a trial of a drug or surgical procedure uncovered a similar high rate, it would be unlikely to be accepted as safe."

The coverage of harms is not aligned with best practice for systematic reviews of healthcare interventions

This extremely narrow coverage of evidence of potential harms of GET and CBT treatments is not only inappropriate in the specific context of ME/CFS, but is also not aligned with the best practice recommendations for systematic reviews. According to the AHRQ Methods Guide, review authors should "gather evidence on harms from a broad range of sources, including observational studies, particularly when clinical trials are lacking; when generalizability is uncertain; or when investigating rare, long-term, or unexpected harms" (Chou, et al, 2008).

An important focus of AMSTAR 2, the most updated AMSTAR instrument for the critical appraisal of systematic reviews of healthcare interventions, is on the inclusion of non-randomised trials that provide real-world observational evidence that informs healthcare interventions (Shea, et al., 2017). This is especially important when it comes to reviewing the harms of such interventions. The AHRQ Methods Guide states that "Observational studies[8] are almost always necessary to assess harms adequately. The exception is when there are sufficient data from RCTs to reliably estimate harm" (AHRQ, 2014, p240). This systematic review acknowledges that this was not the case with the GET and CBT randomized controlled trials, referring to their reporting on harms as "suboptimal," "limited" or "poor" (Chou, et al., 2020, ppii,150,159).

As the 2015 IOM report makes clear, the best evidence on ME/CFS to date has identified PEM as a hallmark symptom of this condition, and CDC concurs (CDC-a). It is critical that the review extensively consider the potential for harms arising from exertion in people with ME/CFS. There is a solid body of relevant research exploring these effects, and it is inappropriate for these studies to have been excluded from the systematic review.

Other sources to consider for harms: exercise used as a provocation

The evidence review excluded numerous observational studies that presented physiological evidence of exercise intolerance in people with ME/CFS in its consideration of sources that may

---

[8] When referring to observational studies, the Methods Guide clarifies that the term "is commonly used to refer to cohort, case-control, and cross-sectional studies, but can refer to a broad range of study designs, including case reports, uncontrolled series of patients receiving surgery or other interventions, and others. All can yield useful information as long as their specific limitations are understood" (AHRQ, 2014, p240).

contextualize harms. For example, two day cardiopulmonary exercise testing (CPET) is used to assess ME/CFS patients for disability and to provoke symptoms for study. Two CPET provocations conducted one day apart show significant differences in people with ME/CFS that do not appear in healthy sedentary controls -- or in people with idiopathic chronic fatigue.

In a study by Keller, et al. (2014), patients with ME/CFS incurred a drop in work at ventilatory threshold (Workload$_{VT}$) of 21.3% on the second of two exercise tests 24 hours apart. The respiratory exchange ratio (RER) was similar on both days, showing patients expended maximum effort for both tests. Reduced VO_2 in people with ME/CFS on two-day CPET has been found in several studies, including those by Jones, et al. (2012) and Keller, et. al. (2014); and Vermeulen & Vermeulen van Eck (2014) also showed differences in VO_2max when people with ME/CFS were compared to sedentary controls. Notable differences in VO_2 on two-day CPET were found between people with multiple sclerosis, people with ME, and healthy controls (Hodges, Nielsen and Baken, 2017). And a recent meta-analysis shows that volume of oxygen consumption and level of workload at peak (VO_2@peak, Workload@peak) and at ventilatory threshold (VO_2@VT, Workload@VT) all drop on day two in two-day CPET in people with ME/CFS, whereas all increase in healthy controls (Lim, et al., 2020). An invasive CPET study by Joseph, et al. (2021) also demonstrated cardiac pre-load failure and poor oxygen extraction in ME/CFS patients.

But perhaps most tellingly, in a study conducted by van Campen and Visser (2021), age- gender- disease-duration- and severity-matched patients with idiopathic chronic fatigue (Fukuda) and patients with one of the sets of criteria that require PEM (ICC) were compared on two-day CPET. While patients with PEM required showed decreased VO_2peak, Workload@peak, VO_2@VT and Workload@VT on day two of their two-day CPET as previously described, the ICF patients showed gains on each measure, similar to healthy controls. This should certainly speak to the generalizability of one set of diagnostic criteria to the other regarding harms of exercise therapies.

In addition, there have been studies conducted on post-exercise immune, neurological, and endocrine changes in people with ME/CFS. There is such an abundance of studies on post-exercise pathology in people with ME/CFS that its omission in consideration of potential harms is conspicuous.

Importantly, in some trials, the number of patient withdrawals were significantly higher within GET groups compared to control groups. For example, the Wearden, et al. (1998) study had a withdrawal rate of 29% of those receiving GET, compared to 17% for the attention control group. In the Windthorst, et al. (2017) study, 4 of the 15 original members in the exercise group dropped out, citing lack of benefit. While included in appendices, these and other examples like it were not discussed in the text of the review.

Potential harms are not isolated to exercise trials, and may be present for those patients who undergo CBT as well. In Jason et al. (2007), 22% of participants in the CBT group reported that treatments made them feel from "worse" to "unchanged." It is a mistake to presume that CBT will *de facto* be a safe therapy, given its record of harms in other diseases and disorders

(Schermuly-Haupt, et al., 2018), and its focus in ME/CFS: to encourage the patient to deny the presence of symptoms and the potential for chronicity in order to become more active.

Exclusion of patient surveys

The evidence review did not include any patient surveys, the results of which show the harms of graded exercise therapy and cognitive behavioral therapy (CBT) for patients with ME/CFS. Most recently, a survey of nearly 2300 people with ME/CFS in the United Kingdom found that over two-thirds who underwent GET alone and three-quarters who underwent GET combined with CBT experienced a deterioration in their physical health (MEAction, 2019). 98.5% of responders reported the hallmark symptom of post-exertional malaise. CBT also had a net negative physical effect on people with ME/CFS, with 26% reporting deterioration and only 16% reporting improvement. Geraghty, et al. (2017) present results of their ME/CFS patient survey, as well as ten others over the 2000-2015 period. They found that patients reported GET provoked negative responses in 54-74% of patients.

Prioritizing identification of harms from patient and clinical expert stakeholders

Systematic review best practice recommends prioritizing using multiple sources of information, including patients, clinical experts and stakeholders, to identify harms that they judge to be of greatest importance (Chou, et al., 2018). The systematic review authors state that they consulted key informants representing clinical, research, or patient perspectives in ME/CFS in the scoping and development of key questions for this review. However, nothing in this review would indicate that there has been any consultation with the patient community or clinical or research experts about the proper identification or prioritization of harms. Engagement on such an important issue should not be given such cursory treatment or considered only at the public comment stage after the review has already been designed and drafted.

The absence of robust engagement of the community and investigation of multiple sources on potential harms of CBT and GET is deeply concerning, given this is not the first time these issues have been raised. The prior AHRQ review in 2014 received criticism for its lack of sufficient evidence on harms from exercise, from studies that discuss post-exertional pathology, and from patient surveys. To continue to exclude this evidence, and to scope the review in such a way as to exclude it, is a methodological departure from AHRQ, Cochrane and AMSTAR II guidance and an abdication of ethical responsibility to patients who may be at risk.

An evidence review of CBT and GET treatments for ME/CFS must include a broader examination of the harms from exertion and specifically how people with PEM could be affected by these treatments. The review authors have not obtained adequate "input from patients, to ensure that outcomes reflect their priorities." Exacerbation of PEM should be considered a

serious harm, as it results "in persistent or significant incapacity or ability to perform normal life functions" (Chou, et al., 2018, p9).

Harms not specified in the original protocol must still be included

It should be underscored that even if harms were not specified in the original protocol, this is not a sufficient excuse for the systematic review not to incorporate them. The AHRQ Methods Guide instructs reviewers that they need to be prepared to add harms to the review that were not specified in the original protocol or identified in the prioritization process (AHRQ, 2014, p11).

Other systematic reviews of ME/CFS interventions are more robust on harms

There is solid precedent for more robust and balanced reporting on potential harms in other systematic reviews of ME/CFS interventions. Wormgoor and Rodenburg (2021) address multiple sources of information beyond just examining the reports of randomized controlled trials. These include case-control studies that confirm concerns regarding exercise programs evaluating response on sub-maximal activity in ME/CFS patients, and clear indications of potential negative patient-reported experiences from a review of eleven patient surveys.

Wormgoor and Rodenburg acknowledge that few randomized controlled trials of GET interventions "reported on the occurrence of adverse events or non-adherence due to intolerance to the intervention. However, in intervention research involving ME/CFS patients with PEM, reporting of adverse effects seems of particular significance; interventions are not necessarily harmless when adverse effects and compliance have not been systematically reported" (Wormgoor and Rodenburg, 2021, p19). In contrast to this CDC systematic review, the authors conclude that "Even though the results of this review did not reveal substantial negative responses, the marginal and doubtful effects, patient-reported experiences and evidence coming from biomedical research strongly suggest an overall reduction in tolerance of physical exertion in ME patients" (Wormgoor and Rodenburg, 2021, p20).

Other authors of systematic reviews on exercise therapies in ME/CFS concur. Ahmed, Mewes and Vrijhoef (2020) conclude in their systematic review that "The findings of this systematic review do not support the claim that CBT and GET are effective treatments for ME/CFS patients" because of the low methodological quality and biases of the studies, which include the poor reporting of harms. They supplement their discussion of harms by acknowledging the important contributions of Kindlon (2011) on the reporting of harms associated with GET and CBT in ME/CFS and Vink and Vink-Niese's (2018) re-analysis of the Cochrane review showing that GET is ineffective and unsafe for ME/CFS. They state that "over the past 20 years patients have consistently reported either no or adverse effects from these interventions."

The UK's NICE evidence review of non-pharmacological interventions for people with ME/CFS (NICE, 2020) utilized a stakeholder committee model which resulted in a more extensive examination of the evidence for each intervention and considered discussion of areas of potential harms by including the perspective of individuals with lived experience of ME/CFS. "The committee considered this evidence along with the clinical effectiveness and qualitative evidence. Given the uncertainty around the health benefits of GET combined with the possibility of harm due to over-exertion, especially when GET is poorly implemented, the committee agreed to not recommend GET." They came to their conclusion by considering both data from randomized trials and more qualitative evidence. "The committee noted that no harms were identified in the clinical evidence but also noted these were rarely included as an outcome and reported. The committee reflected that in contrast harms such as worsening of symptoms were reported in the qualitative evidence and took this into consideration when making recommendations on physical activity and exercise." The NICE committee also considered the potential for harm when CBT was inappropriately delivered by a therapist who did not understand ME/CFS, and recommended that CBT was not to be used as a treatment or cure for ME/CFS (NICE, 2020, p326).

Part IV: Interpretations of Results

Reporting average effects is misleading, given high risk of bias, high heterogeneity, and low strength of evidence

Meta-analyses should be performed when their results can be meaningful; a pooled estimate of effect should achieve more power and precision than that provided at the individual study level (Morton, et al., 2018, p3). The systematic review's meta-analyses for exercise and CBT interventions, however, do not provide such insight. Rather, the average effects reported are unlikely to reflect the impact of these interventions on people with ME/CFS. They do not provide a sensible or meaningful summary because of the high risk of bias in the reviewed trials, the considerable clinical and statistical heterogeneity within and between studies, the focus on subjective primary outcomes, and the low strength of the body of evidence as a whole. Given these serious concerns, the interpretation of results and conclusions derived from meta-analyses are misleading. They are unlikely to convey the true effect that either CBT or GET have on ME/CFS patients as described by CDC: those with PEM.

Reliance on trials with a high risk of bias and subjective primary outcomes

As discussed in Part II: Risk of Bias, subjective outcomes, particularly when in open-label trials, are subject to significant bias (Page, et al., 2016). Response bias leads to overinflated point estimates of effect that cannot be attributed to the exercise and CBT treatments themselves. Creating a pooled estimate from these trials perpetuates a cycle of "garbage in, garbage out".

Primary outcomes in this review are patient-reported fatigue and patient-reported physical function, and this review's meta-analyses focus largely on results flowing from these outcomes. But a body of studies explore how subjective improvements, particularly in psychotherapeutic treatments, may simply reflect response biases or placebo effects (Kindlon, 2015).

Objective measures are necessary as evidence of true effect (Kewley, 2013), but in almost all of the studies included in this review, objective measures, when considered at all, were often not statistically significant. Perhaps for this reason, objective measures were not often reported in the text of study publications. This systematic review reinforces the selective non-reporting of objective measures by de-emphasizing them in their own summaries of results.

Obfuscating the results from objective measures obscures the chance for a more accurate conclusion regarding GET and CBT interventions: that the marginal, temporary improvements in self-reported fatigue and physical functioning following GET and CBT therapies do not translate

to objective improvement (Vink and Vink-Niese, 2018, 2019; Wormgoor and Rodenburg, 2021). Based on objective tests, CBT and GET are ineffective for ME/CFS, even under Oxford and Fukuda criteria.

High clinical heterogeneity

Patient participants were not similar within or between GET and CBT studies. As discussed in Part I: Applicability, using overly-broad ME/CFS diagnostic criteria introduces a serious methodological issue with impact on the validity and reliability of study findings. Systematic review authors note that the Oxford criteria in particular "classifies more patients with ME/CFS compared with more current case definitions, potentially resulting in misclassification and misleading results" (Chou, et al., 2020, p159). One U.S. population study by Baraniuk (2017) found that "around 85% of Oxford CFS subjects were false positives when subsequently assessed using the more rigorous quadrant and Fukuda criteria." The 1994 Fukuda (CDC) criteria, also, comprises a mixed group of fatigued patients, some of whom have post-exertional malaise (PEM), and some of whom do not.

It is likely that exercise and CBT interventions had a variable impact on different populations within these trials, in which almost all used Oxford or Fukuda. Patients with and without PEM likely respond differently to an exercise intervention. A significant body of evidence, including thousands of patient testimonials, indicates that GET and CBT may be harmful to this former cohort. An evidence review by Ahmed, et al. (2020) states: "Information on the occurrence of adverse/unwanted events is essential, as research conducted by Kindlon (2011) and Vink and Vink-Niese (2018) showed that over the past 20 years patients have consistently reported either no or adverse effects from these interventions."

High patient heterogeneity at the study level translates to high clinical heterogeneity in meta-analysis. The AHRQ Methods Guide states that when patient characteristics are not similar enough for an average effect to be meaningful, reviewers should consider not pooling studies (Morton, et al., 2018). If underlying effects are likely similar across subpopulations, some have suggested that pooling is acceptable, but in this evidence review, this is clearly not the case. In the Journal of Evidence-Based Medicine, Sun and Guyatt (2009) write,

> *Meta-analysis does not inherently secure an optimal estimation of effects of health care interventions. The results can be misleading when a meta-analysis fails to maximize the methodological rigor. For instance, pooling of the results of a set of trials that arise from an excessively broad definition of the question in terms of patients, interventions, or the way outcomes are measured may be misleading.*

Within this systematic review's meta-analyses, the high clinical heterogeneity is likely to negatively affect the validity and reliability of pooled results, and render "average effects" useless and misleading. They are potentially harmful to the cohort of patients most obscured by this body of studies: the patients with PEM.

The allegiance effect significantly increases effect sizes

All studies included in this review's meta-analyses but one were conducted by researchers with an allegiance to the biopsychosocial model of ME/CFS, and to CBT and GET as effective interventions for people with ME/CFS. According to the allegiance effect, treatments favored by the investigator "tend to produce the superior outcome" (Westen et al., 2004). This phenomenon has been studied in psychotherapeutic research for several decades.

According to Westen et al. (2004), "Perhaps the best predictors of whether a treatment finds its way to the empirically supported list are whether anyone has been motivated (and funded) to test it and whether it is readily testable in a relatively brief format." This assertion, though humorously stated, has been supported by research. Luborsky, et al. (1999) found that allegiance alone could account for over 69% of the variance in outcome across a large set of studies, consistently yielding higher effect sizes for investigators' preferred treatments. Mediators of this effect, to which most of the reviewed CBT and GET studies are vulnerable, include differences in the way treatment and comparison conditions are designed and implemented, with the comparison conditions being implemented in a weaker form (Vink and Vink-Niese, 2019).

In validating a tool to measure researcher allegiance in psychotherapeutic studies, Yoder, et al. (2019) state that allegiance may be a factor in a study's conclusions "if the author developed the intervention, if the author provides an extensive amount of information about one intervention compared with the other(s), if the author refers to previous research showing superiority of one intervention and if the author advocates the intervention through their writing." They divide potential sources of RA into five categories:
  ● Effectiveness - if the author(s) state the interventions are effective for the diagnosis of interest
  ● Superiority - if the author(s) state the intervention is effective beyond other interventions for the diagnosis of interest
  ● Advocacy - if the author advocates use of the intervention (i.e. instructs other therapists to apply the therapy for people with the diagnosis of interest)
  ● Development/contribution - if the author(s) developed or contributed to the development or enhancement of the intervention for the diagnosis of interest
  ● Methodology - if the author(s) was/were involved in the trial in a way that may have influenced the implementation of the intervention

When all treatments are not presented as equally viable, this can lead to significant response bias. A researcher's enthusiasm can also bias them, unconsciously or not, to overinterpret findings and overlook limitations (Vink and Vink-Niese, 2019). In their re-analysis of Cochrane reviews on GET and CBT, Vink and Vink-Niese (2018, 2019) found that Chalder and Wessely, Sharpe, O'Dowd, Surawy, Fulcher, Powell, Wearden, Wallman, Moss-Morris and White are all known "to have favoured the approach to the illness being tested."

There is abundant textual evidence of these researchers' allegiance to these therapies for ME/CFS, including study materials lauding their success. Materials researchers have created and shared support that they believe CBT and GET for ME/CFS to be effective and superior to other available interventions, and that they train other therapists to perform CBT for people with a ME/CFS diagnosis, fulfilling the Effectiveness, Superiority, and Advocacy criteria for RA (researcher allegiance) (Yoder et al., 2019). In addition, a significant percentage of lead authors in included studies have helped develop or modify CBT, GET, or both for people with ME/CFS, fulfilling the Development/contribution category. Finally, study authors and their trainees have been directly involved with implementation in such a way that could further bias results of the study: for example, by way of researchers coming into contact with the experimental group only, as they did in any study in which the control group was represented by patients on a waiting list, fulfilling the Methodology criteria for RA.

It is important to acknowledge, as per Yoder, et al. (2019), "RA may not be representative of a purposeful attempt to skew results as it is simply human nature to hold beliefs in ways that can compromise objectivity. However, overlooking RA could be considered a methodological issue in psychotherapy research." Researcher allegiance may affect reported results in a variety of ways; however, these effects may be abrogated through the use of clearly-defined, objective measures.

High statistical heterogeneity and the omission of Powell, et al.

Statistical heterogeneity is the variability observed in intervention effects that is due to something other than chance alone (Higgins et al., 2021, ch10). Best practice suggests that even if studies appear clinically and methodologically similar enough to be combined, "the statistical heterogeneity can be so considerable that it does not make sense to combine the results" (Verbeek, et al., 2011). In this review's exercise and CBT meta-analyses, statistical heterogeneity is often so high that reporting the average effects of these combined trials is misleading.

The *AHRQ Methods Guide* reads, "If statistical heterogeneity is very high, the investigators may question whether an "average" effect is really meaningful or useful" (Morton, et al., 2018). In this review's meta-analysis examining the association between graded exercise and self-reported fatigue severity, at I^2 = 85% (Chou et al., 2020, p30), statistical heterogeneity fell into the highest possible category according to the Cochrane handbook (Higgins, et al, 2021, ch10). Reporting on the average effect for that outcome is valueless in and of itself, even disregarding sources of bias and other limitations previously described. But this systematic review uses this average effect to construct one of its "Key Points. "In adults diagnosed with ME/CFS," the review reads, "graded exercise therapy (GET) was associated with decreased fatigue severity...versus inactive controls" (Chou et al., 2020, p18).

The systematic review addresses high statistical heterogeneity through performing subgroup analyses and sensitivity analyses; yet those analyses do not inform the high-level interpretation of results reflected in "Key Points," and carried to the report's conclusions. One of the most

glaring examples of where information derived from sensitivity analyses should (but does not) inform the overall interpretation of results is in the GET trials analysis that excludes the outlier trial, Powell, et al. (2001).

Excluding the Powell, et al. trial reduced statistical heterogeneity. Vitally, it reduced the pooled point estimate so much that one cannot say with confidence that the minimum clinically important difference of an SMD of 0.2 is met. The 95% confidence interval of the estimates for fatigue both at the end of the intervention and at post-intervention follow-up include values that are both below and just above this established minimum, suggesting that there is some likelihood that the improvement in fatigue for the treatment group is not clinically meaningful.

The association between GET and outcomes on function is even more dubious. The pooled estimate suggesting that graded exercise was associated with less severe functional impairment versus 'inactive control therapies' had high statistical heterogeneity and very wide confidence intervals, suggesting significant imprecision. When the Powell, et al. study is excluded, statistical heterogeneity as measured by I^2 decreases significantly. Importantly, the mean difference in function also drops significantly, from 11.73 to 5.89. 5.89 is well below the minimum clinically important difference, established at 10. Given the information collected through this sensitivity analysis, it is dubious to conclude that GET is associated with improved function.

When factoring in the high level of systematic bias pervading these trials in the direction of the treatment intervention, the high statistical heterogeneity, and the low strength of evidence review authors assigned to this body of studies, one can only conclude that GET may or may not reduce fatigue and/or improve function: the evidence is insufficient.

The review author's interpretation of results of CBT vs. inactive control is also vulnerable to clinical and statistical heterogeneity, high bias, and low strength of evidence. And yet these factors do not appear to inform the report's interpretations of results. This is not aligned with the Cochrane Handbook's guidance: "An overall interpretation of meta-analysis results must consider heterogeneity, sources of bias, and impactful subgroup analyses….The interpretation of the true effect needs to be seen in the context of further uncertainty resulting from those concerns" (Schunemann, et al., 2021).

An overall interpretation of meta-analysis results must consider heterogeneity, sources of bias, and utilize impactful subgroup analyses as applicable (Schunemann, et al., 2021). Methods of addressing heterogeneity might include performing a subgroup analysis or meta-regression, and its use in the formation of a meta-analysis's ultimate conclusions. However, meta-analysis ought not be performed if heterogeneity is too high and subgroup analysis is unlikely to be useful due to a limited number of studies per group, as is the case here. The statistical heterogeneity present in these studies should preclude confident statements about GET's success altogether; and authors of the review explicitly acknowledge that this is the case.

Subgroup analyses cannot solve the problem

This evidence review addressed heterogeneity by performing subgroup analyses. In its "Methods" section, review authors write, "We conducted subgroup analyses based on the inactive control type…, ME/CFS case definition, and CBT type…, and evaluated for the subgroup differences with a statistical test" (Chou, et al., 2020, p9).

Unfortunately, these subgroup analyses are neither reliable nor meaningful. First, most included a small number of studies, meaning analyses were likely underpowered. Second, the analyses likely reflect a high risk of bias, since the individual studies are at high risk of bias. An *AHRQ Methods Guide* reads, "Studies rated as having a high risk of bias for the main analysis of benefits or harms will also likely have a high risk of bias for subgroup analysis (Chou, et al., 2008). While some subgroup analyses find significant difference in effects, because the studies are rife with limitations and the analyses are underpowered, it is challenging to know whether those effects are truly the result of the subgroup characteristic, or something else.

Finally, these subgroup analyses are not used to inform the evidence review's interpretation of results or conclusions. As discussed in our comments regarding the outlier trial of Powell, et al., although review authors carry out subgroup analyses, they do not undertake the more difficult and necessary task of deciding how their findings should inform overall interpretations and conclusions.

Conclusions drawn from meta-analyses of both GET vs. inactive controls and CBT vs. inactive controls are misleading

The systematic review concludes that GET and CBT are both beneficial treatments for adults with ME/CFS and are more effective than inactive control therapies. The review comes to these conclusions based upon the meta-analyses it conducts of both GET and CBT randomized controlled treatment trials. While the authors acknowledge certain limits on the strength of these conclusions, they present them as viable nevertheless.

In this Part IV: Interpretations of Results, it should be clear that the utility and reliability of the review's meta-analyses and subgroup analyses should be called into question. Conclusions drawn from the average of small, subjective study results is not meaningful for ME/CFS patients' ultimate health and wellbeing. The subgroup analyses performed do not adequately address these underlying concerns. Therefore the conclusions of the review are misleading and likely put people with ME/CFS at risk of harm.

The review's inclusion of cursory statements of uncertainty about the results do not sufficiently address concerns about the validity of or confidence in the Review's findings. The problems with

the data in the primary studies do not justify meta-analyses of average effects as a strategy for deriving conclusions about clinical interventions for ME/CFS patients.

The results of other systematic reviews, each using a range of approaches, all support that a meta-analysis of average treatment effects of these primary studies cannot alone justify this review's conclusion that GET and CBT are beneficial treatments for people with ME/CFS.

Ahmed, Mewes and Vrijhoef (2020) assessed "the methodological quality of studies on the effectiveness of GET and CBT for ME/CFS patients," as well as, "studied the effectiveness of CBT and GET for ME/CFS patients, analysed the cut-off scores used for outcome measures, identified whether and which harmful events are reported in the literature and evaluated the inclusion criteria of RCTs with regard to including patients with PEM." Their systematic analysis found that "the methodological quality of the included studies was evaluated to be relatively low, as all studies were found to score an unclear or high risk in three to six of the bias categories. Thus, robust evidence to support the effectiveness of CBT and GET is lacking. The included trials showed that GET and CBT mainly were not found to be effective treatments for [ME/CFS]" (Ahmed, Mewes and Vrijhoef, 2020, p11).

Wormgoor and Rodenburg's (2021) systematic review of physiotherapy (i.e., GET) treatments for ME/CFS analyzed the studies and reported their results according to three distinct subgroups: RCTs with diagnostic inclusion criteria without PEM as a criterion (i.e., CF), RCTs with diagnostic inclusion criteria with PEM as an optional criterion (i.e., CFS), and RCTs with diagnostic inclusion criteria with PEM as a required criterion (i.e., ME). Because of the limitations in the heterogeneity of comparison groups, group sizes and follow-up duration in the studies, they did not attempt to compare results and calculate effect sizes across the different treatments and diagnostic groups and instead used a narrative synthesis. They conclude that "Currently, there is no scientific evidence when it comes to effective physiotherapy treatment for ME patients diagnosed with narrow diagnostic criteria sets that include PEM. Findings indicating effectiveness of physiotherapeutic interventions for ME/CFS are mainly based on RCTs involving patients diagnosed with diagnostic criteria that do not require PEM. Possible evidence vanished when diagnostic specificity, outcome objectivity or follow-up time increased."

The CDC systematic review's conclusions — that GET and CBT offer modest improvements for the treatment of ME/CFS — are not justified by the data of the primary studies. Meta-analyses of average treatment effects across these studies — despite at times having the highest possible category of statistical heterogeneity according to the Cochrane handbook, which should abrogate the possibility of a meta-analysis — lead to the reviewers' claim that these widely disputed interventions are appropriate, safe and beneficial to people with ME/CFS.

But the differing conclusions of other systematic reviews, the balance of expert opinion, the repeated concerns from patients, the body of evidence suggesting significant biological dysfunction resulting from exertion, and the reversal of previous CDC and NICE recommendations, should lead readers and reviewers to the conclusion that these interventions are not appropriate in ME/CFS. Even within the restricted scope of this

systematic review, its meta-analyses do not provide robust evidence that would justify the claim that GET and CBT are more effective than inactive controls in improving fatigue, function, and other outcomes in patients with ME/CFS.

References

Ahmed SA, Mewes JC, Vrijhoef HJM (2020). 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. Jour Health Psychology 25(2). Doi: 10.1177/1359105319847261.

AHRQ (2014). Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(14)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality. Available from: https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/cer-methods-guide_overview.pdf.

Althubaiti A (2016). Information bias in health research: definition, pitfalls, and adjustment methods. Journal of multidisciplinary healthcare 9:211-217. Doi: 10.2147/JMDH.S104807.

Atkins D, Chang S, Gartlehner G, et al. (2011). Assessing the Applicability of Studies When Comparing Medical Interventions. Agency for Healthcare Research and Quality. Methods Guide for Comparative Effectiveness Reviews. AHRQ Publication No. 11-EHC019-EF. Available from: https://effectivehealthcare.ahrq.gov/sites/default/files/methods_guide_atkins.pdf.

Baraniuk JN (2017). Chronic Fatigue Syndrome prevalence is grossly overestimated using Oxford criteria compared to Centers for Disease Control (Fukuda) criteria in a U.S. population study. Fatigue 5(4):215-230. Doi: 10.1080/21641846.2017.1353578.

Brurberg KG, Fonus MS, Larun L, et al. (2014). Case definitions for chronic fatigue syndrome/myalgic encephalomyelitis (CFS/ME): a systematic review. BMJ Open 4(2):e003973. Doi: 10.1136/bmjopen-2013-003973.

Burgess M, Chalder T. (2004). Cognitive Behaviour Therapy for CFS/ME. PACE Manual for Participants. ISRCTN54285094. Available from: https://me-pedia.org/images/7/7a/PACE-cbt-participant-manual.pdf.

CDC-a: Centers for Disease Control and Prevention. Myalgic Encephalomyelitis / Chronic Fatigue Syndrome: Symptoms of ME/CFS. Available from: https://www.cdc.gov/me-cfs/symptoms-diagnosis/symptoms.html.

CDC-b: Centers for Disease Control and Prevention. Myalgic Encephalomyelitis / Chronic Fatigue Syndrome: Etiology and Pathophysiology. Available from: https://www.cdc.gov/me-cfs/healthcare-providers/presentation-clinical-course/etiology-pathophysiology.html.

CDC-c: Centers for Disease Control and Prevention. Myalgic Encephalomyelitis / Chronic Fatigue Syndrome: Treating the Most Disruptive Symptoms First and Preventing Worsening of Symptoms. Available from: https://www.cdc.gov/me-cfs/healthcare-providers/clinical-care-patients-mecfs/treating-most-disruptive-symptoms.html.

Chou R, Aranson N, Atkins D, et al. (2008). Methods Guide for Effectiveness and Comparative Effectiveness Reviews: Assessing Harms When Comparing Medical Interventions: Assessing Risk of Bias (Quality) of Harms Reporting. Available from: https://www.ncbi.nlm.nih.gov/books/NBK47098/#cerguideharms.s8.

Chou R, Baker WL, Banez LL, et al. (2018). Prioritization and Selection of Harms for Inclusion in Systematic Reviews. Methods Guide for Comparative Effectiveness Reviews. Available from: https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/prioritization-selection-of-harms.pdf.

Chou R, McDonagh M, Griffin J, Grusing S (2020). Management of Myalgic Encephalomyelitis / Chronic Fatigue Syndrome (ME/CFS): An Updated Systematic Evidence Review (Draft). Pacific Northwest Evidence-based Practice Center, Oregon Health and Science University. Available from: https://www.regulations.gov/document/CDC-2021-0053-0001.

Clinician Coalition: U.S. ME/CFS Clinician Coalition, Myalgic Encephalomyelitis / Chronic Fatigue Syndrome. Treatment Recommendations. Available from: https://mecfscliniciancoalition.org/wp-content/uploads/2021/05/MECFS-Clinician-Coalition-Treatment-Recs-V1.pdf.

Dimmock ME, Goodell C, Lopez-Majano D, et al. (2014). Comments on AHRQ Draft Comparative Effectiveness Review; Diagnosis and Treatment of Myalgic Encephalomyelitis / Chronic Fatigue Syndrome. Available from: https://www.meaction.net/wp-content/uploads/2021/07/Dimmock-Spotila-Comments-on-AHRQ-Evidence-Review.pdf.

Dimmock M, Spotila J (2016). AHRQ Agrees: GET Useless, CBT Ineffective. MEAction. Available from: https://www.meaction.net/2016/08/18/ahrq-agrees-get-useless-cbt-ineffective/.

Fiedorowicz JG, Levenson JL, Leentjens AFG (2021). When is lack of scientific integrity a reason for retracting a paper? A case study. Jour Psychosomatic Research 144(110412). Doi: 10.1016/j.jpsychores.2021.110412.

Friedberg F, Bateman L, Bested AC, et al. (2014). Chronic Fatigue Syndrome Myalgic Encephalomyelitis. Primer for Clinical Practitioners. IACFS/ME. Available from: https://growthzonesitesprod.azureedge.net/wp-content/uploads/sites/1869/2020/10/Primer_Post_2014_conference.pdf.

Geraghty K, Hann M, Kurtev S (2017). Myalgic encephalomyelitis/chronic fatigue syndrome patients' reports of symptom changes following cognitive behavioural therapy, graded exercise therapy and pacing treatments: Analysis of a primary survey compared with secondary surveys. Jour. Health Psychology 24(10):1318-1333. Doi: 10.1177/1359105317726152.

Geraghty K, Jason L, Sunnquist M, et al. (2019). The 'cognitive behavioural model' of chronic fatigue syndrome: Critique of a flawed model. Health Psychol Open 6(1):2055102919838907. Doi: 10.1177/2055102919838907.

Green CR, Cowan P, Elk R, et al. (2015). National Institutes of Health Pathways to Prevention Workshop: Advancing the Research on Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. Annals of Internal Medicine. Doi: 10.7326/M15-0338.

Higgins JPT, Savovic J, Page MJ, et al. (2019). Revised Cochrane risk-of-bias tool for randomized trials (RoB 2). Available from: https://drive.google.com/file/d/19R9savfPdCHC8XLz2iiMvL_71IPJERWK/view.

Higgins JPT, Savovic J, Page MJ, et al. (2021). Chapter 8: Assessing risk of bias in a randomized trial. Cochrane Training. Available from: https://training.cochrane.org/handbook/current/chapter-08.

Higgins JPT, Savovic J, Page MJ, et al. (2021). Chapter 10: Analysing data and undertaking meta-analyses. Cochrane Training. Available from: https://training.cochrane.org/handbook/current/chapter-10#section-10-10-1.

Hodges LD, Nielsen T, Baken D (2017). Physiological measures in participants with chronic fatigue syndrome, multiple sclerosis and healthy controls following repeated exercise: A pilot study. Clinical Physiology and Functional Imaging 38(4):639-644. Doi: 10.1111/cpf.12460.

Hrobjartsson A, Emanuelsson F, Thomsen ASS, et al. (2014). Bias due to lack of patient blinding in clinical trials. A systematic review of trials randomizing patients to blind and nonblind sub-studies. Int J Epidemiol. 43(4):1272-1283. Doi: 10.1093/ije/dyu115.

Jones DEJ, Hollingsworth KG, Jakovljevic DG, et al. (2012). Loss of capacity to recover from acidosis on repeat exercise in chronic fatigue syndrome: a case–control study. European Journal of Clinical Investigation 42(2). Doi: 10.1111/j.1365-2362.2011.02567.x.

Joseph P, Arevalo C, Oliveira RKF, et al. (2021). Insights From Invasive Cardiopulmonary Exercise Testing of Patients With Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. Chest. S0012-3692(21)00256-7. Doi: 10.1016/j.chest.2021.01.082.

Keller BA, Pryor JL, Giloteaux L (2014). Inability of myalgic encephalomyelitis/chronic fatigue syndrome patients to reproduce VO$_2$peak indicates functional impairment. J Transl Med 12:104. Doi: 10.1186/1479-5876-12-104.

Kewley AJ (2013). Does Cognitive Behavioral Therapy or Graded Exercise Therapy Reduce Disability in Chronic Fatigue Syndrome Patients? Objective Measures Are Necessary. Clinical Psychology: Science and Practice 20(3):321-322. Doi: 10.1111/cpsp.12042.

Kindlon T (2011). Reporting of Harms Associated with Graded Exercise Therapy and Cognitive Behavioural Therapy in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. Bull IACFS/ME 19(2):59-11. Available from: https://www.ncf-net.org/library/Reporting%20of%20Harms.htm.

Kindlon TP (2015). Rapid Response to: Tackling fears about exercise is important for ME treatment, analysis indicates. Available from: https://www.bmj.com/content/350/bmj.h227/rr-10.

Kirkham JJ, Dwan KM, Altman DG, et al. (2010). The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. BMJ 340:c365. Doi: 10.1136/bmj.c36.

Lim E-J, Kang E-B, Jang E-S, Son C-G (2020). The Prospects of the Two-Day Cardiopulmonary Exercise Test (CPET) in ME/CFS Patients: A Meta-Analysis. Journal of Clinical Medicine 9(12):4040. Doi: 10.3390/jcm9124040.

Loades ME, Starbuck J, Chapple K (2021). CBT for CFS Therapist Manual. Royal United Hospital, Bath, UK. Available from: https://www.ruh.nhs.uk/patients/services/clinical_depts/paediatric_cfs_me/documents/Resources_for_Clinicians_and_Healthcare_Professionals/CBT_for_CFS_Therapist_Manual.pdf.

Luborsky L, Diguer L, Seligman DA, et al. (1999). The Researcher's Own Therapy Allegiances: A "Wild Card" in Comparisons of Treatment Efficacy. Clinical Psychology: Science and Practice 6:95-106. Doi: 10.1093/clipsy.6.1.95.

MEAction (2019). GET And CBT Are Not Safe For ME – Summary Of Survey Results. Available from: https://www.meaction.net/2019/04/03/get-and-cbt-are-not-safe-for-me-summary-of-survey-results/.

Morton SC, Muran MH, O'Connor E, et al. (2018). Quantitative Synthesis—An Update. Methods Guide for Comparative Effectiveness Reviews. (Prepared by the Scientific Resource Center under Contract No. 290-2012-0004-C). AHRQ Publication No. 18-EHC007- EF. Rockville, MD: Agency for Healthcare Research and Quality. Doi: 10.23970/AHRQEPCMETHGUIDE3.

NAM: National Academy of Medicine, National Academy of Sciences (2015). Beyond Myalgic Encephalomyelitis / Chronic Fatigue Syndrome: Redefining an Illness. Available from: http://www.nationalacademies.org/hmd/Reports/2015/ME-CFS.aspx.

NCBI: National Center for Biotechnology Information (2015). Review: Beyond Myalgic Encephalomyelitis / Chronic Fatigue Syndrome: Redefining an Illness. Available from: https://pubmed.ncbi.nlm.nih.gov/25695122/.

NCBI: National Center for Biotechnology Information (2016). Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome: July 2016 Addendum - Agency for Healthcare Research and Quality (AHRQ). Available from: https://www.ncbi.nlm.nih.gov/books/NBK379582/.

NICE: National Institute for Health and Care Excellence (2020). Myalgic encephalomyelitis (or encephalopathy) / chronic fatigue syndrome: diagnosis and management. [G] Evidence Reviews for the non-pharmacological management of ME/CFS. Available from: https://www.nice.org.uk/guidance/gid-ng10091/documents/evidence-review-7.

Nuesch E, Reichenbach S, Trelle S, et al. (2009). The importance of allocation concealment and patient blinding in osteoarthritis trials: A meta-epidemiologic study. Arthritis Care and Research 61(12):1633-1641. Doi: https://doi.org/10.1002/art.24894.

Page MJ, Higgins JPT, Clayton G, et al. (2016). Empirical Evidence of Study Design Biases in Randomized Trials: Systematic Review of Meta-Epidemiological Studies. PLOS ONE. Doi: 10.1371/journal.pone.0159267.

Powell P, Bentali RP, Nye FJ, Edwards BHT (2001). Randomised controlled trial of patient education to encourage graded exercise in chronic fatigue syndrome. BMJ 322(7283):387-390. Doi: 10.1136/bmj.322.7283.387.

Racaniello V (2015, November 13). An Open Letter to Dr. Richard Horton and the Lancet. Virology Blog. Available from: https://www.virology.ws/2015/11/13/an-open-letter-to-dr-richard-horton-and-the-lancet/.

Rowe PC, Underhill RA, Friedman KJ, et al. (2017). Myalgic Encephalomyelitis/Chronic Fatigue Syndrome Diagnosis and Management in Young People: A Primer. Front Pediatr 5(121):1-44. Doi: 10.3389/fped.2017.00121.

Schermuly-Haupt ML, Linden M, Rush AJ (2018). Unwanted Events and Side Effects in Cognitive Behavior Therapy. Cognitive Therapy and Research 42(3):219-229. Doi: 10.1007/s10608-018-9904-y.

Schultz KF, Chalmers I, Hayes RJ, Altman DG (1995). Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. JAMA 273(5):408-412. Doi: 10.1001/jama.273.5.408.

Schulz KF, Grimes DA (2002). Blinding in randomised trials; hiding who got what. The Lancet 359(9307):696-700. Doi: 10.1016/S0140-6736(02)07816-9.

Schunemann HJ, Vist GE, Higgins JPT, et al. (2021). Chapter 15: Interpreting results and drawing conclusions. Cochrane Training. Available from: https://training.cochrane.org/handbook/current/chapter-15#section-15-3.

Shea BJ, Reeves BC, Wells G, et al. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ 358:j4008. Doi: 10.1136/bmj.j4008.

Smith, MEB (2014). Evidence Report/Technology Assessment Disposition of Comments Report: Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. Pacific Northwest Evidence-based Practice Center. Available from: https://effectivehealthcare.ahrq.gov/sites/default/files/related_files/chronic-fatigue_disposition-comments-2015.pdf.

Smith MEB, Nelson HD, Haney E, et al. (2014). Diagnosis and Treatment of Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS). Evid Rep Technol Assess (Full Rep). 2014 Dec;(219):1-433. Doi: 10.23970/AHRQEPCERTA219.

Sun X, Guyatt G (2009). Meta-analysis of randomized trials for health care interventions: one for all? J Evid Based Med 2(1):53-56. Doi: 10.1111/j.1756-5391.2009.01006.x.

Tack M (2021). Comments on the CDC evidence review on ME/CFS. Available from: https://mecfsskeptic.com/comments-on-the-cdc-evidence-review-on-me-cfs/.

van Campen CLMC, Visser FC (2021). Comparing Idiopathic Chronic Fatigue and Myalgic Encephalomyelitis/Chronic Fatigue Syndrome (ME/CFS) in Males: Response to Two-Day Cardiopulmonary Exercise Testing Protocol. Healthcare 9(6):683. Doi: 10.3390/healthcare9060683.

Verbeek JH, Ruotsalainen J, Hoving JL (2011). Synthesizing study results in a systematic review. Scand J Work Environ Health 38(3):282-290. Doi: 10.5271/sjweh.3201.

Vermeulen RC, Vermeulen van Eck IWG (2014). Decreased oxygen extraction during cardiopulmonary exercise test in patients with chronic fatigue syndrome. Journal of Translational Medicine 12:20. Doi: 10.1186/1479-5876-12-20.

Vink M, Vink-Niese A (2018). Graded exercise therapy for myalgic encephalomyelitis/chronic fatigue syndrome is not effective and unsafe. Re-analysis of a Cochrane review. Health Psychology Open Jul-Dec 2018:1-12. Doi: 10.1177/2055102918805187.

Vink M, Vink-Niese A (2019). Cognitive behavioural therapy for myalgic encephalomyelitis/chronic fatigue syndrome is not effective. Re-analysis of a Cochrane review. Health Psychology Open Jan-Jun 2019:1-23. Doi: 10.1177/2055102919840614.

Viswanathan M, Patnode C, Berkman ND, et al. (2017). Assessing the Risk of Bias in Systematic Reviews of Health Care Interventions. Methods Guide for Comparative Effectiveness Reviews. (Prepared by the Scientific Resource Center under Contract No. 290-2012-0004-C). AHRQ Publication No. 17(18)-EHC036-EF. Rockville, MD: Agency for Healthcare Research and Quality. Doi: 10.23970/AHRQ EPC MATHGUIDE.

Wearden AJ, Riste L, Dowrick C, et al. (2006). Fatigue Intervention by Nurses Evaluation – The FINE Trial. A randomised controlled trial of nurse led self-help treatment for patients in primary care with chronic fatigue syndrome: study protocol. BMC Med 4:9. Doi: 10.1186/1741-7015-4-9.

Wearden AJ, Dowrick C, Chew-Graham C, et al. (2010). Nurse led, home based self help treatment for patients in primary care with chronic fatigue syndrome: randomised controlled trial. BMJ 340:c1777. Doi: 10.1136/bmj.c1777.

Westen D, Novotny CM, Thompson-Brenner H, et al. (2004). The empirical status of empirically supported psychotherapies: assumptions, findings, and reporting in controlled clinical trials. Psych Bull 130(4):631-663. Doi: 10.1037%2F0033-2909.130.4.631.

Williams JB, Popp D, Kobak KA, Detke MJ (2012). The power of expectation bias. European Psychiatry 27(1):1. Doi: 10.1016/S0924-9338(12)74807-1.

Wilshire C, Kindlon T, Matthees A, McGrath S (2017). Can patients with chronic fatigue syndrome really recover after graded exercise or cognitive behavioural therapy? A critical commentary and preliminary re-analysis of the PACE trial. Fatigue: Biomedicine, Health & Behavior 5(1):43-56. Doi: 10.1080/21641846.2017.1259724.

Windthorst P, Mazurak N, Kuske M, et al. (2017). Heart rate variability biofeedback therapy and graded exercise training in management of chronic fatigue syndrome: An exploratory pilot study. Jour Psychosomatic Research 93:6-13. Doi: 10.1016/j.jpsychores.2016.11.014.

Wormgoor MEA, Rodenburg SC (2021). The evidence base for physiotherapy in myalgic encephalomyelitis / chronic fatigue syndrome when considering post-exertional malaise: a systematic review and narrative synthesis. J. Transl. Medicine 19(1):1. Doi: 10.1186/s12967-020-02683-4.

Yoder WR, Karyotaki E, Cristea IA, et al. (2019). Researcher allegiance in research on psychosocial interventions: meta-research study protocol and pilot study. BMJ Open 9(2):e024622. Doi: 10.1136/bmjopen-2018-024622.