

## Review Article

**Cite this article:** Ingman T, Smakowski A, Goldsmith K, Chalder T (2022). A systematic literature review of randomized controlled trials evaluating prognosis following treatment for adults with chronic fatigue syndrome. *Psychological Medicine* **52**, 2917–2929. <https://doi.org/10.1017/S0033291722002471>

Received: 25 November 2021

Revised: 20 June 2022

Accepted: 18 July 2022

First published online: 5 September 2022

**Key words:**

chronic fatigue syndrome; cognitive behavioral therapy; graded exercise therapy; literature review; prognosis

**Author for correspondence:**

Tom Ingman,

E-mail: [tom.b.ingman@kcl.ac.uk](mailto:tom.b.ingman@kcl.ac.uk)

# A systematic literature review of randomized controlled trials evaluating prognosis following treatment for adults with chronic fatigue syndrome

Tom Ingman<sup>1</sup> , Abigail Smakowski<sup>2</sup> , Kimberley Goldsmith<sup>3</sup>   
and Trudie Chalder<sup>4</sup> 

<sup>1</sup>Department of Psychology, Institute of Psychiatry, Psychology and Neuroscience, Addiction Sciences Building, King's College London, London, UK; <sup>2</sup>Persistent Physical Symptoms Clinical Research and Treatment Unit, South London and Maudsley NHS Foundation Trust, London, UK; <sup>3</sup>Department of Biostatistics & Health Informatics, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK and <sup>4</sup>Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

**Abstract**

This systematic review investigated randomized controlled trials evaluating cognitive behavioral therapy (CBT) and graded exercise therapy (GET) for adults with chronic fatigue syndrome (CFS). The objective was to determine prognosis following treatment. Studies were eligible if they were peer-reviewed and investigated treatment at least 12 weeks in duration. Studies were excluded if they used co-morbid diagnoses as entry criteria or if they did not measure fatigue, disability, or functioning. Literature published between 1988 and 2021 was searched using MEDLINE, EMBASE, PsycINFO, and Web of Science. Study quality was assessed using the Effective Public Health Practice Project assessment tool. Outcomes were synthesized when three or more studies reported outcomes obtained from the same validated measurement tool. The review included 15 publications comprising 1990 participants. Following CBT, and at short-term to medium-term follow-up, 44% considered themselves better and 11% considered themselves worse. Following GET, and at post-treatment to short-term follow-up, 43% considered themselves better and 14% considered themselves worse. These outcomes were 8–26% more favorable compared to control conditions. Two-thirds of studies were of moderate quality and the remainder were of weak quality. Limitations of this review relate to the clinical heterogeneity of studies and that most outcomes were self-reported. Results suggest some support for the positive effects of CBT and GET at short-term to medium-term follow-up although this requires further investigation given the inconsistent findings of previous reviews. Findings may not be generalizable to severe CFS. This review was registered with PROSPERO (CRD42018086002).

**Introduction**

Chronic fatigue syndrome (CFS) is a serious illness characterized by persistent and medically unexplained fatigue which is severe enough to result in substantial disability. Other symptoms such as musculoskeletal pain, sleep disturbance, and cognitive dysfunction are common (Collin et al., 2016). Some researchers consider myalgic encephalomyelitis (ME) to be the same disorder, while others consider it a different condition with separate diagnostic criteria (Lim & Son, 2020). The current review uses the term CFS, rather than ME, as this has been operationalized in the literature. The National Institute for Health and Care Excellence (NICE, 2007) estimates the prevalence of CFS in the UK to be 0.2–0.4%.

Prognosis, which is defined as the proportion of subjects who improve or worsen according to a specific outcome, and within a discrete time period, plays a critical role in bridging the gap between research and practice (Kent, Cancelliere, Boyle, Cassidy, & Kongsted, 2020; Moons et al., 2018). Indeed, the percentage change, not mean change, may be the most relevant determinant of outcome in CFS (Schluederberg et al., 1992). In their review of naturalistic, cohort, and intervention studies, Cairns and Hotopf (2005) reported that over time, 39% of patients with CFS showed some improvement, 7% recovered, and 5–20% deteriorated.

Prognosis following treatment is less clear, partly because few systematic reviews have reported it and partly because there is no gold standard treatment. Nonetheless, current literature shows that cognitive behavioral therapy (CBT) and graded exercise therapy (GET) are the most promising treatments, both of which yield improvements in fatigue and functioning (Castell, Kazantzis, & Moss-Morris, 2011; Larun, Brurberg, Odgaard-Jensen, & Price, 2019; Malouff, Thorsteinsson, Rooke, Bhullar, & Schutte, 2008; Marques, De Gucht, Gouveia,

© The Author(s), 2022. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Leal, & Maes, 2015; Price, Mitchell, Tidy, & Hunot, 2008). One review reported that, following CBT, between 40% and 48% of patients showed clinical improvements in fatigue, compared to 26% receiving usual care and 27% receiving other therapies, such as relaxation (Price *et al.*, 2008). In another review, Malouff *et al.* (2008) reported that, at follow-up, 50% of patients receiving CBT were within the 'normal range' on a variety of outcomes.

However, reviews such as these are limited by the synthesis of data obtained from multiple measurement tools, the inclusion of non-randomized studies, which may produce between-group imbalances or biased findings, or studies with a short-term follow-up. The use of non-specialist diagnostic procedures, or broad inclusion criteria (e.g. idiopathic chronic fatigue), may also lead to misclassification and yield findings that do not apply to CFS (Newton, Mabillard, Scott, Hoad, & Spickett, 2010). Further research reviewing randomized treatment trials which employ stricter diagnostic criteria, utilize the same outcome measures, and have a longer follow-up, is therefore required to determine prognosis following treatment.

## Aims

This review aimed to assess randomized controlled trials investigating CBT and GET, compared to active or passive control conditions, to extract dichotomous outcomes, and determine the prognosis of CFS in an adult population.

## Method

### Design

A systematic review (PROSPERO registration: CRD42018086002) was conducted following PRISMA guidelines (Moher, Liberati, Tetzlaff, & Altman, 2009). Ethical approval was not required for a systematic review. A person with lived experience of CFS contributed to this research.

### Study selection criteria

Studies were eligible if they: (1) contained original data from an English-language, peer-review journal; (2) were randomized and controlled; (3) included participants who met at least one case definition for CFS, including the Oxford criteria (Sharpe *et al.*, 1991) which require 6 months of severe fatigue affecting physical and mental functioning, is present over 50% of the time, and is accompanied by symptoms such as myalgia, mood, and sleep disturbance; CDC criteria (Fukuda *et al.*, 1994) which require 6 months of persistent or relapsing fatigue, substantially reduced functioning, and at least four additional symptoms (e.g. sleep disturbance, sore throat, headaches, or post-exercise malaise); or NICE criteria (NICE, 2007), which requires 4 months of persistent or relapsing fatigue, substantially reduced activity, post-exertional malaise and/or fatigue, and at least one of ten symptoms, including those listed above. Criteria stipulating that screening or assessment was conducted by a secondary care medical doctor or psychiatrist was added during screening to increase the probability of an accurate diagnosis; (4) investigated individually or group-based CBT or GET, or sufficient components of these treatments (e.g. for CBT, addressing perpetuating cognitive and behavioral factors, and increasing activity, and for GET, incrementally increasing physical exercise) as well as these

treatments in combination with other interventions not routinely applied (e.g. pharmacotherapy when needed), with treatment at least 12 weeks in duration; and (5) described treatment outcome. Active controls (psycho-social, psychological, and/or pharmacological treatments), passive controls (specialist medical care, relaxation, or flexibility treatments), and non-inferiority studies comparing different CBT or GET protocols were included. There was no maximum follow-up period or upper age limit for inclusion.

Studies were excluded if they: (1) included participants <18 years old; (2) used idiopathic chronic fatigue or co-morbid diagnoses as entry criteria; (3) did not include at least one outcome related to fatigue, disability, functioning, or quality of life; and (4) were economic evaluations, ecological or case-control studies, or cross-over trials without an independent control group.

### Databases and search strategy

Published literature was searched using MEDLINE, EMBASE, PsycINFO, and Web of Science, between 1988, the year of first case definition of CFS, and March 2021. Other databases listed in the PROSPERO protocol were not used as four were considered sufficient. The search was performed using the PICO (patient/intervention/comparator/outcome) model with keywords such as chronic fatigue syndrome, cognitive behavioral or graded exercise therapy, randomized controlled trials and prognosis combined with the 'AND' Boolean operator. Searches in MEDLINE, EMBASE, and PsycINFO were limited to human studies and searches in Web of Science excluded studies of children or adolescents. Online Supplementary eAppendix 1 shows the search strategy for each database.

### Study selection

Authors TI and AS independently reviewed all titles and abstracts for eligibility using the Rayyan citation website (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016), in addition to retrieving and reviewing full texts for papers that met inclusion criteria or those in which eligibility was unclear. TI and AS were blinded to each other's review and disagreements regarding eligibility were resolved by discussion.

### Terminology and definitions

Prognosis referred to dichotomous outcomes relating to the proportion of subjects who had improved or worsened according to a specific outcome (e.g. fatigue), and during a discrete time period following treatment. Outcomes were reported at various time-points and categorized as post-treatment, short-term follow-up (1–6 months post-treatment), medium-term follow-up (7–12 months post-treatment), and long-term follow-up (>12 months post-treatment).

### Data extraction

Extracted data were entered onto a Microsoft Excel (Version 16) spreadsheet by TI and checked for accuracy and completeness by AS. The data extraction sheet template can be found in online Supplementary eAppendix 2. To increase validity and focus, outcomes obtained before or after the primary outcome point, which concerned participant subsets, or which were measured using unvalidated measurement tools, and predictive factors, were not extracted. Fulcher and White (1997) was a cross-over trial and

findings following cross-over were not extracted. To maximize the reliability of outcomes obtained from the global impression-improvement scale (CGI-I; Guy, 1976), outcomes from one study (Jason et al., 2007) which defined improvement as 'better', 'much better', or 'very much better', without describing the raw data to allow comparison with other studies, were not reported. Two studies (Clark et al., 2017; Jason et al., 2007) used a variant of the CGI-I (e.g. a 6-point scale), so were also excluded. Adverse effects reported by three studies (Clark et al., 2017; Janse, Worm-Smeitink, Bleijenberg, Donders, & Knoop, 2018; White et al., 2011) were not extracted, as deterioration or exacerbation, also reported by these studies, were considered more relevant to prognosis at follow-up. Additional data related to participant worsening on the CGI-I were obtained from authors of three papers (Clark et al., 2017; Moss-Morris, Sharon, Tobin, & Baldi, 2005; White et al., 2011) and used in the synthesis reported in Table 4. A summary of this data can be found in Tables 2 and 3 (see superscripts) and a further breakdown is available on request from authors of the current review.

### Study quality and risk of bias

Study quality was assessed using the Effective Public Health Practice Project (EPHPP) quality assessment tool for quantitative studies (Thomas, Ciliska, Dobbins, & Micucci, 2004). This tool was chosen because it has content and construct validity, assesses the psychometrics of data collection tools, and may be more reliable than other tools (Armijo-Olivo, Stiles, Hagen, Biondo, & Cummings, 2012; Minozzi, Cinquini, Gianola, Gonzalez-Lorenzo, & Banzi, 2020). The EPHPP comprises six domains (see online Supplementary eAppendix 4) from which a global rating can be calculated. Studies were penalized for confounding if there were clinically important between-group differences in baseline scores of a main outcome (e.g. fatigue or physical functioning). Studies were coded as 'strong' (no weak ratings), 'moderate' (one weak rating), or 'weak' (two or more weak ratings). Studies were independently rated (TI and AS) and disagreements were resolved by consensus. The EPHPP authors were also contacted to clarify that studies should not be penalized for lacking individual reasons for participant withdrawal (D. Ciliska, personal communication, 15 April 2022), which resulted in five studies originally rated as 'poor' to be re-rated as 'moderate' or 'strong' on this outcome (Fulcher & White, 1997; Powell, Bentall, Nye, & Edwards, 2001; White et al., 2011; White, Goldsmith, Johnson, Chalder, & Sharpe, 2013; Wiborg, van Bussel, van Dijk, Bleijenberg, & Knoop, 2015).

### Statistical analysis and data synthesis

The following is an update and extension of the strategy described in the PROSPERO protocol, which was based on a previous review of naturalistic, cohort, and intervention studies reporting prognosis in CFS (Cairns & Hotopf, 2005). It incorporates methods used in reviews of intervention-based trials (Malouff et al., 2008; Price et al., 2008). Data were analyzed using Microsoft Excel (Version 16) and IBM SPSS Statistics for Macintosh (version 26). Means (and ranges) were used to summarize participant age and treatment follow-up and medians (and ranges) to summarize number of treatment sessions and treatment duration. Descriptive statistics (percentages) were used to report variables related to participant and study characteristics such as gender, diagnostic criteria, type of control condition, and prognostic

outcomes for individual studies. Synthesis was conducted when three or more studies reported outcomes obtained from the same measurement tool. Synthesis of participant improvement involved summing the total number of participants who improved in the relevant (CBT or GET) intervention groups in the selected studies divided by the total number randomized to these conditions to produce a weighted mean percentage. This procedure assumed that participants without outcome data were non-improvers. The same procedure was conducted for improvers in non-CBT or non-GET control conditions. Synthesis of participant worsening involved dividing the number of participants who had worsened by the number of outcome completers only, rather than all those randomized, to avoid underestimating this outcome. Differences in prognosis between intervention and control conditions for each synthesized outcome were calculated by subtracting one from the other and 95% confidence intervals (CIs) for these differences were calculated. Post-hoc sensitivity analysis of global worsening was conducted assuming that those without outcome data had worsened. To assess the robustness of findings, additional analyses were conducted which excluded data from studies of weak quality thereby generating estimates of prognosis based on studies of at least moderate quality. Meta-analyses were not conducted as it was not clear how many outcomes would fulfill criteria for pooling data and therefore it was not possible to specify primary outcomes *a-priori*. Conducting post-hoc meta-analyses may have increased the chance of multiple testing and reporting biases. Furthermore, several previous reviews have assessed the efficacy of CBT and GET using meta-analytic methods (Castell et al., 2011; Larun et al., 2019; Malouff et al., 2008; Marques et al., 2015; Price et al., 2008).

### Unit of analysis

The intervention and control arms of studies investigating one CBT or GET condition were compared. Data were pooled when intervention arms in the same study had more than one CBT or GET condition (i.e. Janse et al. 2018; Powell et al. 2001; Wiborg et al. 2015; Worm-Smeitink et al. 2019) and when both the intervention and control in the same study comprised different CBT conditions (i.e. Tummers, Knoop, & Bleijenberg, 2010; Worm-Smeitink et al. 2019).

## Results

### Description of included studies

After removing duplicates, 3817 studies were identified. After screening, 3767 (99%) were excluded, resulting in 50 full texts. Reasons for exclusion and references to excluded studies can be found in online Supplementary eAppendix 3. Reviewing full texts led to further exclusion of 35 studies, resulting in 15 included publications. White et al. (2011, 2013) reported from the same trial which resulted in 11 publications with a CBT condition (see Table 2), reporting from ten trials, and six studies with a GET condition (see Table 3), reporting from five trials. Figure 1 shows a flow diagram of the selection process.

### Reliability of study selection

Reviewers were initially discrepant on 32 (1%) occasions when screening titles and abstracts for study eligibility. These discrepancies were resolved through discussion.

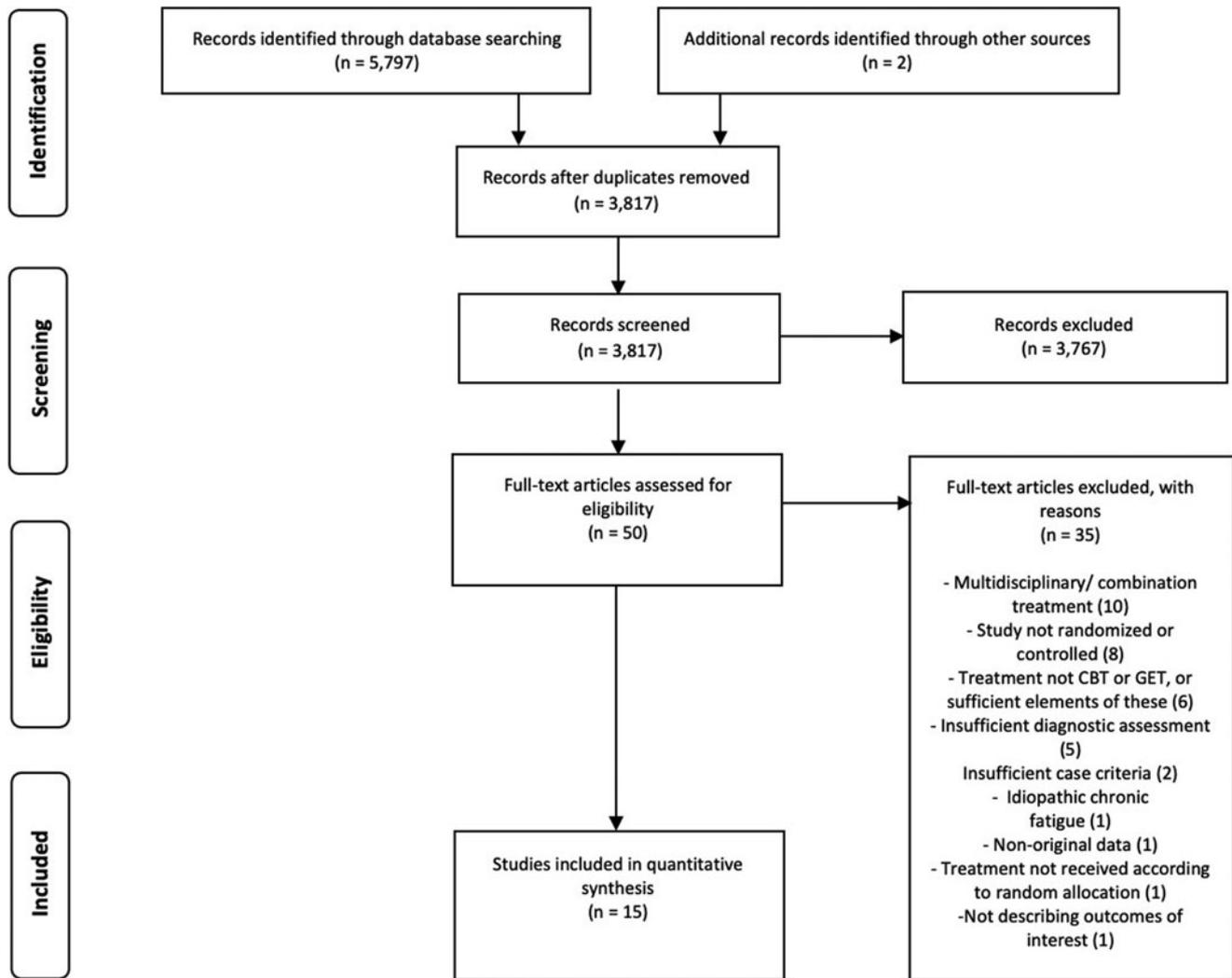


Fig. 1. PRISMA flowchart of study selection.

### Study characteristics

Study characteristics are shown in Tables 2 and 3. Studies were conducted in the UK ( $n = 7$ ), the Netherlands ( $n = 6$ ), the USA ( $n = 1$ ), and New Zealand ( $n = 1$ ).

### Participant characteristics

Eligible studies comprised 1990 participants, excluding 1266 reported in another included study or in an excluded treatment arm. As seen in Tables 2 and 3, a diagnosis of CFS was made using the CDC criteria ( $n = 8$ , 50%), the Oxford criteria ( $n = 5$ , 31%), both the CDC and the Oxford criteria ( $n = 1$ , 7%), or the NICE (2007) criteria ( $n = 1$ , 6%).

### Characteristics of CBT

As shown in Table 2, treatment in six of the ten trials with a CBT condition was face-to-face. Of the remaining four trials, one (Janse et al., 2018) comprised internet-based CBT with higher (fortnightly therapist contact) and lower resource conditions (therapist contact when requested), and one (Knoop, van der Meer, & Bleijenberg, 2008) consisted of self-help manuals; clinician contact in both was mostly remote (e.g. via telephone).

Tummers et al. (2010) and Worm-Smeitink et al. (2019) investigated a stepped-care approach incorporating the treatment protocol in Knoop et al. (2008) and Janse et al. (2018), respectively, followed by face-to-face CBT if required. The control condition in Tummers et al. (2010) and Worm-Smeitink et al. (2019) comprised face-to-face CBT.

Participants were offered a median of 14 sessions (range: 13–16) in eight trials with a face-to-face CBT condition. Participants in White et al. (2011, 2013) were offered a follow-up CBT session and received a median of three specialist medical care sessions. Participants in Sharpe et al. (1996) were offered medical care consultations (frequency not reported). Participants in the higher resource Internet-based cognitive behavioral therapy (iCBT) condition in Janse et al. (2018) were instructed to contact their therapist 12 times. The self-help CBT in Knoop et al. (2008) required participants to contact their therapist every 2 weeks over a 16-week period. The median duration of treatment offered in CBT conditions was 26 weeks (range: 16–52).

### Characteristic of GET

Table 3 shows that three trials (Fulcher & White, 1997; Moss-Morris et al., 2005; White et al., 2011; 2013) investigated

face-to-face GET, with a median of 12 sessions offered (range: 12–14). Participants in White et al. (2011, 2013) were offered a follow-up session and received a median of three specialist medical care sessions. Clark et al. (2017) investigated a 12-week self-help graded exercise treatment comprising four guidance sessions, delivered face-to-face or remotely, plus  $\geq 1$  specialist medical care consultation. Powell et al. (2001) investigated various ‘doses’ of education to encourage a self-managed graded exercise program, comprising between three and ten sessions, delivered face-to-face or remotely, with additional consultations available on request. The median duration of GET treatment was 12 weeks (range: 12–23).

### Characteristics of (non-CBT/GET) comparison conditions

The control condition in five trials (Clark et al., 2017; Moss-Morris et al., 2005; Powell et al., 2001; Sharpe et al., 1996; White et al., 2011; 2013) comprised  $\geq 1$  sessions of specialist or standard medical care (e.g. generic advice or medication), in three trials (Deale, Chalder, Marks, & Wessely, 1997; Fulcher & White, 1997; Jason et al., 2007) comprised relaxation or flexibility treatment (between ten and 13 sessions offered), and in three trials (Janse et al., 2018; Knoop et al., 2008; Wiborg et al., 2015) consisted of a wait list control. The control in one trial (Vos-Vromans et al., 2016) was multidisciplinary rehabilitation treatment, comprising 44.5 h of CBT, body awareness therapy, pacing, mindfulness, social reintegration, and sleep normalization, delivered by a multidisciplinary team of clinicians.

### Outcome equivalence in pooled treatment conditions

There were no significant differences between pooled conditions in primary and secondary outcomes (e.g. fatigue and physical functioning) in trials comparing higher and lower resource or intensity CBT (Janse et al., 2018; Tummers, Knoop, van Dam, & Bleijenberg, 2013; Wiborg et al., 2015; Worm-Smeitink et al., 2019) or higher and lower intensity GET (Powell et al., 2001)

### Outcome measures

Table 1 lists the outcome measures and cut-off scores used by each included study. Fatigue was an outcome in nine studies, with six of these (Janse et al., 2018; Knoop et al., 2008; Tummers et al., 2010; Vos-Vromans et al., 2016; Wiborg et al., 2015; Worm-Smeitink et al., 2019) using the self-report Checklist of Individual Strength (CIS; Vercoulen et al., 1994; range: 8–56; higher = more fatigue) and three (Clark et al. 2017; Deale et al. 1997; White et al. 2011) using the 11-item self-report Chalder Fatigue Questionnaire (CFQ; Chalder et al., 1993) which can be scored bimodally (range: 0–11) or on a Likert scale (range: 0–33), with higher scores indicating more fatigue.

Physical functioning was an outcome in seven studies, all of which used the self-rated Short-Form Health Survey (SF-36) physical functioning subscale (Ware & Sherbourne, 1992). This measures limitations in activities such as ‘walking 100 yards’ (range: 0–100; higher = less limited). General functioning was an outcome in one study (Sharpe et al., 1996) and was measured using the clinician-rated Karnofsky scale (Karnofsky, Abelmann, Craver, & Burchenal, 1948) following an interview with the participant (and cohabitee when possible) about their activities over the previous month (range: 0–100; higher = less limited).

Global change in health was an outcome in six studies, all of which used the self-rated Clinical Global Impression-

Improvement scale (CGI-I; Guy, 1976). This asks respondents to rate change on a 7-point scale from ‘very much better’ to ‘very much worse’ compared to study onset. Authors of studies listed in Table 1 defined improvement as ‘much better’ or ‘very much better’. Study authors defined worsening using various combinations of responses, therefore to maximize available data, ratings of ‘a little worse’, ‘much worse’, and ‘very much worse’ were combined and reported in the current review as ‘global worsening’.

Other prognostic outcomes included disability, mental distress, levels of activity, and employment. Composite measures were used to define ‘recovery’ in two studies (White et al., 2011, 2013) and ‘deterioration’ in a further two studies (Clark et al., 2017; White et al., 2011). Primary outcomes, as defined by study authors, are described in Tables 2 and 3. The abovementioned measurement tools are valid and reliable, except for the reliability of the CGI-I, which is difficult to ascertain given that it captures self-reported change relative to pre-treatment.

### Study quality and risk of bias

The quality rating of ten studies (Clark et al., 2017; Deale et al., 1997; Fulcher & White, 1997; Moss-Morris et al., 2005; Powell et al., 2001; Sharpe et al., 1996; Vos-Vromans et al., 2016; White et al., 2011; 2013; Wiborg et al., 2015) was ‘moderate’ and the remainder ‘weak’ (Tables 2 and 3). Ratings by domain can be seen in online Supplementary eAppendix 4. All studies were penalized for the lack of blinding and five (33%) for not adequately describing drop-out per study group. Jason et al. (2007) was also penalized for confounding given the potential for clinically significant between-group differences in baseline scores of physical functioning which were not controlled for in the (pre-specified) analysis or discussed satisfactorily by authors.

### Main findings

#### Prognosis following CBT

Table 2 shows all outcomes following CBT. Table 4 shows that four outcomes, obtained using three measurement tools, were included in the synthesis: (1) significant improvements in fatigue (at post-treatment to short-term follow-up); (2) significant improvements in physical functioning (at short-term to medium-term follow-up); (3) global improvements (at short-term to medium-term follow-up); and (4) global worsening (at short-term to medium-term follow-up). The synthesis included data from nine (90%) CBT trials. Table 1 shows that the same criteria were used for calculating global improvements and global worsening, but not for calculating improvements in physical functioning. The same cut-off scores were used to calculate improvements in fatigue, however, all but one study (Vos-Vromans et al., 2016) used the reliable change index as additional criteria when calculated this outcome.

Table 4 shows that the proportion of improved participants was higher following CBT, compared to control, for all synthesized outcomes: 23% (95% CI 16–31) more participants improved in fatigue, 14% (95% CI 4–23) more improved in physical functioning, and 20% (95% CI 11–29) more rated themselves as improved. Table 4 also shows that 8% (95% CI 1–14) fewer participants reported global worsening following CBT compared to control. Global worsening following CBT remained lower than in the control condition when it was assumed that participants lost to follow-up had also worsened (18% v. 23%, respectively).

**Table 1.** Criteria for prognostic outcomes following treatment

Reference	Outcome	Criteria
Clark <i>et al.</i> (2017)	Global improvement Clinically significant improvement in fatigue and physical functioning Increase in physical activity Global worsening <sup>a</sup> Serious deterioration	CGI-I ratings of 'much better' or 'very much better' ≥3-point Likert score decrease on the 11-item CFQ and an ≥8-point increase on the SF-36 physical functioning subscale High level of activity in the previous week according to the IPAQ CGI-I ratings of 'worse', 'much worse', or 'very much worse' ≥10-point reduction on the SF-36 physical functioning subscale, CGI-I ratings of 'much worse' or 'very much worse', or active withdrawal due to worsening
Deale <i>et al.</i> (1997)	Global improvement Clinically significant improvement in physical functioning No longer fatigue case No longer psychological case (mood and distress) No longer meeting CFS criteria Global worsening <sup>a</sup>	CGI-I ratings of 'much better' or 'very much better' Score of ≥83 or a ≥50-point increase on the SF-36 physical functioning subscale  Bimodal scores of <4 on the 11-item CFQ Score of <4 on the GHQ  Oxford and CDC criteria CGI-I ratings of 'a little worse', 'much worse', or 'very much worse'
Fulcher and White (1997)	Global improvement Global worsening	CGI-I ratings of 'much better' or 'very much better' CGI-I ratings of 'a little worse', 'much worse', or 'very much worse'
Janse <i>et al.</i> (2018)	Normal range for fatigue Exacerbation in fatigue, physical functioning, disability, and distress	RCI >1.96 and a score of <35 on the CIS fatigue subscale RCI of <-1.96 on the CIS fatigue subscale, SF-36 physical functioning subscale, SIP8, and SCL-90
Jason <i>et al.</i> (2007)	Clinically significant improvement in physical functioning No longer meeting CFS criteria Improved work status	Change score exceeds age-adjusted RCI on the SF-36 physical functioning subscale and follow-up scores within 1 s.d. of the normative value <sup>b</sup> CDC criteria  Working ≥20 h per week or in full-time education
Knoop <i>et al.</i> (2008)	Clinically significant improvement in fatigue	RCI >1.96 and a score of <35 on the CIS fatigue subscale
Moss-Morris <i>et al.</i> (2005)	Global improvement Global worsening <sup>a</sup>	CGI-I ratings of 'much better' or 'very much better' CGI-I ratings of 'a little worse', 'much worse', or 'very much worse'
Powell <i>et al.</i> (2001)	Global improvement Clinically significant improvement in physical functioning	CGI-I ratings of 'much better' or 'very much better' Score of ≥25 or a ≥10-point increase on the SF-36 physical functioning subscale <sup>c</sup>
Sharpe <i>et al.</i> (1996)	Global improvement Clinically significant improvement in functioning Normal range for functioning Improved work status Global worsening	CGI-I ratings of 'much better' or 'very much better' ≥10-point increase on the Karnofsky scale Score of ≥80 on the Karnofsky scale Working or studying CGI-I ratings of 'a little worse', 'much worse', or 'very much worse'
Tummers <i>et al.</i> (2010)	Clinically significant improvement in fatigue	RCI >1.96 and a score of <35 on the CIS fatigue subscale
Vos-Vromans <i>et al.</i> (2016)	Clinically significant improvement in fatigue	Score of <35 on the CIS fatigue subscale
Wiborg <i>et al.</i> (2015)	Clinically significant improvement in fatigue, physical functioning, and disability Recovery	RCI >1.96 and a score of <35 on the CIS fatigue subscale, score of ≥65 on the SF-36 physical functioning subscale, and <700 on the SIP-8 Score of <27 on the CIS fatigue subscale, ≥80 on the SF-36 physical functioning subscale, and <203 on the SIP-8
White <i>et al.</i> (2011)	Global improvement Clinically significant improvement in fatigue and physical functioning Normal range for fatigue and physical functioning Global worsening <sup>a</sup> Serious deterioration	CGI-I ratings of 'much better' or 'very much better' ≥2-point Likert score decrease on the 11-item CFQ and an ≥8-point increase on the SF-36 physical functioning subscale Likert score of ≤18 on the 11-item CFQ and ≥60 on the SF-36 physical functioning subscale  CGI-I ratings of 'a little worse', 'much worse', or 'very much worse' A decrease (between baseline and two consecutive assessments) of ≥20 on the SF-36 physical function subscale, CGI-I ratings of 'much worse' or 'very much worse' at two consecutive assessments, withdrawal after 8 weeks because of feeling worse, or a serious adverse reaction (determined by a masked clinician)
White <i>et al.</i> (2013)	Recovery	Likert score of ≤18 on the 11-item CFQ, ≥60 on the SF-36 physical functioning subscale, not meeting Oxford criteria, and CGI-I ratings of 'much better' or 'very much better'
Worm-Smeitink <i>et al.</i> (2019)	Clinically significant improvement in fatigue	RCI >1.96 and a score of <35 on the CIS fatigue subscale

Note. CGI-I, Clinical Global Impressions-Improvement scale; CFQ, Chalder Fatigue Questionnaire; SF-36, Short-Form Health Survey-36; IPAQ, International Physical Activity Questionnaire; GHQ, General Health Questionnaire; CDC, Centre for Disease Control and Prevention; RCI, reliable change index; CIS, Checklist of Individual Strength; SIP8, Sickness Impact Profile; SCL-90, Symptom Checklist 90; s.d., standard deviation.

<sup>a</sup>Criteria applied in this review but not used by study authors.

<sup>b</sup>Not reported by study authors.

<sup>c</sup>Reported on a 10–30 scale; when converted to the more common 0–100 scale by using the following formula:  $\text{mean}_{\text{new}} = (\text{mean}_{\text{old}} - 10) \times 5$ , a score of 25 = 75 and a 10-point increase = 50 points.

**Table 2.** Characteristics of included CBT studies, their participants, and prognosis

Reference (author/year)	Setting and diagnostic criteria	Intervention and control	Sample size Intervention (#) Control (#)	Population: age (mean years); gender (% female)	Primary outcome and follow-up	Prognosis following CBT treatment(s) (% of participants)	EPHPP global quality rating
Deale et al. (1997)	CFS clinic, UK; Oxford and CDC Criteria	CBT v. relaxation	Intervention: 30 Control: 30	Intervention: 31; 70 Control: 38; 67	Physical functioning; short-term follow-up	63% globally improved; 63% significantly improved in physical functioning; 43% improved in mood; 57% no longer fatigue cases; 50% no longer met CFS criteria; 0% globally worsened	Moderate
Janse et al. (2018)	CFS Clinic, Netherlands; CDC Criteria	iCBT with feedback (i) or iCBT with feedback on demand only (ii) v. WL	Intervention: 80 in both conditions Control: 80	iCBT (i) 37; 68. iCBT (ii) 36; 58 Control: 40; 56	Fatigue; post-treatment	39% in normal range for fatigue; 5% experienced exacerbation in fatigue, 3% in impairment, 5% in physical functioning, and 9% in psychological distress	Weak
Jason et al. (2007)	Hospital, USA; CDC Criteria	CBT v. relaxation	Intervention: 29 Control: 28	Entire sample: 44; 83	None stated; medium-term follow-up	18% <sup>a</sup> significantly improved in physical functioning; 14% no longer met CFS criteria; 17% <sup>a</sup> improved in work status	Weak
Knoop et al. (2008)	CFS clinic, Netherlands; CDC criteria	Guided self-instruction CBT v. WL	Intervention: 84 <sup>p</sup> Control: 85 <sup>b</sup>	See Tummers et al. (2010)	Fatigue, physical functioning, and disability; post-treatment	27% significant improved in fatigue	Weak
Sharpe et al. (1996)	Infectious disease outpatient clinic, UK; Oxford criteria	CBT v. StMC	Intervention 30 Control: 30	Intervention: 34; 60 Control: 38; 77	Functioning; medium-term follow-up	60% globally improved; 73% significantly improved in physical functioning; 73% in normal range for physical functioning; 63% improved in work status; 7% deteriorated and attributed this to treatment; 13% globally worsened	Moderate
Tummers et al. (2010)	CFS clinic, Netherlands; CDC criteria	Self-help CBT v. WL (both conditions followed by face-to-face CBT if needed)	Intervention: 84 <sup>p</sup> Control: 85 <sup>b</sup>	Intervention: 38; 82 Control: 39; 75	Fatigue, physical functioning, and disability; post-treatment	48% significantly improved in fatigue	Weak
Vos-Vromans et al. (2016)	Rehabilitation center, Netherlands; CDC Criteria	MRT v. CBT	Intervention: 62 Control: 60	Intervention: 40; 81 Control: 41; 78	Fatigue and health-related quality of life; short-term follow-up	26% significantly improved in fatigue	Moderate
Wiborg et al. (2015)	CFS outpatient clinic, Netherlands; CDC criteria	CBT group therapy: large CBT group (i) and small CBT group (ii) v. WL	Intervention: CBT (i): 68 CBT (ii): 68 Control: 68	Intervention: CBT (i) 36; 75 CBT (ii) 40; 74 Control: 37; 74	Fatigue, physical functioning and disability; post-treatment	49% significantly improved in fatigue and 38% in all primary outcome measures; 15% recovered	Moderate
White et al. (2011)	Secondary care clinics, UK; Oxford Criteria	CBT v. SpMC	Intervention: 161 Control: 160	Intervention: 39; 80	Fatigue and physical functioning;	38% globally improved; 70% significantly improved in fatigue and 65% in physical functioning; 27% in	Moderate

(Continued)

Table 2. (Continued.)

Reference (author/year)	Setting and diagnostic criteria	Intervention and control	Sample size Intervention (#) Control (#)	Population: age (mean years); gender (% female)	Primary outcome and follow-up	Prognosis following CBT treatment(s) (% of participants)	EPHPP global quality rating
White et al. (2013)	Secondary care clinics, UK; Oxford Criteria	CBT v. SpMC	Intervention: 161 Control: 160	Control: 37; 76  Entire sample: 38; 77	short-term follow-up  Fatigue and physical functioning, short-term follow-up	normal ranges for fatigue and physical functioning; 13% <sup>c</sup> globally worsened; 9% seriously deteriorated  20% recovered	Moderate
Worm-Smeitink et al. (2019)	CFS clinic, Netherlands; CDC criteria	iCBT with feedback (i) or iCBT with feedback on demand only (ii), v. WL (all conditions followed by face-to-face CBT if needed)	Intervention: 121 in both conditions Control: 121	iCBT (i) 37; 64 iCBT (ii) 37; 57 Control 39; 61	Fatigue; post-treatment	45% significantly improved in fatigue	Weak

Note. EPHPP, Effective Public Health Practice Project; CFS, chronic fatigue syndrome; CDC, Centre for Disease Control and Prevention; CBT, cognitive behavioral therapy; iCBT, internet-based cognitive behavioral therapy; WL, waiting list; SpMC, standard medical care; MRT, multidisciplinary rehabilitation treatment; SpMC, specialist medical care.

<sup>a</sup>Study authors did not report if outcomes were based on all participants randomized to treatment, or a subset who completed outcomes only.

<sup>b</sup>Following randomization, one participant from the intervention and one from the control condition were excluded due to medical explanations for their fatigue.

<sup>c</sup>Data obtained from study author.

When excluding three studies of weak quality used in the synthesis of fatigue (Knoop et al., 2008; Tummers et al., 2010; Worm-Smeitink et al., 2019), the proportion of participants who showed significant improvement was 42% following CBT and 28% in the control group (14% difference; 95% CI 0–24), 1% lower and 8% higher compared to the original analysis, respectively. When excluding one study of weak quality used in the synthesis of physical functioning (Jason et al., 2007), the proportion of participants who showed significant improvement was 65% following CBT and 49% in the control group (16% difference; 95% CI 1–26), 6% and 4% higher than in the original analysis, respectively. All data used in the synthesis of global improvement and global worsening were drawn from studies of moderate quality.

### Prognosis following GET

Table 3 shows all outcomes following GET. Table 4 shows that three outcomes, obtained using two outcome measures, were included in the data synthesis: (1) significant improvement in physical functioning (at post-treatment to medium-term follow-up); (2) global improvement (at post-treatment to short-term follow-up); and (3) global worsening (at post-treatment to short-term follow-up). All five GET trials contributed data to synthesis of these outcomes. Table 1 shows that the same criteria were used for outcomes except for physical functioning.

Table 4 shows that the proportion of improved participants was higher following GET, compared to control, for both synthesized outcomes: 23% (95% CI 15–31) more participants receiving GET significantly improved in physical functioning compared to those in the control and 26% (95% CI 19–32) more reported global improvements. Table 4 also shows that 10% (95% CI 4–16) fewer participants reported global worsening following GET compared to the control condition. Global worsening following GET remained lower than in the control condition when participants lost to follow-up were assumed to have worsened (21% v. 28%, respectively). All data used in the synthesis of each of the above outcomes were drawn from GET studies of moderate quality.

## Discussion

This paper reviewed studies reporting the prognosis of CFS following CBT and GET. The review included 15 publications, comprising 11 studies of CBT and six of GET. Results, implications, and limitations will now be discussed.

### Prognosis following CBT

Findings from the synthesis revealed that, at post-treatment to short-term follow-up, fatigue significantly improved in 43% of participants, 23% more than in control conditions. These results are consistent with Price et al. (2008) who found that 40–48% of participants showed clinically significant post-treatment improvements in fatigue, significantly more than in control groups. Findings also showed that, at short-term to medium-term follow-up, 44% of participants considered themselves significantly improved, 20% more than in control groups, and 59% demonstrated clinically significant improvements in physical functioning, 14% more than in control groups. These results are in line with the positive effects of CBT at short-term to medium-term follow-up, although further work is needed to build on these given the inconsistent findings of previous reviews. For example, Price et al. (2008) found that, at short-term to medium-term

**Table 3.** Characteristics of included GET studies, their participants, and prognosis

Reference (author/year)	Setting and diagnostic criteria	Intervention and control	Sample size Intervention (#) Control (#)	Population characteristics: age (mean years); gender (% female)	Primary outcome	Prognosis following GET treatment(s) (% of participants)	EPHPP global quality rating
Clark et al. (2017)	CFS clinic, UK; NICE criteria	Self-help GET v. SpMC	Intervention: 107 Control: 104	Intervention: 38; 82 Control: 39; 76	Fatigue and physical functioning; post-treatment	16% globally improved; 58% significantly improved in fatigue and 41% in physical functioning; 24% increased physical activity; 21% seriously deteriorated; 19% <sup>a</sup> globally worsened	Moderate
Fulcher and White (1997)	CFS clinic, UK; Oxford Criteria	GET v. flexibility and relaxation	Intervention: 33 Control: 33	Entire sample: 37; 74	Global improvement; post-treatment	52% globally improved; 3% globally worsened	Moderate
Moss-Morris et al. (2005)	CFS private general practice, New Zealand; CDC Criteria	GET v. StMC	Intervention: 25 Control: 24	Intervention: 37; 60. Control: 46; 79	Global improvement; post-treatment	48% globally improved; 18% <sup>a</sup> globally worsened	Moderate
Powell et al. (2001)	CFS clinic and an infectious disease outpatient clinic, UK; Oxford criteria	'Doses' of education to encourage GET: minimum (i), telephone (ii), and maximum (iii) conditions v. StMC	Intervention: GET (i) 37 GET (ii) 39 GET (iii) 38 Control: 34	Intervention: GET (i) 34; 76 GET (ii) 32; 85 GET (iii) 33; 82 Control: 34; 71	Fatigue and physical functioning; medium-term follow-up	70% globally improved; 69% significantly improved in physical functioning	Moderate
White et al. (2011)	Secondary care clinic, UK; Oxford Criteria	GET v. SpMC	Intervention: 160 Control: 160	Intervention: 39; 77. Control: 37; 76	Fatigue and physical functioning; short-term follow-up	39% globally improved; 77% significantly improved in fatigue and 68% in physical functioning; 27% in normal ranges for fatigue and physical functioning; 13% <sup>a</sup> globally worsened; 6% seriously deteriorated	Moderate
White et al. (2013)	Secondary care clinics, UK; Oxford Criteria	GET v. SpMC	Intervention: 160 Control: 160	Entire sample: 38; 77	Fatigue and physical functioning; short-term follow-up	20% recovered	Moderate

Note. EPHPP, Effective Public Health Practice Project; CFS, chronic fatigue syndrome; GET, graded exercise therapy; StMC, standard medical care; NICE, National Institute for Health and Clinical Excellence; SpMC, specialist medical care; CDC, Centre for Disease Control and Prevention.

<sup>a</sup>Data obtained from study author.

**Table 4.** Prognosis following CBT, GET, and control conditions

Outcome	CBT			GET		
	Study (N)	Intervention	Control <sup>a</sup>	Study (N)	Intervention	Control
Significant improvement in fatigue severity <sup>b</sup>	5 <sup>1-5</sup>	43 (26–49)	20 (7–49)	–	–	–
Significant improvement in physical functioning <sup>c</sup>	3 <sup>6-8</sup>	59 (18–65)	45 (17–55)	3 <sup>6,10,11</sup>	61 (41–69)	38 (6–55)
Global improvement <sup>d</sup>	3 <sup>6,7,9</sup>	44 (38–63)	24 (23–27)	5 <sup>6,10-13</sup>	43 (16–70)	17 (5–27)
Global worsening <sup>e</sup>	3 <sup>6,7,9</sup>	11 (0–13)	19 (10–20)	4 <sup>6,10,12,13</sup>	14 (3–19)	24 (3–35)

Note. Data are mean percentages and ranges unless specified otherwise.

1: Knoop et al. (2008); 2: Tummers et al. (2010); 3: Wiborg et al. (2015); 4: Worm-Smeitink et al. (2019); 5: Vos-Vromans et al. (2016); 6: White et al. (2011); 7: Deale et al. (1997); 8: Jason et al. (2007); 9: Sharpe et al. (1996); 10: Clark et al. (2017); 11: Powell et al. (2001); 12: Moss-Morris et al. (2005); 13: Fulcher and White (1997).

<sup>a</sup>Excludes conditions in which the control was also CBT.

<sup>b</sup>Measured using the Checklist of Individual Strength.

<sup>c</sup>Measured using the Short-Form Health Survey-36 physical functioning subscale.

<sup>d</sup>Ratings of 'much better' or 'very much better' on the Clinical Global Impressions-Improvement scale.

<sup>e</sup>Ratings of 'a little worse', 'much worse', or 'very much worse' on the Clinical Global Impressions-Improvement scale.

follow-up, CBT led to significant improvements in physical functioning compared to other treatments such as relaxation and counselling, but not compared to treatment as usual. Finally, the synthesis revealed that, at short-term to medium-term follow-up, 11% of participants considered themselves worse, 8% fewer than in control conditions.

### Prognosis following GET

Findings from the data synthesis revealed that at post-treatment to medium-term follow-up, 61% of participants demonstrated clinically significant improvements in physical functioning, 23% more than in control conditions, and at post-treatment to short-term follow-up, 43% considered themselves significantly better, 26% more than in control conditions. These findings are consistent with a previous review of GET (Larun et al., 2019) which reported that treatment may moderately improve physical functioning and increase the number of people who report at least some improvement at post-treatment, although findings regarding outcomes at short, medium, and long-term are mixed, and based on fewer studies.

The synthesis also showed that, at post-treatment to short-term follow-up, 14% of participants considered themselves worse, 10% fewer than in control groups. This is consistent with a review of GET (White & Etherington, 2021) which found no evidence of excess harm with GET by either self-rated deterioration or by withdrawing, although their finding that more GET participants dropped out of trial follow-up in comparison with control interventions requires further investigation.

### Clinical implications

Our analysis indicates that, at post-treatment to medium-term follow-up, CBT and GET result in a better prognosis compared to control conditions. These findings are consistent with previous reviews reporting post-treatment outcomes and may provide some support for the positive effects of CBT and GET at short-term to medium-term follow-up, although this requires further investigation given that a meta-analysis was not conducted and the inconsistent findings of previous reviews. Nonetheless, our findings are consistent with previous findings that CBT and GET yield similar improvements in fatigue and functioning. Commissioners may want to consider this when providing treatment.

### Limitations

A limitation of this review is that one-third of included studies were of weak quality. All studies were penalized for a lack of blinding, which increases the risk of performance or detection bias. However, blinding is not possible in trials of psychotherapeutic treatments while it has been suggested that concerns over bias caused by lack of blinding may be exaggerated (Moustgaard et al., 2020). Five CBT studies were penalized because drop-outs were not adequately described, which can indicate a vulnerability to attrition bias. However, when weak studies were excluded from the analysis, estimates of improvement in fatigue and functioning following CBT remained similar, while all studies used in the synthesis of global improvement and worsening, as well all GET studies, were of moderate quality. This is consistent with previous reviews of CBT and GET which found that the quality of methodological features such as attrition and blinding do not predict effect size for improvements in fatigue and physical functioning (Castell et al., 2011; Malouff et al., 2008; Marques et al., 2015). This suggests that the inclusion of studies rated as weak may not significantly affect estimates of prognosis in the current review.

Another limitation relates to the clinical heterogeneity of included studies, which used different diagnostic criteria, investigated different intensities and doses of treatments, and assessed outcomes at different follow-ups. This may have contributed to between-study differences in prognosis. Indeed, even though they are similar, studies using Oxford criteria for CFS may yield non-significantly higher effect sizes than those using CDC criteria (Castell et al., 2011; Larun et al., 2019). However, all included studies investigated CBT or GET which are underpinned by a bio-psycho-social model of CFS, increasing their comparability. Another contributor to heterogeneity relates to the different thresholds used by study authors to determine improvements in physical functioning following both CBT and GET, and this outcome should therefore be interpreted with caution. The same absolute cut-off scores were used by study authors to calculate improvements in fatigue following CBT, although the reliable change index was used inconsistently as additional criteria and this outcome should therefore also be interpreted with caution.

A further limitation is that most studies reported subjective, self-report measures, which may have increased the risk of observer or detection bias. However, two included studies (Jason et al.,

2007; Sharpe et al., 1996) reported that CBT yielded greater increases in employment than did control conditions. Nonetheless, other research (Deale, Husain, Chalder, & Wessely, 2001; McCrone et al., 2012) has found only modest increases in employment, while subjective and objective measures do not necessarily correlate (King, Beynon, Chalder, Sharpe, & White, 2020). These findings suggest that future trials should obtain a range of outcome measures and investigate potential discrepancies between them.

The risk of review-level bias, and missing relevant articles, was minimized by using four databases to search the literature, in addition to manual and reference list searches. Publication bias was not assessed, and although Castell et al. (2011) found little difference between the effect sizes of published and unpublished CBT trials, Marques et al. (2015) found some evidence of reporting bias for fatigue and physical functioning in their review of interventions with a graded exercise component. Whilst the extent of reporting bias in the current review is unclear, the potential for this bias was minimized by extracting dichotomized outcomes regardless of whether they pertained to the primary outcome analysis, in addition to obtaining unpublished data relating to global worsening in three studies included in the data synthesis.

A final limitation is that treatment in most trials involved at least some face-to-face sessions, thereby excluding less mobile or housebound participants, therefore findings may not be generalizable to severe CFS.

### Future directions

Future trials should report objective outcomes, in addition to self-report measures, which are essential as fatigue is subjective and individuals are best placed to judge how they feel. Further work also needs to investigate the statistical significance of between-treatment differences in prognosis, particularly in the longer term, and whether differences are moderated by diagnostic criteria and treatment intensity. There is a need to further define and measure deterioration and recovery. Further trials, not only of CBT and GET, but also other treatments, are required, particularly for individuals with severe CFS.

### Conclusions

Findings of this review indicate that CBT and GET lead to a more favorable prognosis compared to control groups comprising relaxation, medical care, and wait-list conditions. These findings correspond with existing reviews reporting improvements in post-treatment symptoms; however, previous findings concerning short-term to medium-term follow-up are inconsistent and require further investigation. Further work also needs to improve existing treatments and to explore new treatments, particularly for those with severe CFS.

**Supplementary material.** The supplementary material for this article can be found at <https://doi.org/10.1017/S0033291722002471>

**Acknowledgements.** This paper represents independent research part-funded by the National Institute for Health Research (NIHR) Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London. KG from KCL is supported by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration South London (NIHR ARC South London) at King's College Hospital NHS Foundation Trust. The NIHR had no role in the study design, collection, analysis, or interpretation of the data, writing the

manuscript, or the decision to submit the paper for publication. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The authors thank Donna Ciliska, Lucy Clark, Marcia Tummers, Margreet Worm-Smeitink, Peter White, Richard Bentall, and Rona Moss-Morris for providing additional information about, and clarification regarding, studies in which they were involved. The authors also thank Kim Poole-White for her guidance and advise in drafting this manuscript.

**Conflict of interest.** TC has received ad hoc payments for conducting workshops on evidence-based treatments for persistent physical symptoms, and expenses (i.e. travel and accommodation) for workshops delivered on behalf of the British Association for Behavioural and Cognitive Psychotherapies (BABCP) and the Improved Access to Psychological Therapies (IAPT) service, which is part of the National Health Service (NHS) in the UK. TC has also received grants from the National Institute for Health Research (NIHR), the Health Technology Assessment (HTA) program, the Research for Patient Benefit (RFPB) program, Guy's and St Thomas Charity, the King's Challenge Fund, and is the author of several self-help books on chronic fatigue and has received royalties from these in the past. For other projects, KG has received grants from the National Institute for Health Research (NIHR), UK Research and Innovation (UKRI), the Juvenile Diabetes Foundation, the Stroke Association, and the National Institutes of Health (NIH). TC and KC are the authors of two studies included in this review (White et al., 2011, 2013) and TC is the author of a further included study (Deale et al., 1997). All other authors declare that they have no conflicts of interest.

**Ethical standards.** The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

### References

- Armijo-Olivo, S., Stiles, C. R., Hagen, N. A., Biondo, P. D., & Cummings, G. G. (2012). Assessment of study quality for systematic reviews: A comparison of the Cochrane collaboration risk of bias tool and the effective public health practice project quality assessment tool: Methodological research. *Journal of Evaluation in Clinical Practice*, 18(1), 12–18. <https://doi.org/10.1111/j.1365-2753.2010.01516.x>.
- Cairns, R., & Hotopf, M. (2005). A systematic review describing the prognosis of chronic fatigue syndrome. *Occupational Medicine*, 55(1), 20–31. <https://doi.org/10.1093/occmed/kqi013>.
- Castell, D. B., Kazantzis, N., & Moss-Morris, R. E. (2011). Cognitive behavioral therapy and graded exercise for chronic fatigue syndrome: A meta-analysis. *Clinical Psychology: Science and Practice*, 18(4), 311–324. <https://doi.org/10.1111/j.1468-2850.2011.01262.x>.
- Chalder, T., Berelowitz, G., Pawlikowska, T., Watts, L., Wessely, S., Wright, D., & Wallace, E. P. (1993). Development of a fatigue scale. *Journal of Psychosomatic Research*, 37(2), 147–153. [https://doi.org/10.1016/0022-3999\(93\)90081-p](https://doi.org/10.1016/0022-3999(93)90081-p).
- Clark, L. V., Pesola, F., Thomas, J. M., Vergara-Williamson, M., Beynon, M., & White, P. D. (2017). Guided graded exercise self-help plus specialist medical care versus specialist medical care alone for chronic fatigue syndrome (GETSET): A pragmatic randomised controlled trial. *Lancet (London, England)*, 390(10092), 363–373. [https://doi.org/10.1016/S0140-6736\(16\)32589-2](https://doi.org/10.1016/S0140-6736(16)32589-2).
- Collin, S. M., Nikolaus, S., Heron, J., Knoop, H., White, P. D., & Crawley, E. (2016). Chronic fatigue syndrome (CFS) symptom-based phenotypes in two clinical cohorts of adult patients in the UK and The Netherlands. *Journal of Psychosomatic Research*, 81, 14–23. <https://doi.org/10.1016/j.jpsychores.2015.12.006>.
- Deale, A., Chalder, T., Marks, I., & Wessely, S. (1997). Cognitive behavior therapy for chronic fatigue syndrome: A randomized controlled trial. *American Journal of Psychiatry*, 154(3), 408–414. <https://doi.org/10.1176/ajp.154.3.408>.
- Deale, A., Husain, K., Chalder, T., & Wessely, S. (2001). Long-term outcome of cognitive behavior therapy versus relaxation therapy for chronic fatigue

- syndrome: A 5-year follow-up study. *American Journal of Psychiatry*, 158(12), 2038–2042. <https://doi.org/10.1176/appi.ajp.158.12.2038>.
- Fukuda, K., Straus, S. E., Hickie, I., Sharpe, M. C., Dobbins, J. G., & Komaroff, A. (1994). The chronic fatigue syndrome: A comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. *Annals of Internal Medicine*, 121(12), 953–959. <https://doi.org/10.7326/0003-4819-121-12-199412150-00009>.
- Fulcher, K. Y., & White, P. D. (1997). Randomised controlled trial of graded exercise in patients with the chronic fatigue syndrome. *British Medical Journal*, 314(7095), 1647–1652. <https://doi.org/10.1136/bmj.314.7095.1647>.
- Guy, W. (1976). *ECDEU assessment manual for psychopharmacology*. Rockville, MD: National Institute of Mental Health.
- Janse, A., Worm-Smeitink, M., Bleijenberg, G., Donders, R., & Knoop, H. (2018). Efficacy of web-based cognitive-behavioural therapy for chronic fatigue syndrome: Randomised controlled trial. *British Journal of Psychiatry*, 212(2), 112–118. <https://doi.org/10.1192/bjp.2017.22>.
- Jason, L., Torres-Harding, S., Friedberg, F., Corradi, K., Njoku, M., Donalek, J., ... Papernik, M. (2007). Non-pharmacologic interventions for CFS: A randomized trial. *Journal of Clinical Psychology in Medical Settings*, 14(4), 275–296. <https://doi.org/10.1007/s10880-007-9090-7>.
- Karnofsky, D. A., Abelmann, W. H., Craver, L. F., & Burchenal, J. H. (1948). The use of the nitrogen mustards in the palliative treatment of carcinoma. *Cancer*, 4(1), 636–656. [https://doi.org/10.1002/1097-0142\(194811\)1:4<634::AID-CNCR2820010410>3.0.CO;2-L](https://doi.org/10.1002/1097-0142(194811)1:4<634::AID-CNCR2820010410>3.0.CO;2-L).
- Kent, P., Cancelliere, C., Boyle, E., Cassidy, J. D., & Kongsted, A. (2020). A conceptual framework for prognostic research. *BMC Medical Research Methodology*, 7, 1–13. <https://doi.org/https://doi.org/10.1186/s12874-020-01050-7>.
- King, E., Beynon, M., Chalder, T., Sharpe, M., & White, P. D. (2020). Patterns of daytime physical activity in patients with chronic fatigue syndrome. *Journal of Psychosomatic Research*, 135(October 2019), 110154. <https://doi.org/10.1016/j.jpsychores.2020.110154>.
- Knoop, H., van der Meer, J. W., & Bleijenberg, G. (2008). Guided self-instructions for people with chronic fatigue syndrome: Randomised controlled trial. *British Journal of Psychiatry*, 193(4), 340–341. <https://doi.org/10.1192/bjp.bp.108.051292>.
- Larun, L., Brurberg, K. G., Odgaard-Jensen, J., & Price, J. R. (2019). Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews* 2019(10), CD003200. <https://doi.org/10.1002/14651858.CD003200>.
- Lim, E. J., & Son, C. G. (2020). Review of case definitions for myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS). *Journal of Translational Medicine*, 18(1), 1–10. <https://doi.org/10.1186/s12967-020-02455-0>.
- Malouff, J. M., Thorsteinsson, E., Rooke, S. E., Bhullar, N., & Schutte, N. S. (2008). Efficacy of cognitive behavioral therapy for chronic fatigue syndrome: A meta-analysis. *Clinical Psychology Review*, 28(5), 736–745. <https://doi.org/10.1016/j.cpr.2007.10.004>.
- Marques, M. M., De Gucht, V., Gouveia, M. J., Leal, I., & Maes, S. (2015). Differential effects of behavioral interventions with a graded physical activity component in patients suffering from chronic fatigue (syndrome): An updated systematic review and meta-analysis. *Clinical Psychology Review*, 40, 123–137. <https://doi.org/10.1016/j.cpr.2015.05.009>.
- McCrone, P., Sharpe, M., Chalder, T., Knapp, M., Johnson, A., Goldsmith, K., & White, P. (2012). Adaptive pacing, cognitive behaviour therapy, graded exercise, and specialist medical care for chronic fatigue syndrome: A cost-effectiveness analysis. *PLoS ONE*, 7(8), 1–9. <https://doi.org/10.1371/journal.pone.0040808>.
- Minozzi, S., Cinquini, M., Gianola, S., Gonzalez-Lorenzo, M., & Banzi, R. (2020). The revised Cochrane risk of bias tool for randomized trials (RoB 2) showed low interrater reliability and challenges in its application. *Journal of Clinical Epidemiology*, 126, 37–44. <https://doi.org/10.1016/j.jclinepi.2020.06.015>.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement (The PRISMA Group). *PLoS Medicine*, 6(7): e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
- Moons, K. G., Hooft, L., Williams, K., Hayden, J. A., Damen, J. A., & Riley, R. D. (2018). Implementing systematic reviews of prognosis studies in Cochrane. *Cochrane Database of Systematic Reviews*, 10, ED000129. <https://doi.org/10.1002/14651858.ED000129>.
- Moss-Morris, R., Sharon, C., Tobin, R., & Baldi, J. C. (2005). A randomized controlled graded exercise trial for chronic fatigue syndrome: Outcomes and mechanisms of change. *Journal of Health Psychology*, 10(2), 245–259. <https://doi.org/10.1177/1359105305049774>.
- Moustgaard, H., Clayton, G. L., Jones, H. E., Boutron, I., Jørgensen, L., Laursen, D. L. T., ... Hróbjartsson, A. (2020). Impact of blinding on estimated treatment effects in randomised clinical trials: Meta-epidemiological study. *BMJ*, 368, m358. <https://doi.org/10.1136/bmj.l6802>.
- National Institute for Health and Care Excellence. (2007). *Chronic fatigue syndrome/myalgic encephalomyelitis (or encephalopathy): Diagnosis and management of CFS/ME in adults and children*. (NICE Guideline No. 53). Retrieved from <https://www.nice.org.uk/guidance/CG53>.
- Newton, J. L., Mabillard, H., Scott, A., Hoad, A., & Spickett, G. (2010). The Newcastle NHS chronic fatigue syndrome service: Not all fatigue is the same. *Journal of the Royal College of Physicians of Edinburgh*, 40(4), 303–307. <https://doi.org/10.4997/JRCPE.2010.404>.
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan – A web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 1–10. <https://doi.org/10.1186/s13643-016-0384-4>.
- Powell, P., Bentall, R. P., Nye, F. J., & Edwards, K. H. T. (2001). Randomised controlled trial of patient education to encourage graded exercise in chronic fatigue syndrome. *British Medical Journal*, 322(7283), 387–390. <https://doi.org/10.1136/bmj.322.7283.387>.
- Price, J. R., Mitchell, E., Tidy, E., & Hunot, V. (2008). Cognitive behaviour therapy for chronic fatigue syndrome in adults. *Cochrane Database of Systematic Reviews*, 2008(3), CD001027. <https://doi.org/10.1002/14651858.CD001027.pub2>.
- Schluederberg, A., Straus, S. E., Peterson, P., Blumenthal, S., Komaroff, A., Spring, S. B., ... Buchwald, D. (1992). Chronic fatigue syndrome research definition and medical outcome assessment. *Annals of Internal Medicine*, 117(4), 325–331. <https://doi.org/10.7326/0003-4819-117-4-325>.
- Sharpe, M., Archard, L., Banatvala, J., Borysiewicz, L., Clare, A., David, A., ... White, P. D. (1991). A report-chronic fatigue syndrome: Guidelines for research. *Journal of the Royal Society of Medicine*, 84(2), 118–121. <https://doi.org/10.1177/014107689108400224>.
- Sharpe, M., Hawton, K., Simkin, S., Surawy, C., Hackmann, A., Klimes, I., ... Seagroatt, V. (1996). Cognitive behaviour therapy for the chronic fatigue syndrome: A randomised controlled trial. *British Medical Journal*, 312(7022), 22–26. <https://doi.org/10.1136/bmj.312.7022.22>.
- Thomas, B. H., Ciliska, D., Dobbins, M., & Micucci, S. (2004). A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. *Worldviews on Evidence-Based Nursing*, 1(3), 176–184. <https://doi.org/10.1111/j.1524-475X.2004.04006.x>.
- Tummers, M., Knoop, H., & Bleijenberg, G. (2010). Effectiveness of stepped care for chronic fatigue syndrome: A randomized noninferiority trial. *Journal of Consulting and Clinical Psychology*, 78(5), 724–731. <https://doi.org/10.1037/a0020052>.
- Tummers, M., Knoop, H., van Dam, A., & Bleijenberg, G. (2013). Moderators of the treatment response to guided self-instruction for chronic fatigue syndrome. *Journal of Psychosomatic Research*, 74(5), 373–377. <https://doi.org/10.1016/j.jpsychores.2013.01.007>.
- Vercoulen, J. H. M. M., Swanink, C. M. A., Fennis, J. F. M., Galama, J. M. D., van der Meer, J. W. M., & Bleijenberg, G. (1994). Dimensional assessment of chronic fatigue syndrome. *Journal of Psychosomatic Research*, 38(5), 383–392. [https://doi.org/10.1016/0022-3999\(94\)90099-X](https://doi.org/10.1016/0022-3999(94)90099-X).
- Vos-Vromans, D. C., Smeets, R. J., Huijnen, I. P. J., Koke, A. J. A., Hitters, W. M., Rijnders, L. J. M., ... Knottnerus, J. A. (2016). Multidisciplinary rehabilitation treatment versus cognitive behavioural therapy for patients with chronic fatigue syndrome: A randomized controlled trial. *Journal of Internal Medicine*, 279(3), 268–282. <https://doi.org/10.1111/joim.12402>.
- Ware, J. E. J., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care*, 30(6), 473–483. <https://doi.org/10.1097/00005650-199206000-00002>.
- White, P. D., & Etherington, J. (2021). Adverse outcomes in trials of graded exercise therapy for adult patients with chronic fatigue syndrome. *Journal of Psychosomatic Research*, 147(May), 110533. <https://doi.org/10.1016/j.jpsychores.2021.110533>.

- White, P. D., Goldsmith, K., Johnson, A. L., Chalder, T., & Sharpe, M. (2013). Recovery from chronic fatigue syndrome after treatments given in the PACE trial. *Psychological Medicine*, 43(10), 2227–2235. <https://doi.org/10.1017/S0033291713000020>.
- White, P. D., Goldsmith, K. A., Johnson, A. L., Potts, L., Walwyn, R., DeCesare, J. C., ... Sharpe, M. (2011). Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): A randomised trial. *Lancet (London, England)*, 377(9768), 823–836. [https://doi.org/10.1016/S0140-6736\(11\)60096-2](https://doi.org/10.1016/S0140-6736(11)60096-2).
- Wiborg, J. F., van Bussel, J., van Dijk, A., Bleijenberg, G., & Knoop, H. (2015). Randomised controlled trial of cognitive behaviour therapy delivered in groups of patients with chronic fatigue syndrome. *Psychotherapy and Psychosomatics*, 84(6), 368–376. <https://doi.org/10.1159/000438867>.
- Worm-Smeitink, M., Janse, A., van Dam, A., Evers, A., van der Vaart, R., Wensing, M., & Knoop, H. (2019). Internet-based cognitive behavioral therapy in stepped care for chronic fatigue syndrome: Randomized noninferiority trial. *Journal of Medical Internet Research*, 21(3), 1–15. <https://doi.org/10.2196/11276>.