Cochrane Library

Cochrane Database of Systematic Reviews

Exercise therapy for chronic fatigue syndrome (Review)

Larun L, Brurberg KG, Odgaard-Jensen J, Price JR

Larun L, Brurberg KG, Odgaard-Jensen J, Price JR.
Exercise therapy for chronic fatigue syndrome.
Cochrane Database of Systematic Reviews 2017, Issue 4. Art. No.: CD003200.
DOI: 10.1002/14651858.CD003200.pub7.

www.cochranelibrary.com

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

WILEY

TABLE OF CONTENTS

HEADER . . . 1
ABSTRACT . . . 1
PLAIN LANGUAGE SUMMARY . . . 2
SUMMARY OF FINDINGS FOR THE MAIN COMPARISON . . . 4
BACKGROUND . . . 7
OBJECTIVES . . . 8
METHODS . . . 8
RESULTS . . . 11
Figure 1. . . . 12
Figure 2. . . . 15
Figure 3. . . . 16
DISCUSSION . . . 27
AUTHORS' CONCLUSIONS . . . 29
ACKNOWLEDGEMENTS . . . 29
REFERENCES . . . 29
CHARACTERISTICS OF STUDIES . . . 34
DATA AND ANALYSES . . . 58
Analysis 1.1. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 1 Fatigue (end of treatment). . . . 64
Analysis 1.2. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 2 Fatigue (follow-up). . . . 65
Analysis 1.3. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 3 Participants with serious adverse reactions. . . . 66
Analysis 1.4. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 4 Pain (follow-up). . . . 67
Analysis 1.5. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 5 Physical functioning (end of treatment). . . . 68
Analysis 1.6. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 6 Physical functioning (follow-up). . . . 69
Analysis 1.7. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 7 Quality of life (follow-up). . . . 69
Analysis 1.8. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 8 Depression (end of treatment). . . . 70
Analysis 1.9. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 9 Depression (follow-up). . . . 71
Analysis 1.10. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 10 Anxiety (end of treatment). . . . 72
Analysis 1.11. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 11 Anxiety (follow-up). . . . 73
Analysis 1.12. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 12 Sleep (end of treatment). . . . 74
Analysis 1.13. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 13 Sleep (follow-up). . . . 75
Analysis 1.14. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 14 Self-perceived changes in overall health (end of treatment). . . . 76
Analysis 1.15. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 15 Self-perceived changes in overall health (follow-up). . . . 77
Analysis 1.16. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 16 Health resource use (follow-up) [Mean no. of contacts]. . . . 78
Analysis 1.17. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 17 Health resource use (follow-up) [No. of users]. . . . 79

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.18. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 18 Drop-out. 80
Analysis 1.19. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 19 Subgroup analysis for fatigue. 81
Analysis 2.1. Comparison 2 Exercise therapy versus psychological treatment, Outcome 1 Fatigue at end of treatment (FS; 11 items/0 to 33 points). 82
Analysis 2.2. Comparison 2 Exercise therapy versus psychological treatment, Outcome 2 Fatigue at follow-up (FSS; 1 to 7 points). 82
Analysis 2.3. Comparison 2 Exercise therapy versus psychological treatment, Outcome 3 Fatigue at follow-up (FS; 11 items/0 to 33 points). 83
Analysis 2.4. Comparison 2 Exercise therapy versus psychological treatment, Outcome 4 Participants with serious adverse reactions. 83
Analysis 2.5. Comparison 2 Exercise therapy versus psychological treatment, Outcome 5 Pain at follow-up (BPI, pain severity subscale; 0 to 10 points). 84
Analysis 2.6. Comparison 2 Exercise therapy versus psychological treatment, Outcome 6 Pain at follow-up (BPI, pain interference subscale; 0 to 10 points). 84
Analysis 2.7. Comparison 2 Exercise therapy versus psychological treatment, Outcome 7 Physical functioning at end of treatment (SF-36, physical functioning subscale; 0 to 100 points). 85
Analysis 2.8. Comparison 2 Exercise therapy versus psychological treatment, Outcome 8 Physical functioning at follow-up (SF-36, physical functioning subscale; 0 to 100 points). 86
Analysis 2.9. Comparison 2 Exercise therapy versus psychological treatment, Outcome 9 Depression at end of treatment (HADS depression score; 7 items/21 points). 87
Analysis 2.10. Comparison 2 Exercise therapy versus psychological treatment, Outcome 10 Depression at follow-up (BDI; 0 to 63 points). 87
Analysis 2.11. Comparison 2 Exercise therapy versus psychological treatment, Outcome 11 Depression at follow-up (HADS depression score; 7 items/21 points). 88
Analysis 2.12. Comparison 2 Exercise therapy versus psychological treatment, Outcome 12 Anxiety at end of treatment (HADS anxiety; 7 items/21 points). 88
Analysis 2.13. Comparison 2 Exercise therapy versus psychological treatment, Outcome 13 Anxiety at follow-up (BAI; 0 to 63 points). 89
Analysis 2.14. Comparison 2 Exercise therapy versus psychological treatment, Outcome 14 Anxiety at follow-up (HADS anxiety; 7 items/21 points). 89
Analysis 2.15. Comparison 2 Exercise therapy versus psychological treatment, Outcome 15 Sleep at end of treatment (Jenkins Sleep Scale; 0 to 20 points). 90
Analysis 2.16. Comparison 2 Exercise therapy versus psychological treatment, Outcome 16 Sleep at follow-up (Jenkins Sleep Scale; 0 to 20 points). 90
Analysis 2.17. Comparison 2 Exercise therapy versus psychological treatment, Outcome 17 Self-perceived changes in overall health at end of treatment. 91
Analysis 2.18. Comparison 2 Exercise therapy versus psychological treatment, Outcome 18 Self-perceived changes in overall health at follow-up. 92
Analysis 2.19. Comparison 2 Exercise therapy versus psychological treatment, Outcome 19 Health resource use (follow-up) [Mean no. of contacts]. 93
Analysis 2.20. Comparison 2 Exercise therapy versus psychological treatment, Outcome 20 Health resource use (follow-up) [No. of users]. 94
Analysis 2.21. Comparison 2 Exercise therapy versus psychological treatment, Outcome 21 Drop-out. 95
Analysis 3.1. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 1 Fatigue. 95
Analysis 3.2. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 2 Participants with serious adverse reactions. 96
Analysis 3.3. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 3 Physical functioning. 96
Analysis 3.4. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 4 Depression. 97
Analysis 3.5. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 5 Anxiety. 97
Analysis 3.6. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 6 Sleep. 98
Analysis 3.7. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 7 Self-perceived changes in overall health. 98

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 3.8. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 8 Health resource use (follow-up) [Mean no. of contacts]. 99
Analysis 3.9. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 9 Health resource use (follow-up) [No. of users]. 100
Analysis 3.10. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 10 Drop-out. 101
Analysis 4.1. Comparison 4 Exercise therapy + antidepressant placebo versus antidepressant + exercise placebo, Outcome 1 Fatigue. 101
Analysis 4.2. Comparison 4 Exercise therapy + antidepressant placebo versus antidepressant + exercise placebo, Outcome 2 Depression. 102
Analysis 4.3. Comparison 4 Exercise therapy + antidepressant placebo versus antidepressant + exercise placebo, Outcome 3 Drop-out. 102
Analysis 5.1. Comparison 5 Exercise therapy + antidepressant versus antidepressant + exercise placebo, Outcome 1 Fatigue. 103
Analysis 5.2. Comparison 5 Exercise therapy + antidepressant versus antidepressant + exercise placebo, Outcome 2 Depression. 103
Analysis 5.3. Comparison 5 Exercise therapy + antidepressant versus antidepressant + exercise placebo, Outcome 3 Drop-out. 104
ADDITIONAL TABLES 104
APPENDICES 107
FEEDBACK 111
WHAT'S NEW 133
HISTORY 134
CONTRIBUTIONS OF AUTHORS 134
DECLARATIONS OF INTEREST 134
SOURCES OF SUPPORT 134
DIFFERENCES BETWEEN PROTOCOL AND REVIEW 135
NOTES 136
INDEX TERMS 137

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

[Intervention Review]

Exercise therapy for chronic fatigue syndrome

Lillebeth Larun[1], Kjetil G. Brurberg[2], Jan Odgaard-Jensen[3], Jonathan R Price[4]

[1]Norwegian Insitute of Public Health, Oslo, Norway. [2]Unit for Primary Care and Clinical Procedures, Norwegian Institute of Public Health, Oslo, Norway. [3]Biometrics, Link Medical Research AS, Oslo, Norway. [4]Department of Psychiatry, University of Oxford, Oxford, UK

Contact address: Lillebeth Larun, Norwegian Insitute of Public Health, Postboks 4404 Nydalen, Oslo, N-0403, Norway. Lillebeth.Larun@fhi.no.

Editorial group: Cochrane Common Mental Disorders Group.
Publication status and date: New search for studies and content updated (no change to conclusions), comment added to review, published in Issue 4, 2017.

Citation: Larun L, Brurberg KG, Odgaard-Jensen J, Price JR. Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews* 2017, Issue 4. Art. No.: CD003200. DOI: 10.1002/14651858.CD003200.pub7.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

ABSTRACT

Background

Chronic fatigue syndrome (CFS) is characterised by persistent, medically unexplained fatigue, as well as symptoms such as musculoskeletal pain, sleep disturbance, headaches and impaired concentration and short-term memory. CFS presents as a common, debilitating and serious health problem. Treatment may include physical interventions, such as exercise therapy, which was last reviewed in 2004.

Objectives

The objective of this review was to determine the effects of exercise therapy (ET) for patients with CFS as compared with any other intervention or control.

• Exercise therapy versus 'passive control' (e.g. treatment as usual, waiting-list control, relaxation, flexibility).

• Exercise therapy versus other active treatment (e.g. cognitive-behavioural therapy (CBT), cognitive treatment, supportive therapy, pacing, pharmacological therapy such as antidepressants).

• Exercise therapy in combination with other specified treatment strategies versus other specified treatment strategies (e.g. exercise combined with pharmacological treatment vs pharmacological treatment alone).

Search methods

We searched The Cochrane Collaboration Depression, Anxiety and Neurosis Controlled Trials Register (CCDANCTR), the Cochrane Central Register of Controlled Trials (CENTRAL) and SPORTDiscus up to May 2014 using a comprehensive list of free-text terms for CFS and exercise. We located unpublished or ongoing trials through the World Health Organization (WHO) International Clinical Trials Registry Platform (to May 2014). We screened reference lists of retrieved articles and contacted experts in the field for additional studies

Selection criteria

Randomised controlled trials involving adults with a primary diagnosis of CFS who were able to participate in exercise therapy. Studies had to compare exercise therapy with passive control, psychological therapies, adaptive pacing therapy or pharmacological therapy.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Data collection and analysis

Two review authors independently performed study selection, risk of bias assessments and data extraction. We combined continuous measures of outcomes using mean differences (MDs) and standardised mean differences (SMDs). We combined serious adverse reactions and drop-outs using risk ratios (RRs). We calculated an overall effect size with 95% confidence intervals (CIs) for each outcome.

Main results

We have included eight randomised controlled studies and have reported data from 1518 participants in this review. Three studies diagnosed individuals with CFS using the 1994 criteria of the Centers for Disease Control and Prevention (CDC); five used the Oxford criteria. Exercise therapy lasted from 12 to 26 weeks. Seven studies used variations of aerobic exercise therapy such as walking, swimming, cycling or dancing provided at mixed levels in terms of intensity of the aerobic exercise from very low to quite rigorous, whilst one study used anaerobic exercise. Control groups consisted of passive control (eight studies; e.g. treatment as usual, relaxation, flexibility) or CBT (two studies), cognitive therapy (one study), supportive listening (one study), pacing (one study), pharmacological treatment (one study) and combination treatment (one study). Risk of bias varied across studies, but within each study, little variation was found in the risk of bias across our primary and secondary outcome measures.

Investigators compared exercise therapy with 'passive' control in eight trials, which enrolled 971 participants. Seven studies consistently showed a reduction in fatigue following exercise therapy at end of treatment, even though the fatigue scales used different scoring systems: an 11-item scale with a scoring system of 0 to 11 points (MD -6.06, 95% CI -6.95 to -5.17; one study, 148 participants; low-quality evidence); the same 11-item scale with a scoring system of 0 to 33 points (MD -2.82, 95% CI -4.07 to -1.57; three studies, 540 participants; moderate-quality evidence); and a 14-item scale with a scoring system of 0 to 42 points (MD -6.80, 95% CI -10.31 to -3.28; three studies, 152 participants; moderate-quality evidence). Serious adverse reactions were rare in both groups (RR 0.99, 95% CI 0.14 to 6.97; one study, 319 participants; moderate-quality evidence), but sparse data made it impossible for review authors to draw conclusions. Study authors reported a positive effect of exercise therapy at end of treatment with respect to sleep (MD -1.49, 95% CI -2.95 to -0.02; two studies, 323 participants), physical functioning (MD 13.10, 95% CI 1.98 to 24.22; five studies, 725 participants) and self-perceived changes in overall health (RR 1.83, 95% CI 1.39 to 2.40; four studies, 489 participants). It was not possible for review authors to draw conclusions regarding the remaining outcomes.

Investigators compared exercise therapy with CBT in two trials (351 participants). One trial (298 participants) reported little or no difference in fatigue at end of treatment between the two groups using an 11-item scale with a scoring system of 0 to 33 points (MD 0.20, 95% CI -1.49 to 1.89). Both studies measured differences in fatigue at follow-up, but neither found differences between the two groups using an 11-item fatigue scale with a scoring system of 0 to 33 points (MD 0.30, 95% CI -1.45 to 2.05) and a nine-item Fatigue Severity Scale with a scoring system of 1 to 7 points (MD 0.40, 95% CI -0.34 to 1.14). Serious adverse reactions were rare in both groups (RR 0.67, 95% CI 0.11 to 3.96). We observed little or no difference in physical functioning, depression, anxiety and sleep, and we were not able to draw any conclusions with regard to pain, self-perceived changes in overall health, use of health service resources and drop-out rate.

With regard to other comparisons, one study (320 participants) suggested a general benefit of exercise over adaptive pacing, and another study (183 participants) a benefit of exercise over supportive listening. The available evidence was too sparse to draw conclusions about the effect of pharmaceutical interventions.

Authors' conclusions

Patients with CFS may generally benefit and feel less fatigued following exercise therapy, and no evidence suggests that exercise therapy may worsen outcomes. A positive effect with respect to sleep, physical function and self-perceived general health has been observed, but no conclusions for the outcomes of pain, quality of life, anxiety, depression, drop-out rate and health service resources were possible. The effectiveness of exercise therapy seems greater than that of pacing but similar to that of CBT. Randomised trials with low risk of bias are needed to investigate the type, duration and intensity of the most beneficial exercise intervention.

PLAIN LANGUAGE SUMMARY

Exercise as treatment for patients with chronic fatigue syndrome

Who may be interested in this review?

• People with chronic fatigue syndrome and their family and friends.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

• Professionals working in specialist chronic fatigue services.

• Professionals working in therapeutic exercise.

• General practitioners.

Why is this review important?

Chronic fatigue syndrome (CFS) is sometimes called myalgic encephalomyelitis (ME). Research estimates that between 2 in 1000 and 2 in 100 adults in the USA are affected by CFS. People with CFS often have long-lasting fatigue, joint pain, headaches, sleep problems, and poor concentration and short-term memory. These symptoms cause significant disability and distress for people affected by CFS. There is no clear medical cause for CFS, so people who are affected often deal with misunderstanding of their condition from family, friends and healthcare professionals. National Institute for Health and Care Excellence (NICE) guidelines recommend exercise therapy for individuals with CFS, and a previous review of the evidence suggested that exercise therapy was a promising approach to the treatment. It is thought that exercise therapy can help management of CFS symptoms by helping people gradually reintroduce physical activity into their daily lives.

This review is an update of a previous Cochrane review from 2004, which showed that exercise therapy was a promising treatment for adults with CFS. Since the review, additional studies investigating the effectiveness and safety of exercise therapy for patients with CFS have been published.

What questions does this review aim to answer?

• Is exercise therapy more effective than 'passive' treatments (e.g. waiting list, treatment as usual, relaxation, flexibility)?

• Is exercise therapy more effective than other 'active' therapies (e.g. cognitive-behavioural therapy (CBT), pacing, medication)?

• Is exercise therapy more effective when combined with another treatment than when given alone?

• Is exercise therapy safer than other treatments?

Which studies were included in the review?

We searched databases to find all high-quality studies of exercise therapy for CFS published up to May 2014. To be included in the review, studies had to be randomised controlled trials and include adults over 18 years of age, more than 90% of whom had a clear diagnosis of CFS. We included eight studies with a total of 1518 participants in the review. Seven studies used aerobic exercise therapy such as walking, swimming, cycling or dancing; the remaining study used non-aerobic exercise. Most studies asked participants to exercise at home, between three and five times per week, with a target duration of 5 to 15 minutes per session using different means of incrementation.

What does evidence from the review tell us?

Moderate-quality evidence showed exercise therapy was more effective at reducing fatigue compared to 'passive' treatment or no treatment. Exercise therapy had a positive effect on people's daily physical functioning, sleep and self-ratings of overall health.

One study suggests that exercise therapy was more effective than pacing strategies for reducing fatigue. However exercise therapy was no more effective than CBT.

Exercise therapy did not worsen symptoms for people with CFS. Serious side effects were rare in all groups, but limited information makes it difficult to draw firm conclusions about the safety of exercise therapy.

Evidence was not sufficient to show effects of exercise therapy on pain, use of other healthcare services, or to allow assessment of rates of drop-out from exercise therapy programmes.

What should happen next?

Researchers suggest that further studies should be carried out to discover what type of exercise is most beneficial for people affected by CFS, which intensity is best, the optimal length, as well as the most beneficial delivery method.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

SUMMARY OF FINDINGS FOR THE MAIN COMPARISON [Explanation]

Exercise therapy for chronic fatigue syndrome

Patient or population: males and females over 18 years of age with chronic fatigue syndrome
Intervention: exercise therapy
Comparison: standard care, waiting list or relaxation/flexibility

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Quality of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Control	Exercise				
Fatigue[a]: FS, Fatigue Scale (0 to 11 points) (end of treatment)	Mean fatigue in the control groups was 10.4 points	Mean fatigue in the intervention groups was **6.06 points lower** (6.95 to 5.17 lower)		148 (1 study)	⊕⊕◯◯ **Low**[b,c]	Lower score indicates less fatigue
Fatigue[a]: FS, Fatigue Scale (0 to 33 points) (end of treatment)	Mean fatigue ranged across control groups from 15.3 to 26.3 points	Mean fatigue in the intervention groups was **2.82 points lower** (4.07 to 1.57 lower)		540 (3 studies)	⊕⊕⊕◯ **Moderate**[b]	Lower score indicates less fatigue
Fatigue[a]: FS, Fatigue Scale (0 to 42 points) (end of treatment)	Mean fatigue ranged across control groups from 24.4 to 31.6 points	Mean fatigue in the intervention groups was **6.80 points lower** (10.31 to 3.28 lower)		152 (3 studies)	⊕⊕⊕◯ **Moderate**[b]	Lower score indicates less fatigue
Participants with serious adverse reactions	**Study population**		**RR 0.99** (0.14 to 6.97)	319 (1 study)	⊕⊕⊕◯ **Moderate**[d,e]	
	13 per 1000	**12 per 1000** (2 to 87)				

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Quality of Life (QOL) Scale (16 to 112 points) (follow-up)	Mean QOL score in the control group was 72 points	Mean QOL score in the intervention groups was **9.00 points lower** (19.00 lower to 1.00 higher)		44 (1 study)	⊕○○○ **Very low**[b,f]	Higher score indicates improved QOL
Physical functioning: SF-36 subscale (0 to 100 points) (end of treatment)	Mean physical functioning score ranged from 31.1 to 55.2 points across control groups	Mean physical functioning score in the intervention groups was **13.10 points higher** (1.98 to 24.22 higher)		725 (5 studies)	⊕⊕○○ **Low**[b,g]	Higher score indicates improved physical function
Depression: HADS depression score (0 to 21 points) (end of treatment)	Mean depression score ranged across control groups from 5.2 to 11.2 points	Mean depression score in the intervention groups was **1.63 points lower** (3.50 lower to 0.23 higher)		504 (5 studies)	⊕○○○ **Very low**[b,g,h]	Lower score indicates fewer depressive symptoms
Sleep: Jenkins Sleep Scale (0 to 20 points) (end of treatment)	Mean sleep score ranged across control groups from 11.7 to 12.2 points	Mean sleep score in the intervention groups was **1.49 points lower** (2.95 to 0.02 lower)		323 (2 studies)	⊕⊕○○ **Low**[b,h]	Lower score indicates improved sleep quality
Self-perceived changes in overall health (end of treatment)	**Study population**		**RR 1.83** (1.39 to 2.40)	489 (4 studies)	⊕⊕⊕○ **Moderate**[b]	RR higher than 1 means that more participants in exercise groups reported improvement
	218 per 1000	**399 per 1000** (303 to 523)				
	Medium-risk population					
	238 per 1000	**436 per 1000** (331 to 571)				

Drop-out (end of treatment)	**Study population**		**RR 1.63** (0.77 to 3.43)	843 (6 studies)	⊕⊕◯◯ **Low**[b,g]	RR higher than 1 means that more participants in exercise groups dropped out from treatment
	70 per 1000	**114 per 1000** (54 to 241)				
	Medium-risk population					
	89 per 1000	**145 per 1000** (69 to 305)				

*The basis for the **assumed risk** (e.g. median control group risk across studies) is provided in footnotes. The **corresponding risk** (and its 95% confidence interval) is based on the assumed risk in the comparison group and the **relative effect** of the intervention (and its 95% CI).
CI: Confidence interval; **RR:** Risk ratio.

GRADE Working Group grades of evidence.
High quality: Further research is very unlikely to change our confidence in the estimate of effect.
Moderate quality: Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate.
Low quality: Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate.
Very low quality: We are very uncertain about the estimate.

[a]We choose to present effect estimates as measured on the original scales rather than to transform them to standardised units. As 3 different scoring systems for fatigue were used, the outcome is presented over 3 rows.
[b]Risk of bias (-1): All studies were at risk of performance bias, as they were unblinded.
[c]Inconsistency (-1): shows inconsistencies with other available trials when meta-analysis based on standardised mean differences is performed. Subgroup analyses could not explain variation due to diagnostic criteria, treatment strategy or type of control.
[d]Risk of bias (0): This outcome is unlikely to have been affected by detection or performance bias.
[e]Imprecision (-1): low numbers of events and wide confidence intervals.
[f]Imprecision (-2): very low numbers of participants and wide confidence intervals, which encompass benefit and harm.
[g]Inconsistency (-1): variation in effect size and direction of effect across available studies.
[h]Imprecision (-1): Confidence interval fails to exclude negligible differences in favour of the intervention.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

BACKGROUND

Description of the condition

Chronic fatigue syndrome (CFS) is an illness characterised by persistent, medically unexplained fatigue. Symptoms include severe, disabling fatigue, as well as musculoskeletal pain, sleep disturbance, headaches, and impaired concentration and short-term memory (Prins 2006). Individuals experience significant disability and distress, which may be exacerbated by lack of understanding from others, including healthcare professionals. The term 'myalgic encephalomyelitis (ME)' is often used, but 'CFS' is the term that has been adopted and clearly defined for research purposes, and it will be used in this review. The diagnosis can be made only after all alternative diagnoses have been excluded (Reeves 2003; Reeves 2007); several sets of criteria are currently used to diagnose CFS (Carruthers 2011; Fukuda 1994; NICE 2007; Reeves 2003; Sharpe 1991). The Centers for Disease Control and Prevention (CDC) 1994 diagnostic criteria for CFS (Fukuda 1994) are the most widely cited for research purposes (Fonhus 2011), resulting in prevalence of CFS of between 0.24% (Reyes 2003) and 2.55% (Reeves 2007) among US adults. Practical application of diagnostic criteria may help to explain some of the observed variation in prevalence estimates (Johnston 2013). In practice, most patients visit their local general practitioner (GP) for assessment. A minority of patients may be referred to specialist clinics (e.g. neurology, infectious diseases, psychiatry, endocrinology or general medicine) for exclusion of alternative underlying disorders.

Description of the intervention

Exercise therapy is often included as part of a treatment programme for individuals with CFS. 'Exercise' is defined as "planned structured and repetitive bodily movement done to improve or maintain one or more components of physical fitness" (ACSM 2001); 'therapy' is defined as "treatment intended to relieve or heal a disorder" (Oxford English Dictionary). We define 'exercise therapy' as a "regimen or plan of physical activity designed and prescribed [and] intended to relieve or heal a disorder," and 'therapeutic exercise' or 'exercise therapy' can be described as "planned exercise performed to attain a specific physical benefit, such as maintenance of the range of motion, strengthening of weakened muscles, increased joint flexibility, or improved cardiovascular and respiratory function" (Mosby 2009). Aerobic exercise such as walking, jogging, swimming or cycling is included, along with anaerobic exercise such as strength or stabilising exercises. Graded exercise therapy is characterised by establishment of a baseline of achievable exercise or physical activity, followed by a negotiated, incremental increase in the duration of time spent physically active followed by an increase in intensity (White 2011).

How the intervention might work

Physical activity can improve health and quality of life for patients with chronic disease (Blair 2009). The causal pathway for CFS is unknown; however several hypotheses have been proposed as to why exercise therapy might be a viable treatment. The 'deconditioning model' assumes that the syndrome is perpetuated by reversible physiological changes of deconditioning and avoidance of activity; therefore exercise should improve deconditioning and thus the condition of patients with CFS (Clark 2005; White 2011). However, mediation studies suggest that improved conditioning is not associated with better outcomes (Fulcher 1997; Moss-Morris 2005). Some graded exercise therapy (GET) programmes are designed to gradually reintroduce the patient to the avoided stimulus of physical activity or exercise, which may involve a conditioned response leading to fatigue (Clark 2005; Fulcher 2000; White 2011). Mediation studies suggest that reduced symptom focus may mediate outcomes with GET, consistent with this model (Clark 2005; Moss-Morris 2005). Evidence has also been found for central sensitisation contributing to hyperresponsiveness of the central nervous system to a variety of visceral inputs (Nijs 2011). The most replicated finding in patients with CFS is an increased sense of effort during exercise, which is consistent with this model (Fulcher 2000; Paul 2001). Graded exercise therapy may reduce this extra sense of effort, perhaps by reducing central sensitisation (Fulcher 1997).

Further research is needed to verify these hypotheses, but effective treatments may be discovered without knowledge of the effective pathway or underlying cause.

Why it is important to do this review

The previous Cochrane review (Edmonds 2004) suggested that exercise therapy was a promising treatment but that larger studies were needed to address the safety of this therapy (Edmonds 2004). Such studies have been completed and their findings published, so that the present time is propitious for an updated review. Exercise therapy is often used as treatment for individuals with CFS and is recommended by treatment guidelines (NICE 2007). People with CFS should have the opportunity to make informed decisions about their care and treatment based on robust research evidence. This review will examine the effectiveness of exercise therapy, provided as a stand-alone intervention or as part of a treatment plan. The Cochrane Collaboration has reviewed multiple aspects of treatment for patients with CFS. A review on CBT was published in 2008 (Price 2008), and one on traditional Chinese herbal medicine in 2009 (Adams 2009); also, a protocol on pharmacological treatments was submitted (Hard 2009).

This review, which is an update of a Cochrane review first published in 2004, will update the evidence base that serves as a resource for informed decision making by healthcare personnel and patients. A protocol for an accompanying individual patient data

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

review on chronic fatigue syndrome and exercise therapy has been published (Larun 2014).

OBJECTIVES

The objective of this review was to determine the effects of exercise therapy (ET) for patients with chronic fatigue syndrome (CFS) as compared with any other intervention or control.

- Exercise therapy versus 'passive control' (e.g. treatment as usual, waiting-list control, relaxation, flexibility).

- Exercise therapy versus other active treatment (e.g. cognitive-behavioural therapy (CBT), cognitive treatment, supportive therapy, pacing, pharmacological therapy such as antidepressants).

- Exercise therapy in combination with other specified treatment strategies versus other specified treatment strategies (e.g. exercise combined with pharmacological treatment vs pharmacological treatment alone).

METHODS

Criteria for considering studies for this review

Types of studies

We included randomised controlled trials, as well as cluster-randomised trials and cross-over trials.

Types of participants

We included trials of male and female participants over the age of 18, irrespective of cultures and settings. Investigators currrently have used several sets of criteria to diagnose CFS (Carruthers 2011; Fukuda 1994; NICE 2007; Reeves 2003; Sharpe 1991); therefore we decided to include trials in which participants fulfilled the following diagnostic criteria for CFS or ME.

- Fatigue, or a symptom synonymous with fatigue, was a prominent symptom.
- Fatigue was medically unexplained (i.e. other diagnoses known to cause fatigue such as anorexia nervosa or sleep apnoea could be excluded).
- Fatigue was sufficiently severe to significantly disable or distress the participant.
- Fatigue persisted for at least six months.

We included trials that included participants with disorders other than CFS provided that > 90% of participants had been given a primary diagnosis of CFS based on the criteria discussed above. We included in the analysis of this review trials in which less than 90% of participants had a primary diagnosis of CFS only if data on CFS were reported separately.

Co-morbidity

Studies involving participants with co-morbid physical or common mental disorders were eligible for inclusion only if the co-morbidity did not provide an alternative explanation for fatigue.

Types of interventions

Experimental intervention

Both aerobic and anaerobic interventions aimed at exercising big muscle groups, for example, walking, swimming, jogging and strength or stabilising exercises, could be included. Both individual and group treatment modalities were eligible, but interventions had to be clearly described and supported by appropriate references.

'Exercise therapy' is an umbrella term for the different types of exercise provided; it is based on the American College of Sports Medicine definition (ACSM 2001). We categorised exercise therapies in this review in accordance with descriptions of the interventions provided by individual studies. We prepared a table of Interventions with detailed information on exercise therapy reported by the included studies, as definitions vary across time and context. As a point of reference, we used the following empirical definitions, as derived from descriptions of the interventions.

- Graded exercise therapy (GET): exercise in which the incremental increase in exercise was mutually set.
- Exercise with pacing: exercise in which the incremental increase in exercise was personally set.
- Anaerobic exercise: exercise that requires a high level of exertion, in a brief spurt or short-term in duration by the participant that can be gradually increased over time with practice

We did not impose restrictions with regard to the duration of each treatment session, the number of sessions or the time between sessions. Trials presenting data from one of the following comparisons were eligible for inclusion.

Comparator interventions

- 'Passive control': treatment as usual/waiting-list control/relaxation/flexibility.
    - 'Treatment as usual' comprises medical assessments and advice given on a naturalistic basis. 'Relaxation' consists of techniques that aim to increase muscle relaxation (e.g. autogenic

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

training, listening to a relaxation tape). 'Flexibility' includes stretches performed according to selected exercises given.

- Psychological therapies: cognitive-behavioural therapy (CBT)/cognitive treatment/supportive therapy/behavioural therapies/psychodynamic therapies.
- Adaptive pacing therapy.
- Pharmacological therapy (e.g. antidepressants).

Types of outcome measures

Primary outcomes

1. Fatigue: measured using any validated scale (e.g. Fatigue Scale (FS) (Chalder 1993), Fatigue Severity Scale (FSS) (Krupp 1989)).
2. Adverse outcomes: measured using any reporting system (e.g. serious adverse reactions (SARs) (European Union Clinical Trials Directive 2001)).

Secondary outcomes

3. Pain: measured using any validated scale (e.g. Brief Pain Inventory (Cleeland 1994)).
4. Physical functioning: measured using any validated scale (e.g. Short Form (SF)-36, physical functioning subscale (Ware 1992)).
5. Quality of life (QOL): measured using any validated scale (e.g. Quality of Life Scale (Burckhardt 2003)).
6. Mood disorders: measured using validated instruments (e.g. Hospital Anxiety and Depression Scale (Zigmond 1983)).
7. Sleep duration and quality: measured by self-report on a validated scale, or objectively by polysomnography (e.g. Pittsburgh Sleep Quality Index (Buysse 1989)).
8. Self-perceived changes in overall health: measured by self-report on a validated scale (e.g. Global Impression Scale (Guy 1976)).
9. Health service resource use (e.g. primary care consultation rate, secondary care referral rate, use of alternative practitioners).
10. Drop-outs (any reason).

Timing of outcome assessment

We extracted from all studies data on each outcome for end of treatment and end of follow-up.

Search methods for identification of studies

Electronic searches

The Cochrane Collaboration's Depression, Anxiety and Neurosis (CCDAN) Review Group's Trials Search Coordinator (TSC) searched their Group's Specialized Register (CCDANCTR-Studies and CCDANCTR-References) (all years to 9 May 2014). This register is created from routine generic searches of MEDLINE (1950-), EMBASE (1974-) and PsycINFO (1967-). Details of CCDAN's generic search strategies, used to inform he CCDANCTR can be found on the Group's web site.

The CCDANCTR-Studies Register was searched using the following terms:

Diagnosis = ("Chronic Fatigue Syndrome" or fatigue) and Free Text = (exercise or sport* or relaxation or "multi convergent" or "tai chi")

The CCDANCTR-References Register was searched using a more sensitive list of free-text search terms to identify additional untagged/uncoded references, e.g. fatigue*, myalgic encephalomyelitis*, exercise, physical active* and taiji. Full search strategy listed in Appendix 1.

A complementary search of the following bibliographic databases and international trial registers were also conducted to 9 May 2014 (see Appendix 2):

- SPORTSDiscus (1985 -);
- The Cochrane Central Register of Controlled Trials (CENTRAL, all years -); and
- WHO International Clinical Trials Portal.

Searching other resources

We contacted the authors of included studies and screened reference lists to identify additional published or unpublished data. We conducted citation searches using the Institute for Scientific Information (ISI) Science Citation Index on the Web of Science.

Data collection and analysis

Selection of studies

Two of three review authors (LL, JO-J, KGB) inspected identified studies, using eligibility criteria to select relevant studies. In cases of disagreement, they consulted a third review author (JRP).

Data extraction and management

Melissa Edmonds and Jonatahan R Price independently extracted data from included studies for the 2004 version of this review, and LL and JO-J did so for this review update, using a standardised extraction sheet. They extracted mean scores at endpoint, the standard deviation (SD) or standard error (SE) of these values and the number of participants included in these analyses. When only the SE was reported, review authors converted it to the SD. For dichotomous outcomes, such as drop-outs, we extracted the number of events. We sought clarification from trial authors when necessary from investigators involved in the following trials: Fulcher 1997, Moss-Morris 2005, Wallman 2004, Wearden 2009, Wearden 2010 and White 2011. We resolved disagreement between review authors by discussion.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Main comparisons

- Exercise therapy versus 'passive control'.
- Exercise therapy versus psychological treatment.
- Exercise therapy versus adaptive pacing therapy.
- Exercise therapy versus pharmacological therapy (e.g. antidepressants).
- Exercise therapy as an adjunct to other treatment versus other treatment alone.

Assessment of risk of bias in included studies

Working independently, LL and JO-J, KGB or Jane Dennis (JD) assessed risk of bias using The Cochrane Collaboration risk of bias tool which was published in the most recent version of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). This tool encourages consideration of how the sequence was generated, how allocation was concealed, the integrity of blinding at outcome, the completeness of outcome data, selective reporting and other potential sources of bias. We classified all items in the risk of bias assessment as low risk, high risk or unclear risk by the extent to which bias was prevented.

Measures of treatment effect

Continuous data

For continuous outcomes, we calculated the mean difference (MD) when the same scale was used in a similar manner across studies. When results for continuous outcomes were presented using different scales or different versions of the same scale, we used the standardised mean difference (SMD).

Dichotomous data

For dichotomous outcomes, we expressed effect size in terms of risk ratio (RR).

Unit of analysis issues

Studies with multiple treatment groups

We extracted data from relevant arms of the included studies, and we compared the experimental condition (exercise therapy) versus each individual comparator intervention: 'Passive control' (treatment as usual/waiting-list control/relaxation/flexibility); 'Psychological treatment' (cognitive-behavioural therapy (CBT)/cognitive treatment/supportive therapy/behavioural therapies/psychodynamic therapies); 'Adaptive pacing therapy; and Pharmacological therapy (e.g. antidepressants). This meant that data from the exercise arm could be included in a separate univariate analysis for more than one comparison. We described under Differences between protocol and review planned methods that were found redundant, as we did not include studies requiring their use.

Dealing with missing data

When possible, we calculated missing standard deviations from reported standard errors, P values or confidence limits using the methods described in Chapter 7 (Sections 7.7.3.2 and 7.7.3.3) of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011). We approached trial investigators to obtain other types of missing data.

Assessment of heterogeneity

For this update, we assessed heterogeneity in keeping with the recommendations of the *Cochrane Handbook for Systematic Reviews of Interventions* (I^2 values of 0 to 40%: might not be important; 30% to 60%: may represent moderate heterogeneity; 50% to 90%: may represent substantial heterogeneity; 75% to 100%: show considerable heterogeneity; Higgins 2011). In addition to the I^2 value (Higgins 2003), we present the P value of the Chi^2 test, and we considered the direction and magnitude of treatment effects when making judgements about statistical heterogeneity. We deemed that no analyses were inappropriate as a result of the presence of statistical heterogeneity, as the measures and statistics used have low power and are unstable when based on few and small studies. A P value < 0.1 from the Chi^2 test was used as an indicator of statistically significant heterogeneity because of the low power of provided measures.

Assessment of reporting biases

We planned at the protocol stage to construct funnel plots when sufficient numbers of trials allowed a meaningful presentation, to establish whether other potential biases could be present. Asymmetry of these plots may indicate publication bias, although it also may represent a true relationship between trial size and effect size. We identified an insufficient number of studies to use this approach in the present version of the review (Egger 1997). We considered clinical diversity of the studies as a possible explanation for some of the heterogeneity apparent between studies.

Data synthesis

As the result of expected clinical heterogeneity (slightly different interventions, populations and comparators) among studies, we chose the random-effects model as the default method of analysis because the alternative fixed-effect model assumes that the true treatment effect in each trial is the same, and that observed differences are due to chance.

We performed analyses using Review Manager 5.0.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Subgroup analysis and investigation of heterogeneity

We planned no subgroup analyses a priori. To explore possible differences between studies that used different strategies (e.g. exercise therapy), control conditions and diagnostic criteria, we performed post hoc subgroup analyses. We describe results of these subgroup analyses in the text of the review.

Sensitivity analysis

We planned no sensitivity analyses a priori. To explore the possible impact of our pooling strategy (e.g. the impact of using SMD vs MD), we performed post hoc sensitivity analyses. In addition, we performed sensitivity analyses when studies with outlying results where excluded. We describe results of these sensitivity analyses in the text of the review.

RESULTS

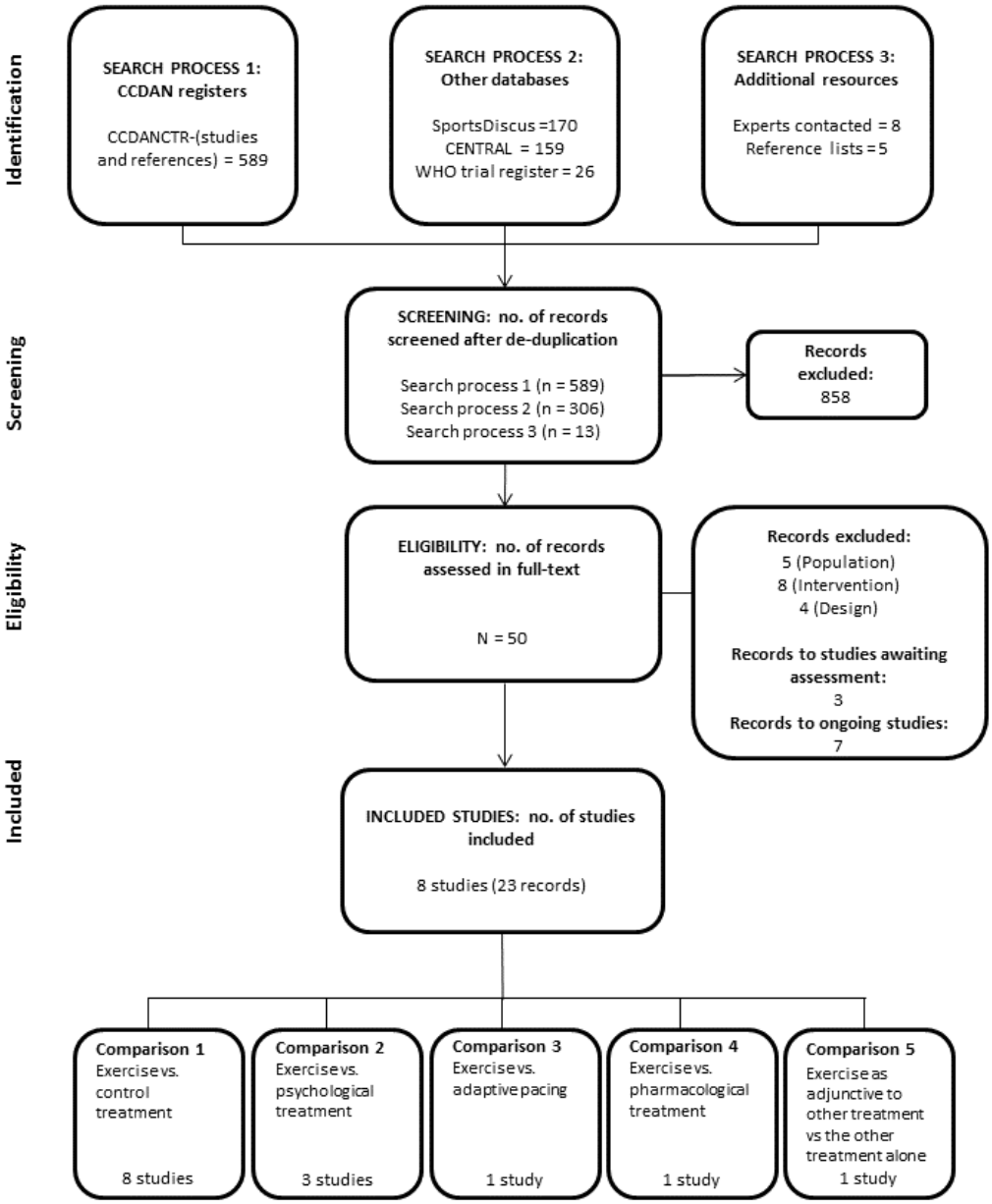
Description of studies

Results of the search

Our searches identified 908 unique records. Of these, we retrieved 50 records and read the full text. Along with the five included studies from the 2004 version of this review (Fulcher 1997; Moss-Morris 2005; Powell 2001; Wallman 2004; Wearden 1998), we have included three additional studies in this update (Jason 2007; Wearden 2010; White 2011; see Figure 1).

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 1. PRISMA flow diagram.



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Included studies

A total of eight studies (Fulcher 1997; Jason 2007; Moss-Morris 2005; Powell 2001; Wallman 2004; Wearden 1998; Wearden 2010; White 2011) met our inclusion criteria for this review (23 reports in all). All included studies were written in English and were published in peer-reviewed journals.

Design

All included studies were described as randomised controlled trials. Three studies included two arms (Fulcher 1997; Moss-Morris 2005; Wallman 2004) comparing exercise versus relaxation/flexibility, waiting list or standard care, respectively.

Four studies had four arms. For Powell 2001, we combined the three intervention arms and used these as comparators versus treatment as usual. We considered two arms (exercise + drug placebo vs exercise placebo + drug placebo) in Wearden 1998 as relevant for this review. For Jason 2007 and White 2011, all four arms were used, as were three arms in Wearden 2010.

The eight studies randomly assigned a total of 1518 participants. Samples included in this review ranged from 49 (Moss-Morris 2005) to 641 participants (White 2011).

Setting

Two studies took place in primary care settings: one in the United Kingdom (Wearden 2010) and one in Australia (Wallman 2004). Two studies were performed in secondary care facilities: one in the United Kingdom (Fulcher 1997) and one in New Zealand (Moss-Morris 2005). One study recruited from a variety of sources but took place at a hospital in the USA (Jason 2007). Three studies were conducted at secondary/tertiary care settings in the United Kingdom (Powell 2001; Wearden 1998; White 2011).

Participants

Three studies used the Centers for Disease Control and Prevention (CDC) 1994 criteria (Fukuda 1994) as inclusion criteria (Jason 2007; Moss-Morris 2005; Wallman 2004), and five (Fulcher 1997; Powell 2001; Wearden 1998; Wearden 2010; White 2011) used the Oxford criteria (Sharpe 1991). Wearden 2010 and White 2011 showed an overlap between Oxford criteria (Sharpe 1991) and London ME criteria (The National Task Force on CFS) of 31% and 51%, respectively. More female than male participants were included (range 71% to 84% when all arms were included), and mean ages across studies were between 33 and 44.6 years (confirmation of age data was requested from a trial investigator in one case (Wallman 2009)). The studies reported median illness duration of between 2.3 and 7 years. All but one study (Wallman 2004) reported depression, which ranged from 18% (Wearden 2010) of those with a depression diagnosis to 39% among participants with a current Axis I disorder (Jason 2007). Three studies did not report work and employment information (Wallman 2004; Wearden 2010; White 2011). Fulcher 1997 and Jason 2007 reported that 39% and 46% of participants were working or studying on at least a part-time basis, 22% of participants in Moss-Morris 2005 were unemployed and were unable to work because of disability and 42% of participants in Powell 2001 were receiving disability pensions (Table 1).

Intervention characteristics

The exercise therapy regimen lasted between 12 and 26 weeks. Seven studies used variations of aerobic exercise therapy such as walking, swimming, cycling or dancing at mixed levels in terms of intensity of the aerobic activity ranging from very low to quite rigorous; the remaining study used anaerobic exercise (Jason 2007). Scheduled therapist meetings could be conducted face-to-face or by telephone and varied from every second week to weekly; some sessions involved talking, and some exercise. Most of the included studies asked participants to exercise at home, most often between three and five times per week, with a target duration of 5 to 15 minutes per session using different means of incrementation (Fulcher 1997; Moss-Morris 2005; Powell 2001; Wallman 2004; Wearden 1998; Wearden 2010; White 2011). Participants were asked to perform self-monitoring by using such tools as heart monitors, the Borg Scale or a diary including an exercise log to measure adherence to treatment (Table 2). Control interventions included treatment as usual, relaxation plus flexibility and a waiting-list control group.

Outcomes

The main outcomes were symptom levels measured by rating scales at end of treatment (12 to 26 weeks) and at follow-up (52 to 70 weeks). Fatigue was measured by the Fatigue Scale (FS) (Chalder 1993) in seven studies (Fulcher 1997; Moss-Morris 2005; Powell 2001; Wallman 2004; Wearden 1998; Wearden 2010; White 2011) and by the Fatigue Severity Scale (FSS) (Krupp 1989) in one study (Jason 2007). Another study (White 2011) reported adverse outcomes according to SAR categories (European Union Clinical Trials Directive 2001).

The Jason 2007 study measured pain using the Brief Pain Inventory (Cleeland 1994). Physical functioning was measured by the SF-36 physical functioning subscale (Ware 1992) in seven studies (Fulcher 1997; Jason 2007; Moss-Morris 2005; Powell 2001; Wearden 1998; Wearden 2010; White 2011). Quality of life was measured by the Quality of Life Scale (QOLS) (Burckhardt 2003) in another study (Jason 2007).

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Seven studies (Fulcher 1997; Jason 2007; Moss-Morris 2005; Powell 2001; Wallman 2004; Wearden 2010; White 2011) reported self-perceived changes in overall health using the Global Impression Scale (Guy 1976).

Of the seven studies that reported mood disorder, six (Fulcher 1997; Powell 2001; Wallman 2004; Wearden 1998; Wearden 2010; White 2011) used the Hospital Anxiety and Depression Scale (HADS) (Zigmond 1983), and one (Jason 2007) used the Beck Depression Inventory (BDI-II) (Beck 1996) and the Beck Anxiety Inventory (BAI) (Hewitt 1993). Three studies (Powell 2001; Wearden 2010; White 2011) measured sleep problems by using a questionnaire (Jenkins 1988), two (Fulcher 1997; Powell 2001) by using the Pittburgh Sleep Quality Index (PSQI) (Buysse 1989).

One study reported health service resource use (White 2011).

Drop-out was calculated by the review authors.

Included studies reported several outcomes in addition to those reported in this review, such as work capacity by oxygen consumption (VO_2), the six-minute walking test and illness beliefs. See Characteristics of included studies for more detailed information.

Ethics approval

Ethics approval was obtained for all listed studies and sponsoring or funding listed.

Excluded studies

Two studies were excluded in 2004, as the diagnoses used were Gulf War veterans' illness (Guarino 2001) and subclinical chronic fatigue (Ridsdale 2004). The study awaiting assessment from 2004 was also excluded (Stevens 1999), as exercise therapy was a minor part of a combination treatment.

The current version excluded 14 studies (Evering 2008; Gordon 2010; Guarino 2001; Nunez 2011; Ridsdale 2004; Ridsdale 2012; Russel 2001; Stevens 1999; Taylor 2004; Taylor 2006; Thomas 2008; Tummers 2012; Viner 2004; Wright 2005). In addition to the two studies excluded from the 2004 version because of the population included (Guarino 2001; Ridsdale 2004), another with the diagnosis of chronic fatigue was excluded (Ridsdale 2012), as were two in which participants were younger than 18 years (Viner 2004; Wright 2005). Along with the one study excluded in 2004 (Stevens 1999), another five studies (Evering 2008; Nunez 2011; Russel 2001; Taylor 2004; Tummers 2012) were excluded in this review update because exercise therapy was a minor part of the intervention. One study was excluded because investigators compared two exercise interventions (Gordon 2010). Two studies were excluded because they were not RCTs (Taylor 2006; Thomas 2008).

Ongoing studies

We identified five ongoing studies in trial registers (Broadbent 2012; Kos 2012; Marques 2012; Vos-Vromans 2008; White 2012).

Studies awaiting classification

Studies identified from searches run to 9 May 2014 were assessed for eligibility and were classified accordingly. Three studies identified in the search are waiting assessment for possible inclusion, as the available information is too sparse for conclusions about eligibility. One abstract seems to refer to an unpublished study (Hatcher 1998), but we have not been able to contact the study authors for clarification. Additionally, two citations refer to studies that are available only in Chinese (Liu 2010; Zhuo 2007). Again, we have not been able to contact the study authors to clarify their relevance, and we have not had the resources to perform translation.

New studies found at this update

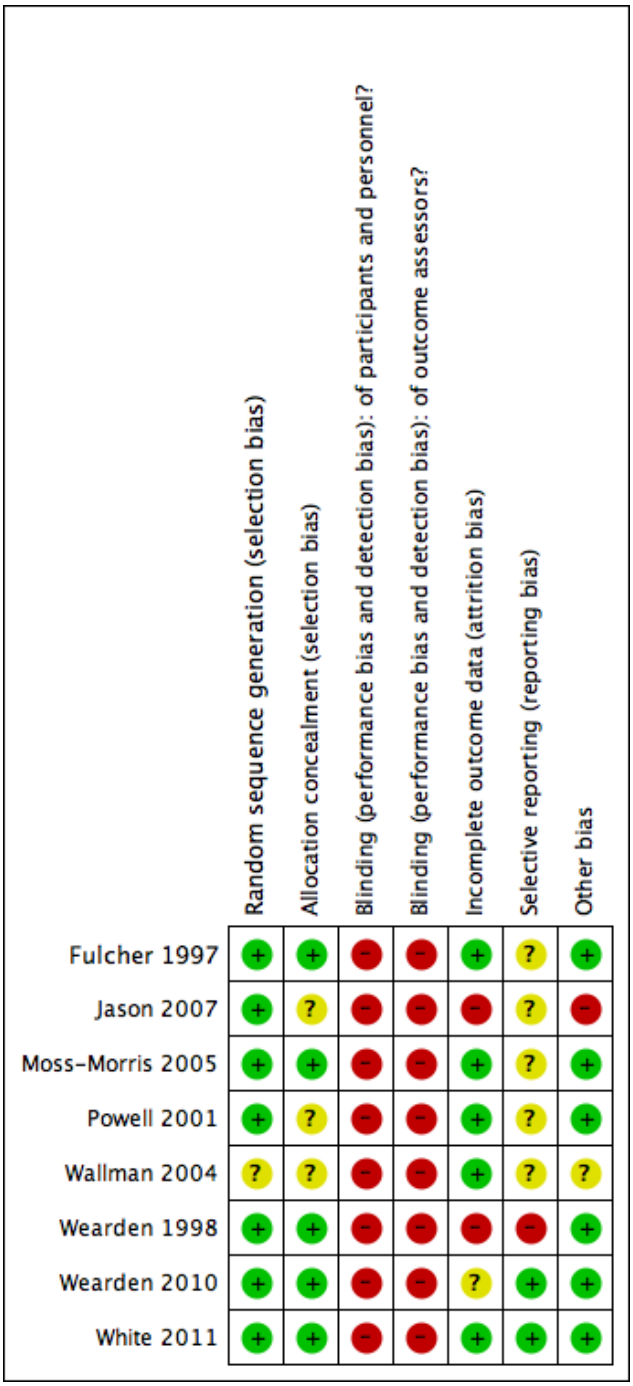
Three new studies have been added in this updated review (Jason 2007; Wearden 2010; White 2011).

Risk of bias in included studies

Summaries of the risk of bias assessments are presented in Figure 2 and Figure 3.

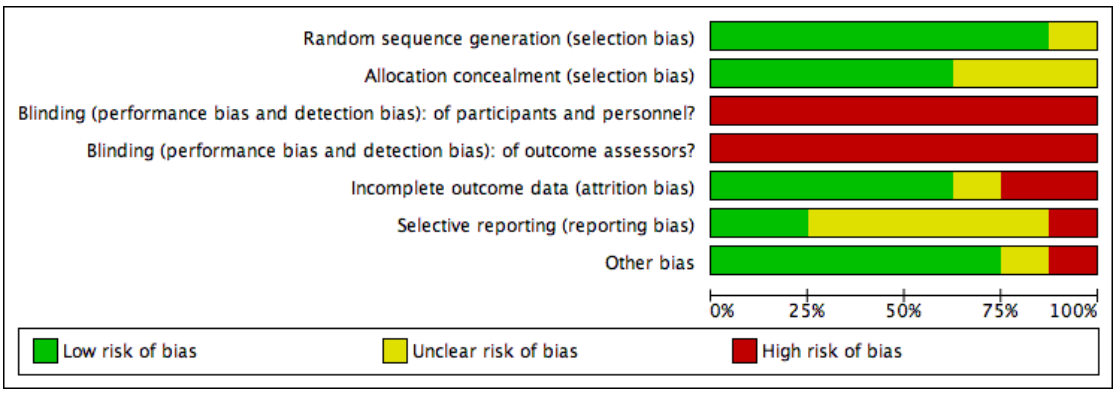
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 2. Risk of bias summary: review authors' judgements about each risk of bias item for each included study.



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Figure 3. Risk of bias graph: review authors' judgements about each risk of bias item presented as percentages across all included studies.



Allocation

All but one of the studies had adequate sequence generation (Wallman 2004). We judged five reported methods of allocation concealment as 'adequate' and found that methods described by the remaining three were unclear (Jason 2007; Powell 2001; Wallman 2004).

Blinding

As the intervention did not allow for blinding of participants or personnel delivering the exercise-based interventions, and as all measures were performed by self-report, blinding was impossible. This inevitably puts the review at some risk of bias, and all of the included studies were rated as having high risk of bias.

Incomplete outcome data

Risk of bias due to incomplete outcomes was low in five of the eight included studies, reflecting the fact that loss to follow-up was low, and that participants who were lost to follow-up were evenly distributed between intervention and control groups (Fulcher 1997; Moss-Morris 2005; Powell 2001; Wallman 2004; White 2011). One trial was associated with unclear risk of attrition bias (Wearden 2010). The drop-out rate in the intervention groups in this trial was relatively high, but most of the participants who dropped out from treatment were still available for follow-up assessments and were analysed within the groups to which they were randomly assigned (Wearden 2010). Two trials were associated with high risk of attrition bias (Jason 2007; Wearden 1998). Wearden 1998 reported large drop-out rates in all intervention groups as compared with control groups, and many participants were lost to follow-up. In Jason 2007, the conservatively defined drop-out rate (i.e. "attending four or fewer sessions or stopping therapy prior to satisfactory completion of therapy") on average was 25%. Study authors used the best linear unbiased predictor to avoid taking missing data into account, but as loss to follow-up for various intervention groups was not reported, we assessed the risk of attrition bias as high for this trial.

Selective reporting

Two studies (Wearden 2010; White 2011) referenced published protocols, and when we checked these against the published results, we found that reporting was adequate. In one study (Wearden 1998), trial investigators reported numerical data for only one subscale (health perception) of the Medical Outcomes Survey (MOS) scale (Ware 1992), for which data favour the intervention group; no numerical data were given for the five other subscales, nor for another scale (anxiety), as data were "similar in trial completers." It was not possible to check the other studies for selective reporting bias; therefore their risk of bias is considered unclear.

Other potential sources of bias

Seven of the eight studies seem to be free of other sources of bias, and one showed a baseline difference across groups for several variables (Jason 2007). These were not discussed when results were presented in the paper. In addition this study had 25 outcome

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

measures; because of this large number, one significant measure would be expected to occur by chance (Jason 2007). Wallman 2004 showed differences between groups for anxiety and mental fatigue at baseline, and this might have influenced the results.

Effects of interventions

See: **Summary of findings for the main comparison**

Exercise therapy versus control

Comparison 1. Exercise therapy versus treatment as usual, relaxation or flexibility

All included studies (Fulcher 1997; Jason 2007; Moss-Morris 2005; Powell 2001; Wallman 2004; Wearden 1998; Wearden 2010; White 2011) contributed data for this comparison.

1.1 Fatigue

Powell 2001 (148 participants) assessed fatigue by dichotomised scoring of an 11-item Fatigue Scale (FS, 0 to 11 points) (Chalder 1993) and reported results clearly in favour of exercise therapy (mean difference (MD) -6.06, 95% confidence interval (CI) -6.95 to -5.17; Analysis 1.1). Three studies (Wallman 2004; Wearden 2010; White 2011) measured fatigue among a total of 540 participants using the same 11-item FS with a different scoring system (0 to 33 points) (Chalder 1993) (Analysis 1.1). The pooled estimate suggests that exercise therapy was significantly more effective than treatment as usual (MD -2.82, 95% CI -4.07 to -1.57) - a result that was not associated with heterogeneity (I^2 = 0%, P value 0.54). Three studies (Fulcher 1997; Moss-Morris 2005; Wearden 1998) with a total of 152 participants measured fatigue using a 14-item FS (0 to 42 points) (Chalder 1993). Pooling shows a significant decrease in fatigue in the exercise group when compared with treatment as usual (MD -6.80 points, 95% CI -10.31 to -3.28), and the analysis was associated with low heterogeneity (I^2 = 20%, P value 0.29).

At follow-up, small strengthening of the effect was observed on the 11-point FS (Chalder 1993) as reported by Powell 2001 (MD -7.13, 95% CI -7.97 to -6.29; 148 participants; Analysis 1.2). Pooling of the two studies (Wearden 2010; White 2011) that measured fatigue on the 33-point scale resulted in almost the same effect estimate at follow-up as at end of treatment (MD -2.87, 95% CI -4.18 to -1.55; 472 participants; Analysis 1.2). The latter analysis was not associated with any unexplained heterogeneity (I^2 = 0%, P value 0.46). Jason 2007 (50 participants) did not report results at end of treatment but showed little or no difference in fatigue between anaerobic exercise and treatment as usual at follow-up, as measured on the Fatigue Severity Scale (FSS) (Krupp 1989) (MD 0.15, 95% CI -0.55 to 0.85; Analysis 1.2).

Sensitivity analysis

Investigating heterogeneity

At end of treatment, fatigue was measured and reported on different scales, and we performed a sensitivity analysis in which all available studies were pooled using an SMD method. This strategy led to a pooled random-effects estimate of -0.68 (95% CI -1.02 to -0.35), but the analysis suffered from considerable heterogeneity (I^2 = 78%, P value < 0.0001; Analysis 1.19). The observed heterogeneity was caused mainly by the deviating results presented in Powell 2001. Exclusion of Powell 2001 gave rise to a pooled SMD of -0.46 (95% CI -0.63 to -0.29) - an estimate that was not associated with heterogeneity (I^2 = 13%, P value 0.33).

At follow-up, the four available studies (Jason 2007; Powell 2001; Wearden 2010; White 2011) measured and reported fatigue on different scales, and we performed a sensitivity analysis in which all available studies were pooled using an SMD method. The pooled SMD estimate is -0.63 (95% CI -1.32 to 0.06), but heterogeneity was extensive (I^2 = 93%, P value < 0.00001). Exclusion of Powell 2001 gave rise to a new pooled SMD of -0.29 (95% CI -0.55 to -0.03) and reduced heterogeneity (I^2 = 46%, P value 0.16).

Subgroup analysis

To explore the possible impact of our pooling strategy (e.g. the impact of pooling studies adhering to different exercise strategies and control conditions), we performed post hoc subgroup analyses within Analysis 1.1 and Analysis 1.2.

Type of exercise

Post hoc subgroup analysis based on treatment strategy could not establish differences (I^2 = 0%, P value 0.60) between studies of graded exercise therapy (Fulcher 1997; Moss-Morris 2005; Powell 2001; Wearden 1998; Wearden 2010; White 2011) and studies testing exercise with self-pacing (Wallman 2004) (SMD -0.71, 95% CI -1.09 to -0.32; I^2 = 82% vs SMD -0.54, 95% CI -1.05 to -0.02, respectively) (Analysis 1.19).

At follow-up, post hoc subgroup analysis resulted in statistically significant subgroup differences (I^2 = 73.7%, P value 0.05) between the three studies (Powell 2001; Wearden 2010; White 2011) comparing graded exercise versus treatment as usual (SMD -0.86, 95% CI -1.67 to -0.05; I^2 = 95%) and Jason 2007, in which anaerobic activity was compared with relaxation (SMD 0.12, 95% CI -0.44 to 0.67).

Type of control

We cannot establish a subgroup difference (I^2 = 0%, P value 0.88) between the five studies with treatment as usual as control (Moss-Morris 2005; Powell 2001; Wearden 1998; Wearden 2010; White 2011) and the two studies prescribing relaxation or flexibility to participants in the control arm (Fulcher 1997; Wallman 2004) (SMD -0.70, 95% CI -1.14 to -0.25 vs SMD -0.65, 95% CI -1.02 to -0.28).

Diagnostic criteria

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

As the use of various diagnostic criteria is often emphasised as particularly important with regard to treatment response, we also performed subgroup analyses based on diagnostic criteria. Comparison of the two studies using 1994 CDC criteria (Moss-Morris 2005; Wallman 2004) and the five studies using the Oxford criteria (Fulcher 1997; Powell 2001; Wearden 1998; Wearden 2010; White 2011) revealed no differences between subgroups (I^2 = 0%, P value 0.84) (SMD -0.73, 95% CI -1.17 to -0.28 vs SMD -0.66, 95% CI -1.09 to -0.24).

1.2 Adverse effects

White 2011 reported two serious adverse reactions (SARs) (European Union Clinical Trials Directive 2001) possibly related to treatment among the 160 participants (i.e. deterioration in mobility and self-care and worse CFS symptoms and function) in the exercise group and two SARs among the 159 participants in the control group (i.e. worse CFS symptoms and function and increased depression and incapacity) (odds ratio (OR) 0.99, 95% CI 0.14 to 7.1; Analysis 1.3). Participants in the Wearden 2010 trial reported no SARs to therapy.

1.3 Pain

Wearden 1998 reported that all treated groups scored similarly on the pain subscale of SF-36 (Ware 1992), but measured values were not reported.

One trial, Jason 2007 (43 participants), assessed pain using the Brief Pain Inventory (Cleeland 1994) at follow-up (Analysis 1.4) and observed an MD of -0.97 (95% CI -2.44 to 0.50) on pain severity and -0.69 on the pain interference subscale (95% CI -2.48 to 1.10). The wide confidence interval implies that the results were inconclusive.

1.4 Physical functioning

Five trials (Fulcher 1997; Moss-Morris 2005; Powell 2001; Wearden 2010; White 2011) with a total of 725 participants assessed physical functioning according to the physical functioning subscale of SF-36 (Ware 1992) at end of treatment. The pooled estimate for these studies (Analysis 1.5) suggests that mean improvement for participants randomly assigned to exercise therapy was 13.10 points higher (95% CI 1.98 to 24.22) than for the treatment as usual group, but heterogeneity was considerable (I^2 = 89%, P value < 0.00001).

Four trials (669 participants) contributed data for evaluation of physical functioning at follow-up (Jason 2007; Powell 2001; Wearden 2010; White 2011). Jason 2007 observed better results among participants in the relaxation group (MD 21.48, 95% CI 5.81 to 37.15). However, results were distorted by large baseline differences in physical functioning between the exercise and relaxation groups (39/100 vs 54/100); therefore we decided not to include these results in the meta-analysis. Pooling of the three remaining trials (621 participants) showed a mean improvement on the SF-36 physical functioning subscale that was 16.33 points higher for exercise than for treatment as usual (95% CI -4.08 to 36.74; Analysis 1.6), but heterogeneity was excessive (I^2 = 96%, P value < 0.00001); therefore little or no difference cannot be ruled out.

Sensitivity analysis

Investigating heterogeneity

Extensive heterogeneity in Analysis 1.5 was largely driven by the remarkably positive effect of exercise therapy reported by Powell 2001. Heterogeneity (I^2) dropped to 52% (P value 0.10) following exclusion of Powell 2001, and the pooled mean difference still showed better improvement for participants in the exercise group (MD 7.37, 95% CI 1.23 to 13.51). The remaining heterogeneity may reflect the large variation in baseline physical functioning observed across studies, ranging from 29.8 (Wearden 2010) to 53.1 (Moss-Morris 2005), but the number of available studies was low; it is therefore difficult to explore this association further.

Also at follow-up, observed heterogeneity was driven by remarkably positive results in favour of exercise as reported by Powell 2001. If Powell 2001 was excluded, heterogeneity dropped to 0% (P value 0.50), and the two remaining trials (Wearden 2010; White 2011) reported a smaller but statistically significant difference in favour of exercise therapy (MD -5.79, 95% CI -10.53 to -1.06).

Subgroup analysis

To explore the possible impact of varying exercise strategies and control conditions, we performed post hoc subgroup analyses within Analysis 1.5 and Analysis 1.6.

Type of exercise

All studies included in Analysis 1.5 and Analysis 1.6 offered graded exercise therapy. Jason 2007 observed better results among participants in the relaxation group than among those in the anaerobic exercise group (MD 21.48, 95% CI 5.81 to 37.15) at follow-up. As stated above, these results were distorted by large baseline differences in physical functioning between exercise and relaxation groups (39 of 100 vs 54 of 100) and were not included in Analysis 1.6.

Type of control

At end of treatment, post hoc subgroup analysis did not establish a subgroup difference (I^2 = 0%, P value 0.92) between the four studies (Moss-Morris 2005; Powell 2001; Wearden 2010; White 2011) using treatment as usual as control (MD -12.96, 95% CI -26.63 to 0.72; I^2 = 92%) and Fulcher 1997, in which relaxation or flexibility was used as a control (MD -13.87, 95% CI -24.31 to -3.43). All studies available for analysis at follow-up adhered to the treatment as usual control condition, hence no sensitivity analyses were performed within Analysis 1.6.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Diagnostic criteria

We found no evidence of subgroup differences (I^2 = 0%, P value 0.91) between one study diagnosing participants according to the 1994 CDC criteria (MD -14.05, 95% CI -27.48 to -0.62; Moss-Morris 2005) and four studies diagnosing participants according to the Oxford criteria (MD -12.92, 95% CI -25.99 to 0.14). All studies available for analysis at follow-up recruited participants in keeping with the Oxford criteria, thus no sensitivity analyses were performed within Analysis 1.6.

1.5 Quality of life

None of the included studies reported quality of life at end of treatment. At follow-up, an estimate of effect suggested improvement towards better quality of life (Burckhardt 2003) among participants in the control group (MD 9.00, 95% CI -1.00 to 19.00; P value 0.08) compared with those given exercise therapy (Jason 2007; Analysis 1.7; 44 participants), but little or no effect cannot be ruled out. This estimate is biased in favour of the control arm because of baseline differences between groups.

1.6.1 Depression

Five studies (Fulcher 1997; Powell 2001; Wallman 2004; Wearden 1998; Wearden 2010) with a total of 504 participants contributed information on depression at end of treatment (12 to 26 weeks), all utilising the depression subscale of the Hospital Anxiety and Depression Scale (HADS) (Zigmond 1983). Pooling study results yielded an estimate of effect that suggested improvement in depression scores among participants allocated to exercise therapy compared with controls (MD 1.6 points, 95% CI -0.23 to 3.5; Analysis 1.8), but the results were highly heterogeneous (I^2 = 84%, P value < 0.0001), and little or no difference cannot be ruled out. At follow-up (Analysis 1.9), Jason 2007 (45 participants) assessed depression using the Beck Depression Inventory (BDI-II) (Beck 1996) and observed no difference in depression scores (MD 3.44, 95% CI -3.00 to 9.88)-an estimate that favours controls because of baseline differences between groups. Three trials reported HADS depression subscale values (Zigmond 1983) at follow-up (Powell 2001; Wearden 2010; White 2011; 609 participants). The pooled estimate of effect suggests that exercise therapy improved depression more than treatment as usual (MD -2.26, 95% CI -5.09 to 0.56), but heterogeneity was considerable (I^2 = 92%, P value < 0.00001), and little or no difference cannot be ruled out.

Sensitivity analysis

Investigating heterogeneity

At end of treatment, Powell 2001 again reported very positive results and contributed greatly to the total heterogeneity. Exclusion of Powell 2001 led to a reduction in observed effect size (MD 0.80, 95% CI -0.21 to 1.82), but heterogeneity was also greatly reduced (I^2 = 36%, P value 0.20).

Also at follow-up, Powell 2001 reported a substantial benefit of exercise therapy compared with results described by the other trials. Exclusion of Powell 2001 from the meta-analysis was associated with a great reduction in heterogeneity, as I^2 dropped from 92% to 9% (P value 0.30). Exclusion of Powell 2001 was also associated with a change in the observed effect estimate (MD -0.77, 95% CI -1.64 to 0.09). Hence, we still see an effect estimate suggesting modest benefit associated with exercise therapy, but little or no difference cannot be ruled out.

Standardised mean difference (SMD)

At longer-term follow-up, depression was measured and reported on different measurement scales; therefore we performed a sensitivity analysis in which all available studies were pooled using an SMD method. The four available studies (Jason 2007; Powell 2001; Wearden 2010; White 2011) yielded a pooled standardised estimate of SMD -0.35 (95% CI -0.93 to 0.23) in an analysis that was associated with considerable heterogeneity (I^2 = 91%, P value < 0.00001).

Subgroup analysis

To explore the possible impact of varying exercise strategies and control conditions, we performed post hoc subgroup analyses within Analysis 1.8 and Analysis 1.9.

Type of exercise

No statistical subgroup differences (I^2 = 0%, P value 0.75) were observed between the four studies offering graded exercise therapy (Fulcher 1997; Powell 2001; Wearden 1998; Wearden 2010) and Wallman 2004, which offered exercise with personal pacing.

At longer-term follow-up, four available studies (Jason 2007; Powell 2001; Wearden 2010; White 2011) provided a pooled standardised estimate of SMD -0.35 (95% CI -0.93 to 0.23) in an analysis that was associated with considerable heterogeneity (I^2 = 91%, P value < 0.00001). Post hoc subgroup analysis resulted in a statistically significant subgroup difference (I^2 = 71.2%, P value 0.06) between the three studies (Powell 2001; Wearden 2010; White 2011) comparing graded exercise therapy versus treatment as usual (SMD -0.53, 95% CI -1.20 to 0.13) and Jason 2007, which compared anaerobic activity versus relaxation (SMD 0.31, 95% CI -0.28 to 0.90).

Type of control

At end of treatment, the post hoc subgroup analysis did not establish a subgroup difference (I^2 = 0%, P value 0.61) between the three studies (Powell 2001; Wearden 1998; Wearden 2010) using treatment as usual as the control (MD -2.01, 95% CI -5.12 to 1.10; I^2 = 91%) and the two studies (Fulcher 1997; Wallman 2004) using relaxation or flexibility as the control (MD -1.05, 95% CI -2.95 to 0.84; I^2 = 59%).

1.6.2 Anxiety

Five trials (Fulcher 1997; Powell 2001; Wallman 2004; Wearden 1998; Wearden 2010) assessed anxiety at end of treatment using the anxiety subscale of the HADS (Zigmond 1983). Three studies

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

(387 participants) reported data in a way that facilitated comparison in a meta-analysis (Powell 2001; Wallman 2004; Wearden 2010), resulting in a pooled MD of -1.48 points (95% CI -3.58 to 0.61; Analysis 1.10). The meta-analysis was associated with heterogeneity (I^2 = 79%, P value 0.008), but some of this heterogeneity can be explained by uncorrected baseline differences in HADS anxiety score in included trials. Wearden 1998 (68 participants) stated that no significant changes were observed on the HADS anxiety score at end of treatment. Fulcher 1997 (58 participants) did not observe changes in median HADS anxiety score in the exercise group, whereas an increase in median HADS anxiety score from 4 to 7 was observed in the control group. However, the difference between exercise and control groups did not reach statistical significance in non-parametric statistical analysis.

Four trials assessed anxiety at longer-term follow-up (52 to 70 weeks; Analysis 1.11). Jason 2007 (45 participants) reported a mean difference on the Beck Anxiety Inventory (BAI) (Beck 1996) of 0.70 points (95% CI -4.52 to 5.92), and the wide confidence interval implies inconclusive results. Three trials (607 participants) assessed follow-up changes in anxiety using the HADS anxiety subscale (Powell 2001; Wearden 2010; White 2011). The pooled MD suggests greater improvement in HADS anxiety score in the exercise group compared with the group given treatment as usual (MD 1.01, 95% CI -0.74 to 2.75), but heterogeneity was considerable (I^2 = 78%, P value 0.01), and little or no difference cannot be ruled out.

Sensitivity analysis

Investigating heterogeneity

At follow-up, Powell 2001 reported very positive results and contributed to increased heterogeneity. Exclusion of Powell 2001 reduced heterogeneity to 63% (P value 0.10), and the pooled MD for White 2011 and Wearden 2010 was reduced to 0.24 (95% CI -1.27 to 1.74).

Standardised mean difference (SMD)

At longer-term follow-up, anxiety was measured and reported on different measurement scales; therefore we performed a sensitivity analysis in which all available studies were pooled using an SMD method. Four available studies (Jason 2007; Powell 2001; Wearden 2010; White 2011) yielded a pooled standardised estimate of SMD -0.17 (95% CI -0.50 to 0.15), but the analysis was associated with heterogeneity (I^2 = 71%, P value 0.02).

Subgroup analysis

To explore the possible impact of varying exercise strategies and control conditions, we performed post hoc subgroup analyses within Analysis 1.10 and Analysis 1.11.

Type of exercise and control

At end of treatment, post hoc subgroup analysis did not establish a subgroup difference (I^2 = 0%, P value 0.64) between the two studies (Powell 2001; Wearden 2010) comparing graded exercise therapy versus treatment as usual (MD -1.22, 95% CI 0.-4.51 to 2.07; I^2 = 88%) and Wallman 2004, which compared exercise with personal pacing versus flexibility and relaxation (MD -2.10, 95% CI -3.86 to -0.34).

At follow-up, four available studies (Jason 2007; Powell 2001; Wearden 2010; White 2011) yielded a pooled standardised estimate of SMD -0.17 (95% CI -0.50 to 0.15), but the analysis was associated with heterogeneity (I^2 = 71%, P value 0.02). We could not establish a statistically significant subgroup difference (I^2 = 0%, P value 0.40) between the three studies (Powell 2001; Wearden 2010; White 2011) comparing graded exercise therapy versus treatment as usual (SMD -0.23, 95% CI -0.61 to 0.16) and Jason 2007, which compared anaerobic activity versus relaxation (SMD 0.08, 95% CI -0.51 to 0.66).

1.7 Sleep

Two trials (Powell 2001; Wearden 2010), with a total of 323 participants, suggested that sleep assessed by the Jenkins Sleep Scale (Jenkins 1988) had improved more among participants in the exercise group at end of treatment (MD -1.49 points, 95% CI -2.95 to -0.02; P value 0.05; Analysis 1.12). Fulcher 1997, with 59 participants at end of treatment, observed a reduction in median sleep score, as assessed by the Pittsburgh Sleep Quality Index, from 7 to 5 in the exercise group, whereas median sleep score remained 6 in the control group; this group difference did not reach statistical significance in non-parametric statistical analysis.

At follow-up, three included trials (Powell 2001; Wearden 2010; White 2011) (610 participants) showed effects in favour of exercise therapy when they were pooled (MD -2.04 points, 95% CI -3.48 to -0.23; P value 0.03; Analysis 1.13), but the three studies showed heterogeneous results: a large positive effect in Powell 2001 (MD -4.05, 95% CI -6.08 to -2.02) and a moderate effect in White 2011 (MD -2.00, 95% CI -3.84 to -0.23), with Wearden 2010 reporting no observed statistically significant differences between the two groups (MD -0.31, 95% CI -1.97 to 1.35).

Subgroup analysis

All available studies compared graded exercise therapy versus treatment as usual. All studies recruited participants according to the Oxford criteria, thus no subgroup analyses were performed within Analysis 1.12 and Analysis 1.13.

1.8 Self-perceived changes in overall health

Seven trials assessed changes in overall health at end of treatment or at follow-up by using a self-rated Global Impression Change Scale with scores ranging from 1 (very much better) to 7 (very

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

much worse). We performed analysis of the numbers of participants reporting improvement. Four trials (523 participants) reported changes in overall health after end of treatment (Fulcher 1997; Moss-Morris 2005; Wallman 2004; Wearden 2010) and consistently showed a larger number of participants with some degree of improvement in the exercise group (RR 1.83, 95% CI 1.39 to 2.40; Analysis 1.14).

Three trials (518 participants) reporting self-perceived changes in overall health at follow-up were more inconsistent (Jason 2007; Powell 2001; White 2011). The point estimate for the risk ratio favoured exercise therapy (RR 1.88, 95% CI 0.76 to 4.64; Analysis 1.15), but the confidence interval implies inconclusive results, and heterogeneity was substantial (I^2 = 85%). Jason 2007 showed no significant differences between exercise and relaxation (RR 0.83, 95% CI 0.44 to 1.56) and White 2011 suggested a positive effect of exercise therapy compared with treatment as usual (RR 1.63, 95% CI 1.16 to 2.29), whereas Powell 2001 indicated a large positive effect for exercise (RR 5.96, 95% CI 2.36 to 15.09).

Subgroup analysis

To explore the potential impact of varying exercise strategies and control conditions, we performed a post hoc subgroup analysis within Analysis 1.14 and Analysis 1.15.

Type of control

At end of treatment, the pooled RR for all available studies was 1.83 (95% CI 1.39 to 2.40; I^2 = 0%) compared with 1.99 (95% CI 1.38 to 2.86; I^2 = 0%) in the treatment as usual subgroup (Moss-Morris 2005; White 2011) and 1.64 (95% CI 1.09 to 2.48; I^2 = 0%) in the relaxation/flexibility subgroup (Fulcher 1997; Wallman 2004). Tests for subgroup differences did not establish differences between the two groups (I^2 = 0%, P value 0.50).

Type of exercise

Three studies offering graded exercise therapy (Fulcher 1997; Moss-Morris 2005; White 2011) tended towards a greater chance of improvement (RR 2.01, 95% CI 1.46 to 2.77) than the study offering exercise with personal pacing (RR 1.43, 95% CI 0.85 to 2.41; Wallman 2004), but statistical tests did not establish a subgroup difference (I^2 = 13.6%, P value 0.28).

At follow-up, the pooled RR for the three available studies was 1.88 (95% CI 0.76 to 4.64) in an analysis associated with extensive heterogeneity (I^2 = 85%, P value 0.001). The post hoc subgroup analysis did not firmly establish a subgroup difference (I^2 = 63%, P value 0.10) between the two studies (Powell 2001; White 2011) comparing graded exercise therapy versus treatment as usual (RR 2.92, 95% CI 0.75 to 11.35; I^2 = 87%) and Jason 2007, which compared anaerobic activity versus relaxation (RR 0.83, 95% CI 0.44 to 1.56).

1.9 Health service resources

Data on health service resources are available for one of the included studies with a total of 320 participants (White 2011). During the 12-month post-randomisation period, participants in the treatment as usual group had a higher mean number of specialist medical care contacts than those allocated to exercise therapy (MD -1.40, 95% CI -1.87 to -0.93; Analysis 1.16). Use of primary care resources (i.e. general practitioner or practice nurse), other doctor contacts (i.e. neurologist, psychiatrist or other specialists), accident and emergency contacts, medication (i.e. hypnotics, anxiolytics, antidepressants or analgesics), contacts with other healthcare professionals (i.e. dentist, optician, pharmacist, psychologist, physiotherapist, community mental health nurse or occupational therapist), inpatient contacts and other contacts with healthcare/social services (e.g. social worker, support worker, nutritionist, magnetic resonance imaging (MRI), computed tomography (CT), electroencephalography (EEG)) did not differ significantly between the two groups (Analysis 1.16; Analysis 1.17)

1.10 Drop-out

Six studies (Fulcher 1997; Moss-Morris 2005; Powell 2001; Wearden 1998; Wearden 2010; White 2011), with a total of 843 participants, reported drop-out rates (Analysis 1.18). The pooled RR for drop-out was 1.63 (95% CI 0.77 to 3.43). The confidence interval implies that these results were inconclusive, and heterogeneity was moderate (I^2 = 50%).

Subgroup analysis

The main analysis pooled studies using treatment as usual (Moss-Morris 2005; Powell 2001; Wearden 1998; Wearden 2010) and studies using flexibility (Fulcher 1997) into the same comparison. The pooled RR for all available studies was 1.63 (95% CI 0.77 to 3.43; I^2 = 50%) compared with 1.77 (95% CI 0.71 to 4.38; I^2 = 61%) in the treatment as usual subgroup and 1.33 (95% CI 0.32 to 5.50) in the flexibility subgroup (Fulcher 1997). Tests for subgroup differences did not establish differences between the two groups (I^2 = 0%, P value 0.74).

Exercise therapy versus other treatments

Comparison 2. Exercise therapy versus psychological treatment

Three trials (Jason 2007; White 2011; Wearden 2010) contributed data to this comparison, which included cognitive-behavioural therapy (CBT) (Jason 2007; White 2011), cognitive therapy treatment (COG) (Jason 2007) and supportive listening (Wearden 2010). We decided not to pool the results in meta-analyses because of clinical and contextual heterogeneity.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

2.1 Fatigue

End of treatment

White 2011 (298 participants) showed little or no difference in fatigue between exercise therapy and CBT (MD 0.20, 95% CI -1.49 to 1.89; Analysis 2.1).

Compared with 97 participants randomly assigned to supportive listening (Wearden 2010), 85 participants in the graded exercise therapy group experienced greater improvement in fatigue (MD -4.03, 95% CI -6.24 to -1.82; P value < 0.001; Analysis 2.1).

Follow-up

Jason 2007 assessed fatigue using a 7-point Fatigue Severity Scale (Krupp 1989) and showed an MD of -0.10 (95% CI -0.79 to 0.59) for anaerobic exercise versus COG (49 participants; Analysis 2.2). The wide confidence interval implies imprecise and inconclusive results.

Wide confidence intervals and imprecise results also apply to the comparison of anaerobic exercise versus CBT as reported by Jason 2007 (49 participants) with an MD of 0.40 (95% CI -0.34 to 1.14; Analysis 2.2). White 2011 compared graded exercise therapy versus CBT (302 participants) by assessing fatigue on a 33-point Fatigue Scale (Chalder 1993) and observed little or no difference between the two groups (MD 0.30, 95% CI -1.45 to 2.05; Analysis 2.3).

Wearden 2010 (182 participants) assessed fatigue on a 33-point Fatigue Scale (Chalder 1993) and reported differences between rehabilitation and supportive listening that favoured graded exercise therapy (MD -2.72, 95% CI -5.14 to -0.30; P value 0.03; Analysis 2.3).

Sensitivity analysis

At follow-up, the available studies (Jason 2007; White 2011) measured and reported fatigue on different scales, and we performed a sensitivity analysis in which the two studies were pooled using an SMD method. The resulting pooled SMD estimate is 0.07 (95% CI -0.13 to 0.28) with no unexplained heterogeneity (I^2 = 0%, P value 0.40).

Subgroup analysis

Post hoc subgroup analysis did not establish a subgroup difference (I^2 = 0%, P value 0.40) between White 2011, which compared graded exercise therapy versus CBT (SMD 0.04, 95% CI -0.19 to 0.26), and Jason 2007, which compared anaerobic activity versus CBT (SMD 0.30, 95% CI -0.26 to 0.86).

2.2 Adverse effects

White 2011 reported the number of serious adverse reactions (SARs) (European Union Clinical Trials Directive 2001) observed in each treatment group (Analysis 2.4). Two adverse reactions possibly related to treatment were observed among the 160 participants in the exercise group (one participant with deterioration in mobility and self-care, and one with worse CFS symptoms and function), and three participants reporting a total of four SARs were described among 161 participants in the CBT group (one incident of self-harm, one incident of low mood with an episode of self-harm, one episode of worsened mood and CFS symptoms and one incident of threatened self-harm). Thus, the observed RR was 0.67 (95% CI 0.11 to 3.96), implying that these results were inconclusive.

Wearden 2010 stated that no participants in the rehabilitation or supportive listening group demonstrated SARs with a probable relation to therapy (Analysis 2.4).

2.3 Pain

Jason 2007 (43 participants) reported differences in pain at follow-up (52 weeks), as assessed by the Brief Pain Inventory (Cleeland 1994). When anaerobic exercise was compared with CBT, results were imprecise for pain severity (MD 0.07, 95% CI -1.52 to 1.66; Analysis 2.5) and for pain interference (MD -0.35, 95% CI -2.29 to 1.59; Analysis 2.6). As the result of baseline differences between groups, these estimates, to some extent, are biased in favour of exercise.

Jason 2007 also compared anaerobic exercise versus COG (44 participants). Here, inconclusive results were observed in pain severity (MD 0.51, 95% CI -0.92 to 1.94; Analysis 2.5) and pain interference (MD 0.39, 95% CI -1.37 to 2.15; Analysis 2.6).

2.4 Physical functioning

End of treatment

White 2011 (298 participants) reported changes in physical functioning between participants randomly assigned to exercise and CBT at end of treatment by using the SF-36 physical functioning subscale (Ware 1992). Scores on this scale range from 0 to 100, and study authors observed little or no difference in physical function between the two groups (MD -1.20, 95% CI -6.30 to 3.90; Analysis 2.7).

Wearden 2010 (181 participants) suggested greater improvement in physical function among participants in the graded exercise therapy group than in the supportive listening group (MD -6.66 point, 95% CI -13.7 to 0.40; P value 0.06; Analysis 2.7), but little or no difference cannot be ruled out.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Follow-up

Both Jason 2007 and White 2011 reported physical function at 52-week follow-up. Whereas White 2011 (302 participants) observed little or no difference between graded exercise therapy and CBT (MD 0.50, 95% CI -4.89 to 5.89; Analysis 2.8), Jason 2007 (46 participants) reported a significant difference favouring CBT (MD 18.92, 95% CI 2.12 to 35.72; Analysis 2.8) when compared with anaerobic exercise. However, results of the latter study are skewed because of uncorrected baseline differences in physical function between the two groups (39 vs 46 points), and this explains some of the observed heterogeneity.

Jason 2007 (47 participants) also compared anaerobic exercise versus COG, suggesting a large difference in favour of COG (MD 21.37, 95% CI 6.61 to 36.13; Analysis 2.8). It should be noted, however, that the latter estimate is probably biased in favour of COG because of uncorrected baseline differences in physical function between the two groups (39 vs 46 points).

Wearden 2010 (171 participants) suggested greater improvement in physical function among participants in the graded exercise therapy than in the supportive listening group (MD -7.55 point, 95% CI -15.57 to 0.47; Analysis 2.8), but little or no difference cannot be ruled out.

2.5 Quality of life

Study authors provided no data.

2.6.1 Depression

End of treatment

In Wearden 2010 (182 participants), graded exercise therapy was associated with greater improvement on the HADS depression subscale (Zigmond 1983) than was seen with supportive listening (MD -1.57, 95% CI -2.74 to -0.40; P value 0.008; Analysis 2.9). We did not identify trials reporting depression for exercise versus CBT or for exercise versus COG at end of treatment.

Follow-up

Jason 2007 assessed depression using the Beck Depression Inventory (BDI-II) (Beck 1996). When comparing anaerobic exercise versus COG (45 participants), study authors saw a trend towards greater improvement among participants in the COG group (MD 5.08, 95% CI -0.77 to 10.93; Analysis 2.10), but little or no difference cannot be ruled out.

Two trials compared exercise therapy versus CBT (Jason 2007; White 2011), with neither showing statistically significant differences between the two groups. Jason 2007 (44 participants) assessed depression using the BDI-II (Beck 1996) and reported imprecise results (MD 2.99, 95% CI -4.37 to 10.35; Analysis 2.10); interpretation of these results is further complicated by baseline differences between groups. On the other hand, White 2011 (287 participants) assessed depression using the HADS depression subscale (Zigmond 1983) and found little or no difference between graded exercise therapy and CBT (MD -0.10, 95% CI -1.00 to 0.80; Analysis 2.11).

Wearden 2010 compared graded exercise therapy and supportive listening. At end of treatment, results favoured exercise, but this effect was not sustained at 70 weeks' follow-up (171 participants; MD -0.79, 95%CI -2.31 to 0.55; Analysis 2.11).

Sensitivity analysis

As depression was measured and reported on two different scales in Jason 2007 and White 2011, we performed a sensitivity analysis in which the two studies were pooled using an SMD method. The resulting pooled SMD estimate is 0.01 (95% CI -0.21 to 0.22) with no unexplained heterogeneity (I^2 = 0%, P value 0.42).

Subgroup analysis

Post hoc subgroup analysis did not establish a subgroup difference (I^2 = 0%, P value 0.42) between White 2011, which compared graded exercise therapy versus CBT (SMD -0.03, 95% CI -0.26 to 0.21) and Jason 2007, which compared anaerobic exercise versus CBT (SMD 0.23, 95% CI -0.36 to 0.83).

2.6.2 Anxiety

End of treatment

Wearden 2010 (182 participants) found little or no difference on the HADS anxiety subscale (Zigmond 1983) between graded exercise therapy and supportive listening (MD -0.48, 95% CI -1.85 to 0.89; Analysis 2.12). We did not identify trials reporting anxiety for exercise therapy versus CBT or for exercise therapy versus COG at end of treatment.

Follow-up

Jason 2007 (45 participants) assessed anxiety using the Beck Anxiety Inventory (BAI) (Beck 1996). When comparing anaerobic exercise versus COG, study authors did not observe statistically significant differences between groups, but results were imprecise (MD 3.15, 95% CI -1.17 to 7.47; Analysis 2.13).

Two trials compared exercise therapy versus CBT (Jason 2007; White 2011), with neither showing statistically significant differences between the two groups. Jason 2007 (44 participants) assessed anxiety using the BAI (Beck 1996), with imprecise and statistically insignificant results (MD 0.66, 95% CI -4.68 to 6.00; Analysis 2.13). White 2011 (287 participants) found little or no difference between graded exercise therapy and CBT using the HADS anxiety subscale (MD 0.30, 95% CI -0.71 to 1.31; Analysis 2.14).

Wearden 2010 (171 participants) did not observe statistically significant differences on the HADS anxiety subscale between graded

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

exercise therapy and supportive listening at 70 weeks (MD -0.08, 95%CI -1.52 to 1.36; Analysis 2.14).

Sensitivity analysis

As depression was measured and reported on two different scales in Jason 2007 and White 2011, we performed a sensitivity analysis in which the two studies were pooled using an SMD method. The resulting pooled SMD estimate is 0.07 (95% CI -0.15 to 0.28) with no unexplained heterogeneity (I^2 = 0%, P value 0.99).

Subgroup analysis

Post hoc subgroup analysis did not establish a subgroup difference (I^2 = 0%, P value 0.99) between White 2011, which compared graded exercise therapy versus CBT (SMD 0.07, 95% CI -0.16 to 0.30) and Jason 2007, which compared anaerobic activity versus CBT (SMD 0.07, 95% CI -0.52 to 0.66).

2.7 Sleep

End of treatment

Wearden 2010 observed that the 83 participants in the graded exercise therapy group experienced greater improvement on the 20-point Jenkins Sleep Scale (Jenkins 1988) as compared with the 97 participants in the supportive listening group (MD -2.46 points, 95% CI -4.01 to -0.91; P value 0.002; Analysis 2.15). We did not identify trials reporting sleep for exercise therapy versus CBT or for exercise therapy versus COG at end of treatment.

Follow-up

White 2011 (287 participants) assessed sleep using the Jenkins Sleep Scale (Jenkins 1988) and found little or no difference between graded exercise therapy and CBT (MD -0.90, 95%CI -2.07 to 0.27; Analysis 2.16). Wearden 2010 (171 participant) also used the Jenkins Sleep Scale and found little or no difference between graded exercise therapy and supportive listening (MD -0.86, 95% CI -2.56 to 0.84; Analysis 2.16).

2.8 Self-perceived changes in overall health

Two trials (Jason 2007; White 2011) assessed changes in overall health by using a self-rated Global Impression Change Scale with scores ranging from 1 (very much better) to 7 (very much worse) (Guy 1976). We performed analysis of the numbers of participants reporting improvement.

End of treatment

White 2011 (320 participants) reported changes in overall health following graded exercise therapy versus CBT, but results were inconclusive (RR 0.96, 95% CI 0.71 to 1.31; Analysis 2.17).

Follow-up

At follow-up, self-perceived changes in overall health were reported by Jason 2007 and White 2011.

For the comparison of COG versus anaerobic exercise, Jason 2007 (50 participants) showed that more participants in the CBT group than in the exercise group tended to report improvement, but little or no difference between CBT and exercise therapy cannot be ruled out (RR 0.63, 95% CI 0.36 to 1.10; Analysis 2.18).

Both Jason 2007 (47 participants) and White 2011 (321 participants) compared exercise therapy versus CBT. Pooling resulted in an RR of 0.71 (95% CI 0.33 to 1.54; Analysis 2.18), implying imprecise and inconclusive results. The meta-analysis was associated with considerable heterogeneity (I^2 = 86%) as the result of inconsistency between effect estimates reported by Jason 2007, which compared anaerobic exercise versus CBT (RR 0.46, 95% CI 0.28 to 0.77), and White 2011, which compared graded exercise therapy versus CBT (RR 1.02, 95% CI 0.77 to 1.35).

2.9 Health service resources

Data on health service resources were provided by one of the included studies with a total of 321 participants (White 2011). During the 12-month post-randomisation period, participants in the CBT group showed lower mean numbers of contacts with neurologist, psychiatrist or other specialists (MD 0.60, 95% CI 0.05 to 1.15; Analysis 2.19) and lower mean numbers of inpatient days (MD 0.80, 95% CI 0.41 to 1.19; Analysis 2.19) when compared with participants in the exercise group. However, these group differences were not seen when data were analysed at a dichotomous level (Analysis 2.20).

2.10 Drop-out

White 2011 (321 participant) reported drop-out from treatment. Drop-out rates were not significantly different between graded exercise therapy and CBT (RR 0.59, 95% CI 0.28 to 1.25; Analysis 2.21), but these results were imprecise and inconclusive because few events were reported.

Wearden 2010 reported that more participants discontinued graded exercise therapy (12 of 92 participants) than supportive listening (7 of 91 participants) (RR 1.70, 95% CI 0.70 to 4.11; Analysis 2.21), but the confidence interval implies that these results were imprecise and inconclusive.

Comparison 3. Exercise therapy versus adaptive pacing therapy

One trial contributed data on 319 participants for this comparison (White 2011).

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

3.1 Fatigue

Fatigue assessed by a 33-point Fatigue Scale (Chalder 1993) improved more among participants allocated to graded exercise therapy than adaptive pacing (MD -2.00, 95% CI -3.57 to -0.43; P value 0.01) when measured at end of treatment (24 weeks; 305 participants). This positive effect was sustained at 52 weeks' follow-up (307 participants; MD -2.50, 95% CI -4.16 to -0.84; P value 0.003; Analysis 3.1).

3.2 Adverse effects

White 2011 reported the number of SARs (European Union Clinical Trials Directive 2001) observed in each treatment group (Analysis 3.2). Two SARs possibly related to treatment were observed among the 160 participants in the graded exercise therapy group (one incident of deterioration in mobility and self-care, and one episode of worse CFS symptoms and function) compared with two in the adaptive pacing group (159 participants) (one incident of suicidal thoughts, and one episode of worsened depression). Thus, results were inconclusive, with an RR of 0.99 (95% CI 0.14 to 6.97).

3.3 Pain

No data were provided.

3.4 Physical functioning

The graded exercise therapy group (150 participants) experienced significant improvement in physical functioning compared with the adaptive pacing group (155 participants) (Analysis 3.3). At end of treatment, participants in the graded exercise therapy group scored a mean of 12.2 points better (95% CI -17.23 to -7.17) on the SF-36 physical functioning subscale (Ware 1992) than those in the adaptive pacing group-a difference that was sustained at 52 weeks' follow-up (307 participants; MD -11.8, 95% CI -17.5 to -6.05).

3.5 Quality of life

No data were provided.

3.6.1 Depression

The change on the HADS depression subscale (Zigmond 1983) at end of treatment was not reported (White 2011). At follow-up, participants in the graded exercise therapy group (144 participants) had improved by a mean of 1.10 points (95% CI -2.09 to -0.11) on the HADS depression subscale when compared with the 149 participants in the pacing group (Analysis 3.4).

3.6.2 Anxiety

White 2011 did not report the change on the HADS anxiety subscale (Zigmond 1983) at end of treatment, and they observed little or no difference between the two groups (293 participants) at 52 weeks (MD -0.40, 95% CI -1.40 to 0.60; Analysis 3.5).

3.7 Sleep

White 2011 did not report change in sleep at end of treatment as assessed by the 20-point Jenkins Sleep Scale (Jenkins 1988). At follow-up, participants in the graded exercise therapy group (144 participants) had improved by a mean of 1.60 points (95% CI -2.70 to -0.50) when compared with the 150 participants in the adaptive pacing group (Analysis 3.6).

3.8 Self-perceived changes in overall health

White 2011 assessed changes in overall health by using a self-rated Global Impression Change Scale with scores ranging from 1 (very much better) to 7 (very much worse) (Guy 1976). Comparisons of the numbers of participants reporting improvement showed that a larger fraction of participants in the graded exercise therapy group experienced improvement at end of treatment (319 participants; RR 1.45, 95% CI 1.02 to 2.07; Analysis 3.7). At follow-up, an estimate of effect that suggested improvement favouring graded exercise therapy was still observed, but little or no effect cannot be ruled out (319 participants; RR 1.31, 95% CI 0.96 to 1.79).

3.9 Health service resources

One of the included studies with a total of 319 participants provided data on health service resources (White 2011). During the 12-month post-randomisation period, participants in the pacing group showed lower mean numbers of contacts with complementary healthcare resources (MD 3.80, 95% CI 1.42 to 6.18; Analysis 3.8), lower mean numbers of contacts with other doctors (neurologist, psychiatrist and other specialists) (MD 0.70, 95% CI 0.14 to 1.26; Analysis 3.8), lower mean numbers of accidents and emergencies (MD 0.50, 95% CI 0.31 to 0.69; Analysis 3.8) and higher mean numbers of inpatient days (MD 1.00, 95% CI 0.46 to 1.54; Analysis 3.8) than were seen among participants in the exercise group. However, these group differences were not seen when data were analysed at a dichotomous level (Analysis 3.9).

3.10 Drop-out

In the PACE trial (White 2011), 10 of the 160 participants in the graded exercise therapy group and 11 of the 160 participants in the adaptive pacing group withdrew, thus the results were inconclusive (RR 0.91, 95% CI 0.40 to 2.08; Analysis 3.10).

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Comparison 4. Exercise therapy versus antidepressants

One trial contributed data on a total of 69 participants to this comparison (Wearden 1998). In this trial, investigators combined graded exercise therapy with antidepressant placebo, and the antidepressant used was fluoxetine.

4.1 Fatigue

Investigators assessed fatigue on a 42-point Fatigue Scale (Chalder 1993; 48 participants) at end of treatment, but the results were inconclusive (MD -1.99, 95% CI -8.28 to 4.30; Analysis 4.1).

4.2 Adverse effects

Study authors provided no data.

4.3 Pain

Study authors provided no data.

4.4 Physical functioning

Study authors provided no data.

4.5 Quality of life

Study authors provided no data.

4.6.1 Depression

Researchers assessed depression among 48 participants at end of treatment using the HADS depression subscale (Zigmond 1983), but they found little or no difference between the exercise and fluoxetine groups (MD 0.15, 95% CI -2.11 to 2.41; Analysis 4.2).

4.6.2 Anxiety

Study authors provided no data.

4.7 Sleep

Study authors provided no data.

4.8 Self-perceived changes in overall health

Study authors provided no data.

4.9 Health service resources

Study authors provided no data.

4.10 Drop-out

Wearden 1998 observed similar drop-out rates in both groups, with 11 drop-outs reported among the 34 participants in the exercise group and 10 drop-outs among the 35 participants in the antidepressant group (RR 1.13, 95% CI 0.55 to 2.31; Analysis 4.3), implying that the results were inconclusive.

Exercise therapy adjunctive to other treatment versus the other treatment alone

Comparison 5. Exercise therapy versus antidepressants plus exercise therapy

One trial contributed data for a total of 68 participants to this comparison (Wearden 1998). In this trial, investigators combined graded exercise therapy with use of the antidepressant fluoxetine.

5.1 Fatigue

Researchers assessed fatigue on a 42-point Fatigue Scale (Chalder 1993; 43 participants) at end of treatment, but the results were inconclusive (MD -3.66, 95% CI -10.41 to 3.09; Analysis 5.1).

5.2 Adverse effects

Study authors provided no data.

5.3 Pain

Study authors provided no data.

5.4 Physical functioning

Study authors provided no data.

5.5 Quality of life

Study authors provided no data.

5.6.1 Depression

Researchers assessed depression at end of treatment among 43 participants using the HADS depression subscale (Zigmond 1983), but the results were inconclusive (MD -0.52, 95% CI -2.68 to 2.14; Analysis 5.2).

5.6.2 Anxiety

Study authors provided no data.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

5.7 Sleep

Study authors provided no data.

5.8 Self-perceived changes in overall health

Study authors provided no data.

5.9 Health service resources

Study authors provided no data.

5.10 Drop-out

Wearden 1998 observed similar drop-out rates in both groups, with 14 drop-outs reported among the 33 participants in the exercise plus antidepressant group, and 10 drop-outs among the 35 participants in the antidepressant group (RR 1.48, 95% CI 0.77 to 2.87; Analysis 5.3). The confidence interval implies that the results were inconclusive.

DISCUSSION

Summary of main results

We have included eight studies including 1518 participants in this review.

When exercise therapy was compared with 'passive control,' fatigue was significantly reduced at end of treatment (Analysis 1.1). Data on serious adverse reactions (SARs) were available from only one trial, and SARs were rare, but too few events were reported to allow any conclusions to be drawn (Analysis 1.3). A positive effect of exercise therapy was observed both at end of treatment and at follow-up with respect to sleep (Analysis 1.12; Analysis 1.13), physical functioning (Analysis 1.5; Analysis 1.6) and self-perceived changes in overall health (Analysis 1.14; Analysis 1.15). For the remaining outcomes, we were not able to draw any conclusions.

When exercise therapy was compared with cognitive-behavioural therapy (CBT), little or no difference in fatigue was noted between the two groups (Analysis 2.1; Analysis 2.2). Serious adverse reactions were rare and were reported at similar rates in the two groups. Events were few; therefore results were too imprecise to allow any conclusions to be drawn (Analysis 2.4). Little or no difference was observed between exercise therapy and CBT for physical functioning (Analysis 2.7; Analysis 2.8), depression (Analysis 2.10; Analysis 2.11), anxiety (Analysis 2.13; Analysis 2.14) and sleep (Analysis 2.16). It was not possible to draw any conclusions regarding pain (Analysis 2.5; Analysis 2.6), self-perceived changes in overall health (Analysis 2.17; Analysis 2.18) or drop-out (Analysis 2.21).

When exercise therapy was compared with pacing, fatigue (Analysis 3.1), physical functioning (Analysis 3.3), depression (Analysis 3.4), sleep (Analysis 3.6) and self-perceived changes in overall health at end of treatment (Analysis 3.7) were significantly better. Data on SARs were available from only one trial, and SARs were rare, but events were too few to allow any conclusions to be drawn (Analysis 3.2). For anxiety, little or no difference between groups was reported (Analysis 3.5).

Overall completeness and applicability of evidence

The evidence base was limited to patients able to participate in exercise therapy, and all studies were conducted in developed countries (Australia, New Zealand, North America and the United Kingdom). Settings varied from primary to tertiary care, which suggests easy generalisation. Most of the outcomes investigated were reported in the included studies, apart from health service resources. Most studies used aerobic exercise, but it would be preferable if we had found studies that used different types of exercise therapy, as this would reflect clinical practice.

Quality of the evidence

Risk of bias across studies was relatively low. We were able to identify pre-published protocols for only two studies (Wearden 2010; White 2011) and have identified a risk of unpublished outcomes. One limitation is that formal blinding of participants and clinicians to treatment arm is not inherently possible in trials of exercise therapy. This increases risk of bias, as instructors' and participants' knowledge of group assignation might have influenced the true effect. In addition, outcomes were measured subjectively (e.g. questionnaires, visual analogue scales), leading to risk that this might increase the outcome estimate. Against this, many patient charities are opposed to exercise therapy for chronic fatigue syndrome (CFS), and this may in contrast reduce the effect. Six of the seven studies reported that investigators used intention-to-treat analysis, but this was done in different ways, which might have influenced the effect estimate. One study (Jason 2007) reported baseline differences, used a best linear unbiased predictor to avoid taking missing data into account and described 25 outcomes, with none stated as primary.

Several methodological challenges have become evident during the review process. An obvious topic of discussion is the between-study variation observed with regard to type of exercise, intensity of exercise and incremental procedures used (Table 2). We acknowledge that an effect of exercise therapy is likely to depend on how training is conducted, thus inclusion of trials using different exercise regimens is likely to introduce some heterogeneity into the analysis. Possibly equally important, the treatment provided to participants in the control group was not uniform across included

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

trials. Whereas the difference between waiting list, relaxation and treatment as usual is rather obvious, it is important to recognise that the actual ingredients of 'treatment as usual' differ widely among the included trials, and this may contribute to variation in observed effect estimates. With regard to participants and their health status, it is important to realise that substantial differences in baseline illness severity were noted, as illustrated by the wide range in baseline physical functioning, depression co-morbidity and illness duration shown in Table 1. Some trials applied narrow selection criteria, whereas others seem to have included more heterogeneous sample populations; these differences might cause variation in the observed effect estimate. Our finding of similar outcomes with different definitions of CFS mitigates this risk.

All potential sources of heterogeneity mentioned above could have contributed to variation in results derived from the aggregate analysis presented in the present review and might have reduced our ability to draw firm conclusions. It is easy to imagine a potential correlation between observed treatment effect and factors such as exercise characteristics, control conditions, participant recruitment strategies, participant characteristics and baseline differences. We aimed to explore these associations in subgroup analyses. However, the number of potential heterogeneity factors is high and the number of available trials is low; therefore we were limited in our ability to explore heterogeneity in a sensible way at the aggregate level.

Potential biases in the review process

The strength of this review lies in its rigorous methods, which include thorough searching for evidence, systematic appraisal of study quality and systematic and well-defined data synthesis. Even though we tried to search as extensively as possible, we may have missed out on eligible trials, such as trials reported only in dissertations or in non-indexed journals.

The table of interventions (Table 2) includes published and unpublished information regarding types of interventions, but not effect estimates. For this updated review, we have not collected unpublished data for our outcomes but have used data from the 2004 review (Edmonds 2004) and from published versions of included articles.

The authors of this review had to make a cutoff regarding what kind of exercise should be included. We decided to exclude traditional Chinese exercise such as Tai Chi and Qigong, but to include pragmatic rehabilitation for which the type of exercise is described as walking, walking stairs, bicycling, dancing or jogging. The cutoff might be contentious, and discussion regarding what type of exercise should be included should be ongoing.

One of the included studies (Powell 2001) is an outlier, reporting very positive results in favour of exercise therapy; we decided post hoc to perform a sensitivity analysis from which Powell 2001 was excluded to learn what the results would be if this study was not included.

Review authors noted potential bias regarding how the comparators in this review were categorised and pooled. We decided to report diverse comparators such as cognitive-behavioural therapy (CBT), cognitive therapy treatment (COG) and supportive therapy together as a single comparator called 'psychological treatments' (however, because of clinical and contextual heterogeneity, we decided not to pool the results in meta-analyses). These different psychological treatments do have similar elements, for example, both CBT and COG use cognitive approaches and goal setting; however they differ in certain respects (e.g. CBT tries to change unhelpful thoughts, while COG aims to accept them (Jason 2007)). Our approach of combining these comparators might be considered contentious, and discussion about what should be lumped together and what should be split into different comparators should be ongoing.

Meta-analysis of individual patient data (IPD) constitutes an alternative approach to meta-analysis of aggregate data. Analysis based on individual patient data in general will enable us to use a wider range of statistical and analytical approaches (Higgins 2011). In particular, by utilising IPD, it is possible to explore the relative importance of the various heterogeneity factors mentioned above more thoroughly, and to ensure that missing data and baseline differences are dealt with in standardised ways. With access to IPD, it is also possible to perform subgroup analyses that have not been previously reported. A project aimed at undertaking IPD analyses of the trials included in the present review has been initiated, and when the IPD analyses are presented, they are likely to shed some new light on the aggregate level analyses presented in the current systematic review.

Agreements and disagreements with other studies or reviews

This review is an updated version of a review that was originally published in 2004 (Edmonds 2004); the revised version offers major additions and changes. According to recent updates provided in the *Cochrane Handbook for Systematic Reviews of Interventions*, we have implemented several methodological improvements, including a thorough risk of bias assessment for all included studies (Higgins 2011). Also, the updated search for literature led to the inclusion of three new trials with a total of 1051 participants (Jason 2007; Wearden 2010; White 2011), thus the number of included participants has more than tripled since the 2004 version. The inclusion of new trials has important implications. First, statistical power has been increased by the addition of new data. Second, the most recent trials offered longer follow-up times; therefore we can provide more clear conclusions about follow-up treatment effects in this update than were provided in the original review. Third, the most recent trials involve comparisons beyond exercise therapy versus treatment as usual, for example, comparisons of exercise therapy versus other active treatment strategies such as CBT and adaptive pacing therapy.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

This update provides valuable additional information when compared with the original review, and results reported in the original review are largely confirmed in this update. Moreover, the results reported here correspond well with those of other systematic reviews (Bagnall 2002; Larun 2011; Prins 2006) and with existing guidelines (NICE 2007). One meta-analysis of CBT and GET suggests that the two treatments are equally efficacious, especially for patients with co-morbid anxiety or depressive symptoms (Castell 2011).

A recent randomised trial comparing quality of life among participants randomly assigned to group CBT plus graded exercise therapy plus conventional pharmacological treatment or exercise counselling plus conventional pharmacological treatment found no differences between the two groups at 12 months' follow-up (Nunez 2011). This trial did not meet our a priori inclusion criteria and was excluded from our review. As the comparison used in Nunez 2011 differs from the comparisons reported in our review, it is difficult to compare the results directly; this comparison was complicated further by the fact that Nunez 2011 did not measure outcomes viewed as primary outcomes in our review. Consequently, our view is that the conclusions presented in our review correspond well with those of other relevant studies and reviews, but further research is needed to explore the considerable heterogeneity observed across available trials.

AUTHORS' CONCLUSIONS

Implications for practice

Encouraging evidence suggests that exercise therapy can contribute to alleviation of some symptoms of CFS, especially fatigue. Exercise therapy seems to perform better than no intervention or pacing and seems to lead to results similar to those seen with cognitive behavioural therapy. Reported results were obtained from patients who were able to participate (not from those too disabled to attend clinics); these results were inconclusive as to type of exercise therapy and showed heterogeneity. Few serious adverse reactions were reported. We think the evidence suggests that exercise therapy might be an effective and safe intervention for patients able to attend clinics as outpatients.

Implications for research

Further randomised controlled studies are needed to clarify the most effective type, intensity and duration of exercise therapy. These studies should report contextual characteristics of the exercise therapy provided, such as deliverer of the intervention, schedule, explanation and materials, supervision and monitoring. It is important that these trials measure health service use alongside the primary outcomes of fatigue and adverse effects, as well as alongside relevant secondary outcomes. Researchers should take care to describe which set of diagnostic criteria they have used and how they operationalised the diagnostic process.

ACKNOWLEDGEMENTS

We would like to thank Peter White and Paul Glasziou for advice and additional information provided. We would also like to thank Kathy Fulcher, Richard Bentall, Alison Wearden, Karen Wallman and Rona Moss-Morris for providing additional information from trials in which they were involved, as well as the CCDAN editorial base for providing support and advice and Sarah Dawson for conducting the searches. In addition, we would like to thank Jane Dennis, Ingvild Kirkehei, Hugh McGuire and Melissa Edmonds for their valuable contributions, and Elisabet Hafstad for assistance with the search.

REFERENCES

References to studies included in this review

Fulcher 1997 ***{published and unpublished data}***

Fulcher KY, White PD. Chronic fatigue syndrome: a description of graded exercise treatment. *Physiotherapy* 1998;**84**(9):223–6.

* Fulcher KY, White PD. Randomised controlled trial of graded exercise in patients with chronic fatigue syndrome. *BMJ* 1997;**314**(7095):1647–52.

White PD, Fulcher KY. A randomised controlled trial of graded exercise in patients with a chronic fatigue. Royal College of Psychiatrists Winter Meeting, Cardiff. 1997.

Jason 2007 ***{published data only}***

Hlavaty LE, Brown MM, Jason LA. The effect of homework compliance on treatment outcomes for participants with myalgic encephalomyelitis/chronic fatigue syndrome. *Rehabilitation Psychology* 2011;**56**(3):212–8.

* Jason L, Torres-Harding S, Friedberg F, Corradi K, Njoku M Donalek J, et al. Non-pharmacologic interventions for CFS: a randomized trial. *Journal of Clinical Psychology in Medical Settings* 2007;**172**:485–90.

Moss-Morris 2005 ***{published data only (unpublished sought but not used)}***

* Moss-Morriss R, Sharon C, Tobin R, Baldi JC. A randomized controlled graded exercise trial for chronic fatigue syndrome: outcomes and mechanisms of change. *Journal of Health Psychology* 2005;**10**(2):245–59.

Powell 2001 ***{published and unpublished data}***

Powell P, Bentall ROP, Nye FJ, Edwards RHT. Patient

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

education to encourage graded exercise in chronic fatigue syndrome: 2-year follow-up of randomised controlled trial. *The British Journal of Psychiatry* 2004;**184**:142–6.
* Powell P, Bentall RP, Nye FJ, Edwards RH. Randomised controlled trial of patient education to encourage graded exercise in chronic fatigue syndrome. *BMJ* 2001;**322** (7283):387–90.

Wallman 2004 *{published and unpublished data}*
Wallman KE, Morton AR, Goodman C, Grove R. Exercise prescription for individuals with chronic fatigue syndrome. *Medical Journal of Australia* 2005;**183**(3):142–3.
* Wallman KE, Morton AR, Goodman C, Grove R, Guilfoyle AM. Randomised controlled trial of graded exercise in chronic fatigue syndrome. *Medical Journal of Australia* 2004;**180**(9):444–8.

Wearden 1998 *{published and unpublished data}*
Appleby L. Aerobic exercise and fluoxetine in the treatment of chronic fatigue syndrome. National Research Register 1995.
Morriss R, Wearden A, Mullis R, Strickland P, Appleby L, Campbell I, et al. A double-blind placebo-controlled treatment trial of fluoxetine and graded exercise for chronic fatigue syndrome (CFS). 8th Congress of the Association of European Psychiatrists, London. 1996.
* Wearden AJ, Morriss RK, Mullis R, Strickland PL, Pearson DJ, Appleby L, et al. Randomised, double-blind, placebo-controlled treatment trial of fluoxetine and graded exercise for chronic fatigue syndrome. *British Journal of Psychiatry* 1998;**178**:485–92.

Wearden 2010 *{published and unpublished data}*
Wearden AJ. Randomised controlled trial of nurse-led self-help treatment for patients in primary care with chronic fatigue syndrome. The FINE trial (Fatigue Intervention by Nurses Evaluation) ISRCTN74156610, 2001. http://www.controlled-trials.com/ISRCTN74156610/ISRCTN74156610 (accessed 2 September 2014).
* Wearden AJ, Dowrick C, Chew-Graham C, Bentall RP, Morriss RK, Peters S, et al. Nurse led, home based self help treatment for patients in primary care with chronic fatigue syndrome: randomised controlled trial. *BMJ* 2010;**340** (1777):1–12. [DOI: 10.1136/bmj.c1777]
Wearden AJ, Dowrick C, Chew-Graham C, Bentall RP, Morriss RK, Peters S, et al. Nurse led, home based self help treatment for patients in primary care with chronic fatigue syndrome: randomised controlled trial. BMJ, rapid response 27 May 2010.
Wearden AJ, Riste L, Dowrick C, Chew-Graham C, Bentall RP, Morriss RK, et al. Fatigue interventions by nurses evaluation-The FINE Trial. A randomised controlled trial of nurse led self-help treatment for patients in primary care with chronic fatigue syndrome: study protocol (ISRCTN74156610). *BMC Medicine* 2006;**4**(9):1–12.

White 2011 *{published data only}*
McCrone P, Sharpe M, Chalder T, Knapp M, Johnson AL, Goldsmith KA, et al. Adaptive pacing, cognitive behaviour therapy, graded exercise, and specialist medical care for chronic fatigue syndrome: a cost-effectiveness analysis. *PLoS ONE* 2012;**7**(7):e40808. [DOI: 10.1371/journal.pone.0040808]
Sharpe MD, Goldsmith KA, Johnson AL, Chalder T, Walker J, White PD. Rehabilitative treatments for chronic fatigue syndrome: long-term follow-up from the PACE trial. *Lancet Psychiatry* 2015:ePub ahead of print. [DOI: 10.1016/S2215-0366(15)00317-X]
White P, Chalder T, McCrone P, Sharpe M. Non-pharmacological management of chronic fatigue syndrome: efficacy, cost effectiveness and economic outcomes in the PACE trial [conference abstract]. Journal of Psychosomatic Research. Proceedings of the 15th Annual Meeting of the European Association for Consultation-Liaison Psychiatry and Psychosomatics, EACLPP and 29th European Conference on Psychosomatic Research, ECPR; 2012 Jun 27-30; Aarhus Denmark. 2012; Vol. 72, issue 6:509.
White PD. A randomised controlled trial of adaptive pacing, cognitive behaviour therapy, and graded exercise, as supplements to standardised specialist medical care versus standardised specialist medical care alone for patients with the chronic fatigue syndrome/myalgic encephalomyelitis or encephalopathy [PACE], 2014. http://www.controlled-trials.com/ISRCTN54285094 (accessed 1 September 2014).
* White PD, Goldsmith KA, Johnson AL, Potts L, Walwyn R, DeCesare JC, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. *The Lancet* 2011; **377**:611–90.
White PD, Goldsmith KA, Johnson AL, et al. on behalf of the PACE Trial Management Group. Supplementary web appendix. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. *The Lancet* 2011;**377**:832–6. [DOI: 10.1016/S0140-6736(11)60096-2]
White PD, Sharpe MC, Chalder T, DeCesare JC, Walwyn R, the PACE Trial Group. Protocol for the PACE trial. A randomised controlled trial of adaptive pacing, cognitive behaviour therapy, and graded exercise as supplements to standardised specialist medical care versus standardised specialist medical care alone for patients with the chronic fatigue syndrome/myalgic encephalomyelitis or encephalopathy. *BMC Neurology* 2007;**7**(6):1–20. [DOI: 10.1186/1471-2377-7-6]

References to studies excluded from this review

Evering 2008 *{unpublished data only}*
Evering RMH. Ambulatory feedback at daily physical activity patterns. A treatment for the chronic fatigue syndrome in the home environment?. Universitet Twente, Netherlands 2013:1–223.
Evering RMH. Optimalization of cognitive behavioral therapy (CBT) for CFS patients in rehabilitation by means of ambulatory activity-based feedback (ABF). trialregister.nl/

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

trialreg/admin/rctview.asp?TC=1513 (accessed 7 May 2013).

Gordon 2010 *{published data only}*

* Gordon BA, Knapman LM, Lubitz L. Graduated exercise training and progressive resistance training in adolescents with chronic fatigue syndrome: a randomized controlled pilot study. *Clinical Rehabilitation* 2010;**24**:1072–9. [DOI: 10.1177/0269215510371429]

Guarino 2001 *{published data only}*

Guarino P, Peduzzi P, Donta ST, Engel CC Jr, Clauw DJ, Williams DA, et al. A multicenter two by two factorial trial of cognitive behavioral therapy and aerobic exercise for gulf war veterans' illnesses: design of a Veterans Affairs cooperative study (CSP #470). *Controlled Clinical Trials* 2001;**22**:31032.

Nunez 2011 *{published data only}*

Nunez M, Fernandez Soles J, Nunez E, Fernandez Huerta JM, Godas Sieso T, Gomez Gil E. Health-related quality of life in patients with chronic fatigue syndrome: group cognitive behavioural therapy and graded exercise versus usual treatment. A randomised controlled trial with 1 year of follow-up. *Clinical Rheumatology* 2011;**30**(3):381–9.

Ridsdale 2004 *{published data only}*

* Risdale L, Darbishire L, Seed T. Is graded exercise better than cognitive behaviour therapy for fatigue? A UK randomized trial in primary care. *Psychological Medicine* 2003;**34**:37–49.

Ridsdale 2012 *{published data only}*

Ridsdale L, Hurley M, King M, McCrone P, Dobalson N. The effect of counselling, graded exercise and usual care for people with chronic fatigue in primary care: a randomized trial. *Psychological Medicine* 2012;**42**:2217–24. [DOI: 10.1017/S0033291712000256]

Sabes-Figuera R, McCrone P, Hurley M, King M, Donaldson AN, Risdale L. Cost-effectiveness of counselling, graded-exercise and usual care for chronic fatigue: evidence from a randomised trial in primary care. *BMC Health Services Reserach* 2012;**12**:264.

Russel 2001 *{unpublished data only}*

Russel V, Gaston AM, Lewin RJP. Atkinson CM, Champion PD. Group rehabilitation for adult chronic fatigue syndrome. Unpublished article 2001.

Stevens 1999 *{published data only}*

Stevens MW. *Chronic Fatigue Syndrome: A Chronobiologically Oriented Controlled Treatment Outcome Study*. San Diego: California School of Professional Psychology, 1999. UMI 9928180]

Taylor 2004 *{published data only}*

Taylor RR. Quality of life and symptom severity for individuals with chronic fatigue syndrome: findings from a randomized clinical trial. *American Journal of Occupational Therapy* 2004;**58**:35–43.

Taylor 2006 *{published data only}*

Taylor RR, Jason LA, Shiraishi Y, Schoeny ME, Keller J. Conservation of resources theory, perceived stress, and chronic fatigue syndrome: outcomes of a consumer-driven rehabilitation program. *Rehabilitation Psychology* 2006;**51**: 157–65.

Taylor RR, Thanawala SG, Shiraishi Y, Schoeny ME. Long-term outcomes of an integrative rehabilitation program on quality of life: a follow-up study. *Journal of Psychosomatic Research* 2006;**61**:835–9.

Thomas 2008 *{published data only}*

Thomas M, Sadlier M, Smith A. A multiconvergent approach to the rehabilitation of patients with chronic fatigue syndrome: a comparative study. *Physiotherapy* 2008; **94**(1):35–42.

Thomas MA, Sadlier MJ, Smith AP. The effect of multi convergent therapy on the psychopathology, mood and performance of chronic fatigue syndrome patients: a preliminary study. *Counselling and Psychotherapy Research* 2006;**6**:91–9.

Tummers 2012 *{published data only}*

Tummers M, Knoop H, van Dam A, Bleijenberg G. Implementing a minimal intervention for chronic fatigue syndrome in a mental health centre: a randomized controlled trial. *Psychological Medicine* 2012;**42**:2205–15. [DOI: 10.1017/S0033291712000232]

Viner 2004 *{published data only}*

Viner R, Gregorowski A, Wine C, Bladen M, Fisher D, Miller M, et al. Outpatient rehabilitative treatment of chronic fatigue syndrome (CFS/ME). *Archives of Disease in Childhood* 2004;**89**(7):615–9. [DOI: 10.1136/adc.2003.035154]

Wright 2005 *{published data only}*

Wright B, Ashby B, Beverley D, Calvert E, Jordan J, Miles J, et al. A feasibility study comparing two treatment approaches for chronic fatigue syndrome in adolescents. *Archives of Disease in Childhood* 2005;**90**(4):369–72. [DOI: 10.1136/adc.2003.046649]

References to studies awaiting assessment

Hatcher 1998 *{unpublished data only}*

Hatcher S. A randomised double-blind placebo controlled trial of dothiepin and graded activity in the treatment of chronic fatigue syndrome. *Personal communication,* 1998.

Liu 2010 *{published data only}*

Liu CZ, Lei B. Effect of Tuina on oxygen free radicals metabolism in patients with chronic fatigue syndrome [Chinese]. *Zhongguo Zhenjiu* 2010;**11**:946–8.

Zhuo 2007 *{published data only}*

Zhuo J-X, Gu L-Y. Relative research on treating chronic fatigue syndrome with gradual exercise. *Journal of Beijing Sport University* 2007;**30**(6):801–3.

References to ongoing studies

Broadbent 2012 *{unpublished data only}*

Broadbent S, Coutts R. The protocol for a randomised controlled trial comparing intermittent and graded exercise

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

to usual care for chronic fatigue syndrome patients. *BMC Sports Science, Medicine & Rehabilitation* 2013;**5**(1):1–6.
Broadbent, S. A pilot study on the effects of intermittent and graded exercise compared to no exercise for optimising health and reducing symptoms in chronic fatigue syndrome (CFS) patients. anzctr.org.au/Trial/Registration/TrialReview.aspx?ACTRN=12612001241820 (accessed 7 May 2013).

Kos 2012 ***{unpublished data only}***

* Kos D, Nijs J. Pacing activity self-management for patients with chronic fatigue syndrome: randomized controlled clinical trial, 2012. clinicaltrials.gov/show/NCT01512342 (accessed 7 May 2013).

Marques 2012 ***{unpublished data only}***

Marques M, De Gucht V, Maes S, Leal I. Protocol for the "four steps to control your fatigue (4-STEPS)" randomised controlled trial: a self-regulation based physical activity intervention for patients with unexplained chronic fatigue. *BMC Public Health* 2012;**12**:202. [DOI: 10.1186/1471-2458-12-202]

Vos-Vromans 2008 ***{unpublished data only}***

* Vos-Vromans D. Is a multidisciplinary rehabilitation treatment more effective than mono disciplinary cognitive behavioural therapy for patients with chronic fatigue syndrome? A multi centre randomised controlled trial [FatiGo, ISRCTN77567702]. http://www.controlled-trials.com/isrctn/pf/77567702 (accessed 7 May 2013). ISRCTN77567702]
Vos-Vromans DCWM, Smeets RJEM, Rijnders LJM, Gorrissen RRM, Pont M, Köke AJA, et al. Cognitive behavioural therapy versus multidisciplinary rehabilitation treatment for patients with chronic fatigue syndrome: study protocol for a randomized controlled trial (FatiGo). *Trials [electronic resource]* 2012;**13**:71.

White 2012 ***{published data only}***

White PD. Therapy guided self-help treatment (GETSET) for patients with chronic fatigue syndrome/myalgic encephalomyelitis: a randomised controlled trial in secondary care. ISRCTN22975026, 2012. http://www.controlled-trials.com/ISRCTN22975026/GETSET (accessed 30 October 2014).

Additional references

ACSM 2001

American College of Sports Medicine. *ACSM's Resource Manual for Guidelines for Exercise Testing and Prescription*. 4th Edition. Baltimore, MD: Lippincott Williams & Wilkins, 2001.

Adams 2009

Adams D, Wu T, Yang X, Tai S, Vohra S. Traditional Chinese medicinal herbs for the treatment of idiopathic chronic fatigue and chronic fatigue syndrome. *Cochrane Database of Systematic Reviews* 2009;**4**:1–16. [DOI: 10.1002/14651858.CD006348.pub2]

Alderson 2004

Alderson P, Green S, Higgins JP, editors. Cochrane Reviewers' Handbook 4.2.2 [updated December 2003]. *The Cochrane Library, Issue 1*. Chichester, UK: John Wiley & Sons Ltd, 2004.

Bagnall 2002

Bagnall AM, Whiting P, Richardson R, Sowden AJ. Interventions for the treatment and management of chronic fatigue syndrome/myalgic encephalomyelitis. *Quality & Safety in Health Care* 2001;**11**(3):284–8.

Beck 1996

Beck AT, Steer RA, Brown GK. Manual for the Beck Depression Inventory-II. *Manual for the Beck Depression Inventory-II*. San Antonio: Psychological Cooperation, 1996.

Blair 2009

Blair SN, Morris JN. Healthy hearts-and the universal benefits of being physically active: physical activity and health. *Annals of Epidemiology* 2009;**19**(4):253–6.

Bland 1997

Bland JM, Kerry SM. Statistics notes. Trials randomised in clusters. *BMJ* 1997;**315**:600.

Burckhardt 2003

Burckhardt CS, Anderson KL. The Quality of Life Scale (QOLS): reliability, validity and utilization. *Health and Quality of Life Outcomes* 2003;**1**:60.

Buysse 1989

Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. *Psyciatric Research* 1989;**28**:193–213.

Carruthers 2011

Carruthers BM, an de Sande MI, Meirleir KL, Klimas NG, Broderick G, Mitchell T, et al. Myalgic encephalomyelitis: international consensus criteria. *Journal of Internal Medicine* 2011;**270**(4):327–38.

Castell 2011

Castell BD, Kazantzis N, Moss-Morris RE. Cognitive behavioral therapy and graded exercise for chronic fatigue syndrome: a meta-analysis. *Clinical Psychology: Science and Practice* 2011;**18**:311–24.

Chalder 1993

Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, et al. Development of a fatigue scale. *Journal of Psychosomatic Research* 1993;**37**(6):147–53.

Clark 2005

Clark LV, White PD. The role of deconditioning and therapeutic exercise in chronic fatigue syndrome (CFS). *Journal of Mental Health* 2005;**14**(3):237–52.

Cleeland 1994

Cleeland CS, Ryan KM. Pain assessment: the global use of the Brief Pain Inventory. *Annals Academy of Medicine Singapore* 1994;**23**:123–38.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Edmonds 2004
Edmonds M, McGuire H, Price J. Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews* 2004;**3**(3):1–28. [DOI: 10.1002/14651858.CD003200.pub2]

Egger 1997
Egger M, Davey-Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal* 1997;**315**:629–34.

Elbourne 2002
Elbourne DR, Altman DG, Higgins JP, Curtin F, Worthington HV, Vail A. Meta-analyses involving cross-over trials: methodological issues. *International Journal of Epidemiology* 2002;**31**:140–9.

European Union Clinical Trials Directive 2001
The Europien Parliament and the Councel of the European Union. DIRECTIVE 2001/20/EC Europien Parliament and the Councel of the European Union of 4 April 2001. Official Journal of the European Communities 2001; Vol. L 121/34. [http://www.eortc.be/services/doc/clinical–eu–directive–04–april–01.pdf]

Fonhus 2011
Fønhus MS, Larun L, Brurberg KG. Diagnostic criteria for chronic fatigue syndrome [Diagnosekriterier for kronisk utmattelsessyndrom. Notat fra Kunnskapssenteret 2011]. Norwegian Knowledge Centre for the Health Services 2011.

Fukuda 1994
Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. The chronic fatigue syndrome: a comprehensive approach to its definition and study. *Annals of Internal Medicine* 1994;**121**(12):953–9.

Fulcher 2000
Fulcher KY, White PD. Strength and physiological response to exercise in patients with chronic fatigue syndrome. *Journal of Neurology Neurosurgery & Psychiatry* 2000;**69**:302–7.

Gulliford 1999
Gulliford MC, Ukoumunne OC, Chinn S. Components of variance and intraclass correlations for the design of community-based surveys and intervention studies: data from the Health Survey for England 1994. *American Journal of Epidemiology* 1999;**149**:924–6.

Guy 1976
Guy W. ECDEU assessment manual for psychopharmacology. *ECDEU Assessment Manual for Psychopharmacology*. Rockville, MD: National Institute of Mental Health, 1976:218–222.

Hard 2009
Hard K, Rickards HE, Haque MS, Ward C. Pharmacological treatments for chronic fatigue syndrome in adults. *Cochrane Database of Systematic Reviews* 2007, Issue 4. [DOI: 10.1002/14651858.CD006788.pub2]

Hewitt 1993
Hewitt PL, Norton GR. The Beck Anxiety Inventory: a psychometric analysis. *Psychological Assessment* 1993;**5**:408–12.

Higgins 2003
Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;**327**(7414):557–60.

Higgins 2011
Higgins JPT, Altman DG, Sterne JAC (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated February 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org. The Cochrane Collaboration.

Jenkins 1988
Jenkins D, Stanton B, Niemcryk S, Rose R. A scale for the estimation of sleep problems in clinical research. *Journal of Clinical Epidemiology* 1988;**41**:313–21.

Johnston 2013
Johnston S, Brenu EW, Staines D, Marshall-Gradisnik S. The prevalence of chronic fatigue syndrome/ myalgic encephalomyelitis: a meta-analysis. *Clinical Epidemiology* 2013;**5**:105–10.

Krupp 1989
Krupp LB, LaRocca NG, Muir-Nash J, Steinberg AD. The fatigue severity scale: application to patients with multiple sclerosis and systemic lupus erythematosus. *Archives of Neurology* 1989;**46**:1121–3.

Larun 2011
Larun L, Malterud K. Exercise therapy for patients with chronic fatigue syndrome [Treningsbehandling ved kronisk utmattelsessyndom]. *Tidsskr Nor Laegeforen* 2011;**138**(8):231–6.

Larun 2014
Larun L, Odgaard-Jensen J, Brurberg KG, Chalder T, Dybwad M, Moss-Morris RE, et al. Exercise therapy for chronic fatigue syndrome (individual patient data). *Cochrane Database of Systematic Reviews* 2014, Issue 4. [DOI: 10.1002/14651858.CD011040]

Moncrieff 2001
Moncrieff J, Churchill R, Drummond C, McGuire H. Development of a quality assessment instrument for trials of treatments for depression and neurosis. *International Journal of Methods in Psychiatric Research* 2001;**10**(3):126–33.

Mosby 2009
Mosby. *Mosby's Medical Dictionary*. 8th Edition. Philadelphia: Elsevier, 2009.

NICE 2007
National Institute for Health and Clinical Excellence. Chronic fatigue syndrome/myalgic encephalomyelitis (or encephalopathy): diagnosis and management of CFS/ME in adults and children, 2007. http://guidance.nice.org.uk/CG53/guidance/pdf/English (last accessed November 2009). London: National Institute for Health and Clinical Excellence.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Nijs 2011
Nijs J, Meeus M, Van Oosterwijck J, Ickmans K, Moorkens G, Hans G, et al. In the mind or the brain? Scientific evidence for central sensitisation in chronic fatigue syndrome. *European Journal of Clinical Investigation* 2011; **42**:203–11. [DOI: 10.1111/j.1365-2362.2011.02575.x]

Oxford English Dictionary
OED Online. December 2014. Oxford University Press. "therapy, n.". http://www.oed.com/view/Entry/200468?redirectedFrom=therapy (accessed January 21, 2015).

Paul 2001
Paul LM, Wood L, Maclaren W. The effect of exercise on gait and balance in patients with chronic fatigue syndrome. *Gait and Posture* 2001;**14**:19–27.

Price 2008
Price JR, Mitchell E, Tidy E, Hunot V. Cognitive behaviour therapy for chronic fatigue syndrome in adults. *Cochrane Database of Systematic Reviews* 2008, Issue 3. [DOI: 10.1002/14651858.CD001027.pub2]

Prins 2006
Prins JB, van den Meer JW, Bleijenberg G. Chronic fatigue syndrome. *Lancet* 2006;**367**:346–55.

Reeves 2003
Reeves WC, Lloyd A, Vernon SD, Klimas N, Jason LA, Bleijenberg G, and the International Chronic Fatigue Syndrome Study Group. Identification of ambiguities in the 1994 chronic fatigue syndrome research case definition and recommendations for resolution. *BMC Health Services Research* 2003;**3**(25):1–9.

Reeves 2007
Reeves WC, Jones JF, Heim C, Hoaglin DC, Boneva RS, Mirrissey M, et al. Prevalence of chronic fatigue syndrome in metropolitan, urban, and rural Georgia. *Population Health Metrics* 2007;**5**:1–10.

Reyes 2003
Reyes M, Nisenbaum R, Hoaglin DC, Unger ER, Emmons C, Randall B, et al. Prevalence and incidence of chronic fatigue syndrome in Wichita, Kansas. *Archives of Internal Medicine* 2003;**163**(13):1530–6.

Sharpe 1991
Sharpe M, Archard L, Banatvala J, Borysiewicz LK, Clare AW, David A, et al. Chronic fatigue syndrome: guidelines for research. *Journal of the Royal Society of Medicine* 1991;**84** (2):118–21.

The National Task Force on CFS
The National Task Force on Chronic Fatigue Syndrome. Report from the National Task Force on Chronic Fatigue Syndrome (CFS), Post Viral Fatigue Syndrome (PVFS), Myalgic Encephalomyelitis (ME). Appendix B. Bristol: Westcare, 1994.

Wallman 2009
Wallman K. Confirmation of ages (means and SDs) of groups in trial. Personal correspondence (email to Lillebeth Larun), 2009 2 November.

Ware 1992
Ware JE, Sherbourne CD. The MOS 36-item short form health survey (SF-36). *Medical Care* 1992;**30**:473–83.

Wearden 2009
Wearden AJ. Raw data to facilitate calculations for meta-analysis. Personal communication (email), 2009 March.

Zigmond 1983
Zigmond AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatrica Scandinavica* 1983;**67** (6):361–70.

* *Indicates the major publication for the study*

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Fulcher 1997

Methods	RCT, 2 parallel arms
Participants	Diagnostic criteria: Oxford Number of participants: N = 66 Gender: 49 (65%) female Age, mean (SD): 37.2 (10.7) years Earlier treatment: NS Co-morbidity: 20 (30%) possible cases of depression (HADS): 30 (45%) on full-dose antidepressant (n = 20) or low-dose tricyclic antidepressants as hypnotics (n = 10) Average illness duration: 2.7 (0.6 to 19) years Work and employment status: 26 (395) working or studying at least part time Setting: secondary care (chronic fatigue clinic in a general hospital of psychiatry) Country: UK
Interventions	Group 1: exercise therapy (12 sessions) with 1 weekly supervised session and 5 home sessions a week, initially lasting between 5 and 15 minutes (n = 33) Group 2: flexibility and relaxation (12 sessions) with 5 home sessions prescribed per week (n = 33)
Outcomes	• Changes in overall health (Global Impression Scale, score between 1 and 7, where 1 = very much better, 4 = no change) • Anxiety and depression (Hospital Anxiety and Depression Scale, HADS) • Fatigue (Fatigue Scale, FS; 14-item questionnaire) • Sleep (Pittsburgh Sleep Quality Index, PSQI) • Physical functioning (Short Form (SF)-36) • Physiological assessments (maximal voluntary contraction of quadriceps, peak oxygen consumption, lactate, heart rate) • Perceived exertion (Borg Scale) Outcomes were assessed at end of treatment (12 weeks)
Notes	No long-term follow-up, as participants who completed the flexibility programme were invited to cross over to the exercise programme afterwards

Risk of bias ***Risk of bias***

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	Quote: "determined by random number tables"
Allocation concealment (selection bias)	Low risk	Quote: "Randomisation was achieved blindly to the psychiatrist and independently of the exercise physiologist by placing the letter E or F in 66 separate blank envelopes. These were then arranged in random order determined by random number tables

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Fulcher 1997 (*Continued*)

		and opened by an independent administrator after baseline tests as each new patient entered the study"
Blinding (performance bias and detection bias) of participants and personnel?	High risk	Not possible to blind participants or personnel (supervisors) to treatment allocation
Blinding (performance bias and detection bias) of outcome assessors?	High risk	Blinding not possible for self-reported measurements (e.g. FS, SF-36)
Incomplete outcome data (attrition bias) All outcomes	Low risk	Quote: "We completed follow up assessments on four of the seven patients who dropped out of treatment and included these data in the intention to treat analysis. Patients with missing data were counted as nonimprovers"
Selective reporting (reporting bias)	Unclear risk	All primary outcomes stated under Methods were reported; however, as the trial protocol is not available, we cannot categorically state that the review is free of selective outcome reporting
Other bias	Low risk	We do not suspect other bias

Jason 2007

Methods	RCT, 4 parallel arms
Participants	Diagnostic criteria: CDC 1994 Number of participants: N = 114 Gender: 95 (83.3%) female Age: 43.8 years Earlier treatment: NS Co-morbidity: 44 (39%) with a current Axis I disorder (depression and anxiety most common). Use of antidepressant not stated Illness duration: > 5 years Work and employment status: 52 (46%) working or studying at least part time, 24% unemployed, 6% retired, 25% on disability Setting: secondary care, but recruitment from different sources Country: USA
Interventions	13 sessions every 2 weeks lasting 45 minutes Group 1: cognitive-behavioural therapy (CBT) aimed at showing participants that activity could be done without exacerbating symptoms (n = 29) Group 2: anaerobic activity therapy (ACT) focused on developing individualised and pleasurable activities accompanied by reinforcement of progress (n = 29) Group 3: cognitive therapy treatment(COG) focused on developing strategies to better tolerance, reduce stress and symptoms and lessen self-criticism (n = 28) Group 4: relaxation treatment (RELAX) introducing several types of relaxation techniques along with expectations of skill practice (n = 28)

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Jason 2007 (*Continued*)

Outcomes	Several outcomes are reported (~25), among others. • Physical functioning (SF-36) • Fatigue (Fatigue Severity Scale, FSS) • Depression (Back Depression Inventory, BDI-II) • Anxiety (Beck Anxiety Inventory, BAI) • Self-efficacy (self-efficacy questionnaire) • Stress (Perceived Stress Scale, PSS) • Pain (Brief Pain Inventory) • Quality of life (Quality of Life Scale) • 6-Minute walking test Outcomes assessed at 12 months' follow-up
Notes	Fidelity ratings and drop-out reported across study arms

Risk of bias ***Risk of bias***

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Quote: "Random assignment was done using a random number generator in statistical software (SPSS version 12)"
Allocation concealment (selection bias)	Unclear risk	Not stated
Blinding (performance bias and detection bias) of participants and personnel?	High risk	Not possible to blind participants or personnel (supervisors) to treatment allocation
Blinding (performance bias and detection bias) of outcome assessors?	High risk	Blinding not possible for self-reported measurements (e.g. FSS, BPI)
Incomplete outcome data (attrition bias) All outcomes	High risk	Quote: "The average dropout rate was 25%, but it was not significantly different per condition." The statistical analysis used, the best linear unbiased predictor, is a way to avoid taking missing data into account
Selective reporting (reporting bias)	Unclear risk	All primary outcomes stated under Methods were reported; however, as the trial protocol is not available, we cannot categorically state that the review is free of selective outcome reporting
Other bias	High risk	Baseline data differences across groups for several important parameters (e.g. physical functioning: ACT group 39.17 (15.65) and RELAX group 53.77 (26.66))

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Moss-Morris 2005

Methods	RCT, 2 parallel arms
Participants	Diagnostic criteria: CDC 1994 Number of participants: N = 49 Gender: 34 (69%) female Age, mean (SD): 40.9 years: 36.7 (11.8) in treatment group and 45.5 (10.5) in control group Earlier treatment: NS Co-morbidity, mean (SD): 14 (29%) possible or probable cases of depression (HADS) . $HADS_{Anxiety}$ 6.72(3.44) in treatment group and 7.17 (3.43) in control group. $HADS_{Depression}$ 5.70 (2.69) in treatment group and 6.70 (0.67) in control group. Use of antidepressant not stated Illness duration, median (range): 3.1 years, 2.67 (0.6 to 20) in treatment group and 5 (0.5 to 45) in control group Work and employment status: 11 (22%) unemployed and unable to work because of disability Setting: specialist CFS general practice Country: New Zealand
Interventions	Group 1: graded exercise therapy (12 weeks), met weekly, final goal 30 minutes for 5 days a week, 70% of VO_2max (n = 25) Group 2: standard medical care provided by a CFS specialist physician (n = 24)
Outcomes	• Changes in overall health (Global Impression Scale, score between 1 and 7, where 1 = very much better, 4 = no change) • Physical function (SF-36 physical function subscale score) • Fatigue (Fatigue Scale, FS) • Activity levels • Cognitive function • Physiological assessments (e.g. maximum aerobic capacity, HR) • Acceptability Outcomes assessed at end of treatment (12 weeks). A self-report questionnaire was distributed at 6 months' follow-up and was returned by 16 exercise participants and 17 control participants
Notes	The exact components involved in 'treatment as usual' are not explained

Risk of bias ***Risk of bias***

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Quote: "...randomised into either treatment or control conditions by means of a sequence of computer generated numbers placed in sealed opaque envelopes by an independent administrator"
Allocation concealment (selection bias)	Low risk	Quote: "placed in sealed opaque envelopes by an independent administrator"

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Moss-Morris 2005 (*Continued*)

Blinding (performance bias and detection bias) of participants and personnel?	High risk	Not possible to blind participants or personnel (supervisors) to treatment allocation
Blinding (performance bias and detection bias) of outcome assessors?	High risk	Blinding not possible for self-reported measurements (e.g. FS, SF-36)
Incomplete outcome data (attrition bias) All outcomes	Low risk	3 of 25 participants (12%) dropped out from exercise treatment. Reasons for drop-out: 1 had to return to the USA, 1 had an injured calf and 1 was not reached at follow-up. 3 of 24 patients (12.5%) in control group did not return follow-up questionnaire at 12 weeks. To determine whether drop-out affected the calculated treatment effect, study authors completed intention-to-treat analysis
Selective reporting (reporting bias)	Unclear risk	All primary outcomes stated under Methods were reported; however, as the trial protocol is not available, we cannot categorically state that the review is free of selective outcome reporting
Other bias	Low risk	We do not suspect other bias

Powell 2001

Methods	RCT, 4 parallel arms
Participants	Diagnostic criteria: Oxford Number of participants: N = 148 Gender: 116 (78%) female Age, mean: 33 years Earlier treatment: NS Co-morbidity: 58 (39%) possible cases of depression (HADS), 27 (18%) used antidepressants Illness duration: 4.3 years Work and employment status: 50 (34%) working, 64 (43%) on disability Setting: secondary/tertiary care Country: UK
Interventions	Group 1: treatment as usual (n = 34) Group 2: exercise therapy + 2 sessions (total 3 hours, n = 37) Group 3: exercise therapy + 7 telephone sessions (total 3.5 hours, n = 39) Group 4: exercise therapy + 7 sessions (total 7 hours, n = 38) Sessions, whether telephone or face-to-face, were used to reiterate the treatment rationale and to discuss problems associated with graded exercise
Outcomes	• Physical functioning (SF-36, subscale physical functioning). Clinical improvement at 1 year predetermined as a score ≥ 25 or an increase from baseline of ≥ 10 on the physical functioning scale (score range, 10 to 30)

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

	• Fatigue (Fatigue Scale, FS; 11 items; scores > 3 indicate excessive fatigue) • Anxiety and depression (Hospital Anxiety and Depression Scale, HADS; score range from 0 to 21 worst) • Sleep (Jenkins Sleep Scale, 4 items; lower scores indicate better outcomes; score range 0 to 20 worst) • Changes in overall health (Global Impression Scale; score between 1 and 7, where 1 = very much better, 4 = no change) • Illness beliefs and experience of treatment (simple questionnaire) Outcomes assessed at 3 (end treatment), 6 and 12 months
Notes	Treatment as usual comprised a medical assessment, advice and an information booklet that encouraged graded activity and positive thinking but gave no explanations for symptoms SF-36 physical functioning subscale is reported on a 10 to 30 scale. We transformed scores from the 10 to 30 scale to the more common 0 to 100 scale by using the following formula: $mean_{new}$ = ($mean_{old}$ - 10) * 5 and SD_{new} = 5 * SD_{old}

Risk of bias ***Risk of bias***

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Quote: "Randomised into four groups by means of a sequence of computer generated random numbers...simple randomisation with stratification for scores on the hospital anxiety and depression scale, 15, using a cut off of 11 to indicate clinical depression"
Allocation concealment (selection bias)	Unclear risk	Quote: "...in sealed numbered envelopes"
Blinding (performance bias and detection bias) of participants and personnel?	High risk	Not possible for this intervention
Blinding (performance bias and detection bias) of outcome assessors?	High risk	Blinding not possible for self-reported measurements (e.g. FS, SF-36)
Incomplete outcome data (attrition bias) All outcomes	Low risk	Quote: "We used an intention to treat analysis. For patients who dropped out of treatment, the last values obtained were carried forward. Complete data were obtained for all patients who completed treatment except for three: two did not complete the questionnaire at three months and one did not complete the questionnaire at one year"
Selective reporting (reporting bias)	Unclear risk	All primary outcomes stated under Methods were reported; however, as the trial protocol is not available, we cannot categorically state that the review is free of selective outcome reporting

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Powell 2001 (*Continued*)

Other bias	Low risk	We do not suspect other bias

Wallman 2004

Methods	RCT, 2 parallel arms
Participants	Diagnostic criteria: CDC 1994 Number of participants: N = 68 Gender: 47 (77%) female Age: 16 to 74 years (average 43.3 (12.7) in the exercise group and 45.7 (12.5) in the control group) Earlier treatment: NS Co-morbidity: possible depression not stated, 16 (26%) used antidepressants Illness duration: no initial difference between groups Work and employment status: not stated Setting: primary care Country: Western Australia
Interventions	Group 1: prescribed exercise therapy, 12 weeks (n = 32) Group 2: flexibility and relaxation, 12 weeks (n = 29)
Outcomes	• Physiological assessments (heart rate, blood pressure at rest and during exercise, lactate and oxygen consumption) • Perceived exertion (Borg Scale, rating of perceived exertion (RPE)) • Energy expenditure (Older Adult Exercise Status Inventory) • Fatigue (Fatigue Scale, FS; 11 items) • Anxiety and depression (Hospital Anxiety and Depression Scale, HADS) • Cognitive function (computerised version of the modified Stroop Color Word Test) • Changes in overall health (Global Impression Scale, score between 1 and 7, where 1 = very much better, 4 = no change) Outcomes assessed at 12 weeks (end of treatment)
Notes	Supplementary HADS data obtained from study authors for first version of this review

Risk of bias ***Risk of bias***

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Unclear risk	Quote: "...patients were randomised (by an independent investigator)"
Allocation concealment (selection bias)	Unclear risk	Not adequately described
Blinding (performance bias and detection bias) of participants and personnel?	High risk	Not possible to blind participants or personnel (supervisors) to treatment allocation

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Wallman 2004 (*Continued*)

Blinding (performance bias and detection bias) of outcome assessors?	High risk	Blinding not possible for self-reported measurements (e.g. FS, SF-36)
Incomplete outcome data (attrition bias) All outcomes	Low risk	2 of 34 (6%) participants in the ET group withdrew: "...for reasons not associated with the study" 5 of 34 (15%) participants in control group withdrew: "for reasons not associated with the study, and a further subject was excluded because her body mass index (44 kg/m^2) prevented her form participating in the exercise test"
Selective reporting (reporting bias)	Unclear risk	All primary outcomes stated under Methods were reported; however, as the trial protocol is not available, we cannot categorically state that the review is free of selective outcome reporting
Other bias	Unclear risk	Baseline data differences between groups for anxiety (7.3 in exercise group vs 8.7 in control group) and mental fatigue (6.3 vs 5.6)

Wearden 1998

Methods	RCT, 4 parallel arms
Participants	Diagnostic criteria: Oxford Number of participants: N = 136 Gender: 97 (71%) female Age, mean (SD): 38.7 (10.8) years Earlier treatment: NS Co-morbidity: 46 (34%) with depressive disorder according to DSM-III-R criteria, use of antidepressant not stated Illness duration: duration of fatigue, median (IQR) 28.0 (39.5) months Work and employment status: 114 (84%) had recently changed occupation Setting: secondary/tertiary care Country: UK
Interventions	Group 1: graded exercise + fluoxetine (n = 33) Group 2: graded exercise + drug placebo, 26 weeks, preferred aerobic exercise 20 minutes at least 3 times per week, up to 75% of participants' functional maximum (n = 34) Group 3: exercise placebo + fluoxetine (n = 35) Group 4: exercise placebo + drug placebo, 26 weeks, offered no specific advice but participants told to do what they felt capable of and to rest when the felt they needed to (n = 34)
Outcomes	• Fatigue (Fatigue Scale, FS; 14 items; 4 or more were used as cutoff to designate caseness) • General health status (Medical Outcome Survey Short-Form Scales, MOS SF-36); measure of general health status on the following 6 scales (cutoff score for poor function in parentheses): physical function (< 83.3), role or occupational function (≤ 50), social

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Wearden 1998 (*Continued*)

	function (≤ 40), pain (≤ 50), health perception (≤ 70) and mental health (≤ 67) • Anxiety or depression (Hospital Anxiety and Depression Scale, HADS; cutoff of 11 or more designated cases) • Psychiatric diagnoses (Clinical Interview Schedule + supplementary questions by psychologist) • Physiological assessments (grip strength and functional work capacity) Outcomes assessed at weeks 12 and 26 (end of treatment)
Notes	Group 4 was used as treatment as usual, as participants were given no specific advice on exercise but were advised to exercise when they felt capable. Supplementary HADS data were obtained from study authors for the first version of this review

Risk of bias ***Risk of bias***

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Quote: "...randomised into a treatment group by computer generated numbers, with groups of 10 to obtain roughly equal numbers"
Allocation concealment (selection bias)	Low risk	Quote: "A list of subject numbers marked with the exercise group for each number was held by the physiotherapist. Pharmacy staff dispensed medication in accordance with the subject number assigned to each subject." The initial assessment was done independently: "All patients were medically assessed by a doctor.. .under the supervision of a consultant physician"
Blinding (performance bias and detection bias) of participants and personnel?	High risk	Quote: "The drug treatment was double blind. The placebo to fluoxetine was a capsule of similar taste and appearance. The placebo to the exercise programme was a review of activity diaries by the physiotherapists"
Blinding (performance bias and detection bias) of outcome assessors?	High risk	Blinding not possible for self-reported measurements (e.g. FS, SF-36)
Incomplete outcome data (attrition bias) All outcomes	High risk	Quote: "Analysis was carried out on an intention to treat basis. When there were missing data at 12 and 26 weeks, scores on the previous assessment were substituted. No data were available on 17 patients for the week 12 assessment, functional work capacity assessments at week 0, seven at week 12 and seven at week 26" Large drop-out rates in all intervention groups
Selective reporting (reporting bias)	High risk	It is clear (p 488) that investigators collected data for all six subscales of the MOS that they used (as well as measures for fatigue, depression and anxiety). Data from fatigue and depression (primary outcomes) are reported numerically. Data from the anxiety scale are said to show 'no significant changes' and are not reported numerically. This is also the case for 5 of the 6 subscales

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Wearden 1998 (*Continued*)

		of the MOS, with the exception of health perceptions, which is significant and favours the intervention group NB: Data for forced work capacity (fwc) were collected by investigators but are not reported in this review
Other bias	Low risk	We do not suspect other bias

Wearden 2010

Methods	RCT, 3 parallel arms
Participants	Diagnostic criteria: Oxford (31% fulfilled London ME criteria) Number of participants: N = 296 Gender: 230 (78%) female Age, mean (SD): 44.6 (11.4) years Earlier treatment: 264 (89%) reported medication during the past 6 months with antidepressant (n = 160) or analgesic (n = 79) Co-morbidity, N (%): 53 (18) had a depression diagnosis, 160 (54) were prescribed antidepressants the last 6 months Illness duration (M): 7 (range from 0.5 to 51.7) years Work and employment status: not stated Setting: primary care Country: UK
Interventions	Group 1: pragmatic rehabilitation, 10 sessions over an 18-week period; graded return to activity designed collaboratively by the participant and the therapist, also focusing on sleep patterns and relaxation exercises to address somatic symptoms of anxiety (n = 95) Group 2: supportive listening, 10 sessions over an 18-week period; listening therapy in which the therapist aims to provide an empathic and validating environment in which patients can freely discuss their prioritised concerns (n = 101) Group 3: general practitioner treatment as usual; GPs were asked to manage their cases as they saw fit, but to not refer participants for systematic psychological therapies for CFS/ME during the 18-week treatment period (n = 100)
Outcomes	• Physical functioning (SF-36 physical functioning subscale, percentage score in which higher scores indicate better outcomes) • Fatigue (Fatigue Scale, FS; 11 items; each item was scored dichotomously on a 4-point scale (0, 0, 1 or 1); total scores of 4 or more designated significant levels of fatigue. Lower scores indicated better outcomes) • Anxiety and depression (Hospital Anxiety and Depression Scale (HADS), depression and anxiety scale; lower scores indicate better outcomes) • Sleep (Jenkins Sleep Scale; 4 items; lower scores indicate better outcomes) Outcomes assessed at 20 weeks (end of treatment) and at 70 weeks (follow-up)
Notes	Economic evaluation of the relative cost-effectiveness of pragmatic rehabilitation and supportive listening when compared with treatment as usual, results of which will be reported separately

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Wearden 2010 (*Continued*)

Risk of bias		*Risk of bias*
Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Quote: "Individual patients were randomly allocated to one of the three treatment arms using computer generated randomised permuted blocks (with randomly varying block sizes of 9, 12, 15, and 18), after stratification on the basis of whether the patient was non-ambulatory (used a mobility aid on most days) and whether the patient fulfilled London ME criteria"
Allocation concealment (selection bias)	Low risk	Quote: "The random allocation was emailed to the trial manager, who assigned each patient a unique study number and notified the designated nurse therapist if the patient had been allocated to a therapy arm"
Blinding (performance bias and detection bias) of participants and personnel?	High risk	Not possible to blind participants or personnel (supervisors) to treatment allocation
Blinding (performance bias and detection bias) of outcome assessors?	High risk	Blinding not possible for self-reported measurements (e.g. FS, SF-36)
Incomplete outcome data (attrition bias) All outcomes	Unclear risk	Number of drop-outs (did not complete treatment): 18/95 (group 1), 17/101 (group 2). Reasons for drop-out: unhappy with randomisation (n = 8), lost contact (n = 8), too busy (n = 7), not benefiting or feeling worse (n = 5), nurse therapist safety concern (n = 2), misdiagnosis (n = 1), received different treatment (n = 1) Loss to follow-up at 20 weeks: 10/95 (group 1), 4/101 (group 2), 8/100 (group 3) Loss to follow-up at 70 weeks: 14/95 (group 1), 11/101 (group 2), 14/100 (group 3)
Selective reporting (reporting bias)	Low risk	All relevant outcomes are reported in accordance with the protocol
Other bias	Low risk	We do not suspect other types of bias

White 2011

Methods	RCT, multi-centre, 4 parallel arms
Participants	Diagnostic criteria: Oxford (56% satisfied London ME criteria) Number of participants: N = 641 Gender: 495 (77%) female Age, mean (SD): 38 (12) years

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

	Earlier treatment: NS Co-morbidity: 219 (34%) with any depressive disorder, 260 (41%) used antidepressants Illness duration: median 32 (IQR 16 to 68) months (GET 35 (18 to 67) and SMC 25 (15 to 57) months) Work and employment status: mean baseline score at the work and social adjustment scale, 27.4 Setting: secondary/tertiary care Country: UK
Interventions	Group 1, specialist medical care (SMC): provided by doctors with specialist experience in CFS. All participants were given a leaflet explaining the illness and the nature of this treatment. Treatment consisted of an explanation of chronic fatigue syndrome, generic advice such as to avoid extremes of activity and rest, specific advice on self-help according to the particular approach chosen by the participant (if receiving SMC alone) and symptomatic pharmacotherapy (especially for insomnia, pain and mood, n = 160) Group 2, adaptive pacing therapy (APT): based on the envelope theory aimed at optimum adaptation to the illness by helping the participant to plan and pace activity to reduce or avoid fatigue, achieve prioritised activities and provide the best conditions for natural recovery. Therapeutic strategies consisted of identifying links between activity and fatigue by using a daily diary, with corresponding encouragement to plan activity to avoid exacerbations, developing awareness of early warnings of exacerbation, limiting demands and stress, regularly planning rest and relaxation and alternating different types of activities, with advice not to undertake activities that demanded more than 70% of participants' perceived energy envelopes. Increased activities were encouraged if participants felt able, and as long as they did not exacerbate symptoms (n = 160) Group 3, cognitive-behavioural therapy (CBT): done on the basis of the fear avoidance theory of CFS. The aim of treatment was to change the behavioural and cognitive factors assumed to be responsible for perpetuation of participants' symptoms and disability. Therapeutic strategies guided participants to address unhelpful cognitions, including fears about symptoms or activities, by testing them through behavioural experiments. These experiments consisted of establishing a baseline of activity and rest and a regular sleep pattern, then making collaboratively planned gradual increases in both physical and mental activity. Participants were helped to address social and emotional obstacles to improvement through problem solving (n = 161) Group 4, graded exercise therapy (GET): done on the basis of deconditioning and exercise intolerance theories of chronic fatigue syndrome. The aim of treatment was to help participants gradually return to appropriate physical activities and reverse deconditioning, thereby reducing fatigue and disability. Therapeutic strategies consisted of establishment of a baseline of achievable exercise or physical activity, followed by a negotiated, incremental increase in the duration of time spent being physically active. Target heart rate ranges were set when necessary to avoid overexertion, which eventually aimed at 30 minutes of light exercise 5 times a week. When this rate was achieved, the intensity and aerobic nature of the exercise (usually walking) were gradually increased in response to participant feedback and with mutual planning (n = 160)
Outcomes	Primary outcomes • Fatigue (Fatigue Scale, FS; Likert scoring 0, 1, 2, 3; range 0 to 33; lowest score is least fatigue) • Physical function (Short Form-36 (SF-36) physical function subscale version 2;

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

	range 0 to 100; highest score is best function) • Safety outcomes (non-serious adverse events, serious adverse events, serious adverse reactions to trial treatments, serious deterioration and active withdrawals from treatment) • Adverse events (i.e. any clinical change, disease or disorder reported, whether or not related to treatment) Secondary outcomes • Changes in overall health (Global Impression Scale, score between 1 and 7, where 1 = very much better, 4 = no change) • Overall disability: work and social adjustment scale • 6-Minute walking test (distance in meters walked) • Sleep (Jenkins Sleep Scale score for disturbed sleep) • Anxiety and depression (Hospital Anxiety and Depression Scale, HADS) • Number of chronic fatigue syndrome symptoms (individual symptoms of postexertional malaise and poor concentration or memory) • Use of health service resources Outcomes assessed at 12 weeks, 24 weeks (end of treatment) and 52 weeks (follow-up)
Notes	

Risk of bias ***Risk of bias***

Bias	**Authors' judgement**	**Support for judgement**
Random sequence generation (selection bias)	Low risk	Quote: "Participants were allocated to treatment groups through the Mental Health and Neuroscience Clinical Trials Unit (London, UK) after baseline assessment and obtainment of consent. A database programmer undertook treatment allocation, independently of the trial team. The first three participants at each of the six clinics were allocated with straightforward randomisation. Thereafter allocation was stratified by centre, alternative criteria for chronic fatigue syndrome and myalgic encephalomyelitis and depressive disorder (major or minor depressive episode or dysthymia), with computer-generated probabilistic minimisation"
Allocation concealment (selection bias)	Low risk	Quote: "Once notified of treatment allocation by the Clinical Trials Unit, the research assessor informed the participant and clinicians"
Blinding (performance bias and detection bias) of participants and personnel?	High risk	Quote: "As with any therapy trial, participants, therapists, and doctors could not be masked to treatment allocation and it was also impractical to mask research assessors. The primary outcomes were rated by participants themselves"

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

White 2011 (*Continued*)

Blinding (performance bias and detection bias) of outcome assessors?	High risk	Quote: "The statistician undertaking the analysis of primary outcomes was masked to treatment allocation"
Incomplete outcome data (attrition bias) All outcomes	Low risk	None found
Selective reporting (reporting bias)	Low risk	Quote: "These secondary outcomes were a subset of those specified in the protocol, selected in the statistical analysis plan as most relevant to this report." Our primary interest is the primary outcome reported in accordance with the protocol, so we do not believe that selective reporting is a problem
Other bias	Low risk	We do not suspect other types of bias

ACT, anaerobic activity therapy.
APT, adaptive pacing therapy.
BAI, Beck Anxiety Inventory.
BDI-II, Beck Depression Inventory.
BPI, Brief Pain Inventory.
CBT, cognitive-behavioural therapy.
CDC, Centers for Disease Control and Prevention.
CFS, chronic fatigue syndrome.
COG, cognitive therapy.
ET, exercise therapy.
FS, Fatigue Scale.
FSS, Fatigue Severity Scale.
GET, graded exercise therapy.
HADS, Hospital Anxiety and Depression Scale.
HR, heart rate.
IQR, interquartile range.
ME, myalgic encephalitis.
MOS, Medical Outcome Survey.
NS, Not stated.
PSQI, Pittsburgh Sleep Quality Index.
PSS, Perceived Stress Scale.
RCT, randomised controlled trial.
RELAX, relaxation treatment.
RPE, rating of perceived exertion.
SD, standard deviation.
SF-36, Short Form 36.
SMC, specialist medical care.
VO_2, oxygen consumption.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Evering 2008	RCT The trial was excluded, as the intervention was feedback on physical activity
Gordon 2010	RCT Compares the relative effectiveness of 2 different types of exercise therapy. Even though this is an interesting question, it was beyond the scope of this version of the review
Guarino 2001	The trial was excluded, as the population was "Gulf War veterans"
Nunez 2011	RCT Combination treatment of which exercise therapy is a minor part
Ridsdale 2004	RCT No clinical diagnosis of chronic fatigue syndrome. Our inclusion criteria state that the duration of fatigue needs to > 6 months, whereas inclusion criteria in Risdale 2004 is > 3 months The trial was excluded, as the intervention did not include exercise: "cognitive behaviour therapy (CBT) with counselling"; the population was "patients with chronic fatigue"
Ridsdale 2012	RCT The trial was excluded, as the population was "people presenting with chronic fatigue in primary care"
Russel 2001	RCT The trial was excluded, as exercise was not the main part of the intervention: "Group rehabilitation (psycho-education, graded exercise, goal setting and pacing, breathing control and challenging unhelpful thoughts)"
Stevens 1999	RCT The PhD was excluded, as exercise was a minor component of the intervention: "conducted to implement the use of sleep hygiene education, biofeedback assisted relaxation and breathing retraining, graded aerobic exercise, and cognitive therapy...."
Taylor 2004	RCT The trial was excluded, as exercise was not the main component of the intervention: "In our program, group topics included activity pacing using the Envelope Theory (Jason et al., 1999), cognitive coping skills training, relaxation and meditation training, employment issues and economic self-sufficiency, personal relationships, traditional and complementary medical approaches, and nutritional approaches"
Taylor 2006	The trial was excluded, as the study used a "cross-sectional design"
Thomas 2008	The trial was excluded, as "between-group comparisons were used." This was a controlled trial, but participants were not randomly assigned
Tummers 2012	RCT The trial was excluded, as interventions included variations of CBT: "additional CBT (stepped care) or regular CBT (care as usual)"

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

(*Continued*)

Viner 2004	The trial was excluded, as the population consisted of "young people (aged 9-17 years) with CFS/ME"
Wright 2005	The trial was excluded, as the population included young people 0 to 19 years of age

Characteristics of studies awaiting assessment *[ordered by study ID]*

Hatcher 1998

Methods	RCT, 2 arms
Participants	Patients with chronic fatigue syndrome
Interventions	Dothiepin and graded activity
Outcomes	Not found
Notes	Not able to identify published paper nor study author

Liu 2010

Methods	RCT, 3 arms, N = 90
Participants	Patients with chronic fatigue syndrome
Interventions	Tuina group Taijiquan (take exercise) group Fluoxetine group
Outcomes	Therapeutic effects and changes in malondialdehyde (MDA) content and in activity of serum superoxide dismutases (SOD) and serum glutathione peroxidase (GSH-Px) were observed
Notes	Published paper does not report outcomes that are relevant for this review Study authors were contacted to clarify whether relevant outcomes were measured, but we are still awaiting response

Zhuo 2007

Methods	RCT, 2 parallel arms
Participants	Patients with chronic fatigue syndrome, N = 70
Interventions	Sports group received gradual exercise Comparison group rested
Outcomes	Fatigue symptoms of chronic fatigue syndrome (CFS), sleeping time; symptoms for ears and eyes, muscle and bone system, nervous system and quality of life

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Notes	Information from English abstract. Waiting for translation

Characteristics of ongoing studies *[ordered by study ID]*

Broadbent 2012

Trial name or title	Pilot study on the effects of intermittent and graded exercise compared with no exercise for optimising health and reducing symptoms in chronic fatigue syndrome (CFS) patients
Methods	Randomised controlled trial, parallel
Participants	Inclusion criteria: medical diagnosis of chronic fatigue syndrome: persistent and disabling, and/or recurring, fatigue lasting longer than 6 months, which does not result from physical exertion and is not alleviated by rest. Other symptoms include muscle weakness and pain, ongoing medical symptoms such as swollen lymph nodes and fever, poor sleep, poor concentration and reduced quality of life Exclusion criteria: diagnosed cardiac and/or respiratory disease; joint or muscle condition/disease other than CFS that is contraindicated for exercise; any mental health condition that may affect exercise participation or safety of participants and researchers Age minimum: 18 years Age maximum: 60 years Gender: both male and female
Interventions	Randomised controlled trial of intermittent exercise training compared with graded exercise and standard care. Graded exercise is the current recommended exercise approach to CFS; it consists of self-paced (e.g. low-intensity) steady state exercise at a constant workload for a short time; as the patient's fitness gradually improves, the length of time and eventually the intensity are increased in a gradual graded manner, provided no adverse symptoms occur. Intermittent or interval exercise consists of short blocks of exercise at low to moderate intensity with a rest interval in between bouts of exercise (e.g. 1 minute of low-intensity cycling, followed by 1 minute of rest, followed by 1 minute of cycling); total time spent exercising can be gradually increased whilst rest or unloaded exercise intervals are maintained. Participants will be randomly allocated to 1 of 3 groups. Each group will consist of 20 participants to provide a power of 80% for the study (based on data from Gordon 2010), with an a prior test used to compute required sample size, given alpha (P value 0. 05), power and effect size for an F test, and looking at ANOVA fixed effects, main effects and interactions (GPower). Volunteers will participate in 3 aerobic exercise sessions (cycling on a cycle ergometer) per week, consisting of the following • Warm-up of 5 minutes of unloaded cycling for both ITE and GE groups • Either a steady state (constant effort) low- to moderate-intensity cycling period (50% VO_2peak, RPE 3 Modified Borg Scale) initially for 10 minutes (GE group) OR an intermittent exercise block of 1 minute of moderate-intensity cycling (60% VO_2peak, RPE 4 to 5) alternated with 1 minute of unloaded or very low-intensity/unloaded cycling (20% to 30% VO_2peak, RPE 1 to 2), totaling 20 minutes • Cool-down of 5 minutes unloaded cycling plus stretching of main muscle groups for both groups Over the 12 weeks of the project, we aim to progress the duration of SS exercise towards 20 minutes, as tolerated by the participant, and to progress ITE participants towards intervals of 2 to 3 minutes of moderate-intensity cycling, alternated with 1-minute intervals of low-intensity cycling, totaling 25 to 30 minutes in duration. All group sessions will be supervised by a member of the research team (consisting of accredited exercise physiologists) with assistance from postgraduate Masters of Clinical Exercise Physiology students, who are studying to become accredited exercise physiologists

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Broadbent 2012 (*Continued*)

	Total intervention duration will be 12 weeks for graded, intermittent and control groups
Outcomes	Improved physiological adaptations to exercise (reduced RPE, heart rate and blood pressure). Rate of perceived exertion (RPE) is assessed using a standard 10-point Borg Scale on which participants are asked how hard they feel they are exercising; heart rate will be measured using a 12-lead ECG during prestudy and poststudy exercise tests, and during exercise sessions, by using a Polar heart rate monitor; blood pressure will be monitored constantly during prestudy and poststudy exercise testing, and during exercise sessions, using a standard sphygmomanometer and an adult-sized cuff and stethoscope Increased lymphocyte function and reduced inflammatory cytokines measured prestudy and poststudy by comparison of immune cell counts, lymphocyte (CD4, CD8, CD19, NK) function and inflammatory cytokines (IFN-λ, IL-1) in both exercise groups and control groups. Cell counts will be measured by full blood count (standard pathology); lymphocyte subsets will be measured by cell count using a FACSCanto flow cytometer (Becton Dickinson); lymphocyte function will be analysed using proliferative assays with flow cytometric fluorescent analysis; and inflammatory cytokines will be assessed using standard ELISA assays Increased VO_2peak, as measured prestudy and poststudy by open circuit spirometry (Sensormedics) metabolic cart and by breath-by-breath analysis. The test protocol is a cycle test starting with a 3-minute warm-up of unloaded cycling, followed by 1-minute increments of 10 watts (W) until a VO_2 plateau is achieved (i.e. VO_2 does not increase, although workload continues to increase and/or RER > 1.15 and/or peak heart rate within 10 beats per minute of age-predicted maximum and/or volitional exhaustion). The test may also be stopped at the request of participants if they feel too fatigued. If a submaximal value is achieved at this stage, a peak VO_2 value can be extrapolated by using a linear regression Reduced fatigue and symptoms (Cummins Fatigue Scale)
Starting date	10/02/2013
Contact information	suzanne.broadbent@scu.edu.au
Notes	http://apps.who.int/trialsearch/Trial.aspx?TrialID=ACTRN12612001241820 http://www.anzctr.org.au/ACTRN12612001241820.aspx

Kos 2012

Trial name or title	Pacing activity self-management for patients with chronic fatigue syndrome: randomized controlled clinical trial
Methods	RCT
Participants	Inclusion criteria • Adults between 18 and 65 years of age • Female gender • Willing to sign informed consent form • Fulfilling 1994 Centers for Disease Control and Prevention criteria for the diagnosis of chronic fatigue syndrome Exclusion criteria • Not fulfilling each of the inclusion criteria listed above
Interventions	Behavioural: pacing Behavioural: relaxation therapy

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Kos 2012 (*Continued*)

Outcomes	Change in score on the Canadian Occupational Performance Measure (COPM) Change in autonomic activity at rest and following 3 activities of daily living Change in CFS Symptom List Change in Checklist of Individual Strength (CIS) Change in subscale scores on the Medical Outcomes Short Form-36 Health Status Survey (SF-36)
Starting date	August 2011
Contact information	Jo.Nijs@vub.ac.be
Notes	http://clinicaltrials.gov/show/NCT01512342

Marques 2012

Trial name or title	Protocol for the "four steps to control your fatigue (4-STEPS)" randomised controlled trial: a self-regulation based physical activity intervention for patients with unexplained chronic fatigue
Methods	Multi-centre, randomised controlled trial (RCT)
Participants	Fulfilling operationalised criteria for idiopathic chronic fatigue (ICF) and for chronic fatigue syndrome (CFS) Patients visiting their physician with a main complaint of unexplained fatigue of at least 6 months' duration are recruited for the study Inclusion criteria: meeting the operationalised criteria for ICF or CFS (CDC criteria); between 18 and 65 years of age; fluent in spoken Portuguese; capacity to provide informed consent Exclusion criteria: presence of a concurrent somatic condition that can explain the fatigue symptoms; severe psychiatric disorders
Interventions	Standard care (SC) or standard care plus a self-regulation based physical activity programme (4-STEPS) In addition to standard care, participants in the intervention group received the 4-STEPS programme consisting of the following • 2 face-to-face individual motivational interviewing (MI) sessions aimed at exploring important health and life goals, increasing participants' motivation and confidence to be physically active and setting a specific personal physical activity goal. The first MI session takes place 1 week after the baseline assessment, and the second MI session takes place 2 weeks after the first. The MI session is delivered by a psychologist with MI training (member of the research team). The duration of the sessions is approximately 1 hour. Details on topics addressed during the MI sessions are presented in Table 1 • 2 brief telephone counselling sessions: Sessions take about 20 minutes and are provided 2 weeks and 6 weeks after the last MI session. Details on topics addressed during the telephone sessions are presented in Table 1 • Self-regulation (SR) booklets: 2 booklets were designed to help patients change their level of physical activity (informational booklet and workbook). The informational booklet was provided at the end of the baseline assessment; the "Step 1" part of the workbook is provided at the first MI session, and parts "Step 2, " "Step 3" and "Step 4" are given during the second MI session. Details on topics addressed in the SR booklets are presented in Table 2 • A pedometer to register physical activity on a daily basis (steps taken) during the 3-month intervention period. Instructions on how to use the pedometer are given during the baseline assessment session (Table 2) • Daily activities record (Table 2): Participants received several daily activity records (physical activities, mental activities and rest). The first daily activity record was given to the participant at the end of the first

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Marques 2012 (*Continued*)

	MI session; participants were asked to fill out the activity record during the time between the first and second MI sessions. This homework assignment aimed to evaluate participants' daily activities management while possibly recognising an erratic pattern of rest and activity (boom and bust cycle). At the end of the second MI session, participants received daily activities records that could be used to monitor changes in daily activity patterns during the subsequent 9 weeks • Leaflet for family: At the end of the first MI session, participants received a leaflet for their partner or significant other to increase social support
Outcomes	The primary outcome was the reduction in perceived fatigue severity, which was assessed by using the Checklist of Individual Strength (CIS-20R). A difference of 7 points between intervention and control groups for the main dimension (the subjective feeling of fatigue subscale) of the CIS-20R was considered to be clinically significant
Starting date	The 4-STEPS RCT started in January 2011
Contact information	Marta Marques: mmarques@ispa.pt
Notes	ISRCTN: ISRCTN70763996 Copied from the published protocol: http://www.biomedcentral.com/1471-2458/12/202

Vos-Vromans 2008

Trial name or title	Is a multi-disciplinary rehabilitation treatment more effective than mono-disciplinary cognitive behavioural therapy for patients with chronic fatigue syndrome? A multi-centre randomised controlled trial
Methods	RCT
Participants	Patients were included if they fulfilled the CDC-94 criteria for CFS and had a score ≥ 40 on the Checklist of Individual Strength (CIS)-fatigue questionnaire. CDC-94 criteria for CFS are as follows • At least 6 months of persistent or recurring fatigue for which no physical explanation was found and that ○ was of new onset, that is to say, it had not been lifelong ○ was not the result of ongoing exertion ○ was not substantially alleviated by rest and ○ severely limited functioning In combination with 4 or more of the following symptoms, persistent or regularly recurring over a period of 6 months and that must not have predated the fatigue • Self-reported impairment in memory or concentration • Sore throat • Tender cervical lymph nodes • Muscle pain • Multi-joint pain • Headache • Unrefreshing sleep • Postexertional malaise lasting 24 hours or longer Additional inclusion criteria for this study follow here • Participants are willing to participate in a treatment that is set up to change behaviour • Participants are between 18 and 60 years of age, of either sex

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

	• Participants can speak, understand and write the Dutch language
Interventions	After intake, participants will be randomly divided into 2 groups: cognitive-behavioural therapy (CBT) and multi-disciplinary rehabilitation therapy (MRT) • Cognitive-behavioural therapy (CBT) CBT is based on process variables of a CFS model. This model shows that high physical attributions will decrease physical activity and increase fatigue and functional impairment. A low level of sense of control over symptoms and focusing on physical sensations have a direct causal effect on fatigue. In CFS precipitating and perpetuating factors are important. The perpetuating factors become the focus of the intervention in CBT. An important subject in the therapy is the balance between activity and rest and the patients' responsibility to see to it. Negative beliefs regarding the symptoms of fatigue, self-expectations or self-esteem are identified and patients are encouraged to challenge them the conventional way. Specific lifestyle changes are encouraged if deemed appropriate. At the end of the therapy relapse prevention is addressed. Patients who are assigned to this group will attend 16 individual therapy sessions of one hour duration, spread out over 6 months with a psychologist or behavioural therapist • Multi-disciplinary rehabilitation therapy (MRT): MRT includes CBT, GET, pacing and body awareness therapy (investigational treatment) ○ CBT: as above ○ Graded exercise therapy (GET): structured and supervised activity management that aims at a gradual but progressive increase in aerobic activities. It is completed by graded activity and graded exercise in which a gradual and progressive increase in physical and mental activities is trained. Activities include activities of daily living and occupational and social or leisure activities ○ Pacing: helps the patient divide energy over the day/week. Eventually patients are encouraged to carry out a gradual increase in physical and mental activity ○ Body awareness therapy: teaches the patient to be aware of healthy physical sensations and to link them in the mind (body mentalisation). Patients are taught to react adequately to disturbances in the balance between daily workload and the capacity to deal with it. The balance between activity and rest is linked to the patient's inner control and to healthy physical sensations MRT includes the following • 2 weeks: observation (2 sessions of 1 hour with psychology, 2 sessions of 1 hour with a social worker, 2 sessions of 1/2 hour with occupational therapy, 2 sessions of 1/2 hour with physiotherapy) • 2 weeks: no therapy • 10 weeks therapy (5 sessions of 1 hour with psychology, 4 sessions of 1 hour with a social worker, 26 sessions of 1/2 hour with physiotherapy and 20 sessions of 1/2 hour with occupational therapy) • 6 weeks: no therapy • 1 session of 1 hour with a social worker (after 6 weeks of no therapy) • 2 sessions of both 1/2 and 1 hour of therapy with the therapist chosen by participants During MRT, a participant sees the physician during rehabilitation 3 times (20 minutes per visit) Total duration of both treatments is 6 months. Duration of follow-up for both treatments is also 6 months
Outcomes	Primary outcomes • Fatigue severity as measured using the Checklist of Individual Strength at baseline, 6 months and 12 months after start of therapy Secondary outcomes • Quality of life as measured using the 36-item Short-Form Health Survey (SF-36) • Psychological well-being as measured using Symptom Check List-90 • Sense of control in relation to CFS complaints as measured using a self-efficacy scale • Somatic attributions as measured using the Causal Attribution List • Mindfulness as measured using the Mindfulness Attention Awareness Scale

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

	• Functional activities (the most important) that a patient wants to improve during treatment as measured using the Patient-Specific Complaints and Goals Questionnaire • Impact of disease on both physical and emotional functioning as measured using the Sickness Impact Profile • Physical activity as measured using the Body Media Sensewear Activity Monitor • Self-rated improvement as measured using 5 questions on the 5- and 10-point Likert scale • Life satisfaction as measured using the Life Satisfaction Questionnaire • Utility as measured using EuroQol 6-D • Treatment expectancy and credibility as measured using the Devilly and Borkovec Questionnaire All outcomes are measured at baseline and at 6 and 12 months after start of therapy. Treatment costs and additional expenses (work-related costs, healthcare and non-healthcare costs) are measured using the Trimbos/iMTA Questionnaire for Costs Associated With Psychiatric Illness; will be measured every month (from baseline until 12 months after start of therapy)
Starting date	27/11/2008 Recruitment status: completed
Contact information	d.vos-vromans@rcbreda.nl
Notes	http://isrctn.org/ISRCTN77567702

White 2012

Trial name or title	Graded Exercise Therapy guided SElf-help Treatment (GETSET) for patients with chronic fatigue syndrome/myalgic encephalomyelitis: a randomised controlled trial in secondary care (GETSET)
Methods	Randomised interventional trial
Participants	Inclusion • Patients attending 2 CFS/ME specialist clinics in London • Patients receiving a diagnosis of CFS/ME from a specialist doctor and going onto a waiting list for clinic treatment • Patients 18 years of age or older • Speak and read English adequately to provide informed consent and read the guided support booklet • Target gender: male and female • Lower age limit: 18 years Exclusion • Not receiving a diagnosis of CFS/ME • Co-morbid condition that requires that exercise be performed only in the presence of a doctor • Younger than age 18 • Active suicidal thoughts
Interventions	Guided support, a copy of the GETSET booklet, a 30-minute consultation face-to-face by Skype or by telephone, 3 further Skype telephone contacts Intervention over 9 weeks: follow-up length: 3 month(s); study entry: single randomisation only
Outcomes	Primary: SF-36 physical function subscale (SF-36PF) measured 12 weeks from randomisation Secondary: Clinical Global Impression Change Scale (CGI) score measured 12 weeks from baseline

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

White 2012 (*Continued*)

Starting date	16/05/2012
Contact information	Prof PD White; p.d.white@qmul.ac.uk
Notes	http://www.controlled-trials.com/ISRCTN22975026/GETSET

ANOVA, analysis of variance.
CFS, chronic fatigue syndrome.
CGI, Clinical Global Impression scale.
CIS, Checklist of Individual Strength.
COPM, Canadian Occupational Performance Measure.
ELISA, enzyme-linked immunosorbent assay.
EuroQol 6-D: Short Form 6-D of the standard measure of health outcomes of the EuroQol Group.
GE, Graded exercise.
ICF, idiopathic chronic fatigue.
IFN, interferon.
IL, interleukin.
ITE, intermittent exercise training.
MI, motivational interviewing.
MRT, multi-disciplinary rehabilitation therapy.
NK, natural killer cell.
RER, respiratory exchange ratio.
RPE, rating of perceived exertion.
SC, standard care.
SS, steady state.
VO_2, oxygen consumption

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

DATA AND ANALYSES

Comparison 1. Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Fatigue (end of treatment)	7		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.1 Fatigue Scale, FS (11 items/0 to 11 points)	1	148	Mean Difference (IV, Random, 95% CI)	-6.06 [-6.95, -5.17]
1.2 Fatigue Scale, FS (11 items/0 to 33 points)	3	540	Mean Difference (IV, Random, 95% CI)	-2.82 [-4.07, -1.57]
1.3 Fatigue Scale, FS (14 items/0 to 42 points)	3	152	Mean Difference (IV, Random, 95% CI)	-6.80 [-10.31, -3.28]
2 Fatigue (follow-up)	4		Mean Difference (IV, Random, 95% CI)	Subtotals only
2.1 Fatigue Scale, FS (11 items/0 to 11 points)	1	148	Mean Difference (IV, Random, 95% CI)	-7.13 [-7.97, -6.29]
2.2 Fatigue Scale, FS (11 items/0 to 33 points)	2	472	Mean Difference (IV, Random, 95% CI)	-2.87 [-4.18, -1.55]
2.3 Fatigue Severity Scale, FSS (9 items/1 to 7 points)	1	50	Mean Difference (IV, Random, 95% CI)	0.15 [-0.55, 0.85]
3 Participants with serious adverse reactions	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
4 Pain (follow-up)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
4.1 Brief Pain Inventory, pain severity subscale (0 to 10 points)	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
4.2 Brief Pain Inventory, pain interference subscale (0 to 10 points)	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
5 Physical functioning (end of treatment)	5		Mean Difference (IV, Random, 95% CI)	Subtotals only
5.1 SF-36, physical functioning subscale (0 to 100 points)	5	725	Mean Difference (IV, Random, 95% CI)	-13.10 [-24.22, -1.98]
6 Physical functioning (follow-up)	3		Mean Difference (IV, Random, 95% CI)	Subtotals only
6.1 SF-36, physical functioning subscale (0 to 100 points)	3	621	Mean Difference (IV, Random, 95% CI)	-16.33 [-36.74, 4.08]
7 Quality of life (follow-up)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
7.1 Quality of Life Scale (16 to 112 points)	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
8 Depression (end of treatment)	5		Mean Difference (IV, Random, 95% CI)	Subtotals only
8.1 HADS, depression score (7 items/21 points)	5	504	Mean Difference (IV, Random, 95% CI)	-1.63 [-3.50, 0.23]
9 Depression (follow-up)	4		Mean Difference (IV, Random, 95% CI)	Subtotals only
9.1 Beck Depression Inventory (0 to 63 points)	1	45	Mean Difference (IV, Random, 95% CI)	3.44 [-1.00, 9.88]

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

9.2 HADS, depression subscale (0 to 21 points)	3	609	Mean Difference (IV, Random, 95% CI)	-2.26 [-5.09, 0.56]
10 Anxiety (end of treatment)	3		Mean Difference (IV, Random, 95% CI)	Subtotals only
10.1 HADS, anxiety score (0 to 21 points)	3	387	Mean Difference (IV, Random, 95% CI)	-1.48 [-3.58, 0.61]
11 Anxiety (follow-up)	4		Mean Difference (IV, Random, 95% CI)	Subtotals only
11.1 Beck Anxiety Inventory (0 to 63 points)	1	45	Mean Difference (IV, Random, 95% CI)	0.70 [-4.52, 5.92]
11.2 HADS, anxiety score (0 to 21 points)	3	607	Mean Difference (IV, Random, 95% CI)	-1.01 [-2.75, 0.74]
12 Sleep (end of treatment)	2		Mean Difference (IV, Random, 95% CI)	Subtotals only
12.1 Jenkins Sleep Scale (0 to 20 points)	2	323	Mean Difference (IV, Random, 95% CI)	-1.49 [-2.95, -0.02]
13 Sleep (follow-up)	3		Mean Difference (IV, Random, 95% CI)	Subtotals only
13.1 Jenkins Sleep Scale (0 to 20 points)	3	610	Mean Difference (IV, Random, 95% CI)	-2.04 [-3.84, -0.23]
14 Self-perceived changes in overall health (end of treatment)	4	489	Risk Ratio (M-H, Random, 95% CI)	1.83 [1.39, 2.40]
15 Self-perceived changes in overall health (follow-up)	3	518	Risk Ratio (M-H, Random, 95% CI)	1.88 [0.76, 4.64]
16 Health resource use (follow-up) [Mean no. of contacts]	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
16.1 Primary care	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
16.2 Other doctor	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
16.3 Healthcare professional	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
16.4 Inpatient	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
16.5 Accident and emergency	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
16.6 Other health/social services	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
16.7 Complementary health care	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
16.8 Standardised medical care	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
17 Health resource use (follow-up) [No. of users]	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
17.1 Primary care	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
17.2 Other doctor	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
17.3 Healthcare professional	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
17.4 Inpatient	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
17.5 Accident and emergency	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
17.6 Medication	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
17.7 Complementary health care	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
17.8 Other health/social services	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
17.9 Standardised medical care	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
18 Drop-out	6	843	Risk Ratio (M-H, Random, 95% CI)	1.63 [0.77, 3.43]
19 Subgroup analysis for fatigue	7	840	Std. Mean Difference (IV, Random, 95% CI)	-0.68 [-1.02, -0.35]
19.1 Graded exercise therapy	6	779	Std. Mean Difference (IV, Random, 95% CI)	-0.71 [-1.09, -0.32]
19.2 Exercise with self-pacing	1	61	Std. Mean Difference (IV, Random, 95% CI)	-0.54 [-1.05, -0.02]

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Comparison 2. Exercise therapy versus psychological treatment

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Fatigue at end of treatment (FS; 11 items/0 to 33 points)	2		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 CBT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
1.2 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Fatigue at follow-up (FSS; 1 to 7 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.1 CT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2.2 CBT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3 Fatigue at follow-up (FS; 11 items/0 to 33 points)	2		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.1 CBT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3.2 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
4 Participants with serious adverse reactions	2		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
4.1 CBT	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
4.2 Suportive listening	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
5 Pain at follow-up (BPI, pain severity subscale; 0 to 10 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
5.1 CBT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
5.2 CT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
6 Pain at follow-up (BPI, pain interference subscale; 0 to 10 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
6.1 CBT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
6.2 CT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
7 Physical functioning at end of treatment (SF-36, physical functioning subscale; 0 to 100 points)	2		Mean Difference (IV, Random, 95% CI)	Totals not selected
7.1 CBT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
7.2 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
8 Physical functioning at follow-up (SF-36, physical functioning subscale; 0 to 100 points)	3		Mean Difference (IV, Random, 95% CI)	Subtotals only
8.1 CBT	2	348	Mean Difference (IV, Random, 95% CI)	7.92 [-9.79, 25.63]
8.2 CT	1	47	Mean Difference (IV, Random, 95% CI)	21.37 [6.61, 36.13]
8.3 Supportive listening	1	171	Mean Difference (IV, Random, 95% CI)	-7.55 [-15.57, 0.47]
9 Depression at end of treatment (HADS depression score; 7 items/21 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
9.1 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
10 Depression at follow-up (BDI; 0 to 63 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
10.1 CT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
10.2 CBT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

11 Depression at follow-up (HADS depression score; 7 items/21 points)	2		Mean Difference (IV, Random, 95% CI)	Totals not selected
11.1 CBT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
11.2 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
12 Anxiety at end of treatment (HADS anxiety; 7 items/21 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
12.1 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
13 Anxiety at follow-up (BAI; 0 to 63 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
13.1 CT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
13.2 CBT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
14 Anxiety at follow-up (HADS anxiety; 7 items/21 points)	2		Mean Difference (IV, Random, 95% CI)	Totals not selected
14.1 CBT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
14.2 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
15 Sleep at end of treatment (Jenkins Sleep Scale; 0 to 20 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
15.1 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
16 Sleep at follow-up (Jenkins Sleep Scale; 0 to 20 points)	2		Mean Difference (IV, Random, 95% CI)	Totals not selected
16.1 CBT	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
16.2 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
17 Self-perceived changes in overall health at end of treatment	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
17.1 CBT	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
18 Self-perceived changes in overall health at follow-up	2		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
18.1 CT	1	50	Risk Ratio (M-H, Random, 95% CI)	0.63 [0.36, 1.10]
18.2 CBT	2	368	Risk Ratio (M-H, Random, 95% CI)	0.71 [0.33, 1.54]
19 Health resource use (follow-up) [Mean no. of contacts]	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
19.1 Primary care	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
19.2 Other doctor	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
19.3 Healthcare professional	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
19.4 Inpatient	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
19.5 Accident and emergency	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
19.6 Other health/social services	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
19.7 Complementary health care	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
19.8 Standardised medical care	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
20 Health resource use (follow-up) [No. of users]	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
20.1 Primary care	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
20.2 Other doctor	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
20.3 Healthcare professional	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
20.4 Inpatient	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

20.5 Accident and emergency	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
20.6 Medication	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
20.7 Complementary health care	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
20.8 Other health/social services	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
20.9 Standardised medical care	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
21 Drop-out	2	Risk Ratio (M-H, Random, 95% CI)	Totals not selected
21.1 CBT	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
21.2 Supportive listening	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]

Comparison 3. Exercise therapy versus adaptive pacing

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Fatigue	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Fatigue Scale, FS (11 items/33 points)-end of treatment	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
1.2 Fatigue Scale, FS (11 items/33 points)-follow-up	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Participants with serious adverse reactions	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3 Physical functioning	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.1 SF-36, physical functioning subscale (0 to 100)-end of treatment	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3.2 SF-36, physical functioning subscale (0 to 100)-follow-up	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
4 Depression	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
4.1 HADS, depression score (7 items/21 points)-follow-up	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
5 Anxiety	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
5.1 HADS, anxiety score (0 to 21 points)-follow-up	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
6 Sleep	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
6.1 Jenkins Sleep Scale (0 to 20 points)-follow-up	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
7 Self-perceived changes in overall health	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
7.1 End of treatment	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
7.2 Follow-up	1		Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
8 Health resource use (follow-up) [Mean no. of contacts]	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
8.1 Primary care	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
8.2 Other doctor	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

8.3 Healthcare professional	1	Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
8.4 Inpatient	1	Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
8.5 Accident and emergency	1	Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
8.6 Other health/social services	1	Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
8.7 Complementary health care	1	Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
8.8 Standardised medical care	1	Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
9 Health resource use (follow-up) [No. of users]	1	Risk Ratio (M-H, Random, 95% CI)	Totals not selected
9.1 Primary care	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
9.2 Other doctor	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
9.3 Healthcare professional	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
9.4 Inpatient	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
9.5 Accident and emergency	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
9.6 Medication	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
9.7 Complementary health care	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
9.8 Other health/social services	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
9.9 Standardised medical care	1	Risk Ratio (M-H, Random, 95% CI)	0.0 [0.0, 0.0]
10 Drop-out	1	Risk Ratio (M-H, Random, 95% CI)	Totals not selected

Comparison 4. Exercise therapy + antidepressant placebo versus antidepressant + exercise placebo

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Fatigue	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Fatigue Scale, FS (14 items/0 to 42 points)-end of treatment	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Depression	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.1 HADS, depression score (7 items/21 points)-end of treatment	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3 Drop-out	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Comparison 5. Exercise therapy + antidepressant versus antidepressant + exercise placebo

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1 Fatigue	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.1 Fatigue Scale, FS (14 items/0 to 42 points)-end of treatment	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
2 Depression	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.1 HADS, depression score (7 items/21 points)-end of treatment	1		Mean Difference (IV, Random, 95% CI)	0.0 [0.0, 0.0]
3 Drop-out	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected

Analysis 1.1. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 1 Fatigue (end of treatment).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 1 Fatigue (end of treatment)

Study or subgroup	Exercise therapy N	Mean(SD)	TAU or relax+flex N	Mean(SD)	Mean Difference IV,Random,95% CI	Weight	Mean Difference IV,Random,95% CI
1 Fatigue Scale, FS (11 items/0 to 11 points)							
Powell 2001	114	4.34 (4.48)	34	10.4 (1.04)		100.0 %	-6.06 [-6.95, -5.17]
Subtotal (95% CI)	**114**		**34**			**100.0 %**	**-6.06 [-6.95, -5.17]**
Heterogeneity: not applicable							
Test for overall effect: Z = 13.29 (P < 0.00001)							
2 Fatigue Scale, FS (11 items/0 to 33 points)							
Wallman 2004	32	11.06 (7.65)	29	15.34 (8.15)		9.9 %	-4.28 [-8.26, -0.30]
Wearden 2010	85	22.78 (8.56)	92	26.27 (7.68)		27.2 %	-3.49 [-5.89, -1.09]
White 2011	150	21.7 (7.1)	152	24 (6.9)		62.9 %	-2.30 [-3.88, -0.72]
Subtotal (95% CI)	**267**		**273**			**100.0 %**	**-2.82 [-4.07, -1.57]**
Heterogeneity: Tau^2 = 0.0; Chi^2 = 1.23, df = 2 (P = 0.54); I^2 =0.0%							
Test for overall effect: Z = 4.41 (P = 0.000010)							
3 Fatigue Scale, FS (14 items/0 to 42 points)							
Fulcher 1997	27	20.96 (9.08)	30	27.5 (7.44)		46.5 %	-6.54 [-10.88, -2.20]
Moss-Morris 2005	22	13.91 (10.88)	21	24.41 (9.69)		27.1 %	-10.50 [-16.65, -4.35]

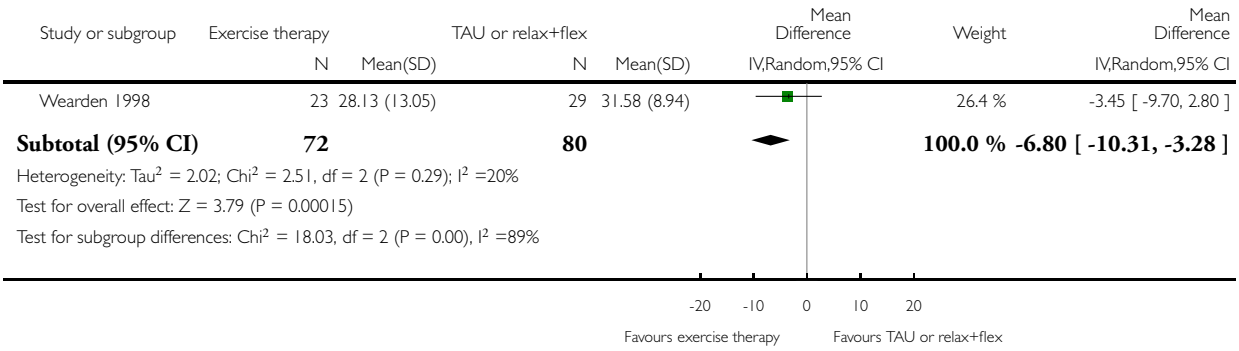
-20 -10 0 10 20

Favours exercise therapy Favours TAU or relax+flex

(*Continued . . .*)

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

(. . . *Continued*)

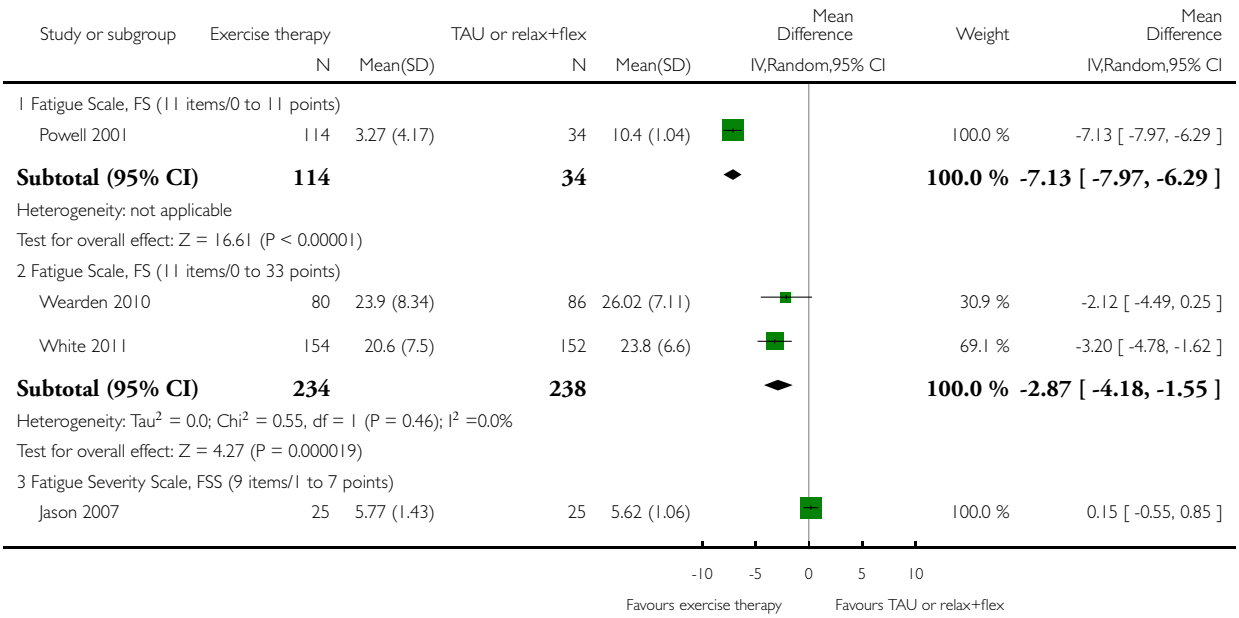


Analysis 1.2. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 2 Fatigue (follow-up).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 2 Fatigue (follow-up)



(*Continued* . . .)

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

(. . . *Continued*)

Study or subgroup	Exercise therapy N	Mean(SD)	TAU or relax+flex N	Mean(SD)	Mean Difference IV,Random,95% CI	Weight	Mean Difference IV,Random,95% CI
Subtotal (95% CI)	**25**		**25**			**100.0 %**	**0.15 [-0.55, 0.85]**
Heterogeneity: not applicable							
Test for overall effect: Z = 0.42 (P = 0.67)							
Test for subgroup differences: Chi^2 = 170.37, df = 2 (P = 0.00), I^2 =99%							



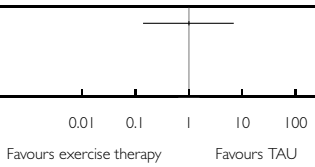
Analysis 1.3. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 3 Participants with serious adverse reactions.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 3 Participants with serious adverse reactions

Study or subgroup	Exercise therapy n/N	TAU or relax+flex n/N	Risk Ratio M-H,Random,95% CI	Risk Ratio M-H,Random,95% CI
White 2011	2/160	2/159		0.99 [0.14, 6.97]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

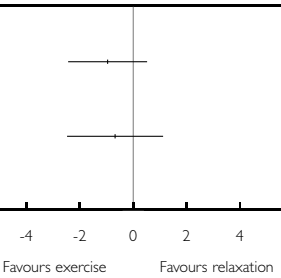
Analysis 1.4. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 4 Pain (follow-up).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 4 Pain (follow-up)

Study or subgroup	Exercise therapy		TAU or relax+flexibility		Mean Difference	Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Random,95% CI	IV,Random,95% CI
1 Brief Pain Inventory, pain severity subscale (0 to 10 points)						
Jason 2007	20	3.63 (2.72)	23	4.6 (2.1)		-0.97 [-2.44, 0.50]
2 Brief Pain Inventory, pain interference subscale (0 to 10 points)						
Jason 2007	20	3.75 (3.14)	23	4.44 (2.79)		-0.69 [-2.48, 1.10]



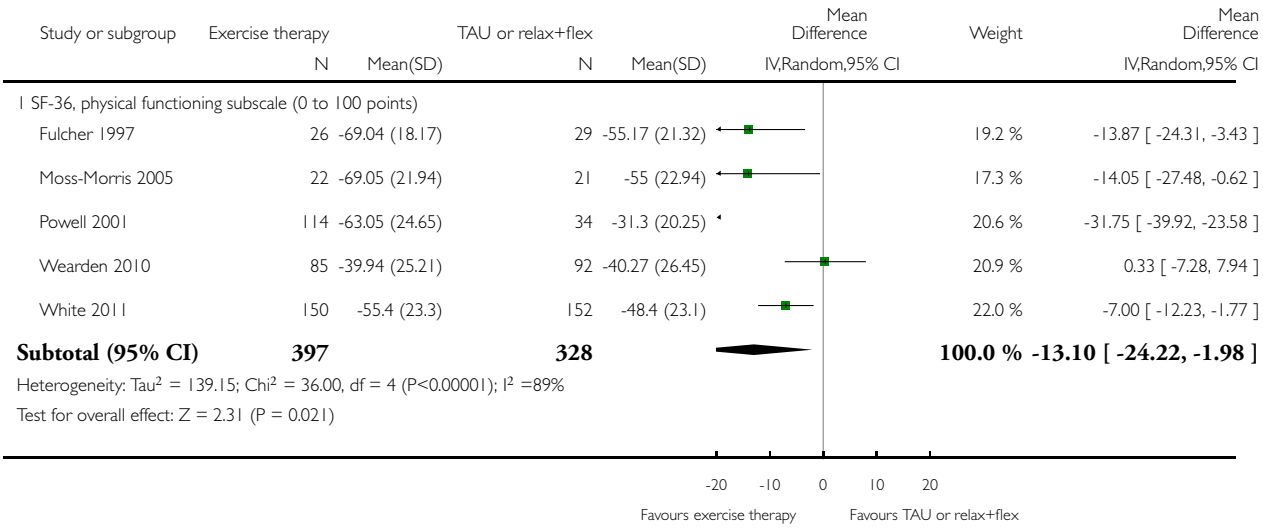
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.5. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 5 Physical functioning (end of treatment).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 5 Physical functioning (end of treatment)



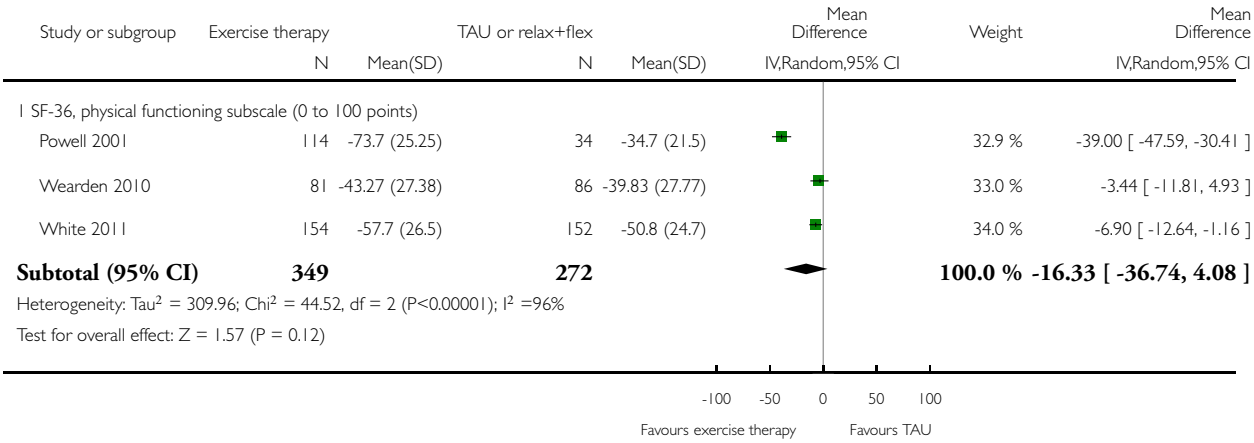
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.6. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 6 Physical functioning (follow-up).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 6 Physical functioning (follow-up)

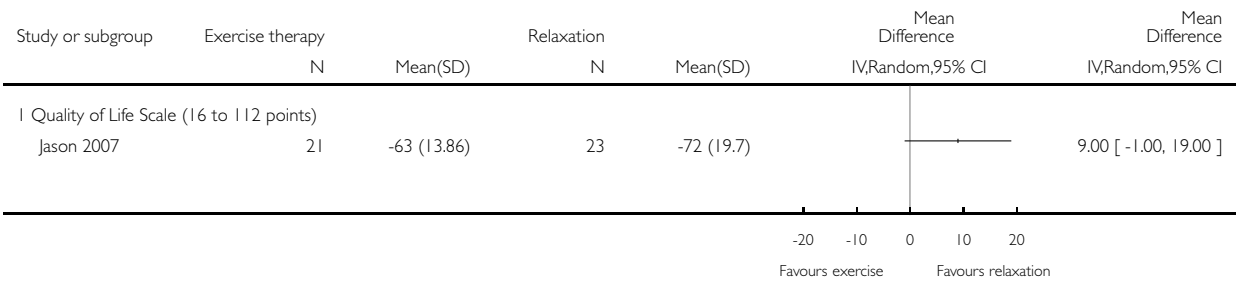


Analysis 1.7. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 7 Quality of life (follow-up).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 7 Quality of life (follow-up)



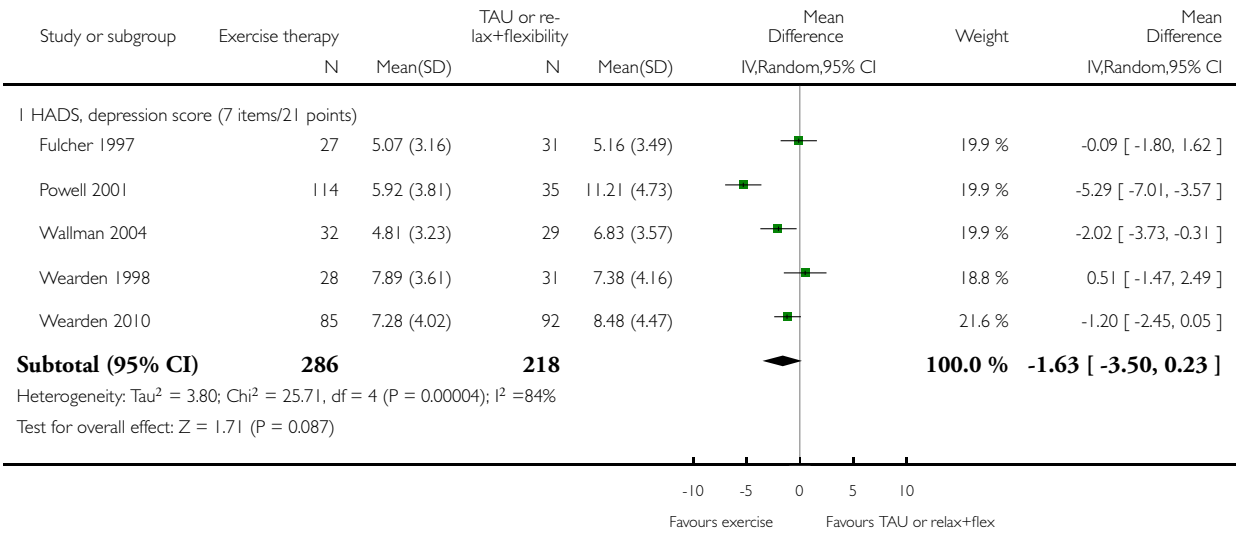
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.8. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 8 Depression (end of treatment).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 8 Depression (end of treatment)



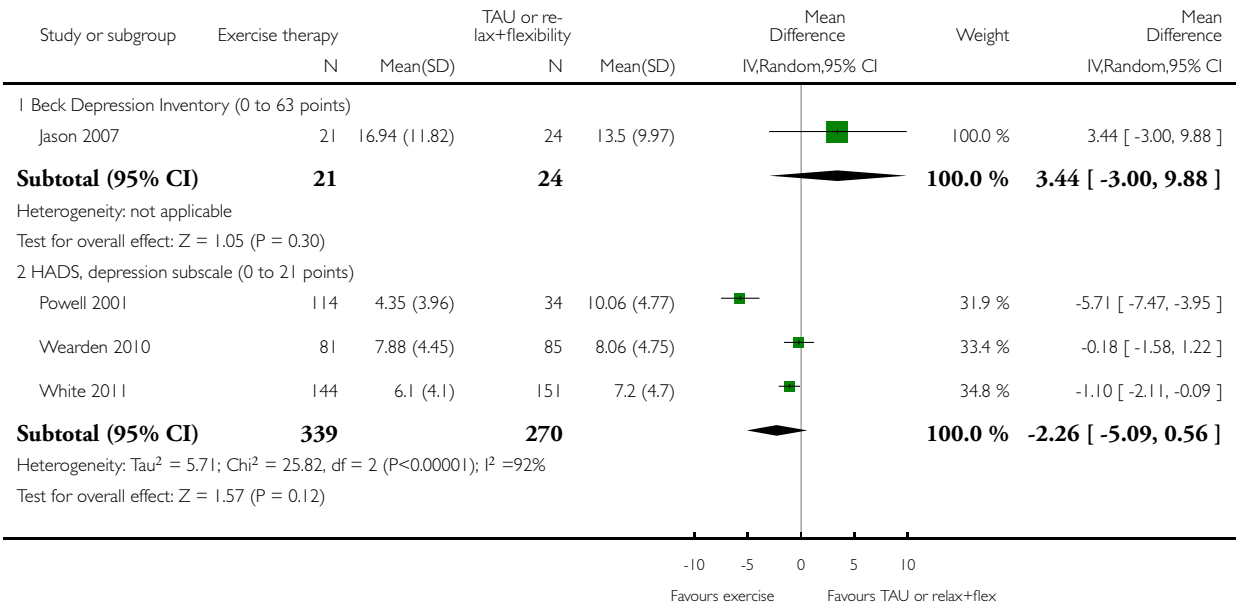
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.9. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 9 Depression (follow-up).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 9 Depression (follow-up)



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.10. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 10 Anxiety (end of treatment).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 10 Anxiety (end of treatment)

Study or subgroup	Exercise therapy N	Mean(SD)	TAU or relax+flex N	Mean(SD)	Mean Difference IV,Random,95% CI	Weight	Mean Difference IV,Random,95% CI
1 HADS, anxiety score (0 to 21 points)							
Powell 2001	114	8.49 (4.51)	35	11.44 (4.84)		32.1 %	-2.95 [-4.75, -1.15]
Wallman 2004 (1)	32	5.7 (3.61)	29	7.8 (3.42)		32.5 %	-2.10 [-3.86, -0.34]
Wearden 2010 (2)	85	9.04 (4.51)	92	8.63 (5.06)		35.4 %	0.41 [-1.00, 1.82]
Subtotal (95% CI)	**231**		**156**			**100.0 %**	**-1.48 [-3.58, 0.61]**

Heterogeneity: $Tau^2 = 2.71$; $Chi^2 = 9.62$, df = 2 (P = 0.01); $I^2 = 79\%$

Test for overall effect: Z = 1.39 (P = 0.16)

Test for subgroup differences: Not applicable



(1) Baseline uncorrected difference in favour of ET (7.3 vs 8.7)

(2) Baseline uncorrected difference in favour of control (11.0 vs 9.7 points)

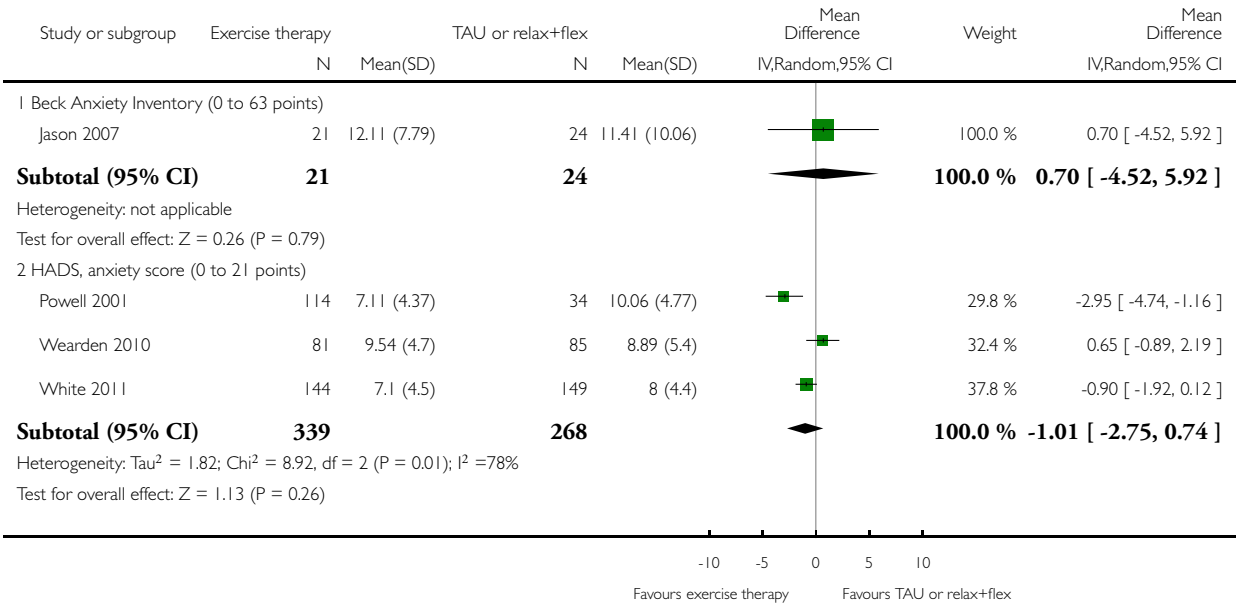
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.11. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 11 Anxiety (follow-up).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 11 Anxiety (follow-up)



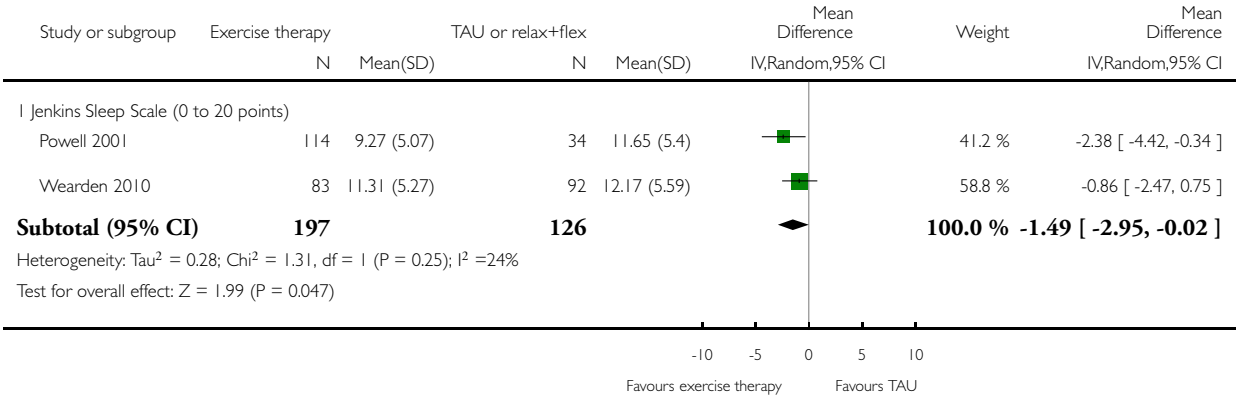
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.12. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 12 Sleep (end of treatment).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 12 Sleep (end of treatment)



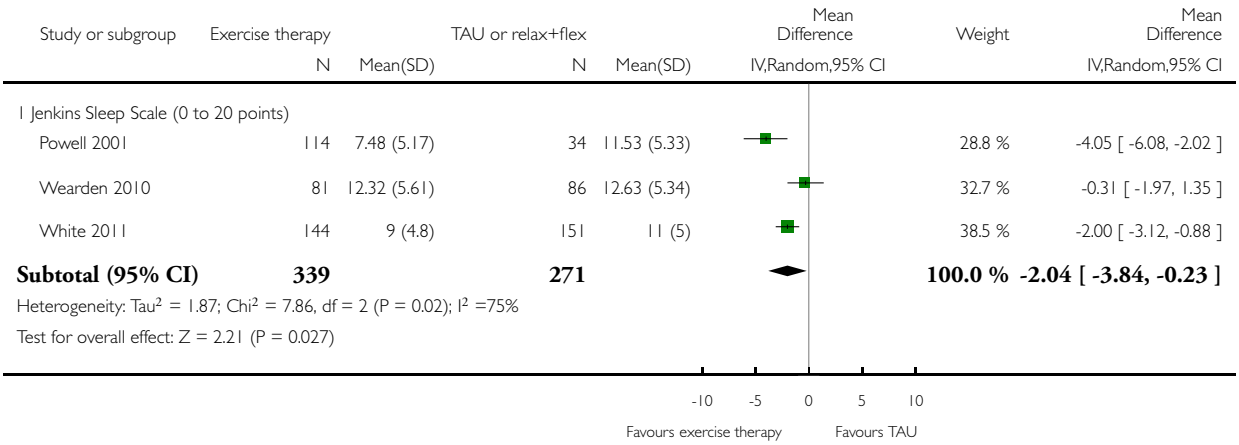
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.13. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 13 Sleep (follow-up).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 13 Sleep (follow-up)



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

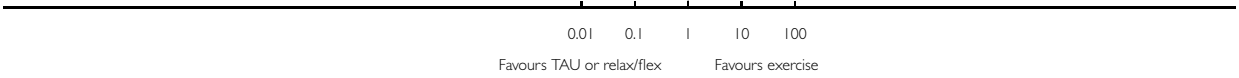
Analysis 1.14. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 14 Self-perceived changes in overall health (end of treatment).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 14 Self-perceived changes in overall health (end of treatment)

Study or subgroup	Experimental n/N	Control n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
Fulcher 1997	16/29	8/30		16.2 %	2.07 [1.05, 4.08]
Moss-Morris 2005	12/25	5/24		9.6 %	2.30 [0.96, 5.56]
Wallman 2004	19/32	12/29		27.6 %	1.43 [0.85, 2.41]
White 2011	54/160	28/160		46.5 %	1.93 [1.29, 2.88]
Total (95% CI)	**246**	**243**		**100.0 %**	**1.83 [1.39, 2.40]**
Total events: 101 (Experimental), 53 (Control)					
Heterogeneity: $Tau^2 = 0.0$; $Chi^2 = 1.32$, df = 3 (P = 0.72); $I^2 = 0.0\%$					
Test for overall effect: Z = 4.33 (P = 0.000015)					
Test for subgroup differences: Not applicable					



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.15. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 15 Self-perceived changes in overall health (follow-up).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 15 Self-perceived changes in overall health (follow-up)

Study or subgroup	Experimental n/N	Control n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
Jason 2007	10/25	12/25		33.7 %	0.83 [0.44, 1.56]
Powell 2001	80/114	4/34		28.3 %	5.96 [2.36, 15.09]
White 2011	62/160	38/160		38.1 %	1.63 [1.16, 2.29]
Total (95% CI)	**299**	**219**		**100.0 %**	**1.88 [0.76, 4.64]**

Total events: 152 (Experimental), 54 (Control)

Heterogeneity: $Tau^2 = 0.53$; $Chi^2 = 13.50$, df = 2 (P = 0.001); $I^2 = 85\%$

Test for overall effect: Z = 1.36 (P = 0.17)

Test for subgroup differences: Not applicable



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

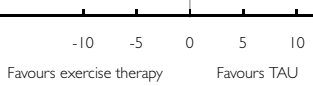
Analysis 1.16. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 16 Health resource use (follow-up) [Mean no. of contacts].

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 16 Health resource use (follow-up) [Mean no. of contacts]

Study or subgroup	Exercise therapy N	Mean(SD)	TAU N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 Primary care						
White 2011	160	6.3 (3.9)	160	7 (4.5)		-0.70 [-1.62, 0.22]
2 Other doctor						
White 2011	160	3.1 (2.9)	160	3.2 (5.6)		-0.10 [-1.08, 0.88]
3 Healthcare professional						
White 2011	160	5.6 (8.3)	160	4.7 (4.7)		0.90 [-0.58, 2.38]
4 Inpatient						
White 2011	160	2.2 (2.4)	160	2.2 (2.3)		0.0 [-0.52, 0.52]
5 Accident and emergency						
White 2011	160	1.6 (1.2)	160	1.8 (1.2)		-0.20 [-0.46, 0.06]
6 Other health/social services						
White 2011	160	7.3 (8.1)	160	7.6 (10.2)		-0.30 [-2.32, 1.72]
7 Complementary health care						
White 2011	160	12.3 (12)	160	10.2 (11.1)		2.10 [-0.43, 4.63]
8 Standardised medical care						
White 2011	160	3.6 (1.4)	160	5 (2.7)		-1.40 [-1.87, -0.93]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

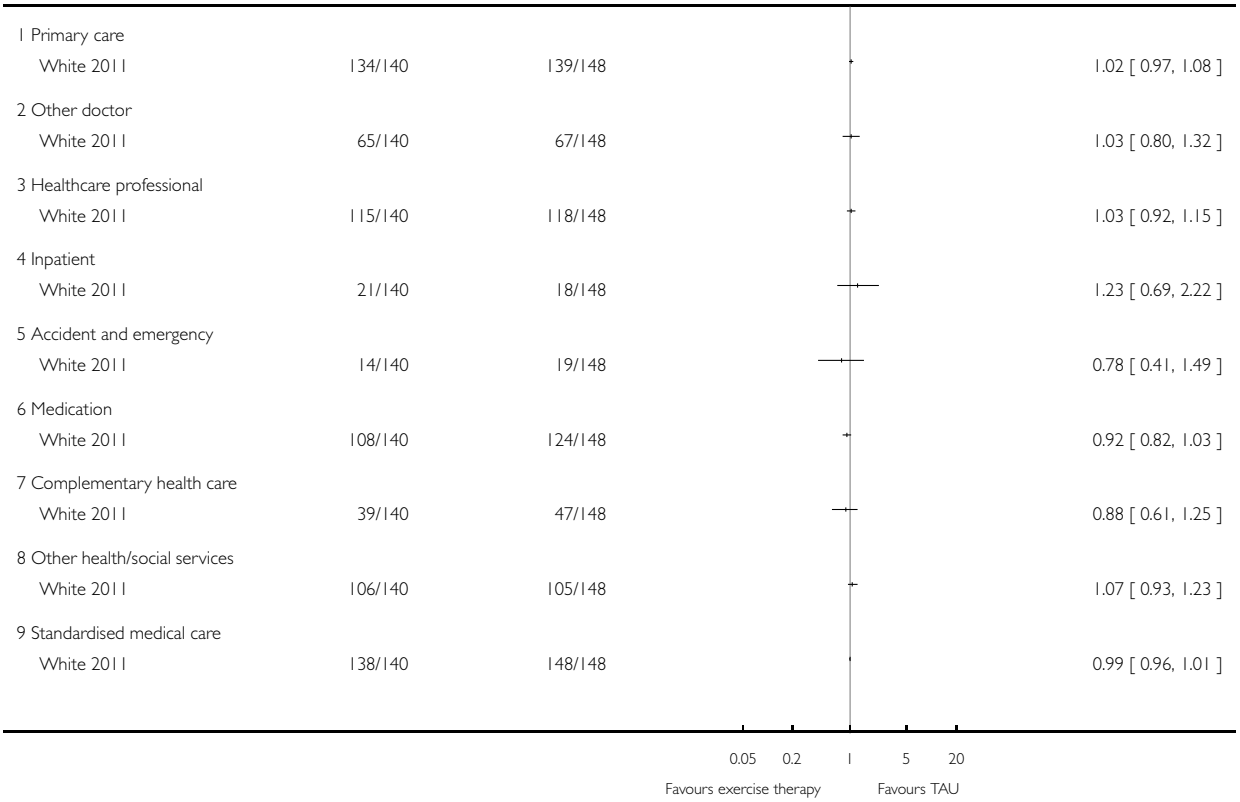
Analysis 1.17. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 17 Health resource use (follow-up) [No. of users].

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 17 Health resource use (follow-up) [No. of users]

Study or subgroup	Exercise therapy n/N	TAU n/N	Risk Ratio M-H,Random,95% CI	Risk Ratio M-H,Random,95% CI
1 Primary care				
White 2011	134/140	139/148		1.02 [0.97, 1.08]
2 Other doctor				
White 2011	65/140	67/148		1.03 [0.80, 1.32]
3 Healthcare professional				
White 2011	115/140	118/148		1.03 [0.92, 1.15]
4 Inpatient				
White 2011	21/140	18/148		1.23 [0.69, 2.22]
5 Accident and emergency				
White 2011	14/140	19/148		0.78 [0.41, 1.49]
6 Medication				
White 2011	108/140	124/148		0.92 [0.82, 1.03]
7 Complementary health care				
White 2011	39/140	47/148		0.88 [0.61, 1.25]
8 Other health/social services				
White 2011	106/140	105/148		1.07 [0.93, 1.23]
9 Standardised medical care				
White 2011	138/140	148/148		0.99 [0.96, 1.01]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.18. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 18 Drop-out.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 18 Drop-out

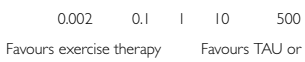
Study or subgroup	Exercise therapy n/N	TAU or relax+flex n/N	Risk Ratio M-H,Random,95% CI	Weight	Risk Ratio M-H,Random,95% CI
Fulcher 1997	4/33	3/33		15.6 %	1.33 [0.32, 5.50]
Moss-Morris 2005	3/25	3/24		14.6 %	0.96 [0.21, 4.30]
Powell 2001	19/114	2/34		15.7 %	2.83 [0.69, 11.56]
Wearden 1998	11/34	5/34		22.7 %	2.20 [0.86, 5.66]
Wearden 2010	12/92	0/100		5.9 %	27.15 [1.63, 452.16]
White 2011	10/160	14/160		25.6 %	0.71 [0.33, 1.56]
Total (95% CI)	**458**	**385**		**100.0 %**	**1.63 [0.77, 3.43]**

Total events: 59 (Exercise therapy), 27 (TAU or relax+flex)

Heterogeneity: $Tau^2 = 0.41$; $Chi^2 = 10.07$, df = 5 (P = 0.07); $I^2 = 50\%$

Test for overall effect: Z = 1.29 (P = 0.20)

Test for subgroup differences: Not applicable



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 1.19. Comparison 1 Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 19 Subgroup analysis for fatigue.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 1 Exercise therapy versus treatment as usual, relaxation or flexibility

Outcome: 19 Subgroup analysis for fatigue

Study or subgroup	Exercise therapy N	Mean(SD)	TAU or relax+flex N	Mean(SD)	Std. Mean Difference IV,Random,95% CI	Weight	Std. Mean Difference IV,Random,95% CI
1 Graded exercise therapy							
Fulcher 1997	27	20.96 (9.08)	30	27.5 (7.44)		12.9 %	-0.78 [-1.32, -0.24]
Moss-Morris 2005	22	13.91 (10.88)	21	24.41 (9.69)		11.4 %	-1.00 [-1.64, -0.36]
Powell 2001	114	4.34 (4.48)	34	10.4 (1.04)		14.9 %	-1.52 [-1.94, -1.10]
Wearden 1998	23	28.13 (13.05)	29	31.58 (8.94)		12.8 %	-0.31 [-0.86, 0.24]
Wearden 2010	85	22.78 (8.56)	92	26.27 (7.68)		16.8 %	-0.43 [-0.73, -0.13]
White 2011	150	21.7 (7.1)	152	24 (6.9)		17.8 %	-0.33 [-0.55, -0.10]
Subtotal (95% CI)	**421**		**358**			**86.6 %**	**-0.71 [-1.09, -0.32]**
Heterogeneity: $Tau^2 = 0.18$; $Chi^2 = 27.87$, df = 5 (P = 0.00004); $I^2 = 82\%$							
Test for overall effect: Z = 3.60 (P = 0.00032)							
2 Exercise with self-pacing							
Wallman 2004	32	11.06 (7.65)	29	15.34 (8.15)		13.4 %	-0.54 [-1.05, -0.02]
Subtotal (95% CI)	**32**		**29**			**13.4 %**	**-0.54 [-1.05, -0.02]**
Heterogeneity: not applicable							
Test for overall effect: Z = 2.05 (P = 0.040)							
Total (95% CI)	**453**		**387**			**100.0 %**	**-0.68 [-1.02, -0.35]**
Heterogeneity: $Tau^2 = 0.15$; $Chi^2 = 27.88$, df = 6 (P = 0.00010); $I^2 = 78\%$							
Test for overall effect: Z = 3.98 (P = 0.000069)							
Test for subgroup differences: $Chi^2 = 0.28$, df = 1 (P = 0.60), $I^2 = 0.0\%$							



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.1. Comparison 2 Exercise therapy versus psychological treatment, Outcome 1 Fatigue at end of treatment (FS; 11 items/0 to 33 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 1 Fatigue at end of treatment (FS; 11 items/0 to 33 points)

Study or subgroup	Exercise Therapy N	Mean(SD)	Psychological treatment N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 CBT						
White 2011	150	21.7 (7.1)	148	21.5 (7.8)		0.20 [-1.49, 1.89]
2 Supportive listening						
Wearden 2010	85	22.78 (8.56)	97	26.81 (6.29)		-4.03 [-6.24, -1.82]



Analysis 2.2. Comparison 2 Exercise therapy versus psychological treatment, Outcome 2 Fatigue at follow-up (FSS; 1 to 7 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 2 Fatigue at follow-up (FSS; 1 to 7 points)

Study or subgroup	Exercise Therapy N	Mean(SD)	Psychological Treatment N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 CT						
Jason 2007	25	5.77 (1.43)	24	5.87 (1.01)		-0.10 [-0.79, 0.59]
2 CBT						
Jason 2007	25	5.77 (1.43)	24	5.37 (1.19)		0.40 [-0.34, 1.14]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

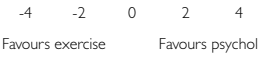
Analysis 2.3. Comparison 2 Exercise therapy versus psychological treatment, Outcome 3 Fatigue at follow-up (FS; 11 items/0 to 33 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 3 Fatigue at follow-up (FS; 11 items/0 to 33 points)

Study or subgroup	Exercise Therapy N	Mean(SD)	Psychological Treatment N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 CBT						
White 2011	154	20.6 (7.5)	148	20.3 (8)		0.30 [-1.45, 2.05]
2 Supportive listening						
Wearden 2010	80	23.9 (8.34)	90	26.62 (7.71)		-2.72 [-5.14, -0.30]



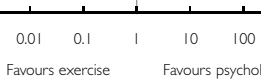
Analysis 2.4. Comparison 2 Exercise therapy versus psychological treatment, Outcome 4 Participants with serious adverse reactions.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 4 Participants with serious adverse reactions

Study or subgroup	Exercise Therapy n/N	Psychological Treatment n/N	Risk Ratio M-H,Random,95% CI	Risk Ratio M-H,Random,95% CI
1 CBT				
White 2011	2/160	3/161		0.67 [0.11, 3.96]
2 Suportive listening				
Wearden 2010	0/85	0/97		Not estimable



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.5. Comparison 2 Exercise therapy versus psychological treatment, Outcome 5 Pain at follow-up (BPI, pain severity subscale; 0 to 10 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 5 Pain at follow-up (BPI, pain severity subscale; 0 to 10 points)

Study or subgroup	Exercise therapy N	Mean(SD)	Psychological Treatment N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 CBT						
Jason 2007	20	3.63 (2.72)	23	3.56 (2.57)		0.07 [-1.52, 1.66]
2 CT						
Jason 2007	20	3.63 (2.72)	24	3.12 (1.96)		0.51 [-0.92, 1.94]



Analysis 2.6. Comparison 2 Exercise therapy versus psychological treatment, Outcome 6 Pain at follow-up (BPI, pain interference subscale; 0 to 10 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 6 Pain at follow-up (BPI, pain interference subscale; 0 to 10 points)

Study or subgroup	Exercise therapy N	Mean(SD)	Psychological Treatment N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 CBT						
Jason 2007	20	3.75 (3.14)	23	4.1 (3.36)		-0.35 [-2.29, 1.59]
2 CT						
Jason 2007	20	3.75 (3.14)	24	3.36 (2.74)		0.39 [-1.37, 2.15]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

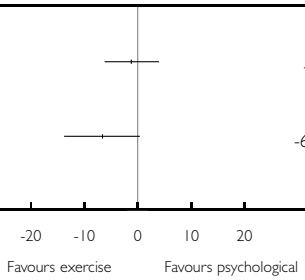
Analysis 2.7. Comparison 2 Exercise therapy versus psychological treatment, Outcome 7 Physical functioning at end of treatment (SF-36, physical functioning subscale; 0 to 100 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 7 Physical functioning at end of treatment (SF-36, physical functioning subscale; 0 to 100 points)

Study or subgroup	Exercise Therapy		Psychological Treatment		Mean Difference	Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Random,95% CI	IV,Random,95% CI
1 CBT						
White 2011	150	-55.4 (23.3)	148	-54.2 (21.6)		-1.20 [-6.30, 3.90]
2 Supportive listening						
Wearden 2010	85	-39.94 (25.21)	96	-33.28 (22.94)		-6.66 [-13.72, 0.40]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

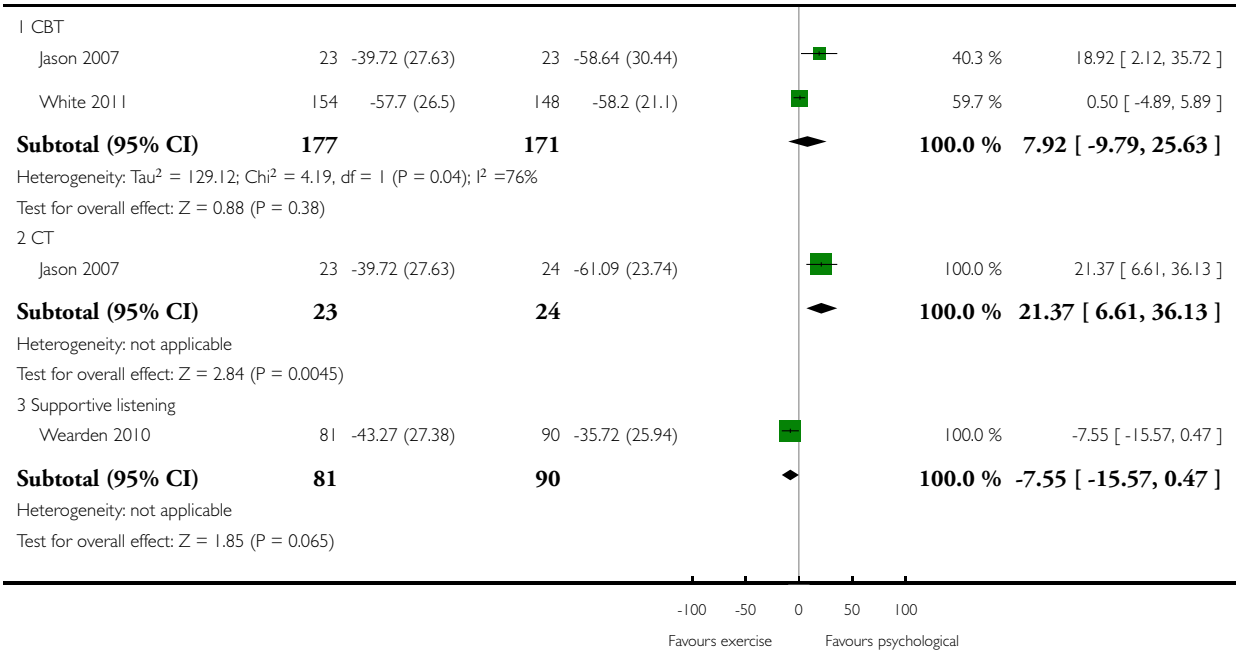
Analysis 2.8. Comparison 2 Exercise therapy versus psychological treatment, Outcome 8 Physical functioning at follow-up (SF-36, physical functioning subscale; 0 to 100 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 8 Physical functioning at follow-up (SF-36, physical functioning subscale; 0 to 100 points)

Study or subgroup	Exercise Therapy		Psychological Treatment		Mean Difference	Weight	Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Random,95% CI		IV,Random,95% CI
1 CBT							
Jason 2007	23	-39.72 (27.63)	23	-58.64 (30.44)		40.3 %	18.92 [2.12, 35.72]
White 2011	154	-57.7 (26.5)	148	-58.2 (21.1)		59.7 %	0.50 [-4.89, 5.89]
Subtotal (95% CI)	**177**		**171**			**100.0 %**	**7.92 [-9.79, 25.63]**
Heterogeneity: $Tau^2 = 129.12$; $Chi^2 = 4.19$, df = 1 (P = 0.04); $I^2 = 76\%$							
Test for overall effect: Z = 0.88 (P = 0.38)							
2 CT							
Jason 2007	23	-39.72 (27.63)	24	-61.09 (23.74)		100.0 %	21.37 [6.61, 36.13]
Subtotal (95% CI)	**23**		**24**			**100.0 %**	**21.37 [6.61, 36.13]**
Heterogeneity: not applicable							
Test for overall effect: Z = 2.84 (P = 0.0045)							
3 Supportive listening							
Wearden 2010	81	-43.27 (27.38)	90	-35.72 (25.94)		100.0 %	-7.55 [-15.57, 0.47]
Subtotal (95% CI)	**81**		**90**			**100.0 %**	**-7.55 [-15.57, 0.47]**
Heterogeneity: not applicable							
Test for overall effect: Z = 1.85 (P = 0.065)							



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

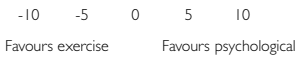
Analysis 2.9. Comparison 2 Exercise therapy versus psychological treatment, Outcome 9 Depression at end of treatment (HADS depression score; 7 items/21 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 9 Depression at end of treatment (HADS depression score; 7 items/21 points)

Study or subgroup	Exercise therapy		Psychological Treatment		Mean Difference	Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Random,95% CI	IV,Random,95% CI
1 Supportive listening						
Wearden 2010	85	7.28 (4.02)	97	8.85 (4.01)		-1.57 [-2.74, -0.40]



Analysis 2.10. Comparison 2 Exercise therapy versus psychological treatment, Outcome 10 Depression at follow-up (BDI; 0 to 63 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 10 Depression at follow-up (BDI; 0 to 63 points)

Study or subgroup	Exercise Therapy		Psychological Treatment		Mean Difference	Mean Difference
	N	Mean(SD)	N	Mean(SD)	IV,Random,95% CI	IV,Random,95% CI
1 CT						
Jason 2007	21	16.94 (11.82)	24	11.86 (7.36)		5.08 [-0.77, 10.93]
2 CBT						
Jason 2007	21	16.94 (11.82)	23	13.95 (13.08)		2.99 [-4.37, 10.35]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.11. Comparison 2 Exercise therapy versus psychological treatment, Outcome 11 Depression at follow-up (HADS depression score; 7 items/21 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 11 Depression at follow-up (HADS depression score; 7 items/21 points)

Study or subgroup	Exercise Therapy N	Mean(SD)	Psychological Treatment N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 CBT						
White 2011	144	6.1 (4.1)	143	6.2 (3.7)		-0.10 [-1.00, 0.80]
2 Supportive listening						
Wearden 2010	81	7.88 (4.45)	90	8.67 (4.51)		-0.79 [-2.13, 0.55]



Analysis 2.12. Comparison 2 Exercise therapy versus psychological treatment, Outcome 12 Anxiety at end of treatment (HADS anxiety; 7 items/21 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 12 Anxiety at end of treatment (HADS anxiety; 7 items/21 points)

Study or subgroup	Exercise Therapy N	Mean(SD)	Psychological Treatment N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 Supportive listening						
Wearden 2010	85	9.04 (4.51)	97	9.52 (4.93)		-0.48 [-1.85, 0.89]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.13. Comparison 2 Exercise therapy versus psychological treatment, Outcome 13 Anxiety at follow-up (BAI; 0 to 63 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 13 Anxiety at follow-up (BAI; 0 to 63 points)

Study or subgroup	Exercise Therapy N	Mean(SD)	Psychological Treatment N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 CT						
Jason 2007	21	12.11 (7.79)	24	8.96 (6.87)		3.15 [-1.17, 7.47]
2 CBT						
Jason 2007	21	12.11 (7.79)	23	11.45 (10.22)		0.66 [-4.68, 6.00]



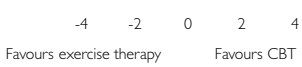
Analysis 2.14. Comparison 2 Exercise therapy versus psychological treatment, Outcome 14 Anxiety at follow-up (HADS anxiety; 7 items/21 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 14 Anxiety at follow-up (HADS anxiety; 7 items/21 points)

Study or subgroup	Exercise Therapy N	Mean(SD)	Psychological Treatment N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 CBT						
White 2011	144	7.1 (4.5)	143	6.8 (4.2)		0.30 [-0.71, 1.31]
2 Supportive listening						
Wearden 2010	81	9.54 (4.7)	90	9.62 (4.87)		-0.08 [-1.52, 1.36]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.15. Comparison 2 Exercise therapy versus psychological treatment, Outcome 15 Sleep at end of treatment (Jenkins Sleep Scale; 0 to 20 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 15 Sleep at end of treatment (Jenkins Sleep Scale; 0 to 20 points)

Study or subgroup	Exercise Therapy N	Mean(SD)	Psychological Treatment N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 Supportive listening						
Wearden 2010	83	11.31 (5.27)	97	13.77 (5.29)		-2.46 [-4.01, -0.91]



Analysis 2.16. Comparison 2 Exercise therapy versus psychological treatment, Outcome 16 Sleep at follow-up (Jenkins Sleep Scale; 0 to 20 points).

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 16 Sleep at follow-up (Jenkins Sleep Scale; 0 to 20 points)

Study or subgroup	Exercise Therapy N	Mean(SD)	Psychological Treatment N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 CBT						
White 2011	144	9 (4.8)	143	9.9 (5.3)		-0.90 [-2.07, 0.27]
2 Supportive listening						
Wearden 2010	81	12.32 (5.61)	90	13.18 (5.71)		-0.86 [-2.56, 0.84]



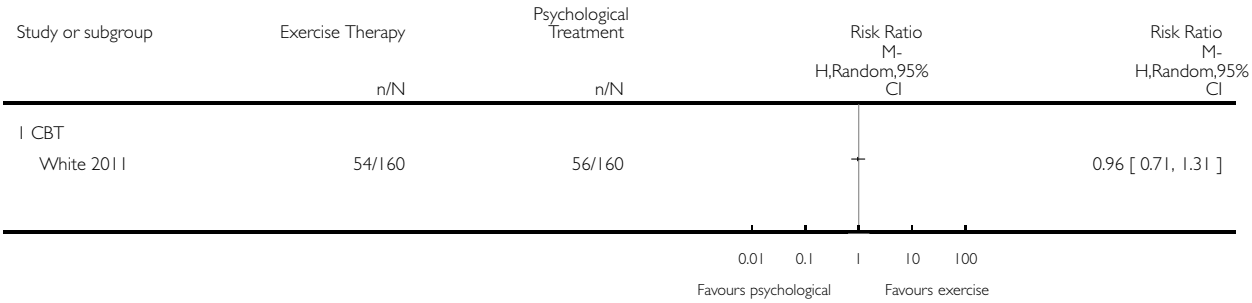
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.17. Comparison 2 Exercise therapy versus psychological treatment, Outcome 17 Self-perceived changes in overall health at end of treatment.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 17 Self-perceived changes in overall health at end of treatment



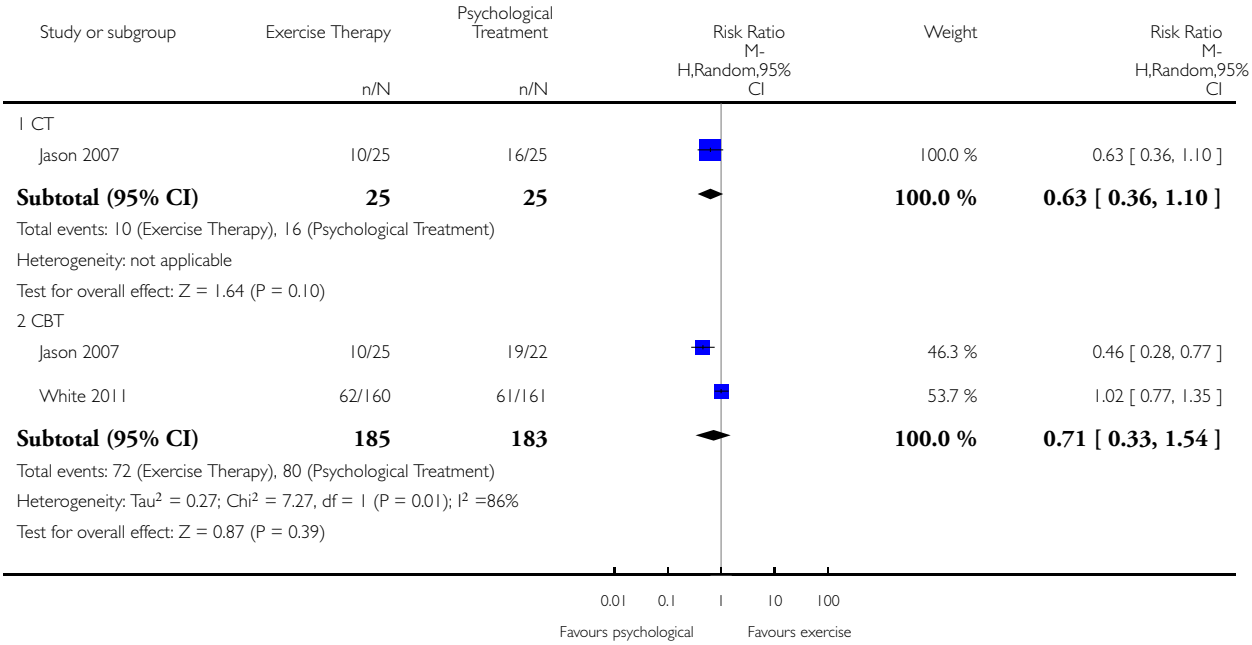
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.18. Comparison 2 Exercise therapy versus psychological treatment, Outcome 18 Self-perceived changes in overall health at follow-up.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 18 Self-perceived changes in overall health at follow-up



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

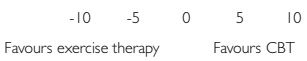
Analysis 2.19. Comparison 2 Exercise therapy versus psychological treatment, Outcome 19 Health resource use (follow-up) [Mean no. of contacts].

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 19 Health resource use (follow-up) [Mean no. of contacts]

Study or subgroup	Exercise therapy N	Mean(SD)	CBT N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 Primary care						
White 2011	160	6.3 (3.9)	161	6.6 (5.6)		-0.30 [-1.36, 0.76]
2 Other doctor						
White 2011	160	3.1 (2.9)	161	2.5 (2)		0.60 [0.05, 1.15]
3 Healthcare professional						
White 2011	160	5.6 (8.3)	161	4.4 (5.9)		1.20 [-0.38, 2.78]
4 Inpatient						
White 2011	160	2.2 (2.4)	161	1.4 (0.7)		0.80 [0.41, 1.19]
5 Accident and emergency						
White 2011	160	1.6 (1.2)	161	1.4 (0.7)		0.20 [-0.02, 0.42]
6 Other health/social services						
White 2011	160	7.3 (8.1)	161	6.3 (9.2)		1.00 [-0.90, 2.90]
7 Complementary health care						
White 2011	160	12.3 (12)	161	10 (14.4)		2.30 [-0.60, 5.20]
8 Standardised medical care						
White 2011	160	3.6 (1.4)	161	3.7 (2.2)		-0.10 [-0.50, 0.30]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.20. Comparison 2 Exercise therapy versus psychological treatment, Outcome 20 Health resource use (follow-up) [No. of users].

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 20 Health resource use (follow-up) [No. of users]

Study or subgroup	Exercise therapy n/N	CBT n/N	Risk Ratio M-H,Random,95% CI	Risk Ratio M-H,Random,95% CI
1 Primary care				
White 2011	134/140	134/145		1.04 [0.98, 1.10]
2 Other doctor				
White 2011	65/140	71/145		0.95 [0.74, 1.21]
3 Healthcare professional				
White 2011	115/140	110/145		1.08 [0.96, 1.22]
4 Inpatient				
White 2011	21/140	16/145		1.36 [0.74, 2.50]
5 Accident and emergency				
White 2011	14/140	22/145		0.66 [0.35, 1.24]
6 Medication				
White 2011	108/140	117/145		0.96 [0.85, 1.08]
7 Complementary health care				
White 2011	39/140	32/145		1.26 [0.84, 1.89]
8 Other health/social services				
White 2011	106/140	110/145		1.00 [0.88, 1.14]
9 Standardised medical care				
White 2011	138/140	145/145		0.99 [0.96, 1.01]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 2.21. Comparison 2 Exercise therapy versus psychological treatment, Outcome 21 Drop-out.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 2 Exercise therapy versus psychological treatment

Outcome: 21 Drop-out

Study or subgroup	Exercise therapy n/N	Psychological Treatment n/N	Risk Ratio M-H,Random,95% CI	Risk Ratio M-H,Random,95% CI
1 CBT				
White 2011	10/160	17/161		0.59 [0.28, 1.25]
2 Supportive listening				
Wearden 2010	12/92	7/91		1.70 [0.70, 4.11]



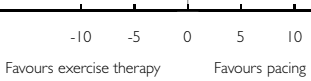
Analysis 3.1. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 1 Fatigue.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 3 Exercise therapy versus adaptive pacing

Outcome: 1 Fatigue

Study or subgroup	Exercise Therapy N	Mean(SD)	Pacing N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 Fatigue Scale, FS (11 items/33 points)—end of treatment						
White 2011	150	21.7 (7.1)	155	23.7 (6.9)		-2.00 [-3.57, -0.43]
2 Fatigue Scale, FS (11 items/33 points)—follow-up						
White 2011	154	20.6 (7.5)	153	23.1 (7.3)		-2.50 [-4.16, -0.84]



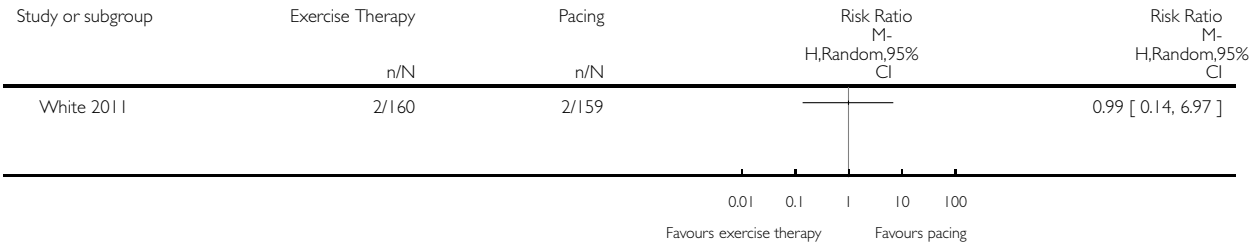
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 3.2. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 2 Participants with serious adverse reactions.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 3 Exercise therapy versus adaptive pacing

Outcome: 2 Participants with serious adverse reactions

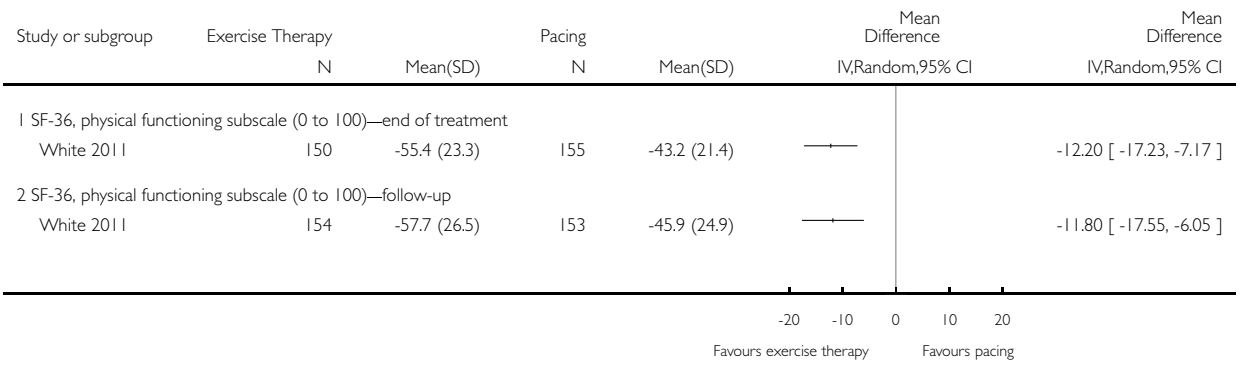


Analysis 3.3. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 3 Physical functioning.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 3 Exercise therapy versus adaptive pacing

Outcome: 3 Physical functioning



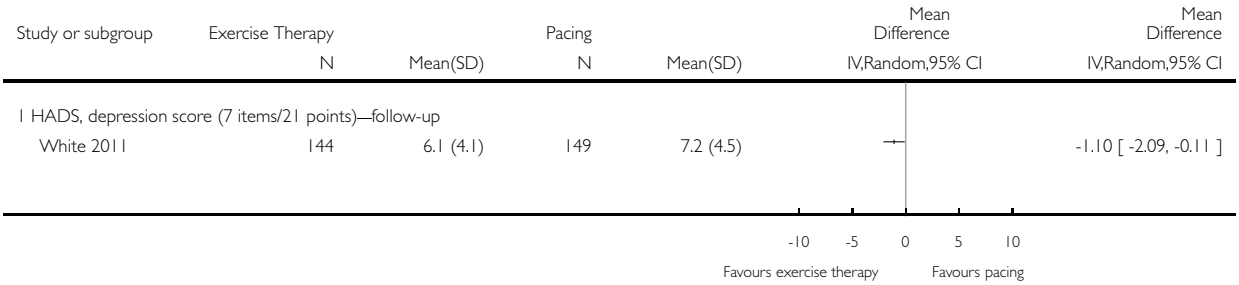
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 3.4. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 4 Depression.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 3 Exercise therapy versus adaptive pacing

Outcome: 4 Depression

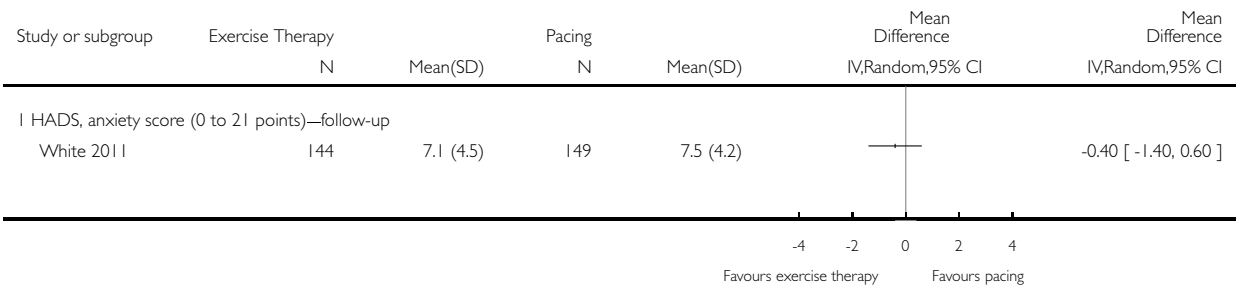


Analysis 3.5. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 5 Anxiety.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 3 Exercise therapy versus adaptive pacing

Outcome: 5 Anxiety



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

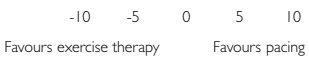
Analysis 3.6. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 6 Sleep.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 3 Exercise therapy versus adaptive pacing

Outcome: 6 Sleep

Study or subgroup	Exercise Therapy N	Mean(SD)	Pacing N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 Jenkins Sleep Scale (0 to 20 points)—follow-up						
White 2011	144	9 (4.8)	150	10.6 (4.8)		-1.60 [-2.70, -0.50]



Analysis 3.7. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 7 Self-perceived changes in overall health.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 3 Exercise therapy versus adaptive pacing

Outcome: 7 Self-perceived changes in overall health

Study or subgroup	Exercise therapy n/N	Pacing n/N	Risk Ratio M-H,Random,95% CI	Risk Ratio M-H,Random,95% CI
1 End of treatment				
White 2011	54/160	37/159		1.45 [1.02, 2.07]
2 Follow-up				
White 2011	62/160	47/159		1.31 [0.96, 1.79]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

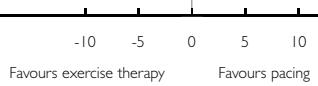
Analysis 3.8. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 8 Health resource use (follow-up) [Mean no. of contacts].

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 3 Exercise therapy versus adaptive pacing

Outcome: 8 Health resource use (follow-up) [Mean no. of contacts]

Study or subgroup	Exercise therapy N	Mean(SD)	Pacing N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 Primary care						
White 2011	160	6.3 (3.9)	161	7.1 (5.7)		-0.80 [-1.87, 0.27]
2 Other doctor						
White 2011	160	3.1 (2.9)	161	2.4 (2.2)		0.70 [0.14, 1.26]
3 Healthcare professional						
White 2011	160	5.6 (8.3)	161	5.3 (7.9)		0.30 [-1.47, 2.07]
4 Inpatient						
White 2011	160	2.2 (2.4)	161	3.2 (2.5)		-1.00 [-1.54, -0.46]
5 Accident and emergency						
White 2011	160	1.6 (1.2)	161	1.1 (0.3)		0.50 [0.31, 0.69]
6 Other health/social services						
White 2011	160	7.3 (8.1)	161	6.3 (6.7)		1.00 [-0.63, 2.63]
7 Complementary health care						
White 2011	160	12.3 (12)	161	8.5 (9.6)		3.80 [1.42, 6.18]
8 Standardised medical care						
White 2011	160	3.6 (1.4)	161	3.6 (1.7)		0.0 [-0.34, 0.34]



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

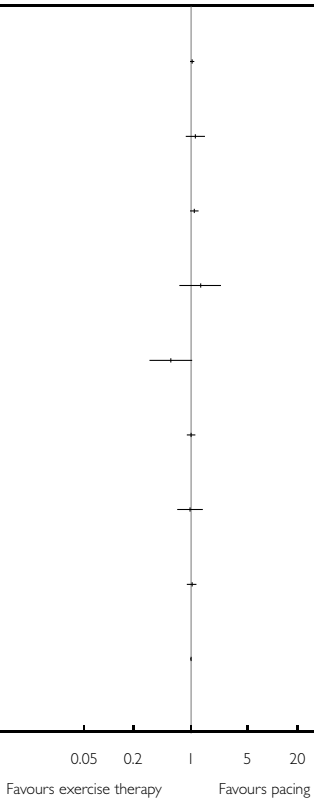
Analysis 3.9. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 9 Health resource use (follow-up) [No. of users].

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 3 Exercise therapy versus adaptive pacing

Outcome: 9 Health resource use (follow-up) [No. of users]

Study or subgroup	Exercise therapy n/N	Pacing n/N	Risk Ratio M-H,Random,95% CI	Risk Ratio M-H,Random,95% CI
1 Primary care				
White 2011	134/140	134/146		1.04 [0.98, 1.11]
2 Other doctor				
White 2011	65/140	60/146		1.13 [0.87, 1.47]
3 Healthcare professional				
White 2011	115/140	109/146		1.10 [0.97, 1.24]
4 Inpatient				
White 2011	21/140	17/146		1.29 [0.71, 2.34]
5 Accident and emergency				
White 2011	14/140	26/146		0.56 [0.31, 1.03]
6 Medication				
White 2011	108/140	112/146		1.01 [0.89, 1.14]
7 Complementary health care				
White 2011	39/140	42/146		0.97 [0.67, 1.40]
8 Other health/social services				
White 2011	106/140	108/146		1.02 [0.89, 1.17]
9 Standardised medical care				
White 2011	138/140	146/146		0.99 [0.96, 1.01]



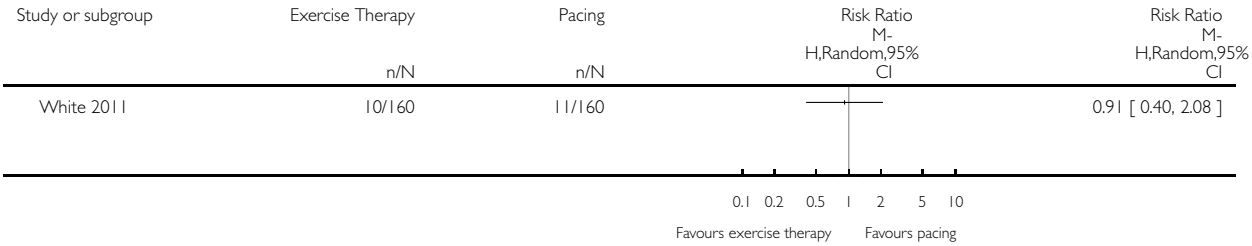
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 3.10. Comparison 3 Exercise therapy versus adaptive pacing, Outcome 10 Drop-out.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 3 Exercise therapy versus adaptive pacing

Outcome: 10 Drop-out

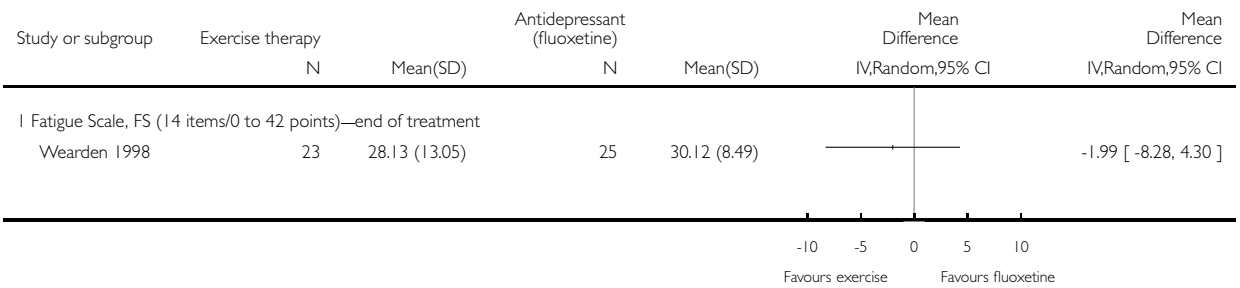


Analysis 4.1. Comparison 4 Exercise therapy + antidepressant placebo versus antidepressant + exercise placebo, Outcome 1 Fatigue.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 4 Exercise therapy + antidepressant placebo versus antidepressant + exercise placebo

Outcome: 1 Fatigue



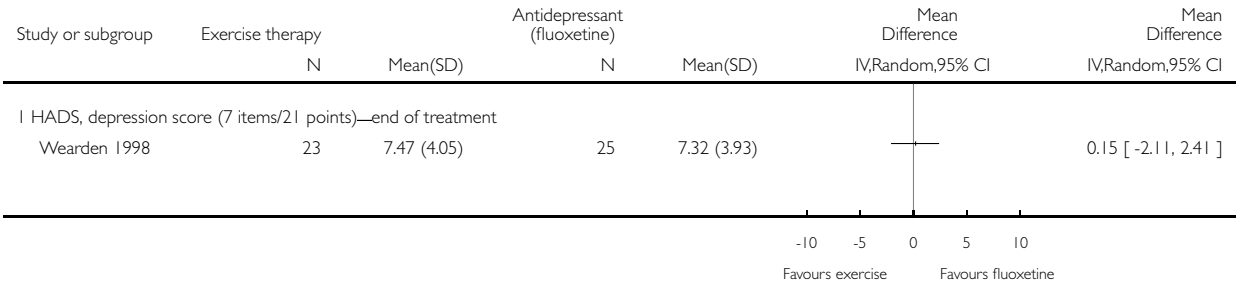
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 4.2. Comparison 4 Exercise therapy + antidepressant placebo versus antidepressant + exercise placebo, Outcome 2 Depression.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 4 Exercise therapy + antidepressant placebo versus antidepressant + exercise placebo

Outcome: 2 Depression

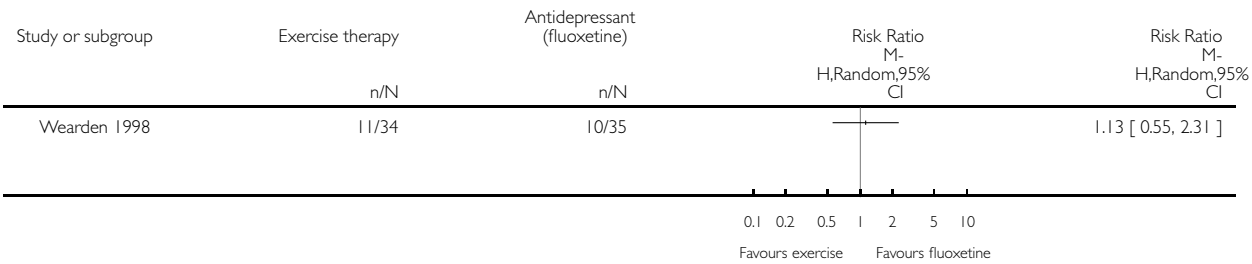


Analysis 4.3. Comparison 4 Exercise therapy + antidepressant placebo versus antidepressant + exercise placebo, Outcome 3 Drop-out.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 4 Exercise therapy + antidepressant placebo versus antidepressant + exercise placebo

Outcome: 3 Drop-out



Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

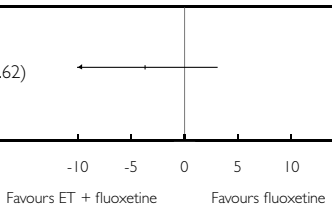
Analysis 5.1. Comparison 5 Exercise therapy + antidepressant versus antidepressant + exercise placebo, Outcome 1 Fatigue.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 5 Exercise therapy + antidepressant versus antidepressant + exercise placebo

Outcome: 1 Fatigue

Study or subgroup	Exercise + Antidepressant N	Mean(SD)	Antidepressant N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 Fatigue Scale, FS (14 items/0 to 42 points)—end of treatment						
Wearden 1998	19	26.26 (12.9)	24	29.92 (8.62)		-3.66 [-10.41, 3.09]



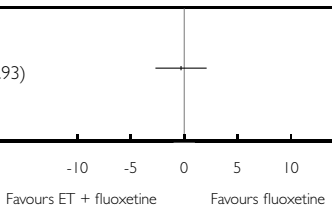
Analysis 5.2. Comparison 5 Exercise therapy + antidepressant versus antidepressant + exercise placebo, Outcome 2 Depression.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 5 Exercise therapy + antidepressant versus antidepressant + exercise placebo

Outcome: 2 Depression

Study or subgroup	Exercise + antidepressant N	Mean(SD)	Antidepressant N	Mean(SD)	Mean Difference IV,Random,95% CI	Mean Difference IV,Random,95% CI
1 HADS, depression score (7 items/21 points)—end of treatment						
Wearden 1998	19	7.05 (4.13)	25	7.32 (3.93)		-0.27 [-2.68, 2.14]



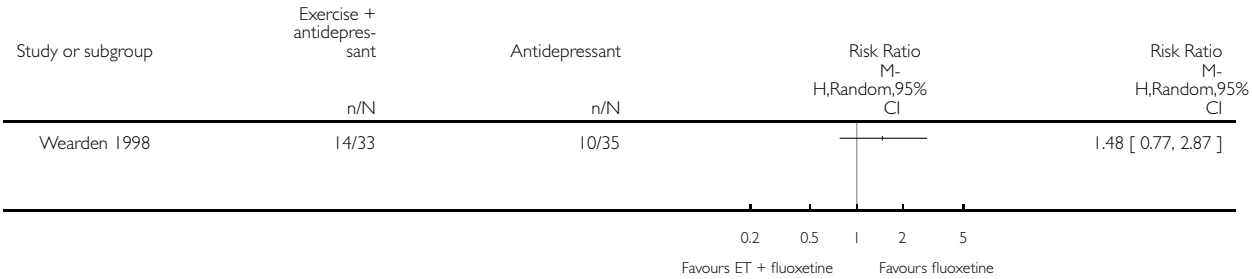
Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Analysis 5.3. Comparison 5 Exercise therapy + antidepressant versus antidepressant + exercise placebo, Outcome 3 Drop-out.

Review: Exercise therapy for chronic fatigue syndrome

Comparison: 5 Exercise therapy + antidepressant versus antidepressant + exercise placebo

Outcome: 3 Drop-out



ADDITIONAL TABLES

Table 1. Study demographics

Study ID	N	Gender	Duration of illness	Depression comorbidity	Use of antidepressants (ADs)	Work and employment status
Fulcher 1997	66	49F/17M 65% female	2.7 years	20 (30%) possible cases of depression (HADS)	30 (45%) on full-dose AD (n = 20) or low-dose AD (n = 10)	26 (39%) working or studying at least part time
Jason 2007	114	95F/19M 83% female	> 5.0 years	44 (39%) with a current Axis I disorder (depression and anxiety most common)	Not stated	52 (46%) working or studying at least part time, 24% unemployed, 6% retired, 25% on disability
Moss-Morris 2005	49	34F/15M 69% female	3.1 years	14 (29%) possible or probable cases of depression (HADS)	Not stated	11 (22%) were unemployed and were unable to work because of disability
Powell 2001	148	116F/32M 78% female	4.3 years	58 (39%) possible or probable cases of depression (HADS)	27 (18%) used AD	50 (34%) were working, 64 (43%) were on disability

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Table 1. Study demographics (*Continued*)

Wallman 2004	61	47F/14M 77% female	Not stated	Not stated	16 (26%) used AD	Not stated
Wearden 1998	136	97F/39M 71% female	2.3 years	46 (34%) with depressive disorder according to DSM-III-R criteria	Not stated	114 (84%) had recently changed occupation
Wearden 2010	296	230F/66M 78% female	7.0 years	53 (18%) had a depression diagnosis	160 (54%) were prescribed AD in the past 6 months	Not stated
White 2011	641	495F/146M 77% female	2.7 years	219 (34%) with any depressive disorder	260 (41%) used AD	Not stated

Table 2. Characteristics of exercise interventions

Study ID	Deliverer of intervention	Explanation and materials	Type of exercise	Schedule therapist	Schedule home	Duration of activity	Initial exercise level	Increment steps	Participant self-monitoring	Criteria for (non)-increment
Fulcher 1997	Exercise physiologist	Verbal explanation of deconditioning and reconditioning	Walking (encouraged to take other modes such as cycling and swimming)	Weekly (1 hour), talking only	5 days/wk	5 to 15 minutes increasing to 30 minutes/d	5 to 15 minutes at 40% of peak O_2 consumption (target HR of resting + 50% of HRR)	Duration increased 1 to 2 minutes per week up to 30 minutes; then intensity increased	Ambulatory heart rate monitors	If increased fatigue, continue at the same level for an extra week
Wearden 1998	Physiotherapist, fitness focus	Minimal explanation; no written materials	Preferred activity (walking/jogging, some did cycling, swimming)	At week 0, 1, 2, 4, 8, 12*, 20, 26*, talking only (*evaluation visits)	3 days/wk	20 minutes	75% of VO_2max from bike test	Intensity increased	Borg Exertion Scale chart, before and after HR	Increase if: 10 beats/min drop post exercise and 2-point drop in Borg Scale score

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Table 2. Characteristics of exercise interventions (*Continued*)

Powell 2001	Senior clinical therapist	Explanations for GET, circadian dysrhythmia, deconditioning, sleep "educational information pack"	Aerobic exercise; own choice but mostly exercise bike	9 face-to-face (1.5 hours each)	Tailored	Tailored to functional abilities	Tailored to functional abilities: "a level which you are capable of doing on a BAD DAY"	Varying daily increase (e.g. "5 second increase each day for the rest of the second week" to 30 minutes twice/d	Duration of exercise	Discouraged, but restart at lower level and rapidly reincrease
Wallman 2004	Single physical therapist	Small laminated Borg Scale and heart rate monitor	Walking/jogging, swimming or cycling	Phone contact every 2 weeks	Every second day	From 5 to 15 minutes, increasing to 30 minutes	Initial exercise duration was between 5 and 15 minutes, and intensity was based on the mean HR value achieved midpoint during submaximal exercise tests	Duration increased by 2 to 5 minutes/2 wk	Heart rate monitoring, Borg Exertion Scale	Keep Borg within 11 to 14. Adjust every 2 weeks. Average peak HR when exercising comfortably at a typical day represents patient's target heart rate (± 3 bpm) for future sessions
Moss-Morris 2005	Health psychology MSc student, researcher	Focused on the "downward spiral of activity reduction, deconditioning"	Walking (but could also do other preferred exercise, e.g. jogging, swimming)	Weekly for 12 weeks, talking only	4 to 5 days/wk	Set collaboratively approx 5 to 15 minutes	HR at 40% of VO_2max	Duration 3 to 5 minutes/wk Intensity increased after 6 weeks 5 bpm/wk	Ambulatory heart rate monitors	If increased fatigue, continue at the same level for an extra week

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Table 2. Characteristics of exercise interventions (*Continued*)

Jason 2007	Registered nurses supervised by exercise physiologist	"Behavioral goals explained, energy system education, redefining exercise"	"individualized, constructive and pleasurable activities"	Every 2 weeks (45 minutes), 13 sessions	3 per week	Tailored	Flexibility tests Strength test (hand grip)	"Gradually increasing anaerobic activity levels"	Self-monitoring daily exercise diary	New targets only after habituation, or if goals achieved for 2 weeks
Wearden 2010	Nurses with 16 half-days of training and supervision	Explanation of physiological symptoms and training in first session	Wide choice: walking, stairs, bicycle, dance, jog	10 sessions over 18 weeks	Several times per day	First 90 minutes, then alternating 60 and 30 minutes	Determined collaboratively with the participant	"Increased very gradually," examples show 50% increase per day	Diary of progress on exercise programme, with note of daily activities	On "bad days," try to do same as day before
White 2011	Exercise therapist/ physiotherapist (8 to 10 days training + ongoing supervision)	142-page manual: benefits of exercise and "how to" of GET; some got pedometers	Wide choice: walking, cycling, swimming, Tai Chi. Aim to build into daily activities	Weekly × 4, then fortnightly; total of 15 sessions	5 to 6 days/wk	Negotiated, goal to get to 30 minutes per session	Test of fitness (step test. and 6-minute walking test), perceived physical exertion, actigraphy data	"20% increases" per fortnight; increase duration to 30 minutes, then increase intensity	Exercise diary + Borg scale + "Use non-symptoms to monitor" and heart rate monitor (for intensity increases)	Do not increase if global increase in symptoms

© 9. March 2012, Paul Glasziou, Bond University, Australia

© 9. March 2012
Bond University

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

APPENDICES

Appendix 1. Search strategy-CCDANCTR-References

CCDANCTR-References Register

(fatigue* or asthenia or "muscular disorder*" or neurasthenia* or "infectious mononucleos*" or "myalgic encephalomyelit*" or "royal free disease*" or lassitude or "muscular weakness*" or "akureyri disease" or "atypical poliomyelitis" or CFIDS or CFS or (chronic and mononucleos*) or "epidemic neuromyasthenia" or "iceland disease" or "post infectious encephalomyelitis" or PVFS or tiredness or adynamia or legasthenia or (perspective and asthenia) or neurataxia or ("muscle strength" and loss) or "muscle* weak*" or "weak* muscle*" or (muscular and insufficiency) or (neuromuscular and fatigue))

and

exercise or "physical fitness" or "physical education" or "physical condition*" or "physical train*" or "physical mobility" or "physical activ*" or "physical exertion" or "physical effort*" or (breathing and (therap* or exercise*)) or (respiration and therap*) or "gi gong" or gigong or *kung or tai or thai or taiji or taijiquan or taichi or walking or yoga or relaxation* or gymnastics or calisthenics or aerobic or danc* or jumping or hopping or running or jogging or ambulat* or "muscle strengthening" or (muscular and (strength or resistance)) or ((weight or weights) and lifting) or weightlifting or "power lifting" or "weight train*" or pilates or stretching or plyometric* or "cardiopulmonary conditioning" or "motion therap*" or "neuromuscular facilitation*" or "movement therap*" or ((recreation or activity) and therap*) or "isometric training" or climbing or cycling or bicycle* or "lifting effort*" or swim* or (training and (technical or course or program*)) or writing or kinesi* or gardening or multiconvergent)

Appendix 2. Other search strategies

SPORTDiscus (EBSCOHost)

1. exp Exercise/
2. exp Exercise Therapy/
3. exp Exercise Movement Techniques/
4. Physical Fitness/
5. exp "Physical Education and Training"/
6. (exercise$ or exercising).tw.
7. ((breathing or respiration) adj (therap$ or exercise$)).tw.
8. (gi gong or gigong).tw.
9. relaxation$.tw.
10. ((tai adj ji) or ((tai or thai) adj chi) or taiji or taijiquan or taichi).tw.
11. walking.tw.
12. yoga.tw.
13. (physical adj (fitness or condition$ or education or training or mobility or activit$ or exertion or effort)).tw.
14. gymnastics.tw.
15. calisthenics.tw.
16. aerobic danc$.tw.
17. danc$.tw.
18. (jumping or hopping).tw.
19. (running or jogging).tw.
20. ambulat$.tw.
21. muscle strengthening.tw.
22. (muscular adj (strength or resistance) adj training).tw.
23. ((weight$1 adj2 lifting) or weightlifting or power lifting or weight training).tw.
24. pilates.tw.
25. stretching.tw.
26. plyometric$.tw.
27. cardiopulmonary conditioning.tw.
28. motion therap$.tw.
29. neuromuscular facilitation$.tw.
30. movement therap$.tw.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

31. ((recreation or activity) adj therap$).tw.
32. gymnastic therap$.tw.
33. isometric training.tw.
34. climbing.tw.
35. cycling.tw.
36. lifting effort$.tw.
37. swimming.tw.
38. writing.tw.
39. technical training.tw.
40. (training adj (course$ or program$)).tw.
41. (training adj (course$ or program$)).tw.
42. kinesi?therap$.tw.
43. gardening.tw.
44. multiconvergent.tw.
45. exp Sports/
46. or/1-45
47. Fatigue Syndrome, Chronic/
48. exp Fatigue/
49. Asthenia/
50. Neurasthenia/
51. chronic fatigue$.tw.
52. fatigue syndrom$.tw.
53. infectious mononucleos$.tw.
54. postviral fatigue syndrome$.tw.
55. chronic fatigue-fibromyalgia syndrome$.tw.
56. myalgic encephalomyelit$.tw.
57. royal free disease$.tw.
58. neurasthenic neuroses.tw.
59. akureyri disease.tw.
60. atypical poliomyelitis.tw.
61. benign myalgic encephalomyelitis.tw.
62. (CFIDS or CFS).tw.
63. (chronic adj4 mononucleos$).tw.
64. epidemic neuromyasthenia.tw.
65. iceland disease.tw.
66. post infectious encephalomyelitis.tw.
67. PVFS.tw.
68. (perspective adj4 asthenia).tw.
69. neurasthenic syndrome$.tw.
70. neurataxia.tw.
71. neuroasthenia.tw.
72. (neuromuscular adj6 fatigue).tw.
73. or/47-72
74. randomized controlled trial.pt.
75. controlled clinical trial.pt.
76. randomi#ed.ab.
77. placebo$.ab.
78. randomly.ab.
79. trial.ab.
80. (clinic$ adj3 (trial$ or study or studies$)).ti,ab.
81. (control$ or prospectiv$ or volunteer$).ti,ab.
82. ((singl$ or doubl$ or tripl$) adj (blind$ or mask$ or dummy)).ti,ab.
83. or/74-82

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

84. (animals not (humans and animals)).sh.
85. 83 not 84
95. 46 and 73 and 85

Cochrane Central Register of Controlled Trials (CENTRAL)

#1 MeSH descriptor Exercise
#2 MeSH descriptor Exercise Therapy
#3 MeSH descriptor Exercise Movement Techniques
#4 MeSH descriptor Physical Fitness
#5 MeSH descriptor Physical Education and Training
#6 exercis*
#7 breathing NEAR/2 (therap* or exercis*)
#8 respiration NEAR/2 (therap* or exercis*)
#9 (gi gong or gigong)
#10 relaxation*
#11 tai or thai or taiji or taijiquan or taichi
#12 walking
#13 yoga
#14 (physical NEAR/2 (fitness or condition* or education or training or mobility or activit* or exertion or effort))
#15 gymnastics
#16 calisthenics
#17 aerobic*
#18 danc*
#19 jumping or hopping
#20 ambulat*
#21 muscle strengthening
#22 (muscular NEAR/2 (strength or resistance))
#23 (weight or weights) NEAR/2 lift*
#24 weightlifting or power lifting or weight training
#25 (Pilates or stretching or plyometric* or cardiopulmonary conditioning or motion therap* or neuromuscular facilitation* or movement therap* or gymnastic therap* or isometric training or climbing or cycling or lifting effort* or swimming or writing) #26 ((recreation or activity) NEAR/2 therap*)
#27 technical training
#28 (training NEAR/2 (course* or program*))
#29 (training adj (course* or program*))
#30 kinesi*
#31 gardening
#32 multiconvergent
#33 MeSH descriptor Sports explode all trees
#34 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 OR #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33)
#35 MeSH descriptor Fatigue Syndrome, Chronic
#36 MeSH descriptor Fatigue
#37 MeSH descriptor Asthenia
#38 MeSH descriptor Neurasthenia
#39 chronic fatigue*
#40 fatigue syndrom*
#41 infectious mononucleos*
#42 postviral fatigue syndrome*
#43 chronic fatigue-fibromyalgia syndrome*
#44 myalgic encephalomyelit*
#45 royal free disease*
#46 neurasthenic neuroses

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

#47 akureyri disease
#48 atypical poliomyelitis
#49 benign myalgic encephalomyelitis
#50 CFIDS or CFS
#51 chronic NEAR/5 mononucleos*
#52 epidemic neuromyasthenia
#53 iceland disease
#54 post infectious encephalomyelitis
#55 PVFS
#56 perspective NEAR/5 asthenia
#57 neurasthenic syndrome*
#58 neurataxia
#59 neuroasthenia
#60 neuromuscular NEAR/6 fatigue
#61 (#35 OR #36 OR #37 OR #38 OR #39 OR #40 OR #41 OR #42 OR #43 OR #44 OR #45 OR #46 OR #47 OR #48 OR #49 OR #50 OR #51 OR #52 OR #53 OR #54 OR #55 OR #56 OR #57 OR #58 OR #59 OR #60)
#62 (#34 AND #61)

International Trial Registers

World Health Organization International Clinical Trials Portal available at http://apps.who.int/trialsearch/, incorporating the following International trials registers/registries.

- Australian New Zealand Clinical Trials Registry
- ClinicalTrials.gov
- EU Clinical Trials Register (EU-CTR)
- International Standard Randomised Controlled Trial Number (ISRCTN)
- Brazilian Clinical Trials Registry (ReBec)
- Chinese Clinical Trial Registry
- Clinical Trials Registry-India
- Clinical Research Information Service-Republic of Korea
- Cuban Public Registry of Clinical Trials
- German Clinical Trials Register
- Iranian Registry of Clinical Trials
- Japan Primary Registries Network
- Pan African Clinical Trial Registry
- Sri Lanka Clinical Trials Registry
- The Netherlands National Trial Register
- Thai Clinical Trials Register (TCTR)

FEEDBACK

Feedback

Summary

The two reviews about chronic fatigue syndrome (CFS) (on exercise and CBT) are important documents in a controversial field. However, they seem to be listed on the website as mental health topics, alongside depression, etc. CFS is not a form of mental illness, although of course individual cases may have a psychological component that can be addressed during treatment. May I suggest that you place them elsewhere, as it is misleading and confusing to include them under the mental health umbrella?

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Reply

Many thanks for your comment on the two Cochrane CFS reviews. Apologies for the delay in responding, I have been on annual leave. We appreciate your observations about the placement of these reviews in *The Cochrane Library*. Feedback on reviews is normally dealt with by the relevant review author, but in this case I am responding, as your query relates more to an organisational issue. These reviews are listed as topics under a mental health heading because, as a result of the psychological component to which you refer, both reviews are supported by a mental health Cochrane group. Similar arrangements are in place for reviews of treatments for other disorders involving a variety of component problems and that as a result do not easily fit within the scope of one Cochrane group. These reviews however can be accessed in a number of different ways, for example, by searching for the specific topic (CFS and associated terminology, exercise and associated terminology, CBT and associated terminology); by searching for the study authors; by looking under subject headings, etc. The subject headings are not really intended as a comment on/guide to the aetiology of an illness, but they sometimes reflect the services involved in management of the condition. I have copied this response to the review authors in case they wish to comment further. Many thanks for your feedback.

Contributors

Cathy Stillman-Lowe (occupation freelance editor/science writer)
cathy.stillman-lowe@care4free.net
Submitter agrees with default conflict of interest statement:
I certify that I have no affiliations with or involvement in any organisation or entity with a financial interest in the subject matter of my feedback.

Types of evidence included, 3 June 2013

Summary

Unfortunately, this review ignores the large body of patient testimony suggesting that many persons with severe myalgic encephalomyelitis have been harmed by graded exercise therapy.
Since it was prepared, the International Consensus Primer and Guidelines for Medical Practitioners have been published.
Current thinking is to stay within your energy envelope. People with ME tend to overdo not underdo what they are capable of....
Care must be taken to NOT encourage them to do too much.
Further many definitions are used for CFS, and this muddies the waters.
I agree with the conflict of interest statement below:
I certify that I have no affiliations with or involvement in any organisation or entity with a financial interest in the subject matter of my feedback.

Reply

Thank you for your comments on this Cochrane Review.
In conducting this review, our aim was to gather and synthesise a specific type of evidence-that reported by randomised controlled trials. We fully accept that patient testimony, particularly that gathered and synthesised by high-quality qualitative research, is invaluable in any clinical area, particularly in an area as challenging for patients and healthcare professionals as CFS-ME. However, this project was not designed to incorporate such evidence.
We do consider the possibility of harm arising from graded exercise therapy by considering reported adverse events. Clearly this is an important issue to consider with any therapeutic intervention. Moreover, in the usual course of any illness, the condition of some patients improves (with or without treatment) and the condition of others worsens (with or without treatment). It is only through the use of randomised controlled trials that the effects (whether beneficial or adverse) of putative treatments can be disentangled reliably from the natural history of illness.
You raise the important point that (some) 'people with ME tend to overdo not underdo what they are capable of.' The critical point is the extent to which patients should be 'encouraged to do more' and the way in which they should be encouraged to do so. These are important research questions. As you know, new randomised evidence is available from the PACE trial, published in 2011 in *Lancet*. Whilst this is a controversial trial, it is an important randomised comparison of graded exercise therapy and 'adaptive pacing.' We look forward to further randomised evidence in due course.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

We also look forward to continuing to work in this clinical area, in the hope that we can advance our understanding of the impact of this treatment approach.

Contributors

Submitter: Adrienne.
Response prepared by Jonathan Price.

Comment 1 of 2, 9 September 2015

Summary

I would first like to thank those involved for their work in preparing this document. Even for those of us who have read the individual Chronic Fatigue Syndrome (CFS) papers it is useful to have the results collated, as well as details regarding the interventions. Also it is interesting to see the results of sensitivity analyses, subgroup analyses, standardised mean differences, etc.
I would like to make a few comments. I'm splitting them into two submissions as the piece had become very long. I've added some loose headings to hopefully make it more readable.
Objective measures
The review assessed the studies as having a high risk of bias regarding blinding, since neither participants nor assessors were blinded. Evidence suggests that subjective outcomes are more prone to bias than objective outcomes when there is no blinding (1). It is thus unfortunate that the review concentrated almost exclusively on subjective measures, failing to include results from nearly all the objective outcome measures that have been published with trials. (The exception was health resource use for which you presented follow-up data from one trial).
I hope objective outcome data can be included in a future revision or edition of this review.
Examples of objective outcomes include: exercise testing (work capacity by oxygen consumption); fitness test/step test; the six minute walking test; employment status; and disability payments.
Adding in these results would allow a more rigorous assessment of the effectiveness and relevance of the therapies, their causal mechanisms, therapeutic compliance, and safety.
On exercise testing, for example, in the PACE Trial (the largest trial in the review) there was no improvement in fitness levels as measured by a step test (2). The fitness data contrasts sharply with the many positive results from subjective self-report measures in the trial, so one is left wondering how much the subjective measures reflect reality.
On another exercise test used in the PACE Trial, the 6 minute walk test, there was a small (mean) increase from 312 metres at baseline to 379 metres at 12 months: this was 35.3 metres more than the "passive" control group when adjustments were made. However, the final result of 379 metres remains very poor compared to the more than 600 metres one would expect from healthy people of a similar age and gender make-up (3,4). By comparison, a group with Class III heart failure walked an average of 402 metres (5). A score of less than 400 metres has been suggested as the level at which somebody should be put on a lung transplant list (6). Such information from objective measures helps to add important context to the subjective measures and restraint to the conclusions that can be drawn from them.
Objective data is also needed to check compliance with a therapy. If patients diligently exercised for 12 months one would expect much better results on fitness and exercise testing than the aforementioned results in the PACE Trial. This is important when considering adverse events and safety: such trials may not give us good information on the safety of complying with such interventions if patients haven't actually complied.
Employment and receipt of disability payments are practical objective measures of general functional capacity so data on them would help establish whether patients can actually do more overall or whether they may just be doing, for example, a little more exercise but have substituted that for doing less in other areas (7,8). Also, CFS patients are sometimes pressured by insurance companies into doing graded exercise therapy (GET) programs so it would be useful to have data collated on employment outcomes to see whether pressure can in any way be justified (9,10). In the PACE Trial, there was no significant improvement in employment measures and receipt of disability payments in the GET group (11). Outside the realm of clinical trials, the quantitative and qualitative data in a major (UK) ME Association survey also found that GET didn't lead to higher levels of employment and lower levels of receipt of disability payments on average (9). Also, extensive external audits were performed of Belgian CFS rehabilitation clinics that treated using cognitive behavioural therapy (CBT) and GET. The main reports are in French and Dutch (12,13), with an English summary

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

available (14) that says, "Employment status decreased at the end of the therapy, from an average of 18.3% of a 38h working week, to 14.9% [...] The percentage of patients living from a sickness allowance increased slightly from 54 to 57%." This contrasts with the average improvements reported in the audit for some symptoms like fatigue.

While data on (self-reported) symptoms like fatigue (one of your two primary outcomes) is interesting, arguably more important to patients is improving their overall level of functioning (and again, objective measures are needed here). Being able to work, for example, despite experiencing a certain level of fatigue would likely be more important for many than being unable to work but having slightly lower levels of fatigue.

An example of how reductions in the reported levels of fatigue may not lead to improvements in functioning can be seen in an analysis of three graded activity-oriented CBT therapy interventions for CFS (15). The analysis showed, compared to controls, there were no improvements in overall activity levels as measured by actometers despite improvements in self-reported fatigue (15). Activity in these trials was assessed using actometers. Another study that exemplifies the problem of focusing too much on fatigue scores after behavioural interventions is a study of CBT in multiple sclerosis (MS) patients with "MS fatigue"(16). The study found that following the intervention, patients with MS reported significantly lower (i.e. better) scores on the Chalder Fatigue Scale (0-33 scoring) than those in a healthy, nonfatigued comparison group! This significant difference was maintained at 3 and 6 months' follow-up. It is difficult to believe that patients with MS fatigue (at baseline) truly subsequently had less fatigue than healthy nonfatigued controls: a much more likely scenario is that undertaking the intervention had led to response biases.

You mention that "many patient charities are opposed to exercise therapy for chronic fatigue syndrome (CFS)". One reason for concern about the way in which exercise programmes are promoted to patients is that they are often based upon models which assume that there is no abnormal physiological response to exercise in the condition, and make unsupported claims to patients. For example, in the FINE trial (Wearden et al., 2010) patient booklet (17), it is boldly asserted that: "Activity or exercise cannot harm you" (p. 49). However, a large number of studies have found abnormal responses to exercise, and the possibility of harm being done simply cannot be excluded on the basis of current evidence (discussed in 4, 18-20)."

Compliance

The review doesn't include any information on compliance. I'm not sure that there is much published information on this but I know there was a measure based on attendance at therapy sessions (which could be conducted over the phone) given for the PACE Trial (3). Ideally, it would be interesting if you could obtain some unpublished data from activity logs, records from heart-rate monitors, and other records to help build up a picture of what exercise was actually performed and the level of compliance. Information on adherence and what exercise was actually done is important in terms of helping clinicians, and indeed patients, to interpret and use the data. I mention patients because patients' own decisions about their behaviour is likely to be affected by the medical information available to them, both within and outside of a supervised programme of graded exercise; unlike with an intervention like a drug, patients can undertake exercise without professional supervision.

"Selective reporting (outcome bias)" and White et al. (2011)

I don't believe that White et al. (2011) (the PACE Trial) (3) should be classed as having a low risk of bias under "Selective reporting (outcome bias)" (Figure 2, page 15). According to the Cochrane Collaboration's tool for assessing risk of bias (21), the category of low risk of bias is for: "The study protocol is available and all of the study's pre-specified (primary and secondary) outcomes that are of interest in the review have been reported in the pre-specified way". This is not the case in the PACE Trial. The three primary efficacy outcomes can be seen in the published protocol (22). None have been reported in the pre-specified way. The Cochrane Collaboration's tool for assessing risk of bias states that a "high risk" of bias applies if any one of several criteria are met, including that "not all of the study's pre-specified primary outcomes have been reported" or "one or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified". In the PACE Trial, the third primary outcome measure (the number of "overall improvers") was never published. Also, the other two primary outcome measures were reported using analysis methods that were not pre-specified (including switching from the bimodal to the Likert scoring method for The Chalder Fatigue Scale, one of the primary outcomes in your review). These facts mean that the "high risk of bias" category should apply.

Thank you for taking the time to read my comments.

Tom Kindlon

Conflict of Interest statement: I am a committee member of the Irish ME/CFS Association and do a variety of unpaid work for the Association.

References:

1. Turner L, Boutron I, Hróbjartsson A, Altman DG, Moher D: The evolution of assessing bias in Cochrane systematic reviews of interventions: celebrating methodological contributions of the Cochrane Collaboration. Syst Rev 2013, 2:79.

2. Chalder T, Goldsmith KA, White PD, Sharpe M, Pickles AR. Rehabilitative therapies for chronic fatigue syndrome: a secondary mediation analysis of the PACE trial. Lancet Psychiatry. 2015;2:141-152.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

3. White PD, Goldsmith KA, Johnson AL, Potts L, Walwyn R, DeCesare JC, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. The Lancet 2011;377:823-36.
4. Kindlon T. Reporting of Harms Associated with Graded Exercise Therapy and Cognitive Behavioural Therapy in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome. Bull IACFS ME. 2011;19:59-111. http://iacfsme.org/ME-CFS-Primer-Education/Bulletins/BulletinRelatedPages5/Reporting-of-Harms-Associated-with-Graded-Exercise.aspx
5. Lipkin DP, Scriven AJ, Crake T, Poole-Wilson PA (1986). Six minute walking test for assessing exercise capacity in chronic heart failure. British Medical Journal 292, 653-5.
6. Kadikar A, Maurer J, Kesten S. The six-minute walk test: a guide to assessment for lung transplantation. J Heart Lung Transplant. 1997 Mar;16(3):313-9.
7. Friedberg F, Sohl S. Cognitive-behavior therapy in chronic fatigue syndrome: is improvement related to increased physical activity? J Clin Psychol. 2009 Feb 11.
8. Friedberg F. Does graded activity increase activity? A case study of chronic fatigue syndrome. Journal of Behavior Therapy and Experimental Psychiatry, 2002, 33, 3-4, 203-21
9. Results and In-depth Analysis of the 2012 ME Association Patient Survey Examining the Acceptability, Efficacy and Safety of Cognitive Behavioural Therapy, Graded Exercise Therapy and Pacing, as Interventions used as Management Strategies for ME/CFS. Gawcott, England. http://www.meassociation.org.uk/2015/05/23959/ Accessed: September 3, 2015
10. Critical Illness - A Dreadful Experience with Scottish Provident. http://forums.moneysavingexpert.com/showthread.php?t=2356683 Accessed: September 4, 2015
11. McCrone P, Sharpe M, Chalder T, Knapp M, Johnson AL, Goldsmith KA, White PD. Adaptive pacing, cognitive behaviour therapy, graded exercise, and specialist medical care for chronic fatigue syndrome: a cost-effectiveness analysis. PLoS One. 2012;7(8):e40808.
12. Rapport d'évaluation (2002-2004) portant sur l'exécution des conventions de rééducation entre le Comité de l'assurance soins de santé (INAMI) et les Centres de référence pour le Syndrome de fatigue chronique (SFC). 2006. http://health.belgium.be/internet2Prd/groups/public/@public/@shc/documents/ie2divers/14926531˙fr.pdf (Starts on page 223.) Accessed September 4, 2015 (French language edition)
13. Evaluatierapport (2002-2004) met betrekking tot de uitvoering van de revalidatieovereenkomsten tussen het Comité van de verzekering voor geneeskundige verzorging (ingesteld bij het Rijksinstituut voor Ziekte- en invaliditeitsverzekering) en de Referentiecentra voor het Chronisch vermoeidheidssyndroom (CVS). 2006. http://health.belgium.be/internet2Prd/groups/public/@public/@shc/documents/ie2divers/14926531.pdf (Starts on page 227.) Accessed September 4, 2015 (Dutch language version)
14. Stordeur S, Thiry N, Eyssen M. Chronisch Vermoeidheidssyndroom: diagnose, behandeling en zorgorganisatie. Health Services Research (HSR). Brussel: Federaal Kenniscentrum voor de Gezondheidszorg (KCE); 2008. KCE reports 88A (D/2008/10.273/58) https://kce.fgov.be/sites/default/files/page˙documents/d20081027358.pdf Accessed September 4, 2015
15. Wiborg JF, Knoop H, Stulemeijer M, Prins JB, Bleijenberg G. How does cognitive behaviour therapy reduce fatigue in patients with chronic fatigue syndrome? The role of physical activity. Psychol Med. 2010; 40:1281-1287.
16. Van Kessel K, Moss-Morris R, Willoughby, Chalder T, Johnson MH, Robinson E, A randomized controlled trial of cognitive behavior therapy for multiple sclerosis fatigue, Psychosom. Med. 2008; 70:205-213.
17. Powell P. FINE Trial Patient Booklet http://www.fine-trial.net/downloads/Patient%20PR%20Manual%20ver9%20Apr05.pdf Accessed September 7, 2015
18. Twisk FNM, Maes M. A review on Cognitive Behavorial Therapy (CBT) and Graded Exercise Therapy (GET) in Myalgic Encephalomyelitis (ME)/Chronic Fatigue Syndrome (CFS): CBT/GET is not only ineffective and not evidence-based, but also potentially harmful for many patients with ME/CFS. Neuro Endocrinol Lett. 2009;30:284-299.
19. Carruthers BM et al. Myalgic Encephalomyelitis - Adult & Paediatric: International Consensus Primer for Medical Practitioners. ISBN 978-0-9739335-3-6 http://www.investinme.org/Documents/Guidelines/Myalgic%20Encephalomyelitis%20International%20Consensus%20Primer%20-2012-11-26.pdf Accessed September 5, 2015
20. Twisk FN. Objective Evidence of Post-exertional "Malaise" in Myalgic Encephalomyelitis and Chronic Fatigue Syndrome. J Sports Med Doping Stud 2015. 5:159. doi: 10.4172/2161-0673.100015
21. Higgins JPT, Green S: Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. Table 8.5.d. The Cochrane Collaboration; 2011. http://handbook.cochrane.org/chapte...a˙for˙judging˙risk˙of˙bias˙in˙the˙risk˙of.htm Accessed: September 5, 2015
22. White PD, Sharpe MC, Chalder T, DeCesare JC, Walwyn R; on behalf of the PACE trial group. Protocol for the PACE trial: A randomised controlled trial of adaptive pacing, cognitive behaviour therapy, and graded exercise as supplements to standardised specialist medical care versus standardised specialist medical care alone for patients with the chronic fatigue syndrome/myalgic encephalomyelitis or encephalopathy. BMC Neurology 2007, 7:6 http://www.biomedcentral.com/1471-2377/7/6 Accessed: September 5, 2015

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Reply

Thank you for reading the review so carefully and for your comments. I have split the answers according to the headings you have used.

Objective measures and compliance

The protocol for this review did not include objective measurements or compliance as outcomes, hence are not included. You make a strong case and including objective measures and compliance should be carefully considered in an update.

Selective reporting (outcome bias)

The Cochrane Risk of Bias tool enables the review authors to be transparent about their judgments, but due to the subjective nature of the process it does not guarantee an indisputable consensus. You particularly mention the risk of bias in the PACE trial regarding not providing pre-specified outcomes however the trial did pre-specify the analysis of outcomes. The primary outcomes were the same as in the original protocol, although the scoring method of one was changed and the analysis of assessing efficacy also changed from the original protocol. These changes were made as part of the detailed statistical analysis plan (itself published in full), which had been promised in the original protocol. These changes were drawn up before the analysis commenced and before examining any outcome data. In other words they were pre-specified, so it is hard to understand how the changes contributed to any potential bias. The relevant paper also alerted readers to all these changes and gave the reasons for them. Overall, we don't think that the issues you raise with regard to the risk of selective outcome bias are such as to suspect high risk of bias, but recognize that you may reach different conclusions than us.

Kind Regards,

Lillebeth Larun

Contributors

Feedback submitted by: Tom Kindlon

Response submitted by: Lillebeth Larun

Comment 2 of 2, 9 September 2015

Summary

Variation in interventions

It would have been useful to have some more information on the "exercise with pacing" intervention tested in the Wallman et al. (2004) trial and how it was distinct from some other exercise interventions tested. The authors say (1): "On days when symptoms are worse, patients should either shorten the session to a time they consider manageable or, if feeling particularly unwell, abandon the session altogether" (p. 143). I don't believe the description given in the review conveys this. In the review, this approach is described as "Exercise with pacing: exercise in which the incremental increase in exercise was personally set." But Wallman et al.'s approach allows patients to decrease as well as increase how much exercise they do on the day. This approach also contrasts with how White (an investigator in two of the trials) has described graded exercise therapy: "if [after increasing the intensity or duration of exercise] there has been an increase in symptoms, or any other adverse effects, they should stay at their current level of exercise for a further week or two, until the symptoms are back to their previous levels" (2). In the PACE Trial manual White co-wrote (3), the GET intervention was guided by the principle that "planned physical activity and not symptoms are used to determine what the participant does" (p. 21); similarly, "it is their planned physical activity, and not their symptoms, that determine what they are asked to do" (p. 20). Compliance data would help us examine which approach patients are actually using: I suspect many patients are in fact doing exercise with pacing even in trials such as the PACE Trial (i.e. when they have increased symptoms, often reducing levels of exercise and sometimes doing no exercise activities at all on that day).

Bimodal versus Likert scoring in Wearden et al. (2010)

I find it odd that the fatigue scores for the Wearden et al. (2010) trial (4) are given in the 0-33 format rather than the 0-11 scoring method. The 0-11 scoring system is what is mentioned as a primary outcome measure in the protocol and is what is reported in the main paper reporting the results (4, 5). It is even what your own report says on p. 44 is the scoring method ("Fatigue Scale, FS; 11 items; each item was scored dichotomously on a 4-point scale [0, 0, 1 or 1]"). This is important because using the scoring method for which you don't report data (0-11), there is no statistically significant difference at the primary outcome point of 70 weeks (5).

Diagnostic criteria

One problem with using these trials as an evidence base, which I don't believe was mentioned, is that all the trials used the Oxford and Fukuda diagnostic criteria (6, 7). Neither of these criteria require patients to have post-exertional malaise (or something similar). Many

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

consider this to be a core symptom of ME/CFS and it is mandatory in most of the other major criteria (8-11). [Aside: The London criteria were assessed in the PACE Trial (12) but they seem to have been operationalised in an unusual way. Ninety seven per cent of the participants who satisfied the (broad) Oxford criteria who didn't have a psychiatric disorder satisfied the definition of M.E. used (13). Ellen Goudsmit, one of the authors of the London criteria, has rejected the way they were used in the PACE Trial (14)]. So this lack of requirement for patients to have post-exertional malaise (or a similar description) means we cannot be sure that the evidence can be generalised to such patients. An independent National Institutes of Health committee this year concluded "continuing to use the Oxford definition may impair progress and cause harm. Therefore, for progress to occur, we recommend that this definition be retired" (15). An Agency for Healthcare Research and Quality review of diagnostic methods this year reached a similar conclusion: "Consensus groups and researchers should consider retiring the Oxford case definition because it differs from the other case definitions and is the least restrictive, probably including individuals with other overlapping conditions" (16). An Agency for Healthcare Research and Quality review of ME/CFS treatments said: "The Oxford CFS case definition is the least restrictive, and its use as entry criteria could have resulted in selection of participants with other fatiguing illnesses or illnesses that resolve spontaneously with time" (17).

Exclusion of some data from analyses due to baseline differences

It seems unfortunate that some data cannot be used due to baseline differences e.g. "Four trials (669 participants) contributed data for evaluation of physical functioning at follow-up (Jason 2007; Powell 2001; Wearden 2010; White 2011). Jason 2007 observed better results among participants in the relaxation group (MD 21.48, 95% CI 5.81 to 37.15). However, results were distorted by large baseline differences in physical functioning between the exercise and relaxation groups (39/100 vs 54/100); therefore we decided not to include these results in the meta-analysis". It would be good if other methods could be investigated (e.g. using baseline levels as covariates) to analyse such data.

Thank you for taking the time to read my comments.

Tom Kindlon

I am a committee member of the Irish ME/CFS Association and do a variety of unpaid work for the Association.

1. Wallman KE, Morton AR, Goodman C, Grove R. Exercise prescription for individuals with chronic fatigue syndrome. Med J Aust. 2005;183:142-3.

2. White P. How exercise can help chronic fatigue syndrome. Pulse: 1998. June 20:86-87.

3. Bavinton J, Darbishire L, White PD -on behalf of the PACE trial management group. Graded Exercise Therapy for CFS/ME (Therapist Manual) http://www.pacetrial.org/docs/get-therapist-manual.pdf

4. Wearden AJ, Riste L, Dowrick C, Chew-Graham C, Bentall RP, Morriss RK, Peters S, Dunn G, Richardson G, Lovell K, Powell P. Fatigue Intervention by Nurses Evaluation--the FINE Trial. A randomised controlled trial of nurse led self-help treatment for patients in primary care with chronic fatigue syndrome: study protocol. [ISRCTN74156610]. BMC Med. 2006 Apr 7;4:9.

5. Wearden AJ, Dowrick C, Chew-Graham C, Bentall RP, Morriss RK, Peters S, Riste L, Richardson G, Lovell K, Dunn G; Fatigue Intervention by Nurses Evaluation (FINE) trial writing group and the FINE trial group. Nurse led, home based self help treatment for patients in primary care with chronic fatigue syndrome: randomised controlled trial. BMJ. 2010 Apr 23;340:c1777. doi: 10.1136/bmj.c1777.

6. Sharpe M, Archard L, Banatvala J, Borysiewicz LK, Clare AW, David A, et al. Chronic fatigue syndrome: guidelines for research. Journal of the Royal Society of Medicine 1991;84 (2):118-21.

7. Fukuda K, Straus SE, Hickie I, et al. The chronic fatigue syndrome: A comprehensive approach to its definition and study. Ann Intern Med. 1994; 121: 953- 959.

8. Carruthers BM, Jain AK, De Meirleir KL, et al. Myalgic Encephalomyelitis/chronic fatigue syndrome: Clinical working case definition, diagnostic and treatments protocols. Journal of Chronic Fatigue Syndrome. 2003; 11: 7-115.

9. Carruthers BM, van de Sande MI, De Meirleir KL, et al. Myalgic encephalomyelitis: International Consensus Criteria. J Intern Med. 2011; 270: 327-338.

10. IOM (Institute of Medicine). Beyond myalgic encephalomyelitis/chronic fatigue syndrome: Redefining an illness. Washington, DC: The National Academies; 2015.

11. National Institute for Health and Clinical Excellence. Chronic fatigue syndrome/myalgic encephalomyelitis (or encephalopathy): diagnosis and management of CFS/ME in adults and children, 2007. http://www.nice.org.uk/guidance/CG53 Accessed September 6, 2015. London: National Institute for Health and Clinical Excellence.

12. White PD, Goldsmith KA, Johnson AL, Potts L, Walwyn R, DeCesare JC, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. The Lancet 2011;377:823-36.

13. Kindlon T. PACE Trial - 97% of the participants who didn't have a psychiatric disorder satisfied the definition of M.E. used. https://listserv.nodak.edu/cgi-bin/wa.exe?A2=ind1106A&L=CO-CURE&P=R2764 Accessed: September 6, 2015

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

14. Ellen Goudsmit on PubMed Commons: http://www.ncbi.nlm.nih.gov/myncbi/ellen m.goudsmit.1/comments/
15. Green CR, Cowan P, Elk R, O'Neil KM, Rasmussen AL. National Institutes of Health Pathways to Prevention Workshop: advancing the research on myalgic encephalomyelitis/chronic fatigue syndrome. Ann Intern Med. 2015; 162:860-5.
16. Haney E, Smith MEB, McDonagh M, Pappas M, Daeges M, Wasson N, et al. Diagnostic methods for myalgic encephalomyelitis/chronic fatigue syndrome: a systematic review for a National Institutes of Health Pathways to Prevention Workshop. Ann Intern Med. 2015; 162:834-40.
17. Smith MEB, Haney E, McDonagh M, Pappas M, Daeges M, Wasson N, et al. Treatment of myalgic encephalomyelitis/chronic fatigue syndrome: a systematic review for a National Institutes of Health Pathways to Prevention Workshop. Ann Intern Med. 2015; 162:841-50.

Reply

Variation in interventions

There is ongoing work to improve descriptions of interventions both in primary studies and in systematic reviews (Scroter 2012, Glasziou 2008). We tried to describe the exercise programs, and differences between them in great detail. We did this both in the tables of study characteristics, and in the Characteristics of exercise intervention table (table 2). We also contacted trial authors to check that the information was correct. We recognize the need for more research to explore which parts of an exercise treatment program that are most essential or most closely correlated to an successful outcome, i.e. the active ingredient.

Bimodal versus Likert scoring in Wearden et al. 2010

To enable pooling of as many studies as possible in a mean difference meta-analyses, we used the 33-scale results reported by Wearden. You suggest that the decision to use the 33-point fatigue scores in our analysis may bias the results because there is no statistically significant difference at the 11-point data at 70 weeks. This statement suggests that there is a statistically significant difference when using the 33-point data, but if you look into analysis 1.2 that is not the case. At 70 week we report MD -2.12 (95% CI -4.49 to 0.25) for the FINE trial, i.e. not statistically significant.

Review authors response: Diagnostic criteria

As the use of various diagnostic criteria is often emphasised as particularly important with regard to treatment response, we performed subgroup analyses based on diagnostic criteria. The availability of relevant trials limits which subgroup analyses are possible to carry out in a systematic review, and hence, we were only able to contrast CDC versus Oxford criteria and found no evidence for a difference. We realize that the role of diagnostic criteria as a possible moderator for the efficacy of exercise receives a lot of attention, and would welcome trials to investigate these matters more thoroughly.

Exclusion of some data from analyses due to baseline differences

In meta-analysis based on aggregated data the authors have to act based on the information that is available from original publications or additional information obtained from the original investigators. As you state, these restrictions may be suboptimal. It is possible to adjust for baseline differences in meta-regression type analyses, but this requires adjustment for dependency between the intervention and control group results from the same trial. As a consequence, three variables (intervention vs control, baseline level, and trial) would have to be accounted for in the analyses. This implies that at least 30 data points will be needed to gain somewhat stable and trustworthy estimates adjusted for baseline levels. Systematic reviews based on individual patient data (IPD) allows for more appropriate processing and standardization of data. We are happy to inform you that we have now received individual patient data from most of the studies included in this review, and that the preparation of an IPD review is in progress.

Scroter S, Glasziou P, Heneghan C. Quality of descriptions of treatments: a review of published randomized trials. BMJ Open 2012: 2e001978 doi:10.1136/bmjopen-2012-001978

Glasziou P, Meats M, Heneghan C, Sheppers S. What is missing in descriptions of treatment in trials and reviews? BMJ 2008;336: 1472 doi: /10.1136/bmj.39590.732037.47

Contributors

Feedback submitted by: Tom Kindlon
Response submitted by: Lillebeth Larun

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Feedback submitted, 16 April 2016

Summary

Query regarding use of post-hoc unpublished outcome data: Scoring system for the Chalder fatigue scale, Wearden, 2010

I would like to highlight what appears to be a discrepancy within the Cochrane review [1] with respect to the analysis of data from Wearden 2010 [2,3].

Throughout the Cochrane review (please see details below), the impression is given that only protocol defined and published data or outcomes were used for the Cochrane analysis of the Wearden 2010 study.

However, this does not appear to be the case and, to the best of my knowledge, instead of using protocol defined or published data, the Cochrane analyses of fatigue for the Wearden 2010 study, appears to have used an alternative unpublished set of data.

The relevant analyses of fatigue in the Cochrane review are: Analyses: 1.1, 1.2, 2.1 and 2.3. Each of these analyses states that the "0,1,2,3" scoring system was used for the Chalder fatigue questionnaire. This scoring system is known as the Likert scoring system and uses a fatigue scale of 033

points.

However, to the best of my knowledge, data or analyses using this scoring system were not proposed in the Wearden 2010 trial protocol [3], and were not included in Wearden 2010 [2], and have not previously been formally (i.e. via peer review) published by Wearden et al. A posthoc informal analysis using this data has been informally released by Wearden et al. as a BMJ Rapid Response comment [4].

In the Cochrane review, the analyses using the 0, 1, 2, 3 scoring system contradict text within the section "Characteristics Of Studies", in relation to Wearden 2010: Under "Outcomes", it is stated that Chalder fatigue was measured using the 0,0,1,1 scoring system using

a scale from 011 points: "Fatigue (Fatigue Scale, FS; 11 items; each item was scored dichotomously on a 4 point scale (0, 0, 1 or 1)". Wearden 2010 prespecified Chalder fatigue questionnaire scores as a primary outcome at 70 weeks, and as a secondary outcome immediately after treatment at 20 weeks. The scoring, in both cases, used the 0,0,1,1 system, with a scale of 011. This scoring system was described both in the trial protocol [3] and the main results paper published in 2010 [2].

The Likert (0, 1, 2, 3) scoring system was neither proposed in the trial protocol, nor formally published, and so the Likert scores should be considered posthoc. Even if it is argued that the Chalder fatigue questionnaire (irrespective of the scoring system) was predefined as a primary outcome measure, data using the Likert scoring system was neither proposed nor published and so the data itself must surely be considered to be posthoc. The outcome analyses using the Likert data must be considered posthoc.

Simply changing a scoring system may, at first glance, appear not to be a significant or major adjustment, however, we do not know what difference it made because a sensitivity analysis has not been published.

I cannot find any explanation within the Cochrane review that explains why the Cochrane review has replaced predefined published data with an unpublished and posthoc set of data.

Is it normal practice for a Cochrane metaanalysis to selectively ignore the predefined primary outcome data for a trial, and to selectively include and analyse posthoc data? I wonder if some clarity could be shed on this situation?

I suggest that the posthoc data are replaced with the original published data. Otherwise, the posthoc data should be clearly labelled

as such and the risk of bias analysis amended accordingly; and an explanation should be included in the review explaining why an apparently adequate predefined set of data has been replaced with an apparent novel set of posthoc data.

Also, I suggest that any discrepancies that I will outline below, should be corrected where necessary; Either the analyses (1.1, 1.2, 2.1 and 2.3) should be amended or the description of the data should be amended so it is not incorrectly labelled as protocol defined and published data with a "low risk" of bias.

Discrepancies within the text of the Cochrane Analysis

Please note that all page numbers used below are pertinent to the current version (version 4) of the Cochrane review in PDF format.

1. On page 28 of the Cochrane review [1], in section "Potential biases in the review process", under the heading "Potential bias in the review process", in relation to the review in general, it is stated that: "For this updated review, we have not collected unpublished data for our outcomes ..." However, as explained above, this is not the case for the Wearden 2010 fatigue data for which unpublished data has been used in the Cochrane analysis.

2. On page 45 of the review, in section "Characteristics Of Studies", specifically in relation to Wearden 2010 [2,3], it is stated that only protocol defined

outcomes were used: "all relevant outcomes are reported in accordance with the protocol". "Selective reporting (reporting bias)" is rated as "low risk". However, as explained above, this is not the case, because the Wearden 2010 fatigue data (used in the Cochrane analysis) was not proposed in the

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

protocol. If the data is posthoc, then the "low risk" category will need to be revised.
3. On page 44 of the review, in section "Characteristics Of Studies", in relation to Wearden 2010 [2,3], under "Outcomes", it is stated that Chalder fatigue was measured using the 0,0,1,1 scoring system using a scale from 011 points: "Fatigue (Fatigue Scale, FS; 11 items; each item was scored dichotomously on a 4 point scale (0, 0, 1 or 1)". Wearden 2010 did indeed use the 0,0,1,1 scoring system for the Chalder fatigue scale: This scoring system was proposed in the trial protocol and published with the main outcome data in Wearden 2010. However, as explained above, this scoring system has not been used in the Cochrane analysis.
4. If figures 2 and 3 also contain discrepancies, after any amendments to the review, then they should be amended accordingly.
There may be other related discrepancies and inaccuracies in the text that I haven't noticed. I thank the Cochrane team in advance for giving this submission careful consideration, and for making amendments to the analysis, and providing explanations, where appropriate. I hope you will agree that clarity, transparency and accuracy in relation to the analysis is paramount.

References:

1. Larun L, Brurberg KG, Odgaard Jensen J, Price JR. Exercise therapy for chronic fatigue syndrome. Cochrane Database Syst Rev. 2016; CD003200.
2. Wearden AJ, Dowrick C, ChewGraham C, et al. Nurse led, home based self help treatment for patients in primary care with chronic fatigue syndrome: randomised controlled trial. BMJ. 2010; 340:c1777.
3. Wearden AJ, Riste L, Dowrick C, et al. Fatigue Intervention by Nurses Evaluation - The FINE Trial. A randomised controlled trial of nurse led selfhelp treatment for patients in primary care with chronic fatigue syndrome: study protocol. BMC Med. 2006; 4:9.
4. Wearden AJ, Dowrick C, ChewGraham C, et al. Fatigue scale. BMJ Rapid Response. 2010. http://www.bmj.com/rapidresponse/2011/11/02/fatiguescale0
(accessed April 16, 2016).

Reply

Dear Robert Courtney
Thank you for your detailed comments on the Cochrane review 'Exercise Therapy for Chronic Fatigue Syndrome'. We have the greatest respect for your right to comment on and disagree with our work. We take our work as researchers extremely seriously and publish reports that have been subject to rigorous internal and external peer review. In the spirit of openness, transparency and mutual respect we must politely agree to disagree.
The Chalder Fatigue Scale was used to measure fatigue. The results from the Wearden 2010 trial show a statistically significant difference in favour of pragmatic rehabilitation at 20 weeks, regardless whether the results were scored bi-modally or on a scale from 0-3. The effect estimate for the 70 week comparison with the scale scored bi-modally was -1.00 (CI-2.10 to +0.11; p =.076) and -2.55 (-4.99 to -0.11; p=.040) for 0123 scoring. The FINE data measured on the 33-point scale was published in an online rapid response after a reader requested it. We therefore knew that the data existed, and requested clarifying details from the authors to be able to use the estimates in our meta-analysis. In our unadjusted analysis the results were similar for the scale scored bi-modally and the scale scored from 0 to 3, i.e. a statistically significant difference in favour of rehabilitation at 20 weeks and a trend that does not reach statistical significance in favour of pragmatic rehabilitation at 70 weeks. The decision to use the 0123 scoring did does not affect the conclusion of the review.
Regards,
Lillebeth Larun

Contributors

Feedback submitted by: Robert Courtney
Response submitted by: Lillebeth Larun

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Feedback submitted, 1 May 2016

Summary

Comment: Assessment of Selective Reporting Bias in White 2011

With reference to the current Cochrane review of exercise therapy for chronic fatigue syndrome [1], I would like to follow-up the discussion between Tom Kindlon and Lillebeth Larun that has been published in the latest version of the full review published in 2016. Kindlon submitted two comments, dated 9 September 2016, and Larun issued a response to each.

Kindlon raised the issue of the study referred to as "White 2011" in the Cochrane review, commonly known as the PACE trial [2]; specifically whether or not the risk of bias for selective reporting of outcomes for the trial has been assessed and categorised appropriately, in terms of Cochrane's guidelines and policies.

In this submission I will make reference to the current "Cochrane Handbook for Systematic Reviews of Interventions" [3], including Table 8.5.d ("Criteria for judging risk of bias in the 'Risk of bias' assessment tool"), which I will refer to as the "Cochrane guidelines".

In his submission, Kindlon said: "I don't believe that White et al. (2011) (the PACE Trial) [...] should be classed as having a low risk of bias under "Selective reporting (outcome bias)"."

In a considered response, Larun concluded: "Overall, we don't think that the issues you raise with regard to the risk of selective outcome bias are such as to suspect high risk of bias, but recognize that you may reach different conclusions than us."

Larun's response to the concerns raised by Kindlon has left me unsure about whether Cochrane's guidelines have been applied appropriately in this case, so I would like to discuss some of the finer details.

Pre-Planned Analysis

I note that the PACE trial's protocol was submitted for publication in 2006 and published in 2007 [4], which was after the trial had commenced in 2005 [2]. This raises the question of whether the protocol itself can be defined as a pre-trial report. Cochrane's glossary of terms defines a "pre-specified" analysis as a "Statistical analyses specified in the trial protocol; that is, planned in advance of data collection." [5] So the Cochrane glossary states that a pre-specified analysis plan, or protocol, must be completed before data collection has commenced.

Other sources, such as the Wiley Encyclopedia of Clinical Trials, also define a pre-planned analysis as that which has been defined before data collection has commenced: "A primary efficacy endpoint needs to be specified before the start of the clinical trial." [6]

To be certain that the PACE trial's analyses were defined before data collection had commenced then we would need an earlier publication such as a trial register [7] or trial identifier, both of which were created and which included definitions of primary endpoints which were different to the trial protocol (see appendix, below, for detailed descriptions). To my knowledge, the Cochrane review, does not discuss these issues.

Nevertheless, the Cochrane guidelines (section 8.14.2) advise using a trial protocol as a guide to determine which trial outcomes were pre-determined: "If the protocol is available, then outcomes in the protocol and published report can be compared".

As the protocol was published after the trial had commenced, it seems certain that any subsequent (i.e. after the protocol had been published) changes to methodology were made after data collection had commenced and were therefore not pre-specified.

Statistical Analysis Plan

Larun states that various changes from the protocol were "made as part of the detailed statistical analysis plan (itself published in full), which had been promised in the original protocol." The protocol did indeed refer to a statistical analysis plan, but the protocol wording suggests to me that no changes from the protocol were planned, but that the statistical analysis plan would simply flesh out the protocol: "A full Analysis Strategy will be developed, independently of looking at the trial database, and before undertaking any analysis. This paper [i.e. the protocol] summarises the analysis plan." There was no suggestion that there would be wholesale changes to primary, secondary or recovery outcomes. But, in any case, even if the investigators' initial intentions had been to make changes after data collection had started, the result would still not be a pre-specified analysis according to the Cochrane glossary of terms [5].

The statistical analysis plan was submitted for publication in 2012 and published in 2013 [8] after the main trial results had been published in 2011 [2], and long after the trial had commenced in 2005, so the statistical analysis plan cannot reasonably be considered to be a priori. Indeed the statistical analysis report itself confirms that the analysis was finalised or approved towards the end of data collection in 2010: "These planned analyses were written with a view to publication and are reproduced almost as they were approved by the Trial Steering Committee (Version 1.2 dated 2 May 2010) prior to database lock."

Larun states that the "changes [to the trial] were drawn up before the analysis commenced and before examining any outcome data. In other words they were pre-specified [...]" However, the latter assertion is not consistent with Cochrane's glossary, which states that pre-specified changes are those defined before data collection has commenced [5].

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Investigators of an open-label trial can potentially gain insights into a trial before formal analysis has been carried out. If changes to a planned methodology are made after a trial has started and/or after data collection has commenced (whether or not the data has been formally analysed) then it is generally accepted that this fails the definition of a "pre-specified" study, which is confirmed by the Cochrane glossary and other sources. Otherwise trial registries and protocols could be drawn up after all data had been collected but before the formal analysis has commenced, and still be described as pre-planned. This would be particularly problematic in open-label trials such as the PACE trial.

Pre-Planned and Unplanned Primary Endpoints

The PACE trial's protocol had proposed three primary efficacy analyses which all had binary outcomes (i.e. a positive or negative outcome for each patient), whereas the final primary analyses were entirely different; they were continuous measures focused on the differences in mean improvements between intervention groups at 52 weeks. So the changes to the protocol were substantial. (See appendix, below, for detailed descriptions.) The PACE trial's three a priori primary efficacy analyses were not included in the final published results, and have never been published and, to my knowledge, no sensitivity analysis has been published for the final published primary analyses.

The PACE trial's published results paper [2] confirmed the unplanned outcome switching, as follows: "We used continuous scores for primary outcomes to allow a more straightforward interpretation of the individual outcomes, instead of the originally planned composite measures (50% change or meeting a threshold score)."

This entirely contradicts Larun's claims that: "the trial did pre-specify the analysis of outcomes" and: "The primary outcomes were the same as in the original protocol".

The Cochrane guidelines give guidance that is specific to the issue of changing a pre-planned analysis for the same set of data and they describe such an action as "selective reporting of analyses using the same data". The guidelines couldn't be more specific that changing a method for analysing the same set of data should be considered selective reporting.

The Cochrane guidelines (8.14.1) state: "Selective reporting of analyses using the same data: There are often several different ways in which an outcome can be analysed. For example, continuous outcomes such as blood pressure reduction might be analysed as a continuous or dichotomous variable, with the further possibility of selecting from multiple cut-points."

Scoring System for Chalder Fatigue

A change from the protocol, that will have had a direct impact on the Cochrane analysis, was the scoring system used for the Chalder fatigue scale. The PACE trial protocol proposed two self-report questionnaires as tools to use for the primary endpoint analyses: one was the Chalder fatigue questionnaire, and the other was the Short Form 36 (SF-36) physical function subscale. The scoring system for the Chalder fatigue questionnaire was pre-defined as a bimodal scoring system (i.e. a score of 0 or 1 for each response to the 11 questions, giving a fatigue scale of 0-11). However, after data collection had commenced, the decision was to made to change to a continuous scoring system, known as the Likert system (i.e. a score of 0 ,1, 2, or 3 for each response to the 11 questions, giving a fatigue scale of 0-33). This change was made after the PACE trial's nominal 'sister trial', known as the FINE trial, had completed its analysis of very similar type of data using both the bimodal and Likert scoring systems. The FINE trial investigators had found no significant effect for their primary endpoint when using the bimodal scoring system for Chalder fatigue [9] but determined a significant effect using the Likert system in an informal post-hoc analysis [10]. The FINE trial has published its raw data, as part of the PLoS One data sharing commitment, and an informal analysis has shown that there may potentially be other significant differences in some outcomes, when changing from bimodal to Likert scoring [11].

With regards to risk of selective reporting bias specifically in relation to the PACE trial, the Cochrane review states: "Our primary interest is the primary outcome reported in accordance with the protocol, so we do not believe that selective reporting is a problem."

However, the Likert scoring system for the Chalder fatigue questionnaire is clearly labelled as a secondary outcome in the PACE trial protocol, and not a primary outcome. The protocol specifically lists "Chalder Fatigue Questionnaire Likert scoring (0,1,2,3)" as a "secondary outcome" only. This contradicts the above statement in the review (i.e. "our primary interest is the primary outcome reported in accordance with the protocol"), and it contradicts Larun's implied assertion that the Likert scoring system was "pre-specified" for use as a primary outcome measure.

So, to reiterate, the Likert scoring system, that the Cochrane review has described as a primary outcome measure, is specifically described as only a secondary measure in the PACE trial's protocol. It could not be more specific.

The change in questionnaire scoring methods is more than just a technicality, and may have made a significant difference to the trial's outcomes [11]. The rationale for the change (i.e. "to more sensitively test our hypotheses of effectiveness") may or may not be justified, and the change may or may not be beneficial in terms of better understanding treatment effects, but the fact remains that it was not part of a pre-specified trial plan.

Considering the issues discussed above, the analysis of the Chalder fatigue scores in the Cochrane review should undoubtedly, in my opinion, be considered an unplanned analysis and labelled as such.

Sensitivity Analysis

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

The Cochrane review focused on the mean differences between intervention groups, and whether there was a statistically significant effect, which is the same analysis as the PACE trial's final published outcomes.
Analyses of the PACE trial data using the pre-planned methods have not been published and, to my knowledge, a sensitivity analysis for the (unplanned) final outcomes has neither been published in the PACE trial literature nor the Cochrane review, so it is impossible for the reader to have insight into the impact of the changes.
In terms of what should be done when only post-hoc data is available the Cochrane guidelines (section 8.14.2) advise that a sensitivity analysis should be published: "It is not generally recommended to try to 'adjust for' reporting bias in the main meta-analysis. Sensitivity analysis is a better approach to investigate the possible impact of selective outcome reporting (Hutton 2000, Williamson 2005a)."
It would be helpful if this guideline was adhered to.

Assessment of Risk of Reporting Bias

As well as those outlined above, various other important outcomes in the trial were changed dramatically, such as the recovery analysis, which was reported in a separate publication [12]. Also, the pre-defined 'clinically important difference' was dropped, and was replaced with a 'clinically useful difference' which had an entirely different definition. There were too many deviations from the protocol in the final analyses to list them all in detail here.
The Cochrane guidelines (8.14.2) state: "The assessment of risk of bias due to selective reporting of outcomes should be made for the study as a whole, rather than for each outcome. Although it may be clear for a particular study that some specific outcomes are subject to selective reporting while others are not, we recommend the study-level approach because it is not practical to list all fully reported outcomes in the 'Risk of bias' table."
The Cochrane review currently designates the risk of reporting bias for the PACE trial as "low risk": Under the subheading "Characteristics of included studies" and under: "Selective reporting (reporting bias)", White 2011 is designated as "Low risk". This designation is repeated elsewhere in the review, such as the "Risk of bias summary" in Figure 2.
Kindlon pointed out that the Cochrane guidelines (Table 8.5.d) set out the criteria for the judgement of high risk of reporting bias as follows:
"Any one of the following:

- Not all of the study's pre-specified primary outcomes have been reported;
- One or more primary outcomes is reported using measurements, analysis methods or subsets of the data (e.g. subscales) that were not pre-specified;
- One or more reported primary outcomes were not pre-specified (unless clear justification for their reporting is provided, such as an unexpected adverse effect);
- One or more outcomes of interest in the review are reported incompletely so that they cannot be entered in a meta-analysis;
- The study report fails to include results for a key outcome that would be expected to have been reported for such a study."

I consider the trial to meet at least the first three of these requirements for a high risk of bias. However, in the response to Kindlon, Larun says: "Overall, we don't think that the issues you raise with regard to the risk of selective outcome bias are such as to suspect high risk of bias, but recognize that you may reach different conclusions than us."
I find this claim impossible to square with the Cochrane risk of bias tool, for which, in my opinion, the PACE trial unambiguously meets at least three high risk criteria when only one is required to label the study as high risk.
The Cochrane guidelines advise that the bias risk of a study should be assessed by taking into account the study as a whole; and as all of the PACE trial's main published analyses, including the primary analyses and the recovery analysis, were not pre-planned, this suggests that the Cochrane report should have labelled the trial as having a high risk of reporting bias, according to my interpretation of the Cochrane guidelines.
I request that a revaluation is carried out, with reference to the Cochrane guidelines.

Definition of a Primary Endpoint

In his submission to Cochrane, Kindlon explained that the three pre-planned primary endpoint analyses were abandoned in favour of novel analyses in the final trial analysis: "The three primary efficacy outcomes can be seen in the published protocol" and "None have been reported in the pre-specified way.
I find Larun's response to Kindlon to be confusing and unsatisfactory. As far as my understanding goes, the response does not seem to take account of the Cochrane guidelines. Larun acknowledges that the scoring system for one of the primary outcome assessment tools, was changed: "the scoring method of one was changed", and she acknowledges that the trial's primary endpoint analyses were changed: "the analysis of assessing efficacy also changed from the original protocol."
However, Larun seems to contradict this by saying: "The [final] primary outcomes were the same as in the original protocol [...]"

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Larun's latter statement contradicts the published results which state that the "originally planned" "primary outcomes" were switched: "We used continuous scores for primary outcomes [...] instead of the originally planned composite measures (50% change or meeting a threshold score)." [2]

The primary endpoints (i.e. criteria to judge a successful outcome) were defined in precise detail rather than simply being described as 'fatigue' and 'physical function'. Instead, a specific primary endpoint efficacy analysis was defined which included a required threshold for a positive outcome in fatigue and function at 52 weeks. Also, the questionnaire and scoring method was defined for each primary endpoint.

As the primary endpoint analyses were changed then I would argue that the primary outcomes were substantially changed.

A "primary efficacy endpoint" has been described as "a clinical or laboratory outcome measured in an individual after randomization that allows one to test the primary hypothesis and provides the means of assessing whether a therapy is effective compared with its control." [6]

An example of "Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed" is given in the Consort guidelines: "Example-"The primary endpoint with respect to efficacy in psoriasis was the proportion of patients achieving a 75% improvement in psoriasis activity from baseline to 12 weeks as measured by the PASI [psoriasis area and severity index] Additional analyses were done on the percentage change in PASI scores and improvement in target psoriasis lesions" [13]

In the PACE trial the primary objectives were to compare CBT and GET against SMC. To effectively achieve this comparison, a specific primary analysis was provided, as the primary endpoint, to determine a successful outcome. The results for the pre-planned primary endpoints have not been released.

Conclusion

Larun says: "The Cochrane Risk of Bias tool enables the review authors to be transparent about their judgments, but due to the subjective nature of the process it does not guarantee an indisputable consensus."

I accept that assessment of bias can be subjective but, as I have outlined above, the issues relating to the PACE trial seem clear-cut, according to the Cochrane guidelines, which give very specific advice in relation to the type of the changes that we see here. I do not accept Larun's suggestion that this is a nuanced or subjective evaluation. PACE seems to fail at least the first three criteria in the 'high risk' category of the Cochrane risk tool for reporting bias.

The changes to the PACE trial's primary outcomes had the effect of lowering the threshold for a positive outcome and therefore portraying the interventions in a more positive light. A major purpose of a trial protocol is to avoid bias that can potentially arise through selective reporting. Avoidable bias does a disservice for the medical and patient communities and I would expect Cochrane to be rigorous in pointing out potential bias, and discussing the implications of bias, labelling bias correctly and including unbiased data where possible or including a sensitively analysis where possible. Indeed, this is what the Cochrane guidelines advise, and it is what the public expect of Cochrane. I feel that these issues have been neglected in this specific instance, and a reader of the Cochrane review in isolation would be unaware of any of the issues discussed above, relating to the PACE trial.

I ask for a reassessment and revaluation of this review in relation to the PACE trial and risk of bias.

Many thanks, in advance, for your careful consideration of these issues.

Appendix

PACE trial: protocol-defined primary endpoints - trial protocol [4].

Three Primary Endpoints.

"Primary outcome measures - Primary efficacy measures"

1. "The 11 item Chalder Fatigue Questionnaire measures the severity of symptomatic fatigue, [27] and has been the most frequently used measure of fatigue in most previous trials of these interventions. We will use the 0,0,1,1 item scores to allow a possible score of between 0 and 11. A positive outcome will be a 50% reduction in fatigue score, or a score of 3 or less, this threshold having been previously shown to indicate normal fatigue. [27]"

2. "The SF-36 physical function sub-scale [29] measures physical function, and has often been used as a primary outcome measure in trials of CBT and GET. We will count a score of 75 (out of a maximum of 100) or more, or a 50 % increase from baseline in SF-36 sub-scale score as a positive outcome. A score of 70 is about one standard deviation below the mean score (about 85, depending on the study) for the UK adult population. [51, 52]"

3. "Those participants who improve in both primary outcome measures will be regarded as overall improvers."

PACE trial: post-hoc primary endpoints - main results paper [2].

The difference between mean changes in fatigue and physical function across intervention groups at 52 weeks, using an effect size to assess the efficacy of interventions.

PACE trial: pre-specified primary endpoints -Trial Registry [7]

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

"Endpoints/primary outcome(s)
1. The 11 item Chalder fatigue questionnaire, using categorical item scores to allow a categorical threshold measure of "abnormal" fatigue with a score of 4 having been previously shown to indicate abnormal fatigue.
2. The SF-36 physical function sub-scale, counting a score of 75 (out of a maximum of 100) or more as indicating normal function."

References:

1. Larun L, Brurberg KG, Odgaard-Jensen J, Price JR. Exercise therapy for chronic fatigue syndrome. Cochrane Database Syst Rev. 2016; CD003200.
2. White PD, Goldsmith KA, Johnson AL, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. Lancet 2011; 377:823-36.
3. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. http://handbook.cochrane.org/front˙page.htm (accessed 19 April 2016).
4. White PD, Sharpe MC, Chalder T, et al. Protocol for the PACE trial: a randomised controlled trial of adaptive pacing, cognitive behaviour therapy, and graded exercise, as supplements to standardised specialist medical care versus standardised specialist medical care alone for patients with the chronic fatigue syndrome/myalgic encephalomyelitis or encephalopathy. BMC Neurol. 2007; 7:6.
5. Glossary. Cochrane Community Archive. https://community-archive.cochrane.org/glossary/5#term82 (accessed 20 April 2016)
6. Follmann DA. Primary Efficacy Endpoint. Wiley Encyclopedia of Clinical Trials. 2007.
7. Trial Registry. BioMed Central. Internet Archive. http://web.archive.org/web/20050524130106/ http://www.controlled-trials.com/mrct/trial/CHRONIC FATIGUE SYNDROME/1042/40645.html (accessed 29 April 2016)
8. Walwyn R, Potts L, McCrone P, et al. A randomised trial of adaptive pacing therapy, cognitive behaviour therapy, graded exercise, and specialist medical care for chronic fatigue syndrome (PACE): statistical analysis plan. Trials 2013; 14:386.
9. Wearden AJ, Dowrick C, Chew-Graham C, et al. Nurse led, home based self help treatment for patients in primary care with chronic fatigue syndrome: randomised controlled trial. BMJ. 2010; 340:c1777.
10. Wearden AJ, Dowrick C, Chew-Graham C, et al. Fatigue scale. BMJ Rapid Response. 2010. http://www.bmj.com/rapid-response/2011/11/02/fatigue-scale-0 (accessed 21 Feb 2016).
11. Carter S. Exploring changes to PACE trial outcome measures using anonymised data from the FINE trial. PubMed Commons 2016. http://www.ncbi.nlm.nih.gov/pubmed/23363640#cm23363640˙14248 (accessed 20 Feb 2016).
12. White PD, Goldsmith K, Johnson AL, Chalder T, Sharpe M. Recovery from chronic fatigue syndrome after treatments given in the PACE trial. Psychol Med. 2013; 43:2227-35.
13. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010; 340:c869.

----- ----- -----

I do not have any affiliation with or involvement in any organisation with a financial interest in the subject matter of my comment

Reply

Dear Robert Courtney

Thank you for your detailed comments on the Cochrane review 'Exercise Therapy for Chronic Fatigue Syndrome'. We have the greatest respect for your right to comment on and disagree with our work. We take our work as researchers extremely seriously and publish reports that have been subject to rigorous internal and external peer review. In the spirit of openness, transparency and mutual respect we must politely agree to disagree.

Cochrane reviews aim to report the review process in a transparent way, for example, are reasons for the risk of bias stated. We do not agree that Risk of Bias for the Pace trial (White 2011) should be changed, but have presented it in a way so it is possible to see our reasoning. We find that we have been quite careful in stating the effect estimates and the certainty of the documentation. We note that you read this differently.

Regards,
Lillebeth

Contributors

Feedback submitted by: Robert Courtney
Response submitted by: Lillebeth Larun

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Feedback submitted, 3 June 2016

Summary

Comment: concerns regarding the use of unplanned primary outcomes in the Cochrane review

Summary

In this submission, I will discuss the details and implications of unplanned revisions to the Cochrane review's protocol, specifically changes to the primary outcomes. I will raise concerns about the clarity with which the changes to the protocol have been explained in the review and I will question the justification given for switching the primary outcomes. I will compare the details of the pre-specified primary outcomes with the unplanned (revised) primary outcomes. I will explore how the protocol revisions have impacted the overall conclusions of the review, and how some review outcomes have been misrepresented in the main discussions. I will also briefly discuss potential biases involved in reviewing open-label studies that use self-report outcomes, and how such biases may potentially have affected the review's outcomes. Finally, I will discuss what I believe is: a lack of clarity in how the review has discussed and portrayed outcomes, and; a lack of depth in how potential biases have been considered and explored.

I will conclude by asking the reviewers to reassess the review, including the decision to switch the primary outcomes, with a view to improving clarity, rigour and accuracy. I specifically ask the reviewers to:

1. Amend the review as per the Cochrane guidelines (i.e. "every effort should be made to adhere to a predetermined protocol"), and revert to the pre-planned primary analyses; and

2. Clearly and unambiguously explain that all but one health indicator (i.e. fatigue, physical function, overall health, pain, quality of life, depression, and anxiety, but not sleep) demonstrated a non-significant outcome for pooled treatment effects at follow-up for exercise therapy versus passive control; and

3. Include a rigorous assessment of how the potential for bias may have affected outcomes.

Introduction

After detailed scrutiny of the current version of the Cochrane review of exercise therapy for chronic fatigue syndrome (version 4, dated 7 February 2016) [1], I have noticed that the primary outcomes of the review have not been reported as per the pre-specified review plan, but that unplanned (revised) primary analyses have been published in the place of the pre-specified analyses. (By 'unplanned', I refer to revisions to the methodology that were not pre-specified in the review's protocol.) The switching of primary outcomes (from pre-specified to unplanned analyses) is not mentioned in the main discussions, conclusion, or abstract, and is not explicitly explained anywhere in the review. I had to carry out a detailed inspection of the review to understand exactly what had been changed.

At the very end of the full version of the review, a section titled "[d]ifferences between protocol and review" explains the deviations from the protocol:

"[...] in the protocol it is stated, "where results for continuous outcomes were presented using different scales or different versions of the same scale, we used standardised mean differences (SMDs)." We realise that the standardised mean difference (SMD) is much more difficult to conceptualise and interpret than the normal mean difference (MD); therefore we decided to report both MDs and SMDs in the Results section. In general, MDs are reported in the main Results section, whereas SMDs are supplied under the "Sensitivity and subgroup analysis" subheading."

Although the above quote isn't explicit in referring to the primary outcomes, it explains the nature of, and rationale for, the unplanned changes to the review's primary outcomes. The only reason given for changing the pre-specified outcomes was that "the standardised mean difference (SMD) is much more difficult to conceptualise and interpret than the normal mean difference". No evidence is provided to support this assertion, and it appears to be an assumption about the readers' ability to interpret outcomes.

The outcomes of the review's pre-specified primary analyses are outlined in the analysis section, but are only mentioned briefly (i.e. only one or two sentences are used to explain each outcome), and the pre-specified outcomes are not discussed in the review's main discussions, abstract or conclusions. The pre-specified analyses have been relegated to the status of "sensitivity analyses", and it is not explicitly explained that these sensitivity analyses are the pre-specified primary analyses. It is easy for a reader to overlook these important outcomes and to misunderstand their significance. I am concerned that most readers will be unaware of these changes to the primary outcomes and of the significance of the changes to the protocol.

I consider the changes to have significantly altered the fundamental design, the main outcomes, and overall interpretation of the review.

Primary Outcomes

I would like to take this opportunity to explain the details of the changes to the primary outcomes to the reader, to the best of my understanding. The review compares exercise therapy with a passive control (e.g. treatment as usual), which is the focus of this submission. Outcomes for exercise therapy compared with other interventions (e.g. cognitive-behavioural therapy, supportive therapy, and pacing) are also included in the review, but are not central to the concerns of this submission and will not be discussed further. The

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

review uses two primary outcome measures (fatigue and adverse outcomes) but adverse outcomes are not relevant to this submission. A primary analysis at both end of treatment (12 to 26 weeks) and at follow-up (52 to 70 weeks) is carried out. This submission focuses on primary analyses in relation to fatigue only, for exercise therapy versus passive control only.

The protocol defined two pre-specified primary analyses (one at end of treatment and one at follow-up) that were to determine the pooled treatment effects of all eligible studies on fatigue. The two analyses were to determine a standardised mean difference (SMD) for the pooled studies.

An unplanned decision was later made to relegate these pre-specified primary analyses to the status of sensitivity analyses and to replace them with two unplanned analyses which assessed the same studies but by a different statistical method. The unplanned analyses (1.1 and 1.2) do not provide an overall (pooled) treatment effect but provide mean differences in a number of sub-analyses of studies grouped together based on the specific tool or scoring method used to measure fatigue.

The two pre-specified primary analyses are published as sensitivity analysis 1.19 (fatigue at end of treatment) and another analysis (fatigue at follow-up) which has not been designated a numerical identifier. To reiterate; these two analyses provide the pooled standardised mean difference for fatigue for all eligible studies. Analysis 1.19 was included within the comprehensive set of tables published in the review, however, the follow-up analysis (which demonstrated a non-significant outcome) was (uniquely for primary outcomes) omitted from the set of tables (i.e. it was not published as a table) but was only briefly outlined under the subheadings: "Sensitivity analysis" > "Investigating heterogeneity" (see appendix, below, for quote). As this analysis is not mentioned elsewhere in the review, and is only mentioned in one sentence, it is easy to miss.

To clarify; the unplanned analyses assess the same studies as the pre-specified analyses, but only the pre-specified analyses indicate the overall treatment effect for all eligible studies pooled together.

The outcomes of the two pre-specified analyses, using a pooled standardised mean difference (SMD) for all eligible studies, were that exercise therapy (versus passive control) at end of treatment (i.e. analysis 1.19) had a significant positive treatment effect (SMD: -0.68; 95% CI -1.02 to -0.35), whereas at follow-up the treatment effect was not significant (SMD: -0.63; 95% CI -1.32 to 0.06).

The unplanned primary analysis 1.1 (fatigue at end of treatment) includes three separate sub-analyses which all demonstrate a positive treatment effect, whereas unplanned analysis 1.2 (fatigue at follow-up) had mixed outcomes with two out of three sub-analyses demonstrating a significant treatment effect.

So, to reiterate, the pre-specified primary analyses demonstrate that exercise therapy (versus passive control) had a significant pooled treatment effect on fatigue at end of treatment, but no significant effect at follow-up. Whereas the unplanned (revised) analyses demonstrate significant treatment effects at end of treatment but mixed outcomes at follow-up.

The fact that unplanned analysis 1.2 (fatigue at follow-up) did not consistently demonstrate significant treatment effects is not explained with clarity in the main discussions of the review. For example, the outcomes are described as follows: "Moderate-quality evidence showed exercise therapy was more effective at reducing fatigue compared to 'passive' treatment or no treatment." (See the appendix, below, for more quoted examples.)

The main discussions in the review also fail to inform the reader that the pooled treatment effect on fatigue (compared to a passive control), for all eligible studies at follow-up, demonstrated a lack of significance, as per the pre-specified primary analysis.

All Outcomes at Follow-Up

Despite the limitations associated with self-report measures [2], physical function (a secondary outcome in the review) is widely considered a useful measure for demonstrating severity of illness and functional changes in outcomes for chronic fatigue syndrome [3,4,5]. It may be an especially helpful measure when assessing exercise therapy because exercise therapy is designed specifically to address physical function or tolerance to exercise, or both [6]. It seems reasonable to expect physical function to improve after a course of exercise therapy in chronic fatigue syndrome patients, if the therapy is clinically beneficial. The review reports that exercise therapy (when compared to passive control) has a positive effect on self-report physical function at end of treatment (analysis 1.5), but this effect is not sustained and there was no significant treatment effect at follow-up (see analysis 1.6).

There was also no significant effect on self-perceived overall health at follow-up (see analysis 1.15). Indeed, if we consider all of the health-related pre-specified (primary and secondary) outcomes for the review, for exercise therapy versus passive control, then with the exception only of sleep, all the indicators of health (i.e. fatigue, physical function, overall health, pain, quality of life, depression, and anxiety), showed no significant treatment effects at follow-up. (The remaining measures were: serious adverse reactions to treatment; drop-outs; and 'health resource use' for which a pooled effect size was not provided but which demonstrated non-significant differences between intervention arms in all but one of the sub-analyses.) This means that only sleep had a significant positive treatment outcome, at follow-up, as per the pre-specified health indicators, for exercise therapy versus passive control.

Put simply, apart from sleep, all the pooled analyses demonstrate that there were no significant health benefits from exercise therapy at follow-up.

These outcomes present a significantly different picture to the impression given by the review authors in their main discussions, abstract, conclusions and summaries wherein, for example, outcomes in general, including fatigue, physical function and overall health, are

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

described as being broadly positive (e.g. it is stated that "patients with CFS may generally benefit and feel less fatigued following exercise therapy" and: "Exercise therapy had a positive effect on people's daily physical functioning, sleep and self-ratings of overall health.") Furthermore, some specific erroneous information has been included in the main text to support the review authors' interpretation; i.e. the main discussion erroneously describes both physical function and self-rated overall health as indicating a positive treatment effect at follow-up, when in fact the outcomes (i.e. analyses 1.6 and 1.15) were not significant. The reviewers erroneously assert that: "A positive effect of exercise therapy was observed both at end of treatment and at follow-up with respect to [...] physical functioning (Analysis 1.5; Analysis 1.6) and self-perceived changes in overall health (Analysis 1.14; Analysis 1.15)." (See appendix, below, for full quote.)

The non-significant outcomes seen in all but one of the pre-specified health indicators at follow-up (exercise vs passive control) were not discussed or explored in the discussions of the Cochrane review. I find this omission disappointing because the information would help to inform patients and clinicians of the ongoing treatment effects that they might realistically expect from behavioural therapies such as exercise therapy. I believe that the review would be more robust and helpful if it accurately highlighted and adequately explored these issues in the main discussions.

The health outcomes at follow-up would currently be completely lost on a reader who did not scrutinise the individual analyses of the review but relied upon the abstract or main discussions.

Bias Inherent in Open-Label Studies

Another issue that I believe is not explored with careful consideration is the possible implications relating to a review of purely open-label studies; i.e. the possibility that any initial positive treatment effects broadly seen in this review at end of treatment, may entirely, or to some degree, reflect biases inherent in trial methodologies that are unable to blind patients, therapists or trial investigators to the treatment arm. The review itself explains that formal blinding "is not inherently possible in trials of exercise therapy" and that this "increases risk of bias, as instructors' and participants' knowledge of group assignation might have influenced the true effect." The trend in this review towards non-significant effects, after treatment has ended, may lend strength to a concern that the initial self-report treatment effects are transient and may be the result of various inherent methodological biases in open-label trials that use self-report outcome measures [2,7,8]. Potential methodological biases in open-label trials using self-report outcomes may be, for example: inadequate control conditions; self-reporting bias; therapist allegiance; and/or unplanned changes to trial methodology [7,9].

Readers might be interested to note that, for White 2011, which was the largest trial included in the review, the follow-up data used in the review was at 52 weeks [10] but further follow-up data has also been published, at a median of 2.5 years after randomisation, which demonstrated no significant differences between intervention arms for the primary outcomes [11].

Summary of Outcomes

In summary, the pre-specified primary analyses for fatigue were to assess the pooled standardised mean differences. However, the reviewers then made a post-hoc decision to replace these analyses for which the only rationale provided was an assumption that a standardised mean difference is supposedly "more difficult to conceptualise and interpret". When all of the eligible studies are pooled, as per the pre-specified plan, the pooled treatment effect at follow-up is not significant. However, the promotion of the unplanned analyses has allowed the lack of a significant pooled treatment effect at follow-up to be overlooked and dismissed in the main analyses and discussions, to the point where the main discussions could be interpreted to indicate that the treatment effects for fatigue were entirely positive (see appendix, below, for quotes).

I question whether this is an appropriate level of clarity compared to what is expected from a Cochrane review. Cochrane reviews have a reputation of providing transparent, uncomplicated, straightforward and reliable explanations of complex and rigorous analyses, whereas this review has: used unplanned primary outcomes without a robust or evidence-based reason for switching outcomes; provided just one sentence to explain the changes to the pre-specified primary outcomes; omitted a crucial sensitivity analysis from the tables section; has not reflected the entire range of outcomes in the abstract, conclusions or main discussions; and has inaccurately described outcomes at follow-up for physical function and overall health.

Justification for Switching Primary Outcomes

The reason given for switching the primary outcomes in the review is: "We realise that the standardised mean difference (SMD) is much more difficult to conceptualise and interpret than the normal mean difference (MD) [...]".

However, it is questionable whether the reason given for switching the primary outcomes justifies such an unplanned fundamental change in the methodology of the review; no justification is given as to why the reviewers believe that readers would find it easier to interpret the mean scores of a range of disparate fatigue questionnaires, in a series of sub-analyses, rather than a single standardised mean difference for a pooled analysis of eligible studies. It is not clear to me why it is assumed that a variety of separate fatigue scales should be easier to understand and interpret than a single standardised mean difference. As the changes to the protocol have had the effect of changing the primary outcomes at follow-up, this means it would be desirable to provide a well-reasoned case to deviate from the protocol and switch the primary outcomes.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

The claim with regards to interpretability raises the question of why standardised mean differences are adequate for other Cochrane studies, but not this particular study. Cochrane has not adopted a policy of avoiding using standardised mean differences; instead the Cochrane guidelines (section 12.6) encourage their use [12]. So this decision appears to be a novel post-hoc decision specific for this study.

The Cochrane guidelines (section 12.6.1) actually suggest that ordinary mean differences can be difficult to interpret: "The units of such outcomes [i.e. mean differences] may be difficult to interpret, particularly when they relate to rating scales." [12] The guidelines (section 12.6.1) acknowledge that there may be difficulties in interpreting standardised mean differences: "Without guidance, clinicians and patients may have little idea how to interpret results presented as SMDs." The guidelines do not favour one method over another in general, but describe how each may be used for specific purposes; if one wishes to provide an overall treatment effect for studies that use different measures to measure the same construct, then the standardised mean difference is a standard tool which is used widely in Cochrane reviews and other research. The guidelines suggest that "[t]here are several possibilities for re-expressing [standardised means differences] in more helpful ways".

Implications Related to Changing Trial Protocols and Outcome Switching

The unplanned changes to the review make it vulnerable to potential bias or accusations of bias; conscious or unconscious personal or professional preferences have the potential to affect post-hoc decisions with respect to methodology. Even if investigators are scrupulous in the rigour of their decision making, unexpected biases have the potential to creep into unplanned decisions, which is an issue that factors into the reasons why pre-trial plans (e.g. trial registers and protocols) have become widespread [7], and are used for Cochrane reviews [12].

The Cochrane Guidebook for reviewers (Section: 2.1; "Rationale for protocols") [12] explains: "Post hoc decisions made when the impact on the results of the research is known, such as excluding selected studies from a systematic review, are highly susceptible to bias and should be avoided."

In the same paragraph, the guidelines also state: "While every effort should be made to adhere to a predetermined protocol, this is not always possible or appropriate."

However, it seems that, in this case, every effort was not made to adhere to the protocol because the unplanned changes seem to be based on preference rather then necessity, and the pre-planned analyses have not been shown to be inferior, inadequate or inappropriate.

Conclusion

I find the changes to the protocol to be of particular concern for the following reasons:

1. The final primary analyses are unplanned and have replaced adequate, and arguably more helpful, pre-specified analyses;
2. The rationale provided for the changes was neither robust nor evidence-based but was based upon an assumption;
3. The changes have significantly altered the main outcomes and affected the interpretation of the review (i.e. changed one of the two main outcomes from an insignificant treatment effect to an inconsistent but broadly positive effect); and
4. The pre-planned analysis for fatigue at follow-up, has been omitted from the tables section of the review which, as far as I understand, is a unique omission for the primary outcomes.

For the sake of simplicity, rigour, and transparency, I ask the review team to reassess the review, including the decision to switch the primary outcomes, and to:

1. Amend the review as per the guidelines in the Cochrane Guidebook quoted above (i.e. "every effort should be made to adhere to a predetermined protocol"), and to revert to the pre-planned primary analyses; and
2. Clearly and unambiguously explain that all but one health indicator (i.e. fatigue, physical function, overall health, pain, quality of life, depression, and anxiety, but not sleep) demonstrated a non-significant outcome for pooled treatment effects at follow-up for exercise therapy versus passive control; and
3. Include a rigorous assessment of how the potential for bias may have affected outcomes of the open-label studies in this review, with consideration of the use of self-report measures in open-label studies.

--- ---

Appendix

Relevant Quotes from the Review

Differences between protocol and review

"[...] in the protocol it is stated, "where results for continuous outcomes were presented using different scales or different versions of the same scale, we used standardised mean differences (SMDs)." We realise that the standardised mean difference (SMD) is much more difficult to conceptualise and interpret than the normal mean difference (MD); therefore we decided to report both MDs and SMDs in the Results section. In general, MDs are reported in the main Results section, whereas SMDs are supplied under the "Sensitivity and subgroup analysis" subheading."

Quotes detailing the outcomes of the pre-specified analyses:

Effects of interventions > Exercise therapy versus treatment as usual, relaxation or flexibility > Sensitivity analysis

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Fatigue, End of Treatment

"At end of treatment, fatigue was measured and reported on different scales, and we performed a sensitivity analysis in which all available studies were pooled using an SMD method. This strategy led to a pooled random-effects estimate of -0.68 (95% CI -1.02 to -0.35), but the analysis suffered from considerable heterogeneity (I^2 = 78%, P value < 0.0001; Analysis 1.19). The observed heterogeneity was caused mainly by the deviating results presented in Powell 2001. Exclusion of Powell 2001 gave rise to a pooled SMD of -0.46 (95% CI -0.63 to -0.29) - an estimate that was not associated with heterogeneity (I^2 = 13%, P value 0.33)."

Fatigue, Follow-up

"At follow-up, the four available studies (Jason 2007; Powell 2001; Wearden 2010; White 2011) measured and reported fatigue on different scales, and we performed a sensitivity analysis in which all available studies were pooled using an SMD method. The pooled SMD estimate is -0.63 (95% CI -1.32 to 0.06), but heterogeneity was extensive (I^2 = 93%, P value < 0.00001)."

Quotes from main discussion sections of review re effectiveness of graded exercise (compared with passive control) on fatigue

Abstract > Authors' conclusions

"Patients with CFS may generally benefit and feel less fatigued following exercise therapy, and no evidence suggests that exercise therapy may worsen outcomes."

Plain language summary > What does evidence from the review tell us?

"Moderate-quality evidence showed exercise therapy was more effective at reducing fatigue compared to 'passive' treatment or no treatment."

Discussion > Summary of main results

"When exercise therapy was compared with 'passive control,' fatigue was significantly reduced at end of treatment (Analysis 1.1)."

Quotes selectively reporting secondary outcomes

Abstract > Authors' conclusions

"A positive effect with respect to sleep, physical function and self-perceived general health has been observed, but no conclusions for the outcomes of pain, quality of life, anxiety, depression, drop-out rate and health service resources were possible."

Plain language summary > What does evidence from the review tell us?

"Exercise therapy had a positive effect on people's daily physical functioning, sleep and self-ratings of overall health."

Erroneous reporting of outcomes for physical function and overall health at follow-up

Discussion > Summary of main results

"A positive effect of exercise therapy was observed both at end of treatment and at follow-up with respect to sleep (Analysis 1.12; Analysis 1.13), physical functioning (Analysis 1.5; Analysis 1.6) and self-perceived changes in overall health (Analysis 1.14; Analysis 1.15)."

--- ---

References

1. Larun L, Brurberg KG, Odgaard-Jensen J, Price JR. Exercise therapy for chronic fatigue syndrome. Cochrane Database Syst Rev. 2016; CD003200.

2. Kindlon TP. Objective measures found a lack of improvement for CBT & GET in the PACE Trial: subjective improvements may simply represent response biases or placebo effects in this non-blinded trial. BMJ Rapid Response 2015. http://www.bmj.com/content/350/bmj.h227/rr-10 (accessed May 18, 2016).

3. Buchwald D, Pearlman T, Umali J, Schmaling K, Katon W. Functional status in patients with chronic fatigue syndrome, other fatiguing illnesses, and healthy individuals. Am J Med. 1996;101:364-70.

4. Cook DB, Lange G, DeLuca J, Natelson BH. Relationship of brain MRI abnormalities and physical functional status in chronic fatigue syndrome. Int J Neurosci. 2001;107:1-6.

5. Crawley E, Sterne JA. Association between school absence and physical function in paediatric chronic fatigue syndrome/myalgic encephalopathy. Arch Dis Child. 2009;94:752-6.

6. Bavinton J, Darbishire L, White PD. PACE manual for therapists; graded exercise therapy for CFS/ME. 2004. Internet. http://www.wolfson.qmul.ac.uk/images/pdfs/5.get-therapist-manual.pdf (accessed May 18, 2016).

7. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010; 340:c869.

8. Wilshire CE. Re: Tackling fears about exercise is important for ME treatment, analysis indicates. BMJ Rapid Response 2015. http://www.bmj.com/content/350/bmj.h227/rr-7 (accessed May 18, 2016).

9. Van de Mortel TF. Faking it: social desirability response bias in self-report research. Australian Journal of Advanced Nursing, The. 2008;25:40.

10. White PD, Goldsmith KA, Johnson AL, et al. Comparison of adaptive pacing therapy, cognitive behaviour therapy, graded exercise therapy, and specialist medical care for chronic fatigue syndrome (PACE): a randomised trial. Lancet 2011; 377:823-36.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

11. Sharpe M, Goldsmith KA, Johnson AL, et al. Rehabilitative treatments for chronic fatigue syndrome: long-term follow-up from the PACE trial. Lancet Psychiatry 2015; 2:1067-74.
12. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions. Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Internet. http://handbook.cochrane.org (accessed May 12, 2016).
--- ---
I do not have any affiliation with or involvement in any organisation with a financial interest in the subject matter of my comment.

Reply

Dear Robert Courtney
Thank you for your ongoing and detailed scrutiny of our review. We have the greatest respect for your right to comment on and disagree with our work, but in the spirit of openness, transparency and mutual respect we must politely agree to disagree.
Presenting health statistics in a way that makes sense to the reader is a challenge. Statistical illiteracy is - according to Girgerenzer and co-workers - common in patients, journalists, and physicians (1). With this in mind we have presented the results as mean difference (MD) related to the relevant measurement scales, for example Chalder Fatigue Scale, as well as standardised mean difference (SMD). The use of MD enables the reader to transfer the results to the relevant measurement scale directly and judge the effect in relation to the scale. We disagree that presenting MD and SMD rather than SMD and MD is an important change, and we disagree with the claim that the analysis based on MD and SMD are inconsistent. This has been discussed as part of the peer-review process. Confidence intervals are probably a better way to interpret data that P values when borderline results are found (2). Interpreting the confidence intervals, we find it likely that exercise with its SMD on -0.63 (95% CI -1.32 to 0.06) is associated with a positive effect. Moreover, one should also keep in mind that the confidence interval of the SMD analysis are inflated by the inclusion of two studies that we recognize as outliers throughout our review. Absence of statistical significance does not directly imply that no difference exists.
All the included studies reported results after the intervention period and this is the main results. The results at different follow-up times are presented in the text, but we have only included data available at the last search date, 9 may 2014. When the review is updated, a new search will be conducted to find new, relevant follow up data and new studies. As a general comment, it is often challenging to analyse follow-up data gathered after the formal end of a trial period. There is always a chance that participants may receive other treatments following the end of the trial period, a behaviour that will lead to contamination of the original treatment arms and challenge the analysis.
Cochrane reviews aim to report the review process in a transparent way, which enables the reader to agree or disagree with the choices made. We do not agree that the presentation of the results should be changed. We note that you read this differently.
Regards,
Lillebeth Larun
1. Girgerenzer G, Gaissmaier W, Kurtz-Milcke E, Schwartz LM, Woloshin S. Helping Doctors and Patients Make Sense of Health Statistics. Pyschological Science in the Public Interest, 2008;8:(2):53-96. http://www.psychologicalscience.org/journals/pspi/pspi˙8˙2˙article.pdf.
2. Hackshaw A and Kirkwood A. Interpreting and reporting clinical trials with results of borderline significance. *BMJ* 2011;343:d3340 doi: 10.1136/bmj.d3340

Contributors

Feedback submitted by: Robert Courtney
Response submitted by: Lillebeth Larun

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Feedback submitted, 12 May 2016

Summary

Comment: A query regarding the way outcomes for physical function and overall health have been described in the abstract, conclusion and discussions of the review.

I would like to query the way that the outcomes for both physical function and overall health have been reported in the abstract, conclusion and in the main discussion section of the current version (version 4) of the Cochrane review by Larun et al., dated 7 February 2016 [1].

The abstract, conclusion and main discussion section unambiguously indicate that there was a positive treatment effect on both physical function and overall health, in relation to exercise therapy compared to a passive control.

For example, with respect to exercise therapy versus passive control, the "authors' conclusions" in the abstract state without qualification that: "A positive effect with respect to sleep, physical function and self-perceived general health has been observed[...]". Another section of the review ("What does evidence from the review tell us?") asserts that: "Exercise therapy had a positive effect on people's daily physical functioning, sleep and self-ratings of overall health." The "summary of main results" unequivocally states that: "A positive effect of exercise therapy was observed both at end of treatment and at follow-up with respect to sleep (Analysis 1.12; Analysis 1.13), physical functioning (Analysis 1.5; Analysis 1.6) and self-perceived changes in overall health (Analysis 1.14; Analysis 1.15)." (Please see the appendix, below, to read these quotes in full.)

However, upon careful consideration of the relevant analyses, it seems that there were not consistent positive treatment effects for either physical function or overall health in relation to exercise therapy versus passive control. Instead, for both of these variables, there was a significant treatment effect only at end of treatment, but not at follow-up.

The relevant analyses are 1.5 (end of treatment) and 1.6 (follow-up) for self-report physical function, and 1.14 (end of treatment) and 1.15 (follow-up) for self-report overall health.

Analysis 1.5 assessed the pooled treatment effect on physical function at end of treatment for all eligible studies, and demonstrates a significant effect. Analysis 1.6 used the same criteria but at follow-up, and demonstrates that there was not a significant effect for physical function at follow-up.

Analysis 1.14 assessed the pooled treatment effect on overall health at end of treatment for all eligible studies, and demonstrates a significant effect. Analysis 1.15 used the same criteria but at follow-up, and demonstrates that there was not a significant effect for overall health at follow-up.

The lack of a significant treatment effect at follow-up is clearly illustrated by analyses 1.6 and 1.15.

These outcomes are also confirmed in the analysis section of the review where, in relation to the difference between exercise therapy versus passive control, for physical function at follow-up, it is confirmed that: "[...] little or no difference cannot be ruled out." And for overall health at follow-up, it is confirmed that "the confidence interval implies inconclusive results".

I believe that these outcome are not reflected accurately in the abstract, the main discussions or the conclusions of the review; specifically the extracts that are quoted above and in the appendix below. For example, the "summary of main results" specifically claims that positive treatment effects are demonstrated by analyses 1.6 and 1.15, but these analyses actually demonstrate an absence of significant treatment effects. The discussion claims: "A positive effect of exercise therapy was observed [...] at follow-up with respect to [...] physical functioning ([...]Analysis 1.6) and self-perceived changes in overall health ([...]Analysis 1.15)."

It is generally understood that a "positive" treatment effect equates to a significant effect, and I believe that the Cochrane text should reflect this, or at least clarify that the term "positive effect" is being used to indicate a lack of significance.

It is likely that many readers will not read the full report or scrutinise each individual analysis but will read only the abstract, main discussions or conclusions, so I believe it is important for the discussions to carefully and accurately reflect the outcomes of the analyses.

Cochrane has a reputation for upholding the highest standards including with respect to explaining outcomes in accurate and straightforward language. With this in mind, I request that the Cochrane review team kindly review the apparent disparities described above and amend the text of the discussions and conclusions where appropriate, in order to reflect the lack of a significant treatment effect for physical function and overall health at follow-up with respect to exercise therapy versus passive control.

Appendix

Quotes from the review:

Abstract > Authors' conclusions

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

"Patients with CFS may generally benefit and feel less fatigued following exercise therapy, and no evidence suggests that exercise therapy may worsen outcomes. A positive effect with respect to sleep, physical function and self-perceived general health has been observed, but no conclusions for the outcomes of pain, quality of life, anxiety, depression, drop-out rate and health service resources were possible."
What does evidence from the review tell us?
"Moderate-quality evidence showed exercise therapy was more effective at reducing fatigue compared to 'passive' treatment or no treatment. Exercise therapy had a positive effect on people's daily physical functioning, sleep and self-ratings of overall health."
Summary of main results
"[...] A positive effect of exercise therapy was observed both at end of treatment and at follow-up with respect to sleep (Analysis 1.12; Analysis 1.13), physical functioning (Analysis 1.5; Analysis 1.6) and self-perceived changes in overall health (Analysis 1.14; Analysis 1.15)."

Reference
1. Larun L, Brurberg KG, Odgaard-Jensen J, Price JR. Exercise therapy for chronic fatigue syndrome. Cochrane Database Syst Rev. 2016; CD003200.

I do not have any affiliation with or involvement in any organisation with a financial interest in the subject matter of my comment

Reply

Thank you for your ongoing and detailed scrutiny of our review. We have the greatest respect for your right to comment on and disagree with our work, but in the spirit of openness, transparency and mutual respect we must (again) politely agree to disagree.
All the included studies reported results after the intervention period and this is the main result. The results at different follow-up times are presented in the text. It can be noted that the quality of the evidence is higher for the end-of-treatment time point because more trials are included, and hence, we do not agree that it is wrong to give higher weight to these results in the abstract. Additionally, it is often challenging to analyse follow-up data gathered after the formal end of a trial period. There is always a chance that participants may receive other treatments following the end of the trial period, a behaviour that will lead to contamination of the original treatment arms and challenge the analysis.
Cochrane reviews aim to report the review process in a transparent way, which enables the reader to agree or disagree with the choices made. We do not agree that the presentation of the results should be changed. We note that you read this differently.

Contributors

Feedback submitted by: Robert Courtney
Response submitted by: Lillebeth Larun

WHAT'S NEW

Last assessed as up-to-date: 9 May 2014.

Date	Event	Description
21 June 2016	Feedback has been incorporated	Feedback has been added, along with the author's response.
1 February 2016	Feedback has been incorporated	Feedback has been added along with the author's response.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

(*Continued*)

20 November 2014	New citation required but conclusions have not changed	Four new studies have been added in this update, and the conclusion strengthens results reported in the 2004 version of the review
2 October 2014	New search has been performed	This review has been updated with newer methodology, and new studies have been incorporated

HISTORY

Protocol first published: Issue 3, 2001

Review first published: Issue 3, 2004

Date	Event	Description
1 November 2008	Amended	This review has been converted to the new review format
25 May 2004	New search has been performed	The protocol for this review has undergone post hoc alteration based on feedback from referees. The following sections have been altered: Types of interventions; Search strategy; Methods of the review
8 May 2004	New citation required and conclusions have changed	Substantive amendments have been made

CONTRIBUTIONS OF AUTHORS

LL, KGB, JO-J: checked trials for inclusion.

LL, KGB, JO-J: extracted data for the update.

LL, JO-J, KGB: analysed data for the update.

LL, JO-J, JRP, KGB: wrote the update.

DECLARATIONS OF INTEREST

LL: nothing to declare.

KGB: nothing to declare.

JO-J: nothing to declare.

JRP: nothing to declare.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

SOURCES OF SUPPORT

Internal sources

- University of Oxford Department of Psychiatry, UK.
- Norwegian Knowledge Centre for Health Services, Norway.

External sources

- No sources of support supplied

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

Changes made to the original review are stated below.

Objectives have been changed from '(1) To systematically review all randomised controlled trials of exercise therapy for adults with CFS, and (2) To investigate the relative effectiveness of exercise therapy alone or as part of a treatment plan' in the 2004 version to 'The objective of this review was to determine the effects of exercise therapy (ET) for patients with chronic fatigue syndrome (CFS) as compared with any other intervention or control' in this update.

Comparisons have been changed from: '(1) Exercise therapy versus treatment as usual or relaxation plus flexibility, (2) Exercise therapy versus pharmacotherapy (fluoxetine), (3) Exercise therapy alone versus exercise therapy plus pharmacotherapy (fluoxetine) and (4) Exercise therapy alone versus exercise therapy plus patient education' in the 2004 version to the following in this update.

- '"Passive control": treatment as usual/waiting-list control/relaxation/flexibility.

    - "Treatment as usual" comprises medical assessments and advice given on a naturalistic basis. "Relaxation" consists of techniques that aim to increase muscle relaxation (e.g. autogenic training, listening to a relaxation tape). "Flexibility" includes stretches performed according to selected exercises given.

- Psychological therapies: cognitive-behavioural therapy (CBT)/cognitive treatment/supportive therapy/behavioural therapies/psychodynamic therapies.

- Adaptive pacing therapy.

- Pharmacological therapy (e.g. antidepressants).'

We have revised and reordered the list of secondary outcomes for clarity and have added self-reported changes in overall health as a new outcome, while moving adverse effects from a secondary outcome to a primary outcome.

We have updated the methods according to recommendations provided in the 2011 version of the *Cochrane Handbook for Systematic Reviews of Interventions*. For the first version of this review (2004), assessment of methodological quality was conducted according to contemporary criteria of the handbook of The Cochrane Collaboration (Alderson 2004).The adequacy of allocation concealment was rated as adequate (A), unclear (B) or inadequate (C) or as not used (D), and the CCDAN Quality Rating System (Moncrieff 2001) was applied. For this update, we reextracted data on risk of bias to comply with current recommendations, and we used concealment of allocation as the main quality criterion for included studies.

To explore possible differences between studies using different treatment strategies, control conditions and diagnostic criteria, we decided to perform post hoc subgroup analyses when applicable. We also performed post hoc subgroup analyses excluding Powell 2001, as the results reported in this trial seem to have introduced considerable heterogeneity into the analysis. Moreover, in the protocol it is stated, "where results for continuous outcomes were presented using different scales or different versions of the same scale, we used standardised mean differences (SMDs)." We realise that the standardised mean difference (SMD) is much more difficult to conceptualise and interpret than the normal mean difference (MD); therefore we decided to report both MDs and SMDs in the Results section. In general, MDs are reported in the main Results section, whereas SMDs are supplied under the "Sensitivity and subgroup analysis" subheading.

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Planned methods not used in this review

Cluster trials

Studies often employ 'cluster randomisation' (such as randomisation by clinician or practice), but analysis and pooling of clustered data pose problems. First, study authors often fail to account for intraclass correlation in clustered studies, leading to a 'unit of analysis' error (Bland 1997) whereby P values are spuriously low, confidence intervals unduly narrow and statistical significance overestimated. This causes type I errors (Bland 1997; Gulliford 1999).

No cluster RCTs were identified in this version of the review. Should such studies be identified in future updates, we will use the following methodological approach. When clustering has not been accounted for in primary studies, we will present data in a table, with a (*) symbol to indicate the presence of a probable unit of analysis error. We will seek to contact first authors of studies to obtain intraclass correlation co-efficients for their clustered data and to adjust for this by using accepted methods (Gulliford 1999). When clustering is incorporated, we will present the data as if from a parallel-group randomised study, but adjusted for the clustering effect. We will additionally exclude such studies in a sensitivity analysis.

If cluster studies are appropriately analysed by taking into account intraclass correlation co-efficients and relevant data documented in the report, synthesis with other studies will be possible using the generic inverse variance technique.

Cross-over trials

A major concern of cross-over trials is the potential for carry-over effect. This occurs when an effect (e.g. pharmacological, physiological, psychological) of treatment in the first phase is carried over to the second phase. As a consequence of entry to the second phase, participants can differ systematically from their initial state despite a wash-out phase. For the same reason, cross-over trials are not appropriate when the condition of interest is unstable (Elbourne 2002). As both effects are very likely in CFS/ME, randomised cross-over studies were eligible but only when data up to the point of first cross-over were used. Data from the subsequent (second) period of the cross-over trial were not considered for analysis.

Studies with multiple treatment groups

Multiple dose groups

Some studies may address the effects of different levels of supervision and follow-up with regards to the exercise intervention and the comparator (e.g. sessions for designing exercise therapy, sessions for designing exercise therapy and planned telephone contacts, sessions for designing exercise therapy and seven face-to-face treatment sessions, usual care). Should we identify trials that take this approach in future updates, we will adopt the following approach. For dichotomous outcomes, we will sum up the sample sizes and the numbers of people with events across all intervention groups. For continuous outcomes, means and standard deviations will be combined using the methods described in Chapter 7 (Section 7.7.3.8) of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011).

Multiple medications

Some studies may combine several interventions with one comparison group. Should we identify trials of this nature in future updates, we will analyse the effects of each intervention group versus placebo separately, but we will divide up the total number of participants in the placebo group. In the case of continuous outcomes, the total number of participants in the placebo group again will be divided up, but means and standard deviations will be left unchanged (see Chapter 16, Section 16.5.4, in Higgins 2011).

Methods intended for future reviews

If future updates identify a number of studies that enable reporting at different time points, this should be done for example at end of treatment, at short-term follow-up (zero to six months), at medium-term follow-up (seven to 12 months) and at long-term follow-up (over 12 months).

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

NOTES

A protocol for an accompanying individual patient data review on chronic fatigue syndrome and exercise therapy has been published (Larun 2014).

INDEX TERMS

Medical Subject Headings (MeSH)

Cognitive Therapy; Depression [therapy]; Exercise Therapy [adverse effects; *methods]; Fatigue Syndrome, Chronic [psychology; *therapy]; Health Status; Quality of Life; Randomized Controlled Trials as Topic; Sleep Wake Disorders [therapy]

MeSH check words

Adult; Humans

Copyright © 2017 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.