



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 2019, Issue 10. Art. No.: CD003200.
DOI: [10.1002/14651858.CD003200.pub8](https://doi.org/10.1002/14651858.CD003200.pub8).

www.cochranelibrary.com

TABLE OF CONTENTS

ABSTRACT	1
PLAIN LANGUAGE SUMMARY	2
SUMMARY OF FINDINGS	4
BACKGROUND	17
OBJECTIVES	17
METHODS	17
RESULTS	20
Figure 1.	21
Figure 2.	24
Figure 3.	25
DISCUSSION	34
AUTHORS' CONCLUSIONS	36
ACKNOWLEDGEMENTS	37
REFERENCES	38
CHARACTERISTICS OF STUDIES	43
DATA AND ANALYSES	58
Analysis 1.1. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 1: Fatigue (end of treatment)	62
Analysis 1.2. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 2: Fatigue (follow-up)	62
Analysis 1.3. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 3: Participants with serious adverse reactions	63
Analysis 1.4. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 4: Pain (follow-up) ..	63
Analysis 1.5. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 5: Physical functioning (end of treatment)	63
Analysis 1.6. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 6: Physical functioning (follow-up)	64
Analysis 1.7. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 7: Quality of life (follow-up)	64
Analysis 1.8. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 8: Depression (end of treatment)	64
Analysis 1.9. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 9: Depression (follow-up)	65
Analysis 1.10. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 10: Anxiety (end of treatment)	65
Analysis 1.11. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 11: Anxiety (follow-up)	65
Analysis 1.12. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 12: Sleep (end of treatment)	66
Analysis 1.13. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 13: Sleep (follow-up)	66
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)	66
Analysis 1.15. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 15: Self-perceived changes in overall health (follow-up)	67
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)	67
Analysis 1.17. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 17: Health resource use (follow-up) (No. of users)	68
Analysis 1.18. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 18: Dropout	68
Analysis 1.19. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 19: Sensitivity analysis for fatigue (end of treatment)	69
Analysis 1.20. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 20: Sensitivity analysis for fatigue (follow-up)	69

Analysis 1.21. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 21: Subgroup analysis for fatigue (end of treatment)	70
Analysis 1.22. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 22: Subgroup analysis for fatigue (follow-up)	70
Analysis 1.23. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 23: Sensitivity analysis for depression (follow-up)	71
Analysis 1.24. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 24: Sensitivity analysis for anxiety (follow-up)	71
Analysis 2.1. Comparison 2: Exercise therapy versus psychological treatment, Outcome 1: Fatigue at end of treatment (FS; 11 items/0 to 33 points)	75
Analysis 2.2. Comparison 2: Exercise therapy versus psychological treatment, Outcome 2: Fatigue at follow-up (SMD)	75
Analysis 2.3. Comparison 2: Exercise therapy versus psychological treatment, Outcome 3: Participants with serious adverse reactions	75
Analysis 2.4. Comparison 2: Exercise therapy versus psychological treatment, Outcome 4: Pain at follow-up (Brief Pain Inventory, pain severity subscale; 0 to 10 points)	76
Analysis 2.5. Comparison 2: Exercise therapy versus psychological treatment, Outcome 5: Pain at follow-up (Brief Pain Inventory, pain interference subscale; 0 to 10 points)	76
Analysis 2.6. Comparison 2: Exercise therapy versus psychological treatment, Outcome 6: Physical functioning at end of treatment (SF-36, physical functioning subscale; 0 to 100 points)	76
Analysis 2.7. Comparison 2: Exercise therapy versus psychological treatment, Outcome 7: Physical functioning at follow-up (SF-36, physical functioning subscale; 0 to 100 points)	77
Analysis 2.8. Comparison 2: Exercise therapy versus psychological treatment, Outcome 8: Quality of life (follow-up)	77
Analysis 2.9. Comparison 2: Exercise therapy versus psychological treatment, Outcome 9: Depression at end of treatment (HADS depression score; 7 items/21 points)	77
Analysis 2.10. Comparison 2: Exercise therapy versus psychological treatment, Outcome 10: Depression at follow-up (SMD) ...	78
Analysis 2.11. Comparison 2: Exercise therapy versus psychological treatment, Outcome 11: Anxiety at end of treatment (HADS anxiety; 7 items/21 points)	78
Analysis 2.12. Comparison 2: Exercise therapy versus psychological treatment, Outcome 12: Anxiety at follow-up (SMD)	78
Analysis 2.13. Comparison 2: Exercise therapy versus psychological treatment, Outcome 13: Sleep at end of treatment (Jenkins Sleep Scale; 0 to 20 points)	78
Analysis 2.14. Comparison 2: Exercise therapy versus psychological treatment, Outcome 14: Sleep at follow-up (Jenkins Sleep Scale; 0 to 20 points)	79
Analysis 2.15. Comparison 2: Exercise therapy versus psychological treatment, Outcome 15: Self-perceived changes in overall health at end of treatment	79
Analysis 2.16. Comparison 2: Exercise therapy versus psychological treatment, Outcome 16: Self-perceived changes in overall health at follow-up	79
Analysis 2.17. Comparison 2: Exercise therapy versus psychological treatment, Outcome 17: Health resource use (follow-up) (Mean no. of contacts)	80
Analysis 2.18. Comparison 2: Exercise therapy versus psychological treatment, Outcome 18: Health resource use (follow-up) (No. of users)	81
Analysis 2.19. Comparison 2: Exercise therapy versus psychological treatment, Outcome 19: Dropout	81
Analysis 3.1. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 1: Fatigue	84
Analysis 3.2. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 2: Participants with serious adverse reactions .	84
Analysis 3.3. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 3: Physical functioning	84
Analysis 3.4. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 4: Depression	84
Analysis 3.5. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 5: Anxiety	85
Analysis 3.6. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 6: Sleep	85
Analysis 3.7. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 7: Self-perceived changes in overall health	85
Analysis 3.8. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 8: Health resource use (follow-up) (Mean no. of contacts)	86
Analysis 3.9. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 9: Health resource use (follow-up) (No. of users)	87
Analysis 3.10. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 10: Dropout	87
Analysis 4.1. Comparison 4: Exercise therapy versus antidepressant, Outcome 1: Fatigue	88
Analysis 4.2. Comparison 4: Exercise therapy versus antidepressant, Outcome 2: Depression	88
Analysis 4.3. Comparison 4: Exercise therapy versus antidepressant, Outcome 3: Dropout	88

Analysis 5.1. Comparison 5: Exercise therapy + antidepressant versus antidepressant, Outcome 1: Fatigue	89
Analysis 5.2. Comparison 5: Exercise therapy + antidepressant versus antidepressant, Outcome 2: Depression	89
Analysis 5.3. Comparison 5: Exercise therapy + antidepressant versus antidepressant, Outcome 3: Dropout	89
ADDITIONAL TABLES	90
APPENDICES	94
FEEDBACK	99
WHAT'S NEW	123
HISTORY	123
CONTRIBUTIONS OF AUTHORS	126
DECLARATIONS OF INTEREST	126
SOURCES OF SUPPORT	126
DIFFERENCES BETWEEN PROTOCOL AND REVIEW	126
NOTES	128
INDEX TERMS	129

[Intervention Review]

Exercise therapy for chronic fatigue syndrome

Lillebeth Larun¹, Kjetil G Brurberg¹, Jan Odgaard-Jensen², Jonathan R Price³

¹Division for Health Services, Norwegian Institute of Public Health, Oslo, Norway. ²Medicinrådet, København Ø, Denmark. ³Department of Psychiatry, University of Oxford, Oxford, UK

Contact: Lillebeth Larun, lillebeth.larun@fhi.no, lblarun@gmail.com.

Editorial group: Cochrane Common Mental Disorders Group.

Publication status and date: Edited (no change to conclusions), published in Issue 3, 2021.

Citation: Larun L, Brurberg KG, Odgaard-Jensen J, Price JR. Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews* 2019, Issue 10. Art. No.: CD003200. DOI: [10.1002/14651858.CD003200.pub8](https://doi.org/10.1002/14651858.CD003200.pub8).

Copyright © 2019 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Editorial note: A statement from the Editor in Chief about this review and its planned update is available at <https://www.cochrane.org/news/cfs>

ABSTRACT

Background

Chronic fatigue syndrome (CFS) or myalgic encephalomyelitis (ME) is a serious disorder characterised by persistent postexertional fatigue and substantial symptoms related to cognitive, immune and autonomous dysfunction. There is no specific diagnostic test, therefore diagnostic criteria are used to diagnose CFS. The prevalence of CFS varies by type of diagnostic criteria used. Existing treatment strategies primarily aim to relieve symptoms and improve function. One treatment option is exercise therapy.

Objectives

The objective of this review was to determine the effects of exercise therapy for adults with CFS compared with any other intervention or control on fatigue, adverse outcomes, pain, physical functioning, quality of life, mood disorders, sleep, self-perceived changes in overall health, health service resources use and dropout.

Search methods

We searched the Cochrane Common Mental Disorders Group controlled trials register, CENTRAL, and SPORTDiscus up to May 2014, using a comprehensive list of free-text terms for CFS and exercise. We located unpublished and ongoing studies through the World Health Organization International Clinical Trials Registry Platform up to May 2014. We screened reference lists of retrieved articles and contacted experts in the field for additional studies.

Selection criteria

We included randomised controlled trials (RCTs) about adults with a primary diagnosis of CFS, from all diagnostic criteria, who were able to participate in exercise therapy.

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) or standardised mean differences (SMDs). To facilitate interpretation of SMDs, we re-expressed SMD estimates as MDs on more common measurement scales. We combined dichotomous outcomes using risk ratios (RRs). We assessed the certainty of evidence using GRADE.

Main results

We included eight RCTs with data from 1518 participants.

Exercise therapy for chronic fatigue syndrome (Review)

Copyright © 2019 The Cochrane Collaboration. Published by John Wiley & Sons, Ltd.

Exercise therapy lasted from 12 weeks to 26 weeks. The studies measured effect at the end of the treatment and at long-term follow-up, after 50 weeks or 72 weeks.

Seven studies used aerobic exercise therapies 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, and one study used anaerobic exercise. Control groups consisted of passive control, including treatment as usual, relaxation or flexibility (eight studies); cognitive behavioural therapy (CBT) (two studies); cognitive therapy (one study); supportive listening (one study); pacing (one study); pharmacological treatment (one study) and combination treatment (one study).

Most studies had a low risk of selection bias. All had a high risk of performance and detection bias.

Exercise therapy compared with 'passive' control

Exercise therapy probably reduces fatigue at end of treatment (SMD -0.66, 95% CI -1.01 to -0.31; 7 studies, 840 participants; moderate-certainty evidence; re-expressed MD -3.4, 95% CI -5.3 to -1.6; scale 0 to 33). We are uncertain if fatigue is reduced in the long term because the certainty of the evidence is very low (SMD -0.62, 95% CI -1.32 to 0.07; 4 studies, 670 participants; re-expressed MD -3.2, 95% CI -6.9 to 0.4; scale 0 to 33).

We are uncertain about the risk of serious adverse reactions because the certainty of the evidence is very low (RR 0.99, 95% CI 0.14 to 6.97; 1 study, 319 participants).

Exercise therapy may moderately improve physical functioning at end of treatment, but the long-term effect is uncertain because the certainty of the evidence is very low. Exercise therapy may also slightly improve sleep at end of treatment and at long term. The effect of exercise therapy on pain, quality of life and depression is uncertain because evidence is missing or of very low certainty.

Exercise therapy compared with CBT

Exercise therapy may make little or no difference to fatigue at end of treatment (MD 0.20, 95% CI -1.49 to 1.89; 1 study, 298 participants; low-certainty evidence), or at long-term follow-up (SMD 0.07, 95% CI -0.13 to 0.28; 2 studies, 351 participants; moderate-certainty evidence).

We are uncertain about the risk of serious adverse reactions because the certainty of the evidence is very low (RR 0.67, 95% CI 0.11 to 3.96; 1 study, 321 participants).

The available evidence suggests that there may be little or no difference between exercise therapy and CBT in physical functioning or sleep (low-certainty evidence) and probably little or no difference in the effect on depression (moderate-certainty evidence). We are uncertain if exercise therapy compared to CBT improves quality of life or reduces pain because the evidence is of very low certainty.

Exercise therapy compared with adaptive pacing

Exercise therapy may slightly reduce fatigue at end of treatment (MD -2.00, 95% CI -3.57 to -0.43; scale 0 to 33; 1 study, 305 participants; low-certainty evidence) and at long-term follow-up (MD -2.50, 95% CI -4.16 to -0.84; scale 0 to 33; 1 study, 307 participants; low-certainty evidence).

We are uncertain about the risk of serious adverse reactions (RR 0.99, 95% CI 0.14 to 6.97; 1 study, 319 participants; very low-certainty evidence).

The available evidence suggests that exercise therapy may slightly improve physical functioning, depression and sleep compared to adaptive pacing (low-certainty evidence). No studies reported quality of life or pain.

Exercise therapy compared with antidepressants

We are uncertain if exercise therapy, alone or in combination with antidepressants, reduces fatigue and depression more than antidepressant alone, as the certainty of the evidence is very low. The one included study did not report on adverse reactions, pain, physical functioning, quality of life, sleep or long-term results.

Authors' conclusions

Exercise therapy probably has a positive effect on fatigue in adults with CFS compared to usual care or passive therapies. The evidence regarding adverse effects is uncertain. Due to limited evidence it is difficult to draw conclusions about the comparative effectiveness of CBT, adaptive pacing or other interventions. All studies were conducted with outpatients diagnosed with 1994 criteria of the Centers for Disease Control and Prevention or the Oxford criteria, or both. Patients diagnosed using other criteria may experience different effects.

PLAIN LANGUAGE SUMMARY

Exercise as treatment for adults with chronic fatigue syndrome

What is the aim of this review?

People with chronic fatigue syndrome have long-lasting fatigue, joint pain, headaches, sleep problems, poor concentration and short-term memory. These symptoms cause significant disability and distress. We wanted to find out whether exercise therapy can help people with chronic fatigue syndrome (myalgic encephalomyelitis).

Key messages

People who have exercise therapy probably have less fatigue at the end of treatment than those who receive more passive therapies. We are uncertain if this improvement lasts in the long term. We are also uncertain about the risk of serious side effects from exercise therapy.

What was studied in the review?

We explored whether exercise therapy can reduce chronic fatigue syndrome symptoms. We searched for studies comparing the effect of exercise therapy with treatment as usual or other therapies.

What are the main results of the review?

We found eight studies with 1518 participants. The studies compared participants who received exercise therapy to participants who received treatment as usual or more active treatments such as cognitive behavioural therapy.

Participants had exercise therapy for 12 weeks to 26 weeks. The studies measured the effect of the therapy at the end of the treatment and also long term, after 50 or 72 weeks. Participants exercised at different levels of intensity using variations of aerobic exercising such as walking, swimming or cycling.

Exercise therapy compared to treatment as usual or relaxation

Participants who have exercise therapy probably have less fatigue at the end of treatment, and they may have moderately better physical functioning. We are uncertain if these improvements last long term because we are very uncertain about the evidence.

Participants who have exercise therapy may have slightly better sleep, both at the end of treatment and long term.

We are uncertain about the risk of serious side effects and the effects of exercise therapy on pain, quality of life, and depression. This is because we lack evidence or because we are very uncertain about the evidence.

Exercise therapy compared to cognitive behavioural therapy

Exercise therapy may make little or no difference to participants' fatigue at end of treatment or in the long term. Exercise therapy may make little or no difference to participants' physical functioning at end of treatment, but the long-term effect on physical functioning is uncertain.

No studies looked at the effect of exercise therapy on depression at the end of treatment, but it probably has little or no long-term effect.

We are uncertain about the risk of side effects. We are also uncertain about the effects on pain, quality of life, or sleep. This is because we lack evidence or because we are very uncertain about the evidence.

Exercise therapy compared to adaptive pacing (living within limits)

Participants who have exercise therapy may have slightly less fatigue and depressive symptoms and slightly better physical functioning and sleep at the end of treatment and long term than participants who have adaptive pacing.

We are uncertain about the risk of serious side effects. We are also uncertain about the effect on quality of life or pain. This is because we lack evidence or we are very uncertain about the evidence.

Exercise therapy compared to antidepressants

We are uncertain if exercise therapy is better than antidepressants at reducing fatigue. We are also uncertain of its effect on depression, side effects, pain, physical functioning, quality of life or sleep. This is because we lack evidence or we are very uncertain about the evidence.

Why is this review important?

Exercise therapy is recommended by treatment guidelines and often used as treatment for people with chronic fatigue syndrome. People with chronic fatigue syndrome should have the opportunity to make informed decisions about their care and treatment based on robust research evidence and whether exercise therapy is effective, either as a stand-alone intervention or as part of a treatment plan.

It is important to note that the evidence in this review is from people diagnosed with 1994 criteria of the Centers for Disease Control and Prevention or the Oxford criteria. People diagnosed using other criteria may experience different effects.

SUMMARY OF FINDINGS

Summary of findings 1. Exercise therapy versus control for chronic fatigue syndrome

Exercise therapy versus control for chronic fatigue syndrome

Patient or population: men and women aged over 18 years with chronic fatigue syndrome

Intervention: exercise therapy

Comparison: usual care, waiting list or relaxation/flexibility

Setting: outpatient/primary care

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Control	Exercise				
<p>Fatigue Measured at end of treatment (12-26 weeks)</p> <p>Measured with 3 different versions of the Chalder Fatigue Scale (0-11; 0-33, or 0-42 points).</p> <p>Low score means less fatigue</p>	See comment	<p>SMD 0.66 lower (1.01 lower to 0.31 lower)</p>		840 (7 studies)	⊕⊕⊕⊖ Moderate a,b	<p>Exercise therapy probably reduces fatigue after 12-26 weeks</p> <p>Estimate expressed in standardised units (SMD)^c, and corresponds to a 3.4-point reduction when re-expressed on the Chalder Fatigue Scale (0-33 points, 0 indicates no fatigue)</p> <p>SMD is reduced to -0.44 if Powell 2001 is excluded from the analysis.</p>
<p>Fatigue Measured after 52-70 weeks</p> <p>Measured with different versions of the Chalder Fatigue Scale (0-11, or 0-33 points) or the Fatigue Severity Scale (1-7 points).</p> <p>Low score means less fatigue</p>	See comment	<p>SMD 0.62 lower (1.32 lower to 0.07 higher)</p>		670 (4 studies)	⊕⊕⊕⊖ Very low a,d,e	<p>The effect of exercise therapy on fatigue after 52-70 weeks is uncertain</p> <p>Estimate expressed in standardised units (SMD)^c, and corresponds to a 3.2-point reduction when re-expressed on the Chalder Fatigue Scale (0-33 points, 0 indicates no fatigue)</p>

					SMD is reduced to -0.27 if Powell 2001 is excluded from the analysis.	
Participants with serious adverse reactions	Study population		RR 0.99 (0.14 to 6.97)	319 (1 study)	⊕⊕⊕⊕ Very low f,g,h	The impact of exercise therapy on serious adverse reactions is uncertain
Measured after 52 weeks	13 per 1000	12 per 1000 (2 to 87)				
Measured according to European Union Clinical Trials Directive by recording the number of serious reactions						
Pain	-	-	-	-	-	None of the studies looked at pain at end of treatment
Measured at end of treatment						
Pain intensity	Mean pain score in the control group was 3.63 points	Mean pain score in the exercise group was 0.97 points lower (2.44 lower to 0.50 higher)		43 (1 study)	⊕⊕⊕⊕ Very low i,j	The effect of exercise therapy on pain after 52 weeks is uncertain
Measured after 52 weeks						
Measured with the Brief Pain Inventory subscale, 0-10						
Lower score means less pain						
Physical functioning	Mean physical functioning score in the control group ranged from 31-55 points	Mean physical functioning score in the exercise group was 13.10 points higher (1.98 higher to 24.22 higher)		725 (5 studies)	⊕⊕⊕⊕ Low a,k	Exercise therapy may moderately improve physical functioning after 12-24 weeks
Measured at end of treatment; 12-24 weeks						
Measured with SF-36 physical functioning subscale, 0-100 points						
Higher score means better function						
Physical functioning	Mean physical functioning score in the control group ranged from 35-51 points	Mean physical functioning score in the exercise group was 16.33 points higher (36.74 higher to 4.08 lower)		621 (3 studies)	⊕⊕⊕⊕ Very low a,l	The effect of exercise therapy on physical functioning after 52-70 weeks is uncertain
Measured after 52-70 weeks						
Measured with SF-36 physical functioning subscale, 0-100 points						
Higher score means better function						
Quality of Life (QoL)	-	-	-	-	-	None of the studies looked at QoL at end of treatment

Measured at end of treatment					
Quality of Life (QoL)	Mean QoL score in the control group was 72 points	Mean QoL score in the exercise group was 9.00 points lower (19.00 lower to 1.00 higher)	44 (1 study)	⊕⊕⊕⊕ Very low a,m	The effect of exercise therapy on QoL is uncertain
Measured after 52 weeks					
Measured with the Quality of Life Scale, 16-112 points.					
High score means better QoL					
Depression	Mean depression score in control group ranged from 5.2 to 11.2 points	Mean depression score in the exercise group was 1.63 points lower (3.50 lower to 0.23 higher)	504 (5 studies)	⊕⊕⊕⊕ Very low a,n,o	The effect of exercise therapy on depression after 12-26 weeks is uncertain
Measured at end of treatment; 12-26 weeks					
Measured with the HADS depression score, 0-21 points. Low score means fewer symptoms					
Depression	See comment	SMD 0.35 lower (0.93 lower to 0.23 higher)	654 (4 studies)	⊕⊕⊕⊕ Very low a,n,o	The effect of exercise therapy on depression after 52 weeks is uncertain
Measured after 52-70 weeks					
Measured with HADS depression score, 0-21 points, and Beck Depression Inventory-II, 0-63 points.					Estimate expressed in standardised (SMD) units ^c , and corresponds to a 1.4-point reduction when re-expressed on the HADS depression scale (0-21 points)
Low score means fewer symptoms					
Sleep	Mean sleep score in control group ranged from 11.7-12.2 points	Mean sleep score in the exercise group was 1.49 points lower (2.95 lower to 0.02 lower)	323 (2 studies)	⊕⊕⊕⊕ Low a,n	Exercise therapy may slightly improve sleep quality after 12-26 weeks
Measured at end of treatment, 12-26 weeks					
Measured with Jenkins Sleep Scale, 0-20 points					
Low score means better sleep					
Sleep	Mean sleep score in control group ranged from 11.0-12.6 points	Mean sleep score in the exercise group was 2.04 points lower	610 (3 studies)	⊕⊕⊕⊕ Low a,n	Exercise therapy may slightly improve sleep quality after 52-70 weeks
Measured after 52-70 weeks					
Measured with Jenkins Sleep Scale, 0-20 points.					
Low score means better sleep					

(3.84 lower to 0.23 lower)

*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; **HADS:** Hospital Anxiety and Depression Scale; **QoL:** quality of life; **RR:** risk ratio; **SF-36:** Short Form 36; **SMD:** standardised mean difference

GRADE Working Group grades of evidence

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

^aRisk of bias (certainty downgraded by -1): all studies were at risk of performance bias, as they were unblinded.

^bInconsistency (certainty not downgraded): we chose not to downgrade because all studies gave the same direction and because the observed heterogeneity (80%) was mainly caused by a single outlier. The estimate remains consistent with a non-zero effect size (SMD -0.44; 95% CI -0.63 to -0.24) also when the outlier is excluded.

^cInterpretation of standardised mean difference: less than 0.41 = small; between 0.40 and 0.70 = moderate, and over 0.70 = large effect size.

^dImprecision (certainty downgraded by -1): variation in effect size across studies and confidence intervals ranging from a large positive effect to little or no difference.

^eInconsistency (certainty downgraded by -1): large heterogeneity, and a standardised mean difference that changes from -0.62 (moderate effect size) to -0.27 (small effect size) when Powell 2001 is excluded.

^fRisk of bias (certainty not downgraded 0): this outcome is unlikely to have been affected by detection or performance bias.

^gImprecision (certainty downgraded by -2): low numbers of events and wide confidence intervals.

^hThis available trial was not sufficiently powered to detect differences this outcome.

ⁱRisk of bias (certainty downgraded by -2): unblinded study with large baseline differences between groups.

^jImprecision (certainty downgraded by -1): single study with limited number of participants and confidence interval ranging from a positive effect to little or no difference.

^kImprecision/inconsistency (certainty downgraded by -1): the confidence interval ranges from a large positive to a small benefit. There is variation in the effect size across available studies, but the heterogeneity is in part caused by a single outlier. When excluding the outlier, the pooled estimate is reduced to (mean difference -7.27, 95% CI -13.51 to -1.23).

^lImprecision/inconsistency (certainty downgraded by -2): the confidence interval ranges from a large positive to a small negative effect. There is variation in the effect size across available studies, but the heterogeneity is caused by a single outlier, but when excluding the outlier, the pooled estimate is reduced from a moderate to a slight benefit (mean difference -5.79, 95% CI -10.53 to -1.06).

^mImprecision (certainty downgraded by -2): very low number of participants and wide confidence intervals encompassing potential harmful effects as well as little or no difference.

ⁿImprecision (certainty downgraded by -1): wide confidence interval encompassing benefits and little or no difference.

^oInconsistency (certainty downgraded by -1): there is large variation in the magnitude and the direction of the effect estimate across available studies.

Summary of findings 2. Exercise therapy versus psychological treatment for chronic fatigue syndrome

Exercise therapy versus psychological treatment for chronic fatigue syndrome

Patient or population: men and women aged over 18 years with chronic fatigue syndrome

Intervention: exercise therapy

Comparison: cognitive-behaviour therapy (CBT)

Setting: outpatient/primary care

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	CBT	Exercise				
<p>Fatigue</p> <p>Measured at end of treatment, 24 weeks</p> <p>Measured with Chalder Fatigue Scale, 0-33 points</p> <p>Low score means less fatigue</p>	<p>Mean fatigue score in the CBT group was 21.5 points</p>	<p>Mean fatigue score in the exercise group was 0.20 higher (1.49 lower to 1.89 higher)</p>		<p>298 (1 study)</p>	<p>⊕⊕⊕⊖ Low ^{a,b}</p>	<p>Exercise therapy may make little or no difference to fatigue after 24 weeks</p>
<p>Fatigue</p> <p>Measured after 52 weeks</p> <p>Measured with Chalder Fatigue Scales (0-33 points) or Fatigue Severity Scale (1-7 points)</p> <p>Low score means less fatigue</p>	<p>See comment</p>	<p>SMD 0.07 higher (0.13 lower to 0.28 higher)</p>		<p>351 (2 studies)</p>	<p>⊕⊕⊕⊖ Moderate ^a</p>	<p>Exercise therapy probably makes little or no difference to fatigue after 52 weeks</p> <p>Estimate expressed in standardised units (SMD)^c. SMD of 0.07 corresponds to MD of 0.5 points when re-expressed on the Chalder Fatigue Scale (0-33 points)</p>
<p>Participants with serious adverse reactions</p> <p>Measured after 52 weeks</p> <p>Measured according to European Union Clinical Trials Directive by recording the number of serious reactions</p>	<p>Study population</p> <p>19 per 1000 13 per 1000 (2 to 75)</p>			<p>RR 0.67 (0.11 to 3.96)</p> <p>321 (1 study)</p>	<p>⊕⊕⊖⊖ Very low ^{d,e,f}</p>	<p>The impact of exercise therapy on serious adverse reactions is uncertain</p>
<p>Pain intensity</p> <p>End of treatment</p>	-	-	-	-	-	<p>No studies looked at pain at end of treatment</p>

<p>Pain intensity</p> <p>Measured after 52 weeks</p> <p>Measured with the Brief Pain Inventory subscale, 0-10 Low score means less pain</p>	<p>Mean pain score in the CBT group was 3.56 points</p>	<p>Mean pain score in the exercise group was 0.07 points higher (1.52 lower to 1.66 higher)</p>	<p>43</p> <p>(1 study)</p>	<p>⊕○○○</p> <p>Very low g,h</p>	<p>The effect of exercise therapy on pain intensity after 52 weeks is uncertain</p>
<p>Physical functioning</p> <p>Measured at end of treatment, 24 weeks</p> <p>Measured with SF-36 physical functioning subscale, 0-100 points</p> <p>High score means better physical functioning</p>	<p>Mean physical functioning score in the CBT group was 54.2 points</p>	<p>Mean physical functioning score in the exercise group was 1.20 points higher (3.90 lower to 6.30 higher)</p>	<p>298</p> <p>(1 study)</p>	<p>⊕⊕○○</p> <p>Low a,b</p>	<p>Exercise therapy may make little or no difference to physical functioning after 24 weeks</p>
<p>Physical functioning</p> <p>Measured after 52 weeks</p> <p>Measured with SF-36 physical functioning subscale, 0-100 points</p> <p>High score means better physical functioning</p>	<p>Mean physical functioning score in the CBT group was 58.2 points</p>	<p>Mean physical functioning score in the exercise group was 7.92 points higher (9.79 lower to 25.63 higher)</p>	<p>348</p> <p>(2 studies)</p>	<p>⊕○○○</p> <p>Very low a,i</p>	<p>The effect of exercise therapy on physical functioning after 52 weeks is uncertain</p>
<p>Quality of Life (QoL)</p>	-	-	-	-	<p>No studies looked at QoL at end of treatment</p>
<p>Quality of Life (QoL)</p> <p>Measured after 52 weeks</p> <p>Measured on the Quality of Life Scale, 16-112 points High score means better quality of life</p>	<p>Mean QOL score in the CBT group was 69 points</p>	<p>Mean QOL score in the exercise group was 6.10 points lower (15.87 lower to 3.67 higher)</p>	<p>44</p> <p>(1 study)</p>	<p>⊕○○○</p> <p>Very low b,g</p>	<p>The effect of exercise therapy on quality of life after 52 weeks is uncertain</p>
<p>Depression</p>	-	-	-	-	<p>No studies looked at depression at end of treatment</p>
<p>Depression</p> <p>Measured after 52 weeks</p>	<p>See comment</p>	<p>SMD 0.01 higher (0.21 lower to 0.22 higher)</p>	<p>331</p> <p>(2 studies)</p>	<p>⊕⊕⊕○</p> <p>Moderate a</p>	<p>Exercise therapy probably makes little or no difference to depression after 52 weeks</p>

HADS depression score (0-21 points) or Beck Depression Inventory-II (0-63 points) Low score means fewer symptoms					Estimate expressed in standardised units (SMD). SMD of 0.01 corresponds to MD of 0.4 points when re-expressed on HADS Depression (0-21 points)
Sleep	-	-	-	-	No studies looked at this outcome at end of treatment
Sleep Measured after 52 weeks Jenkins Sleep Scale, 0-20 points Low score means better sleep	Mean sleep score in CBT group was 9.9 points.	Mean sleep score in the exercise group was 0.9 points lower (2.07 lower to 0.27 higher)	287 (1 study)	⊕⊕⊕⊕ Low ^{a,b}	Exercise therapy may make little or no difference to sleep after 52 weeks

*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; **HADS:** Hospital Anxiety and Depression Scale; **QoL:** quality of life; **RR:** risk ratio; **SF-36:** Short Form 36; **SMD:** standardised mean difference

GRADE Working Group grades of evidence

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

^aRisk of bias (certainty downgraded by -1): all studies were at risk of performance bias, as they were unblinded.

^bImprecision (certainty downgraded by -1): single study and/or limited number of participants.

^cRe-expressed standardised mean difference: less than 0.41 = small; between 0.40 and 0.70 = moderate and over 0.70 = large effect size.

^dRisk of bias (certainty not downgraded): this outcome is unlikely to have been affected by detection or performance bias.

^eImprecision (certainty downgraded by -2): low numbers of events and wide confidence intervals.

^fThe only available trial was not powered to detect differences this outcome.

^gRisk of bias (certainty downgraded by -2): unblinded study with large baseline differences between groups.

^hImprecision (certainty downgraded by -2): single study with very few participants and confidence interval ranging from a positive effect to little or no difference.

ⁱImprecision/inconsistency (certainty downgraded by -2): heterogeneity between the two available studies causes a confidence interval that ranges from a benefit of exercise to a large benefit in favour of cognitive behavioural therapy.

Summary of findings 3. Exercise therapy versus adaptive pacing therapy for chronic fatigue syndrome

Exercise therapy versus adaptive pacing therapy for chronic fatigue syndrome

Patient or population: men and women aged over 18 years with chronic fatigue syndrome

Intervention: exercise therapy
Comparison: adaptive pacing
Setting: outpatient/primary care

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Adaptive pacing	Exercise				
<p>Fatigue</p> <p>Measured at end of treatment, 24 weeks</p> <p>Measured with Chalder Fatigue Scale, 0-33 points</p> <p>Low score means less fatigue</p>	Mean fatigue score was 23.7 points	Mean fatigue score in the exercise group was 2.00 lower (3.57 lower to 0.43 lower)		305 (1 study)	⊕⊕⊕⊕ Low a,b	Exercise therapy may slightly reduce fatigue after 12-26 weeks
<p>Fatigue</p> <p>Measured at end of treatment, 52 weeks</p> <p>Measured with Chalder Fatigue Scale, 0-33 points</p> <p>Low score means less fatigue</p>	Mean fatigue score was 23.1 points	Mean fatigue score in the exercise group was 2.50 lower (4.16 lower to 0.84 lower)		307 (1 study)	⊕⊕⊕⊕ Low a,b	Exercise therapy may slightly reduce fatigue after 52 weeks
<p>Participants with serious adverse reactions</p> <p>Measured after 52 weeks</p> <p>Measured according to European Union Clinical Trials Directive by recording the number of serious reactions</p>	<p>Study population</p> <p>13 per 1000 12 per 1000 (2 to 87)</p>		RR 0.99 (0.14 to 6.97)	319 (1 study)	⊕⊕⊕⊕ Very low c,d,e	The impact of exercise therapy on serious adverse reactions is uncertain
<p>Pain</p> <p>End of treatment and long term</p>	-					No studies looked at pain
<p>Physical functioning</p> <p>Measured at end of treatment, 24 weeks</p>	Mean physical functioning score was 43.2 points	Mean physical functioning score in the exercise group was 12.20		305 (1 study)	⊕⊕⊕⊕ Low a,b	Exercise therapy may slightly improve physical functioning after 24 weeks

Measured with SF-36 physical functioning subscale, 0-100 points High score means better physical functioning		points higher (7.17 higher to 17.23 higher)			
Physical functioning Measured at end of treatment, 52 weeks Measured with SF-36 physical functioning subscale, 0-100 points High score means better physical functioning	Mean physical functioning score was 45.9 points	Mean physical functioning score in the exercise group was 11.80 points higher (6.05 higher to 17.55 higher)	307 (1 study)	⊕⊕⊕⊕ Low a,b	Exercise therapy may slightly improve physical functioning after 52 weeks
Quality of Life (QOL) End of treatment and long term	-				No studies looked at quality of life
Depression Measured at end of treatment	-				No studies looked at depression at end of treatment
Depression Measured after 52 weeks HADS depression score, 0-21 points Low score means fewer symptoms	Mean depression score was 7.2 points	Mean depression score in the exercise group was 1.10 points lower (2.09 lower to 0.11 lower)	293 (1 study)	⊕⊕⊕⊕ Low a,b	Exercise therapy may slightly reduce depression after 52 weeks
Sleep Measured at end of treatment	-				No studies looked at sleep at end of treatment
Sleep Measured after 52 weeks Jenkins Sleep Scale, 0-20 points Low score means better sleep	Mean sleep score was 10.6 points	Mean sleep score in the exercise group was 1.60 points lower (2.70 lower to 0.50 lower)	294 (1 study)	⊕⊕⊕⊕ Low a,b	Exercise therapy may slightly improve sleep after 52 weeks

*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; **HADS:** Hospital Anxiety and Depression Scale; **QoL:** quality of life; **RR:** risk ratio; **SF-36:** Short Form 36; **SMD:** standardised mean difference

GRADE Working Group grades of evidence.

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

^aRisk of bias (-1): all studies were at risk of performance bias, as they were unblinded.

^bImprecision (-1): single study, low numbers of events or wide confidence intervals.

^cRisk of bias (0): this outcome is unlikely to have been affected by detection or performance bias.

^dImprecision (-2): single study and very wide confidence intervals.

^eThe only available trial was not powered to detect differences this outcome.

Summary of findings 4. Exercise therapy versus antidepressants for chronic fatigue syndrome

Exercise therapy versus antidepressants for chronic fatigue syndrome

Patient or population: men and women aged over 18 years with chronic fatigue syndrome

Intervention: exercise therapy

Comparison: antidepressant (fluoxetine)

Setting: outpatient

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Antidepressant	Exercise				
Fatigue Measured at end of treatment, 26 weeks Measured with Chalder Fatigue Scale, 0-42 Low score means less fatigue	Mean fatigue score was 30.2 points	Mean fatigue score in the exercise group was 1.99 lower (8.28 lower to 4.30 higher)		48 (1 study)	⊕⊕⊕⊕ Very low ^{a,b}	The effect of exercise therapy is uncertain
Fatigue Long term	No available data for this outcome					No studies looked at fatigue at long term



Serious adverse reactions End of treatment and long term	No available data for this outcome				No studies looked at serious adverse reactions
Pain End of treatment and long term	No available data for this outcome				No studies looked at pain
Physical functioning End of treatment and long term	No available data for this outcome				No studies looked at physical functioning
Quality of Life (QOL) End of treatment and long term	No available data for this outcome				No studies looked at quality of life
Depression Measured at end of treatment, 26 weeks Measured with HADS depression score, 0-21 points Low score means fewer symptoms	Mean depression score was 7.32 points	Mean depression score in the exercise group was 0.15 points higher (2.41 higher to 2.11 lower)	48 (1 study)	⊕⊕⊕⊕ Very low ^{a,b}	The effect of exercise therapy on depression is uncertain
Depression Long term	No available data for this outcome				No studies looked at depression
Sleep End of treatment and long term	No available data for this outcome				No studies looked at sleep

*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; **HADS:** Hospital Anxiety and Depression Scale; **QoL:** quality of life; **RR:** risk ratio; **SF-36:** Short Form 36; **SMD:** standardised mean difference

GRADE Working Group grades of evidence.

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

^aRisk of bias (certainty downgraded by -2): risk of performance and attrition bias.

^bImprecision (certainty downgraded by -2): confidence interval encompass potential benefits and harms. One study with few participants.

Summary of findings 5. Exercise therapy plus antidepressants versus antidepressants alone for chronic fatigue syndrome

Exercise therapy plus antidepressants versus antidepressants alone for chronic fatigue syndrome

Patient or population: men and women aged over 18 years with chronic fatigue syndrome

Intervention: exercise therapy + antidepressant

Comparison: antidepressant alone (fluoxetine)

Setting: outpatient

Outcomes	Illustrative comparative risks* (95% CI)		Relative effect (95% CI)	Number of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Assumed risk	Corresponding risk				
	Antidepressant	Exercise + antidepressant				
<p>Fatigue Measured at end of treatment, 26 weeks Measured with Chalder Fatigue Scale, 0-42 points Low score means less fatigue</p>	Mean fatigue score in comparison group was 29.92 points	Mean fatigue score in the intervention group was 3.66 lower (10.41 lower to 3.09 higher)		43 (1 study)	⊕⊕⊕⊕ Very low ^{a,b}	The effect of exercise therapy is uncertain
<p>Fatigue Long term</p>	No available data for this outcome					No studies looked at fatigue at long term
<p>Serious adverse reactions End of treatment and long term</p>	No available data for this outcome					No studies looked at serious adverse reactions
<p>Pain End of treatment and long term</p>	No available data for this outcome					No studies looked at pain
<p>Physical functioning End of treatment and long term</p>	No available data for this outcome					No studies looked at physical functioning

Quality of Life (QoL) End of treatment and long term	No available data for this outcome				No studies looked at quality of life
Depression Measured at end of treatment, 26 weeks HADS depression score, 0-21 points Lowest score is least symptoms	Mean depression score in comparison group was 7.32 points	Mean depression score in the exercise group was 0.27 points lower (2.68 lower to 2.14 higher)	44 (1 study)	⊕⊕⊕⊕ Very low a,b	The effect of exercise therapy is uncertain
Depression Long term	No available data for this outcome				No studies looked at depression long term
Sleep End of treatment and long term	No available data for this outcome				No studies looked at sleep

*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; **HADS:** Hospital Anxiety and Depression Scale; **QoL:** quality of life; **RR:** risk ratio; **SF-36:** Short Form 36; **SMD:** standardised mean difference

GRADE Working Group grades of evidence.

High certainty: we are very confident that the true effect lies close to that of the estimate of the effect.

Moderate certainty: we are moderately confident in the effect estimate; the true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different.

Low certainty: our confidence in the effect estimate is limited; the true effect may be substantially different from the estimate of the effect.

Very low certainty: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the estimate of effect.

^aRisk of bias (certainty downgraded by -2): risk of performance and attrition bias.

^bImprecision (certainty downgraded by -2): confidence interval encompass potential benefits and harms. One study with few participants.

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 we use it in this review. Clinicians diagnose CFS only after they have excluded all alternative diagnoses (Reeves 2003; Reeves 2007), and several sets of diagnostic criteria are available (Carruthers 2011; Fukuda 1994; NICE 2007; Reeves 2003; Sharpe 1991). The Centers for Disease Control and Prevention (CDC) diagnostic criteria for CFS (Fukuda 1994), are the most widely used for research purposes (Fonhus 2011). Their application results in prevalence of CFS of between 0.24% (Reyes 2003), and 2.55% (Reeves 2007), among adults in the USA. Difference in the application of diagnostic criteria may explain some of the observed variation in prevalence (Johnston 2013). In clinical practice, most patients visit their local general practitioner (GP) for initial assessment and management. The GP will refer some patients to secondary care specialist clinics, including neurology, infectious diseases, psychiatry, endocrinology, and general medicine for exclusion of possible 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". '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).

The comparator interventions are passive treatments: treatment as usual, relaxation and/or flexibility or active therapies: psychological, adaptive pacing therapy (living within limits) or pharmacological.

How the intervention might work

Physical activity can improve health and quality of life for people with chronic disease (Blair 2009). Several hypotheses have been proposed as to why exercise therapy might be a treatment for CFS.

- The 'deconditioning model' assumes that the syndrome is perpetuated by reversible physiological changes of deconditioning and avoidance of activity, and that therefore physical activity (exercise) should reduce deconditioning and facilitate recovery (Clark 2005; White 2011). However, mediation studies suggest that improved conditioning is not necessarily associated with better outcomes (Fulcher 1997; Moss-Morris 2005).
- Some graded exercise therapy 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 graded exercise therapy, consistent with this model (Clark 2005; Moss-Morris 2005).
- Evidence has also been found for central sensitisation contributing to hyper-responsiveness of the central nervous system to a variety of visceral inputs (Nijs 2011). The most replicated finding in people 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 confirm the actual causal mechanism or mechanisms. However, effective treatments for any disorder may be discovered and confirmed without knowledge of cause.

Why it is important to do this review

The previous Cochrane Review, suggested that exercise therapy was a promising treatment but that larger studies were needed to address the safety of this therapy (Edmonds 2004). Larger studies have now been completed and their findings published. Exercise therapy is recommended by treatment guidelines (NICE 2007), and often used as treatment for individuals with CFS. 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.

Cochrane has published one more review on treatment for people with CFS; a CBT review published in 2008 (Price 2008).

OBJECTIVES

The objective of this review was to determine the effects of exercise therapy for adults with CFS compared with any other intervention or control on fatigue, adverse outcomes, pain, physical functioning, quality of life, mood disorders, sleep, self-perceived changes in overall health, health service resources use and dropout.

METHODS

Criteria for considering studies for this review

Types of studies

We included randomised controlled trials (RCTs), cluster-RCTs and randomised cross-over trials.

Types of participants

We included studies of male and female participants over the age of 18 years, irrespective of cultures and settings (e.g. primary, secondary or tertiary care). Researchers have used different sets of criteria to diagnose CFS ([Carruthers 2011](#); [Fukuda 1994](#); [NICE 2007](#); [Reeves 2003](#); [Sharpe 1991](#)), and we therefore decided to include studies 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 studies with participants with disorders other than CFS provided that more than 90% of the participants had a primary diagnosis of CFS according to the criteria specified above. We included studies in which less than 90% of participants had a primary diagnosis of CFS only when data were reported separately for participants with CFS.

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

Exercise therapy is an umbrella term for different types of exercise provided with therapeutic intent based on the American College of Sports Medicine definition ([ACSM 2001](#)). We therefore included any experimental intervention, aerobic and anaerobic, aimed at exercising large muscle groups. This included walking, swimming, jogging and strength or stabilising exercises. We included both individual and group treatment modalities. Interventions had to be clearly described and supported by appropriate references.

We categorised exercise therapies in accordance with descriptions of the interventions provided by individual studies. We prepared a table of interventions with detailed information on the specific exercise therapy used in the included studies. As a point of reference, we used the following empirical definitions.

- Graded exercise therapy: exercise in which the incremental increase in exercise was defined by discussion between participant and therapist
- Exercise with pacing: exercise in which the incremental increase in exercise was defined by the participant alone
- Anaerobic exercise: exercise requiring a high level of exertion for a short period of time, which may be gradually increased with training.

We did not impose restrictions on the duration of each treatment session, the number of treatment sessions, or the time between treatment sessions.

Studies presenting data from one of the following comparisons were eligible for inclusion.

Comparator interventions

- Passive control
 - '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 in a particular routine.
- Psychological therapies: CBT/cognitive therapy/supportive therapy/behavioural therapies/psychodynamic therapies
- Adaptive pacing therapy
- Pharmacological therapy

Types of outcome measures

Primary outcomes

- Fatigue: measured using any validated scale (e.g. Fatigue Scale ([Chalder 1993](#)), Fatigue Severity Scale ([Krupp 1989](#)))
- Adverse effects: measured using any reporting system (e.g. serious adverse reactions ([European Union Clinical Trials Directive 2001](#)))

Secondary outcomes

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

A Cochrane Information Specialist (CIS) with the Common Mental Disorders Group searched their specialised controlled trials register (CCMDCTR-Studies and CCMDCTR-References) from inception to 9 May 2014. This register was created from routine generic searches (for all conditions within the scope of the Group) of MEDLINE (1950-), Embase (1974-) and PsycINFO (1967-). The (weekly) generic searches included subject headings and text-

words for *chronic fatigue syndrome*. Details of the full [generic search strategies](#) used to inform the CCMDCTR can be found on the Group's website.

- The CIS searched the CCMDCTR-Studies Register using the following controlled vocabulary terms: Diagnosis = ("Chronic Fatigue Syndrome" or fatigue) and Free Text = (exercise or sport* or relaxation or "multi convergent" or "tai chi")
- The CIS searched the CCMDCTR-References Register 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](#).

[Note. The Cochrane Common Mental Disorders Group (CCMD) was previously called the *Cochrane Collaboration Depression, Anxiety and Neurosis* (CCDAN) review group. It changed name in 2015 and the re-naming of the specialised register to 'CCMDCTR' reflects this change.]

The following bibliographic databases and international trials registers were also searched to 9 May 2014 (see [Appendix 2](#)):

- Cochrane Central Register of Controlled Trials (CENTRAL, all years to 2014, issue 4) via the Cochrane Library;
- SPORTSDiscus (1985 to 9 May 2014);
- WHO [International Clinical Trials Portal](#) (9 May 2014).

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 JRP 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 studies reported only the SE, review authors converted it to the SD. For dichotomous outcomes, such as dropouts, we extracted the number of events. We sought clarification from authors of the following studies: [Fulcher 1997](#), [Moss-Morris 2005](#), [Wallman 2004](#), [Wearden 1998](#), [Wearden 2010](#) and [White 2011](#). We resolved disagreement between review authors by discussion.

Main comparisons

- Exercise therapy versus 'passive control'
- Exercise therapy versus psychological treatment
- Exercise therapy versus adaptive pacing therapy
- Exercise therapy versus pharmacological therapy

- 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, as described in the *Cochrane Handbook for Systematic Reviews of Interventions* ([Higgins 2011a](#)). This tool encourages consideration of how studies generated the randomisation sequence, how they concealed allocation, 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 and minimal important differences

For continuous outcomes, we calculated the mean difference (MD) when the same scale was used in a similar manner across studies. When studies presented results for continuous outcomes using different scales or different versions of the same scale, we used the standardised mean difference (SMD). For comparison, we also re-expressed SMD estimates using familiar instruments as described in Chapter 12 (Section 12.6.4) of the *Cochrane Handbook for Systematic Reviews of Interventions* ([Schünemann 2011](#)). We adhered to the recommendation in the *Cochrane Handbook for Systematic Reviews of Interventions* to base the conversion from SMDs to MDs on a standard deviation from a representative observational study, and we used the standard deviations reported in [Crawley 2013](#) for this purpose.

Clinical studies and meta-analysis can detect small differences in outcomes with little or no importance to individual participants. Moreover, the interpretation of what is considered an important difference may vary between patients, researchers and clinical experts ([Wyrwich 2007](#)). We therefore identified research literature to help quantify minimal important differences (MID) for important outcome measures. For fatigue, one study among people with systemic lupus erythematosus ([Goligher 2008](#)), reported a threshold around 2.3 points for a minimally important change on the 33-point Chalder Fatigue Scale, an effect size that corresponds to an SMD of about 0.36 ([Goligher 2008](#)).

Studies in people with rheumatoid arthritis or chronic heart disease suggest that the threshold for MID on the physical functioning subscale of SF-36 can be set around 7 points ([Ward 2014](#); [Wyrwich 2007](#)). Studies based on data from patients with chronic obstructive pulmonary disease have also investigated MID for HADS and suggest MIDs around 1.5 points for the HADS anxiety and the HADS depression scale ([Puhan 2008](#); [Smid 2017](#)). We did not detect studies that established a common MID for the Jenkins Sleep Scale, but decided to view a 20% change in sleep scores as a clinically important difference.

Dichotomous data

We expressed dichotomous effect sizes 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. We compared the experimental condition (exercise therapy) with each individual comparator intervention: passive control (treatment as usual/waiting-list control/relaxation/flexibility); psychological treatment (CBT/cognitive therapy/supportive therapy/behavioural therapies/psychodynamic therapies); adaptive pacing therapy; and pharmacological therapy (e.g. antidepressants). This meant that we could include data from the exercise arm in a separate univariate analysis for more than one comparison. We describe under [Differences between protocol and review](#) planned methods that were redundant and not used, 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 intervals 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 2011b](#)). We approached study authors to obtain other types of missing data.

Assessment of heterogeneity

We assessed heterogeneity in accordance 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; [Deeks 2011](#)). In addition to the I^2 value ([Higgins 2003](#)), we present the P value of the χ^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. We used a P value less than 0.1 from the χ^2 test 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 studies allowed a meaningful presentation,

to establish whether reporting biases could be present ([Egger 1997](#)). Asymmetry in funnel plots may indicate publication bias. We identified an insufficient number of studies to use this approach in the present version of the review. We considered clinical heterogeneity of the studies as a possible explanation for some of the heterogeneity in the results.

Data synthesis

We expected some clinical heterogeneity (slightly different interventions, populations and comparators) among studies, and we therefore chose the random-effects model as the default method of analysis. The alternative, the 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 ([Review Manager 2014](#)).

Subgroup analysis and investigation of heterogeneity

We planned no subgroup analyses a priori. To explore possible differences between studies using different exercise strategies, control conditions and diagnostic criteria, we performed post hoc subgroup analyses. We describe the 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, for example the impact of presenting results in terms of SMD or MD, we performed post hoc sensitivity analyses. We also performed sensitivity analyses to investigate the impact of individual studies on overall estimates and heterogeneity measures. We describe results of these sensitivity analyses in the text of the review.

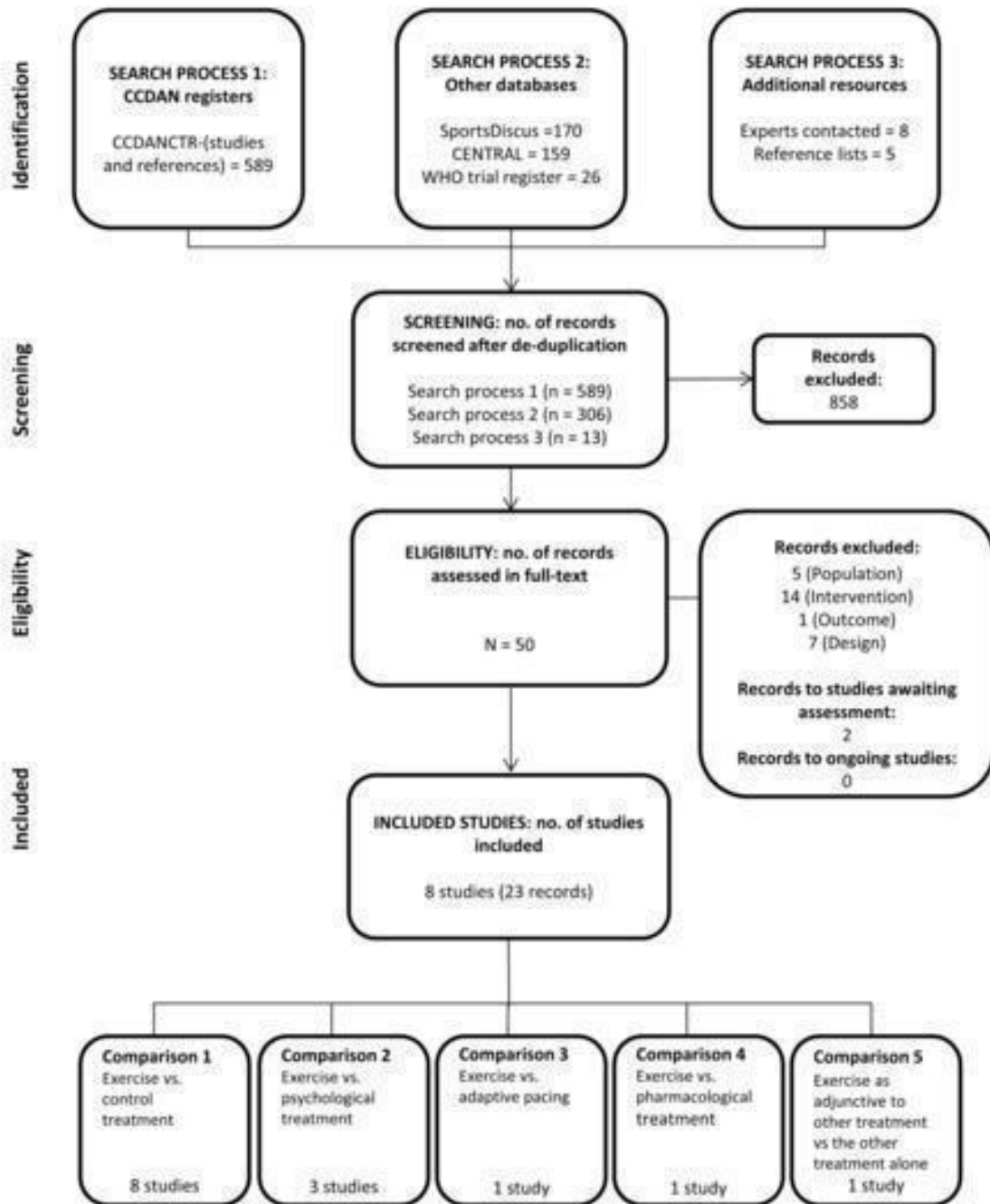
RESULTS

Description of studies

Results of the search

Our searches identified 908 unique records. Of these, we retrieved and read the full text of 50 records. 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 included three newer studies in this update ([Jason 2007](#); [Wearden 2010](#); [White 2011](#); see [Figure 1](#)).

Figure 1. Flow diagram



Included studies

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 in a total of 23 reports. All reports of the included studies were written in English and published in peer-reviewed journals. The eight studies randomly assigned a total of 1518 participants with sample sizes ranging between 49 (Moss-Morris 2005), and 641 participants (White 2011).

Design

All included studies were RCTs. Three studies included two arms and compared exercise versus relaxation/flexibility, waiting list or usual care (Fulcher 1997; Moss-Morris 2005; Wallman 2004). Wearden 2010 had three arms, and four studies had four arms (Jason 2007; Powell 2001; Wearden 1998; White 2011). We used data from all study arms in each study. Regarding Powell 2001, we combined the three interventions into one common intervention group compared with treatment as usual.

Setting

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

Participants

Demographic data for participants are reported in Table 1. Briefly, 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). Mean ages across studies were between 33.0 and 44.6 years. The studies reported median illness durations between 2.3 and 7 years. Depression ranged from 18% of those with a depression diagnosis (Wearden 2010) to 39% among participants with a current Axis I disorder (Jason 2007). Two studies did not report work and employment information (Wallman 2004; Wearden 2010). Fulcher 1997 and Jason 2007 reported that 39% and 46% of the participants were working or studying on at least a part-time basis. In comparison, 22% of participants in Moss-Morris 2005 were unemployed and unable to work because of disability, whereas 43% of participants in Powell 2001 received disability pensions.

Intervention characteristics

Characteristics of the exercise therapy interventions are reported in detail in Table 2. Briefly, the specific duration of the exercise therapy regimen varied from 12 weeks to 26 weeks. Seven studies used variations of aerobic exercise therapy, with levels of intensity ranging from HR at 40% of VO₂max to HR at 75% of VO₂max (Table 2). One study used anaerobic exercise (Jason 2007). Scheduled therapist meetings were conducted face-to-face or by telephone, and varied from every second week to weekly. Some sessions

involved talking, and others involved supervised exercise. Most of the included studies encouraged 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 an exercise diary to measure adherence to treatment (Table 2). Control interventions included treatment as usual, relaxation, flexibility and waiting-list controls. Comparator interventions included CBT, adaptive pacing and antidepressants.

Outcomes

Outcomes for each study are described in detail in the table [Characteristics of included studies](#).

The main outcomes were symptom levels measured by rating scales at end of exercise therapy (12 to 26 weeks) and at follow-up (52 to 70 weeks). Fatigue was measured by the Fatigue Scale (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 (Krupp 1989), in Jason 2007. One study (White 2011), reported adverse outcomes according to serious adverse reactions categories (European Union Clinical Trials Directive 2001).

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

Six studies (Fulcher 1997; Jason 2007; Moss-Morris 2005; Powell 2001; Wallman 2004; 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 HADS (Zigmond 1983), and one (Jason 2007), used the Beck Depression Inventory (Beck 1996), and the Beck Anxiety Inventory (Hewitt 1993). Three studies (Powell 2001; Wearden 2010; White 2011), measured sleep problems by using a questionnaire (Jenkins 1988), and one (Fulcher 1997), by using the Pittsburgh Sleep Quality Index (Buysse 1989).

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

The review authors calculated dropout.

The included studies reported several outcomes in addition to those reported in this review, such as work capacity by oxygen consumption (VO₂), the six-minute walking test and illness beliefs.

Ethics approval

All the included studies listed sponsorship or sources of funding, and reported that they had obtained ethics approvals.

Excluded studies

As described in [Characteristics of excluded studies](#), the current review excluded a total of 20 studies for the following reasons.

- Five studies did not include our target participants; three due to diagnostic criteria (Guarino 2001; Ridsdale 2004; Ridsdale 2012), and two because the participants were younger than 18 years (Viner 2004; Wright 2005).
- We excluded seven studies because exercise therapy was a minor part of the intervention (Evering 2008; Kos 2012; Nunez 2011; Russel 2001; Stevens 1999; Taylor 2004; Tummers 2012).
- Three studies only compared active exercise interventions (Broadbent 2012; Gordon 2010; Vos-Vromans 2008).
- Two studies did not report relevant data (Hatcher 1998 Liu 2010).
- Three studies were not RCTs (Taylor 2006; Thomas 2008; Zhuo 2007).

Ongoing studies

We are not aware of any relevant ongoing studies.

Studies awaiting classification

Two studies that were ongoing when we ran our search for literature in May 2014 (Marques 2012; White 2012), are now published. The publications based on these studies (Clarke 2017; Marques 2015), need to be assessed for eligibility next time this review is updated.

New studies found at this update

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

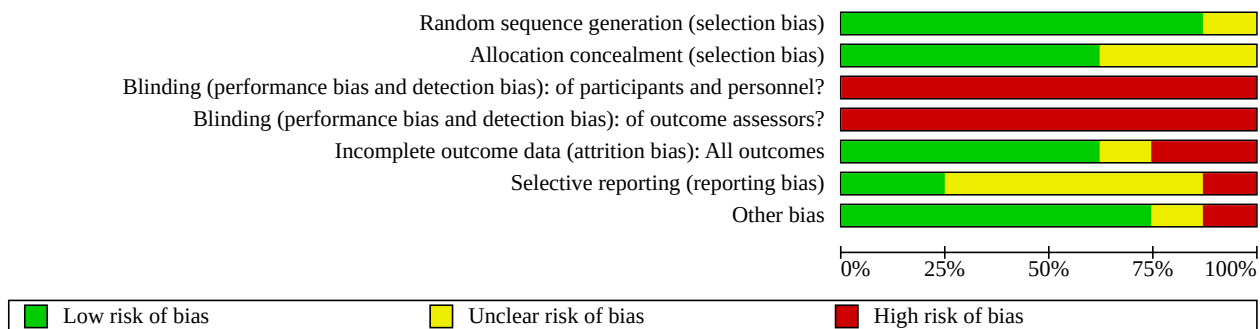
Risk of bias in included studies

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

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

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding (performance bias and detection bias): of participants and personnel?	Blinding (performance bias and detection bias): of outcome assessors?	Incomplete outcome data (attrition bias): All outcomes	Selective reporting (reporting bias)	Other bias
Fulcher 1997	+	+	-	-	+	?	+
Jason 2007	+	?	-	-	-	?	-
Moss-Morris 2005	+	+	-	-	+	?	+
Powell 2001	+	?	-	-	+	?	+
Wallman 2004	?	?	-	-	+	?	?
Wearden 1998	+	+	-	-	-	-	+
Wearden 2010	+	+	-	-	?	+	+
White 2011	+	+	-	-	+	+	+

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



Allocation

Seven of the eight studies had adequate sequence generation and were assessed to low risk of bias, whereas Wallman 2004 was assessed as having unclear risk of bias because the sequence generation was not described in sufficient detail. Five studies reported adequate methods of allocation concealment, i.e. low risk of bias. In three of the studies, we judged the risk of bias to be unclear because the allocation concealment was not described in sufficient detail (Jason 2007; Powell 2001; Wallman 2004).

Blinding

The intervention does not allow blinding of the participants or the staff delivering the exercise-based interventions. As all measures were performed by self-report, we rated all included studies as having high risk of performance and detection 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 study was judged to be at unclear risk of attrition bias (Wearden 2010). The dropout rate in the intervention groups was relatively high, but most of the participants who dropped out from treatment were available for follow-up assessments and were analysed within the groups to which they had been randomly assigned (Wearden 2010). Two studies were associated with high risk of attrition bias (Jason 2007; Wearden 1998). Wearden 1998 reported large dropout rates in all intervention groups, and many participants were lost to follow-up.

Selective reporting

Wearden 2010 and White 2011 referenced published protocols. We checked these against the published results, and found that reporting was adequate and that the risk of bias was low. Wearden 1998 was judged as being at high risk of reporting bias because study 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 other subscales or for anxiety, as data were "similar in trial completers." It was not possible to

check the other studies for selective reporting bias; therefore we considered their risk of bias unclear.

Other potential sources of bias

For six of the eight included studies, we did not suspect other sources of bias, and the risk of bias was assessed as low. Wallman 2004 showed differences between groups for anxiety and mental fatigue at baseline that may have influenced the results, and the risk of bias was therefore judged as unclear. Jason 2007 showed large baseline differences across groups for several variables, and as the consequences of these differences were not discussed satisfactorily in the paper, the risk of bias was assessed to high.

Effects of interventions

See: **Summary of findings 1** Exercise therapy versus control for chronic fatigue syndrome; **Summary of findings 2** Exercise therapy versus psychological treatment for chronic fatigue syndrome; **Summary of findings 3** Exercise therapy versus adaptive pacing therapy for chronic fatigue syndrome; **Summary of findings 4** Exercise therapy versus antidepressants for chronic fatigue syndrome; **Summary of findings 5** Exercise therapy plus antidepressants versus antidepressants alone for chronic fatigue syndrome

Exercise therapy versus control

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

All eight included studies (Fulcher 1997; Jason 2007; Moss-Morris 2005; Powell 2001; Wallman 2004; Wearden 1998; Wearden 2010; White 2011), contributed data for this comparison, but not for all the outcomes. Jason 2007 only reported data after 52 weeks' follow-up, and is not included in meta-analysis of end of treatment data.

1.1 Fatigue

There is moderate-certainty evidence that exercise therapy was probably more effective than control in reducing fatigue at end of treatment (SMD -0.66, 95% CI -1.01 to -0.31; 7 studies, 840 participants; Analysis 1.1). We pooled data using SMD methods because the available studies measured fatigue using different approaches with the Fatigue Scale (Chalder 1993). Briefly, two studies (Powell 2001; Wearden 2010), assessed fatigue by dichotomised scoring of an Chalder's 11-item Fatigue Scale (0 to 11 points; Chalder 1993), and five studies measured fatigue using

the same scale but with a scoring system ranging from 0 to 33 points (Wallman 2004; White 2011) or from 0 to 42 (Fulcher 1997; Moss-Morris 2005; Wearden 1998). If the pooled SMD estimate is re-expressed on the 33-point Chalder Fatigue Scale, it corresponds to an MD of -3.4 points (95% CI -5.3 to -1.6). The analysis suffered from considerable heterogeneity ($I^2 = 80\%$, $P < 0.0001$) that we explored in sensitivity analysis.

Only very low-certainty evidence is available for evaluating the effect of exercise therapy on fatigue at longer-term follow up (SMD -0.62, 95% CI -1.32 to 0.07; 4 studies, 670 participants; Analysis 1.2). We used SMD because two studies (Powell 2001; Wearden 2010), assessed fatigue by dichotomised scoring of an 11-item Fatigue Scale (Chalder 1993), one study used a scoring system from 0 to 33 points (White 2011), and Jason 2007 reported fatigue measured on the Fatigue Severity Scale (Krupp 1989). If the pooled SMD estimate is re-expressed on the 33-point Chalder Fatigue Scale, it corresponds to an MD of -3.2 points (95% CI -6.9 to 0.4). The analysis suffered from extensive heterogeneity ($I^2 = 94\%$, $P < 0.0001$) that was explored in sensitivity analysis.

Sensitivity analysis

- Investigating heterogeneity
 - The meta-analysis of fatigue at end of treatment was associated with considerable heterogeneity (Analysis 1.1). The observed heterogeneity was caused mainly by the deviating results presented in Powell 2001. Exclusion of Powell 2001 from the meta-analysis gave rise to a smaller, but still moderately large SMD of -0.44 (95% CI -0.63 to -0.24). This estimate was not associated with heterogeneity ($I^2 = 26\%$, $P = 0.24$). The exclusion of other studies from the analysis had minimal impact on the total estimate or on heterogeneity measures.
 - The meta-analysis of fatigue at follow-up was also associated with heterogeneity (Analysis 1.2). Exclusion of Powell 2001 from the meta-analysis resulted in a smaller SMD of -0.27 (95% CI -0.54 to 0.00) and reduced heterogeneity ($I^2 = 49\%$, $P = 0.16$). For comparison, exclusion of White 2011, Wearden 2010 and Jason 2007 led to pooled estimates of (SMD -0.68; 95% CI -1.86 to 0.49; $I^2 = 96\%$), (SMD -0.76; 95% CI -1.80 to 0.29; $I^2 = 95\%$) and (SMD -0.85; 95% CI -1.67 to -0.03; $I^2 = 95\%$) respectively.
- Mean difference or standardised mean difference
 - The included studies measured fatigue using different reporting scales, and we performed a sensitivity analysis in which the results were presented on the original reporting scale (Analysis 1.19; Analysis 1.20).

Subgroup analysis

To explore the possible impact of our pooling strategy such as 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. We summarise the results below.

- Type of exercise
 - Post hoc subgroup analysis based on treatment strategy did not establish differences ($I^2 = 0\%$, $P = 0.66$) between studies of graded exercise therapy (Fulcher 1997; Moss-Morris 2005; Powell 2001; Wearden 1998; Wearden 2010; White 2011), and studies of exercise with self-pacing ((Wallman 2004), SMD

-0.68, 95% CI -1.08 to -0.28; $I^2 = 84\%$ versus SMD -0.54, 95% CI -1.05 to -0.02 respectively; Analysis 1.21).

- At follow-up, post hoc subgroup analysis did not result in statistically significant subgroup differences ($I^2 = 72.6\%$, $P = 0.06$, Analysis 1.22) between the three studies (Powell 2001; Wearden 2010; White 2011), comparing graded exercise versus treatment as usual (SMD -0.85, 95% CI -1.67 to -0.03; $I^2 = 95\%$) and Jason 2007, who compared anaerobic activity versus relaxation (SMD 0.12, 95% CI -0.44 to 0.67).
- Type of control
 - We could not establish a subgroup difference ($I^2 = 0\%$, $P = 0.97$) between the five studies with treatment as usual as control (Moss-Morris 2005; Powell 2001; Wearden 1998; Wearden 2010; White 2011), and Fulcher 1997 and Wallman 2004, who used an active control of relaxation or flexibility (SMD -0.66, 95% CI -1.13 to -0.20 versus SMD -0.65, 95% CI -1.02 to -0.28); analysis not shown.
- Diagnostic criteria
 - The use of various diagnostic criteria is often emphasised as relevant to treatment response. We therefore performed subgroup analyses based on diagnostic criteria (analyses not shown). There was little or no difference between subgroups ($I^2 = 0\%$, $P = 0.76$) in our 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; SMD -0.73, 95% CI -1.17 to -0.28 versus SMD -0.63, 95% CI -1.07 to -0.19).

1.2 Adverse effects

One study (White 2011), reported the rate of serious adverse reactions between exercise therapy and treatment as usual (RR 0.99, 95% CI 0.14 to 6.97; 1 study, 319 participants; Analysis 1.3). We defined serious adverse reactions according to the European Union Clinical Trials Directive 2001. White 2011 observed two serious adverse reactions (i.e. deterioration in mobility and self-care, and worse CFS symptoms and function) among 160 participants in the exercise group, and two (i.e. worse CFS symptoms and function, and increased depression and incapacity) among the 159 participants in the control group (Analysis 1.3). Wearden 2010 reported no serious adverse reactions in either group. The confidence interval remains wide due to few events in all intervention groups, and therefore the effect of exercise therapy on serious adverse reactions remains uncertain (very low-certainty evidence).

1.3 Pain

Wearden 1998 reported that the exercise group and control group scored similarly on the pain subscale of SF-36 (Ware 1992), but did not report actual data and this is therefore very low-certainty evidence.

Analysis 1.4 presents analysis based on 43 participants from one study (Jason 2007), which assessed pain after 52 weeks using the Brief Pain Inventory (scale: 0 to 10 points; Cleeland 1994), and observed an MD of -0.97 (95% CI -2.44 to 0.50) on pain severity and MD -0.69 (95% CI -2.48 to 1.10) on the pain interference subscale. The evidence is very low certainty, and hence we are uncertain whether exercise therapy affects pain.

1.4 Physical functioning

At end of treatment there is low-certainty evidence suggesting that exercise therapy may improve physical functioning more than passive control (MD -13.10, 95% CI -24.22 to -1.98; 5 studies, 725 participants; [Analysis 1.5](#)). The five available studies ([Fulcher 1997](#); [Moss-Morris 2005](#); [Powell 2001](#); [Wearden 2010](#); [White 2011](#)), assessed physical functioning according to the physical functioning subscale of SF-36 (scale: 0 to 100 points; [Ware 1992](#)). The meta-analysis was associated with considerable heterogeneity that we explored in sensitivity analysis ($I^2 = 89%$, $P < 0.00001$).

Only very low-certainty evidence is available from three studies ([Powell 2001](#); [Wearden 2010](#); [White 2011](#)), for evaluating the effects of exercise therapy on physical functioning after follow-up of 52 to 70 weeks (MD -16.33, 95% CI -36.74 to 4.08; 3 studies, 621 participants; [Analysis 1.6](#)). In addition to the three studies already mentioned, [Jason 2007](#) observed better results among participants in the relaxation group (MD 21.48, 95% CI 5.81 to 37.15). The latter results were distorted by very large baseline differences in physical functioning between the exercise and relaxation groups (39/100 versus 54/100), and we therefore decided not to include these results in the meta-analysis.

Sensitivity analysis

- Investigating heterogeneity
 - The heterogeneity in [Analysis 1.5](#) was largely driven by the remarkably positive effect of exercise therapy reported by [Powell 2001](#). Heterogeneity (I^2 statistic) dropped to 52% ($P = 0.10$) following exclusion of [Powell 2001](#) (analysis not shown). 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 be associated with the large variation in baseline physical functioning that ranged from 29.8 ([Wearden 2010](#)), to 53.1 ([Moss-Morris 2005](#)). When excluding [Wearden 2010](#) and [Powell 2001](#), the I^2 statistic dropped to zero and the effect estimate became statistically significant (MD -8.99, 95% CI -13.41 to -4.58). Exclusion of [White 2011](#) had limited impact on the I^2 statistic, but the pooled estimate changed from MD -13.10 (95% CI -24.22 to -1.98) to MD -14.83 (95% CI -30.33 to 0.67).
 - The remarkably positive result reported by [Powell 2001](#) also introduced heterogeneity into the meta-analysis of follow-up data ([Analysis 1.6](#)). When we excluded [Powell 2001](#) (analysis not shown), heterogeneity dropped to 0% ($P = 0.50$), and the two remaining studies ([Wearden 2010](#); [White 2011](#)), reported a smaller but statistically significant pooled estimate in favour of exercise therapy (MD -5.79, 95% CI -10.53 to -1.06). Exclusion of [White 2011](#) or [Wearden 2010](#) had limited impact on heterogeneity measures, and yielded pooled estimates of MD -21.21 (95% CI -56.05 to 13.64) and MD -22.78 (95% CI -54.24 to 8.67), respectively.

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](#). The results are summarised below.

- 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 versus 54 of 100), and we did not include them in [Analysis 1.6](#).

- Type of control
 - At end of treatment, post hoc subgroup analysis did not establish a subgroup difference ($I^2 = 0%$, $P = 0.92$), 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 the one study [Fulcher 1997](#), using relaxation or flexibility as a control (MD -13.87, 95% CI -24.31 to -3.43). Analysis is not shown.
 - All studies available for analysis at follow-up adhered to the treatment-as-usual control condition, hence we did not perform any sensitivity analyses within [Analysis 1.6](#).
- Diagnostic criteria
 - We found no evidence of subgroup differences ($I^2 = 0%$, $P = 0.91$) 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 the four studies diagnosing participants according to the Oxford criteria (MD -12.92, 95% CI -25.99 to 0.14; [Fulcher 1997](#); [Powell 2001](#); [Wearden 2010](#); [White 2011](#)). Analysis is not shown.
 - All studies available for analysis at follow-up recruited participants in keeping with the Oxford criteria, thus we did not perform any subgroup analyses within [Analysis 1.6](#).

1.5 Quality of life

None of the included studies reported quality of life at end of treatment. Very low-certainty evidence looked at quality of life at 52 weeks' follow-up, but due to the very low certainty of the evidence, we are uncertain whether exercise therapy affects quality of life at long-term follow-up. The very low-certainty evidence ([Analysis 1.7](#)), was based on 43 participants from one study ([Jason 2007](#)), which observed an MD of 9.00 (95% CI -1.00 to 19.00). The estimate is biased in favour of the control arm because of baseline differences between groups. [Jason 2007](#) measured quality of life on the Quality of Life Scale, consisting of 16 items answered on a scale of 1 to 7 ([Burckhardt 2003](#)).

1.6.1 Depression

Only very low-certainty evidence is available for the assessment of depression at end of treatment, and we are therefore uncertain whether exercise therapy affects depression at end of treatment. The very low-certainty evidence was based on 504 participants from five studies ([Fulcher 1997](#); [Powell 2001](#); [Wallman 2004](#); [Wearden 1998](#); [Wearden 2010](#)). All studies measured symptoms using the depression subscale of the HADS, (scale: 0 to 21 points; [Zigmond 1983](#)), and resulted in a pooled MD of -1.63 (95% CI -3.50 to 0.23) in an analysis that was associated with considerable heterogeneity ($I^2 = 84%$, $P < 0.0001$; [Analysis 1.8](#)).

Only very low-certainty evidence is available for the assessment of depression at 52 to 70 weeks' follow-up, and we are therefore uncertain whether exercise therapy may have an impact on depression at follow-up. The very low-certainty evidence was based on 654 participants from four studies ([Jason 2007](#); [Powell 2001](#); [Wearden 2010](#); [White 2011](#)). [Jason 2007](#) used the Beck Depression Inventory ([Beck 1996](#)), and three studies ([Powell 2001](#);

Wearden 2010; White 2011), used HADS depression subscale values (Zigmond 1983). The four studies yielded a pooled SMD of -0.35 (95% CI -0.93 to 0.23) in an analysis that was associated with considerable heterogeneity ($I^2 = 91\%$, $P < 0.00001$; Analysis 1.9).

Sensitivity analysis

- Investigating heterogeneity
 - At end of treatment, Powell 2001 reported very positive results and contributed greatly to the total heterogeneity. Exclusion of Powell 2001 (analysis not shown) led to a reduction in observed effect size (MD 0.80 , 95% CI -0.21 to 1.82), but heterogeneity was also greatly reduced ($I^2 = 36\%$, $P = 0.20$).
 - At follow-up, Powell 2001 reported a substantial benefit of exercise therapy compared with results described by the other studies. Jason 2007 reported results in favour of the control condition, but these results were impaired by baseline differences between the groups. Exclusion of Powell 2001 from the meta-analysis was associated with a drop in the I^2 statistic from 71% to 36% and SMD -0.04 (95% CI -0.28 to 0.20). Simultaneous exclusion of Jason 2007 led to a pooled SMD -0.05 (95% CI -0.37 to 0.27). Exclusion of White 2011 or Wearden 2010 had limited impact on heterogeneity measures, and resulted in pooled estimates of SMD -0.16 (95% CI -0.70 to 0.38) and SMD -0.29 (95% CI -0.66 to 0.07) respectively. Analyses are not shown.
- Mean difference or standardised mean difference
 - At longer-term follow-up, studies used different measurement scales to measure and report depression. We performed a sensitivity analysis in which all available studies were presented using the original reporting scale (Analysis 1.23). Jason 2007 (45 participants) reported a mean difference on the Beck Depression Inventory (Beck 1996), of 3.44 points (95% CI -3.00 to 9.88). Three studies (609 participants) assessed follow-up changes in depression using the HADS depression subscale (Powell 2001; Wearden 2010; White 2011), yielding a pooled MD of -2.26 points (95% CI -5.09 to 0.56) with considerable heterogeneity ($I^2 = 92\%$, $P < 0.00001$).

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. The results are summarised below.

- Type of exercise
 - We did not observe any statistical subgroup differences ($I^2 = 0\%$, $P = 0.75$) between the four studies offering graded exercise therapy (Fulcher 1997; Powell 2001; Wearden 1998; Wearden 2010), and the one study offering exercise with personal pacing (Wallman 2004). Analysis not shown.
 - 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 (not shown) associated with considerable heterogeneity ($I^2 = 91\%$, $P < 0.00001$). Post hoc subgroup analysis showed that it is uncertain whether there is a subgroup difference ($I^2 = 71.2\%$, $P = 0.06$) 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 the one study (Jason 2007), comparing 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 ($I^2 = 0\%$, $P = 0.61$) 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 ; $I^2 = 59\%$). Analysis is not shown.

1.6.2 Anxiety

Only very low-certainty evidence is available for the assessment of anxiety at end of treatment, and we are therefore uncertain whether exercise therapy affects anxiety at end of treatment. Five studies (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), but only three studies (387 participants) reported data in a way that facilitated comparison in a meta-analysis (Powell 2001; Wallman 2004; Wearden 2010). The meta-analysis yielded a pooled estimate of MD -1.48 points (95% CI -3.58 to 0.61 ; Analysis 1.10). The meta-analysis was associated with substantial heterogeneity ($I^2 = 79\%$, $P = 0.008$), but some of this heterogeneity can be explained by uncorrected baseline differences in HADS anxiety score in the included studies. Wearden 1998 (68 participants) stated that they observed no significant changes 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 they did observe an increase in median HADS anxiety score from 4 to 7 in the control group.

Only very low-certainty evidence is available for the assessment of anxiety at 52 to 70 weeks' follow-up, and we are therefore uncertain whether exercise therapy may have an impact on anxiety at follow-up. The very low-certainty evidence was based on 652 participants from four studies (Jason 2007; Powell 2001; Wearden 2010; White 2011). Jason 2007 used Beck Anxiety Inventory (Hewitt 1993), and three studies (Powell 2001; Wearden 2010; White 2011), used HADS anxiety subscale values (Zigmond 1983). The four studies yielded a pooled SMD of -0.17 (95% CI -0.50 to 0.15) in an analysis that was associated with considerable heterogeneity ($I^2 = 71\%$, $P = 0.02$; Analysis 1.11).

Sensitivity analysis

- Investigating heterogeneity
 - At follow-up, Powell 2001 reported a substantial benefit of exercise therapy compared with results described by the other studies. Jason 2007 reported results in favour of the control condition, but these results were impaired by baseline differences between the groups. Exclusion of Powell 2001 from the meta-analysis (analysis not shown) was associated with a drop in I^2 statistic values from 91% to 43% ($P = 0.17$) and SMD -0.09 (95% CI -0.34 to 0.17). Simultaneous exclusion of Jason 2007 led to a further drop in the I^2 statistic ($I^2 = 14\%$, $P = 0.28$), and a pooled SMD -0.17 (95% CI -0.37 to 0.03). Exclusion of White 2011 or Wearden 2010 (analysis not shown) had limited impact on heterogeneity, and resulted in

pooled estimates of SMD -0.38 (95% CI -1.35 to 0.60) and SMD -0.45 (95% CI -1.30 to 0.40) respectively.

- Mean difference or standardised mean difference
 - Studies used different measurement scales to measure and report anxiety at longer-term follow-up, therefore we performed a sensitivity analysis in which all available studies were presented using the original reporting scale (Analysis 1.24). Jason 2007 (45 participants) reported a mean difference on the Beck Anxiety Inventory (Hewitt 1993), of 0.70 points (95% CI -4.52 to 5.92). Three studies (607 participants) assessed follow-up changes in anxiety using the HADS anxiety subscale (Powell 2001; Wearden 2010; White 2011), yielding a pooled MD of -1.01 points (95% CI -2.75 to 0.74) with considerable heterogeneity ($I^2 = 78\%$, $P = 0.01$).

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. The results are summarised below.

- Type of exercise and control
 - At end of treatment, post hoc subgroup analysis (not shown) did not establish a subgroup difference ($I^2 = 0\%$, $P = 0.64$) between the two studies (Powell 2001; Wearden 2010), comparing graded exercise therapy versus treatment as usual (MD -1.22 , 95% CI -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 (not shown) was associated with substantial heterogeneity ($I^2 = 71\%$, $P = 0.02$). We could not establish a statistically significant subgroup difference ($I^2 = 0\%$, $P = 0.40$) 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 the one study (Jason 2007), comparing anaerobic activity versus relaxation (SMD 0.08 , 95% CI -0.51 to 0.66).

1.7 Sleep

Low-certainty evidence showed that sleep may improve slightly following exercise therapy at end of treatment (MD -1.49 points, 95% CI -2.95 to -0.02 ; 2 studies, 323 participants; Analysis 1.12). The available studies (Powell 2001; Wearden 2010), assessed sleep using Jenkins Sleep Scale (Jenkins 1988), four domains with scores from 0 to 5, which gives a score from 0 to 20. In addition, Fulcher 1997 (59 participants) observed a reduction in median sleep score, as assessed by the Pittsburgh Sleep Quality Index (score 0 to 21, where lower scores denote a healthier sleep quality), from 7 to 5 in the exercise group, whereas median sleep score remained 6 in the control group, although this group difference did not reach statistical significance in non-parametric statistical analysis.

Low-certainty evidence showed that exercise therapy may lead to a slight improvement in sleep at 52 to 70 weeks' follow-up (MD -2.04 points, 95% CI -3.48 to -0.23 ; 3 studies, 610 participants; Analysis 1.13). The meta-analysis was associated with considerable heterogeneity ($I^2 = 75\%$, $P = 0.02$) that we explored in sensitivity analysis.

Sensitivity analysis

We explored the heterogeneity by excluding specific studies from the meta-analysis of follow-up data (analysis not shown). Exclusion of Powell 2001, Wearden 2010 and White 2011 resulted in pooled estimates of MD -1.27 (95% CI -2.91 to 0.37), MD -2.84 (95% CI -4.82 to -0.87), and MD -2.13 (95% CI -5.80 to 1.53), respectively.

Subgroup analysis

All available studies compared graded exercise therapy versus treatment as usual. All studies recruited participants according to the Oxford criteria, thus we did not perform any subgroup analyses within Analysis 1.12 and Analysis 1.13.

1.8 Self-perceived changes in overall health

There is moderate-certainty evidence that exercise therapy probably increases the number of people who report at least some degree of improvement in self-perception of overall health at end of treatment (RR 1.83, 95% CI 1.39 to 2.40; 4 studies, 489 participants, Analysis 1.14). The four available studies (Fulcher 1997; Moss-Morris 2005; Wallman 2004; Wearden 2010), assessed changes in overall health at end of treatment by using a self-rated Global Impression Change Scale with scores ranging from 1 (very much better) to 7 (very much worse). The meta-analysis was not associated with heterogeneity.

Only very low-certainty evidence is available for the assessment of self-perception of overall health at 52 weeks' follow-up, and we are therefore uncertain whether exercise therapy may have an impact on self-perception of overall health at follow-up. Three studies (518 participants) were available for analysis, but they showed highly heterogeneous results yielding a pooled RR estimate of 1.88 (95% CI 0.76 to 4.64; Analysis 1.15).

Sensitivity analysis

- Investigating heterogeneity
 - End-of-treatment data were not associated with heterogeneity (Analysis 1.14), and exclusion of single studies had limited impact on the pooled estimate. Briefly, exclusion of Fulcher 1997, Moss-Morris 2005, Wallman 2004 and White 2011 resulted in RRs of 1.79 (95% CI 1.32 to 2.41), 1.78 (95% CI 1.34 to 2.38), 2.01 (95% CI 1.46 to 2.77) and 1.75 (95% CI 1.20 to 2.54), respectively (analyses not shown).
 - The meta-analysis of follow-up data was associated with considerable heterogeneity ($I^2 = 85\%$, $P = 0.001$; Analysis 1.15). Exclusion of Jason 2007, Powell 2001 and White 2011 led to some changes in pooled estimates and heterogeneity measures (RR 2.92, 95% CI 0.75 to 11.35; $I^2 = 87\%$; RR 1.23, 95% CI 0.64 to 2.36; $I^2 = 71\%$; RR 2.18, 95% CI 0.24 to 19.75; $I^2 = 94\%$), respectively (analyses not shown).

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. The results of the subgroup analysis are summarised below.

- 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 ($I^2 = 0\%$, $P = 0.50$), analyses not shown.

- 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 ($I^2 = 13.6\%$, $P = 0.28$). Analyses not shown.
 - 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 ($I^2 = 85\%$, $P = 0.001$). The post hoc subgroup analysis (not shown) did not firmly establish a subgroup difference ($I^2 = 63\%$, $P = 0.10$) 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 exercise group had a lower mean number of specialist medical care contacts than those allocated to treatment as usual (MD -1.40, 95% CI -1.87 to -0.93; Analysis 1.16). A variety of other health care resource use metrics did not differ significantly between the two groups (Analysis 1.16; Analysis 1.17), including use of primary care resources (e.g. GP or practice nurse), other doctor contacts (e.g. neurologist, psychiatrist or other specialists), accident and emergency contacts, medication (e.g. hypnotics, anxiolytics, antidepressants or analgesics), contacts with other healthcare professionals (e.g. 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)).

During the 12-month post-randomisation period, participants in the exercise group had a lower mean number of specialist medical...?

1.10 Dropout

Only very low-certainty evidence was available for the assessment of dropout during treatment, and we are therefore uncertain whether exercise therapy has an impact on dropout rate. Six studies (Fulcher 1997; Moss-Morris 2005; Powell 2001; Wearden 1998; Wearden 2010; White 2011), reported dropout rates yielding a pooled RR estimate of 1.63 (95% CI 0.77 to 3.43; 6 studies, 843 participants; Analysis 1.18) with moderate heterogeneity ($I^2 = 50\%$).

Sensitivity analysis

The main analysis was associated with moderate heterogeneity ($I^2 = 50\%$, $P = 0.07$), but changed when we excluded individual studies from the analysis (analyses not shown). Exclusion of White 2011 or Wearden 2010 altered the pooled estimate and I^2 statistic value to RR 2.11 (95% CI 0.99 to 4.50; $I^2 = 28\%$) and RR 1.30 (95% CI

0.75 to 2.25; $I^2 = 16\%$), respectively. Exclusion of other studies had only limited impact on the pooled estimates and heterogeneity measures.

Subgroup analysis

The main analysis pooled studies using either treatment as usual (Moss-Morris 2005; Powell 2001; Wearden 1998; Wearden 2010), or 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 ($I^2 = 0\%$, $P = 0.74$). Analyses not shown.

Exercise therapy versus other treatments

Comparison 2. Exercise therapy versus psychological treatment

Two studies (Jason 2007; White 2011), contributed data to the main comparison of exercise therapy versus psychological treatment, using cognitive-behavioural therapy (CBT), described below. We also briefly describe results comparing exercise therapy versus cognitive therapy (Jason 2007), and supportive listening (Wearden 2010).

2.1 Fatigue

End of treatment

There is low-certainty evidence that exercise therapy may lead to little or no difference in fatigue compared to CBT at end of treatment (MD 0.20, 95% CI -1.49 to 1.89; 1 study, 298 participants; Analysis 2.1). One study (White 2011), provided data for this comparison.

Regarding other comparisons, Wearden 2010 reported that exercise therapy was associated with greater improvement in fatigue than supportive listening (MD -4.03, 95% CI -6.24 to -1.82; 1 study; 182 participants; Analysis 2.1).

Follow-up

Moderate-certainty evidence suggests that exercise therapy probably leads to little or no difference in fatigue compared to CBT after 52 weeks (SMD 0.07, 95% CI -0.13 to 0.28; 2 studies, 351 participants; Analysis 2.2). Jason 2007 assessed fatigue using a 7-point Fatigue Severity Scale (Krupp 1989), whereas White 2011 assessed fatigue on a 33-point Fatigue Scale (Chalder 1993), and hence we pooled results using SMD.

Regarding other comparisons, 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 cognitive therapy (49 participants). Wearden 2010 (182 participants) assessed fatigue on a 33-point Fatigue Scale (Chalder 1993), and reported differences between exercise and supportive listening that were in favour of graded exercise therapy (MD -2.72, 95% CI -5.14 to -0.30; 1 study, 182 participants).

Subgroup analysis

Post hoc subgroup analysis did not establish a subgroup difference ($I^2 = 0\%$, $P = 0.40$) 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

One study (White 2011), reported the rate of serious adverse reactions between exercise therapy and CBT (RR 0.67, 95% CI 0.11 to 3.96; 1 study, 321 participants; Analysis 2.3). White 2011 defined serious adverse reactions according to European Union Clinical Trials Directive 2001 and observed two serious adverse reactions (i.e. deterioration in mobility and self-care, and worse CFS symptoms and function) among 160 participants in the exercise group, while three participants in the CBT group reported four serious adverse reactions (i.e. 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). The confidence interval remains wide due to few events in all intervention groups, and it therefore the effect of exercise therapy on serious adverse reactions remains uncertain (very low-certainty evidence).

Wearden 2010 stated that no participants in the exercise or supportive listening group demonstrated serious adverse reactions with a probable relation to therapy (Analysis 2.3).

2.3 Pain

No studies reported pain at end of treatment. Very low-certainty evidence from one study (Jason 2007), reported pain at 52 weeks' follow-up. Hence we are uncertain whether exercise therapy affects pain. The very low-certainty evidence (Analysis 1.4), was based on 43 participants from one study (Jason 2007), assessing pain using the Brief Pain Inventory (Cleeland 1994). When comparing CBT and exercise, Jason 2007 observed an MD of 0.07 (95% CI -1.52 to 1.66; Analysis 2.4) for pain severity and MD -0.35 (95% CI -2.29 to 1.59; Analysis 2.5) for pain interference.

Jason 2007 also compared exercise versus cognitive therapy (44 participants). The estimates were MD 0.51 (95% CI -0.92 to 1.94; Analysis 2.4) for pain intensity and MD 0.39 (95% CI -1.37 to 2.15; Analysis 2.5) for pain interference.

2.4 Physical functioning

End of treatment

There is low-certainty evidence that exercise therapy may have little or no impact on physical functioning at end of treatment when compared to CBT (MD -1.20, 95% CI -6.30 to 3.90; 1 study, 298 participants; Analysis 2.6). White 2011 assessed physical functioning using the SF-36 physical functioning subscale (Ware 1992).

With regard to other comparisons, Wearden 2010 compared physical functioning between exercise therapy and supportive listening (MD -6.66, 95% CI -13.7 to 0.40; 1 study, 181 participants; Analysis 2.6).

Follow-up

Only very low-certainty evidence is available for the comparison between exercise and CBT at 52 weeks' follow-up, and we are therefore uncertain whether exercise therapy may have more or less impact than CBT on physical function at follow-up. The very low-certainty evidence was based on 348 participants from two studies (Jason 2007; White 2011), yielding a pooled estimate

of MD 7.92 (95% CI -9.79 to 25.63; scale 0-100 points; Analysis 2.7). 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 2.7), Jason 2007 (46 participants) reported a significant difference favouring CBT (MD 18.92, 95% CI 2.12 to 35.72; Analysis 2.7), when compared with anaerobic exercise. However, results of the latter study are skewed because of unadjusted baseline differences in physical functioning between the two groups (39 versus 46 points), and this explains some of the observed heterogeneity.

With regard to other comparisons, Jason 2007 (47 participants) compared anaerobic exercise versus cognitive therapy (MD 21.37, 95% CI 6.61 to 36.13; Analysis 2.7). The latter estimate is probably biased in favour of cognitive therapy because of uncorrected baseline differences in physical functioning between the two groups (39 versus 46 points). Wearden 2010 (171 participants) suggested greater improvement in physical functioning among participants in the graded exercise therapy than in the supportive listening group (MD -7.55 points, 95% CI -15.57 to 0.47; Analysis 2.7).

2.5 Quality of life

None of the included studies reported quality of life at end of treatment. Very low-certainty evidence looked at quality of life at 52 weeks' follow-up, but due to the very low certainty of the evidence, we are uncertain whether exercise therapy affects quality of life at long-term follow-up. The very low-certainty evidence (Analysis 2.8), was based on 44 participants from one study (Jason 2007), which observed an MD of -6.1 (95% CI -15.9 to 3.7). Jason 2007 measured quality of life on the Quality of Life Scale, consisting of 16 items answered on a scale of 1 to 7 (Burckhardt 2003).

2.6.1 Depression

End of treatment

We did not identify any studies reporting this outcome for the comparison of exercise therapy versus CBT at end of treatment. With regard to other comparisons, Wearden 2010 reported that graded exercise therapy was associated with greater improvement on the HADS depression subscale (0 to 21 points; Zigmond 1983), than supportive listening (MD -1.57, 95% CI -2.74 to -0.40; 1 study, 182 participants; Analysis 2.9).

Follow-up

Moderate-certainty evidence shows that there is probably little or no difference in depression between exercise therapy and CBT at 52 weeks' follow-up (SMD 0.01, 95% CI -0.21 to 0.22, 2 studies, 331 participants; Analysis 2.10). Jason 2007 (44 participants) assessed depression using the Beck Depression Inventory (Beck 1996), whereas White 2011 (287 participants) assessed depression using the HADS depression subscale (Zigmond 1983). We therefore pooled the results using SMD, and the meta-analysis was not associated with heterogeneity ($I^2 = 0%$, $P = 0.42$).

With regard to other comparisons, Jason 2007 also compared anaerobic exercise versus cognitive therapy, and reported a trend towards greater improvement among participants in the cognitive therapy group (MD 5.08, 95% CI -0.77 to 10.93; 45 participants). Wearden 2010 compared graded exercise therapy and supportive listening without finding clear differences between the groups (MD -0.79, 95% CI -2.31 to 0.55; 171 participants).

Subgroup analysis

Post hoc subgroup analysis did not establish a subgroup difference ($I^2 = 0\%$, $P = 0.42$) 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

We did not identify any studies reporting this outcome for the comparison of exercise versus CBT at end of treatment. With regard to other comparisons, [Wearden 2010](#) reported that graded exercise therapy was associated with greater improvement on the HADS anxiety subscale ([Zigmond 1983](#)), than supportive listening (MD -0.48 , 95% CI -1.85 to 0.89 ; 182 participants; [Analysis 2.11](#)).

Follow-up

Moderate-certainty evidence shows that there is probably little or no difference in anxiety between exercise therapy and CBT at 52 weeks' follow-up (SMD 0.07 , 95% CI -0.15 to 0.28 , 2 studies, 331 participants; [Analysis 2.12](#)). [Jason 2007](#) (44 participants) assessed anxiety using the Beck Anxiety Inventory ([Hewitt 1993](#)), whereas [White 2011](#) (287 participants) assessed anxiety using the HADS anxiety subscale ([Zigmond 1983](#)); we therefore pooled the results using SMD. The meta-analysis was not associated with heterogeneity ($I^2 = 0\%$, $P = 0.99$).

With regard to other comparisons, [Jason 2007](#) also compared anaerobic exercise versus cognitive therapy using the Beck Anxiety Inventory ([Hewitt 1993](#)), without detecting differences between the groups (MD 3.15 , 95% CI -1.17 to 7.47 ; 45 participants), but we considered the certainty of the evidence very low. [Wearden 2010](#) compared graded exercise therapy and supportive listening without finding differences (MD -0.08 , 95% CI -1.52 to 1.36 ; 171 participants).

Subgroup analysis

Post hoc subgroup analysis did not establish a subgroup difference ($I^2 = 0\%$, $P = 0.99$) 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

We did not identify any studies reporting this outcome for the comparison of exercise therapy versus CBT at end of treatment. With regard to other comparisons, [Wearden 2010](#) reported greater improvement on the 20-point Jenkins Sleep Scale ([Jenkins 1988](#)), for participants who received exercise therapy than for those who received supportive listening (MD -2.46 points, 95% CI -4.01 to -0.91 ; 180 participants; [Analysis 2.13](#)).

Follow-up

Low-certainty evidence suggests that there may be little or no difference in sleep between graded exercise therapy and CBT (MD -0.90 , 95% CI -2.07 to 0.27 ; 1 study, 287 participants; [White 2011](#); [Analysis 2.14](#)), using the Jenkins Sleep Scale ([Jenkins 1988](#)). With regard to other comparisons, [Wearden 2010](#) 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 ; 1 study, 171 participants; [Analysis 2.14](#)).

2.8 Self-perceived changes in overall health

Two studies ([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

There is low-certainty evidence that there may be little or no difference between exercise therapy and CBT in the number of participants who reported some degree of improvement at end of treatment (RR 0.96 , 95% CI 0.71 to 1.31 ; 1 study, 320 participants; [Analysis 2.15](#)).

Follow-up

There is only very low-certainty evidence available for the comparison of exercise therapy versus CBT at 52 weeks' follow-up. Hence, we are uncertain whether there is a difference in self-reported improvement between exercise therapy and CBT. Two studies were available for this comparison ([Jason 2007](#); [White 2011](#)). The meta-analysis was associated with considerable heterogeneity ($I^2 = 86\%$), and yielded a pooled estimate of RR 0.71 (95% CI 0.33 to 1.54 ; 2 studies, 368 participants; [Analysis 2.16](#)).

For the comparison of cognitive therapy versus anaerobic exercise, [Jason 2007](#) showed that more participants in the cognitive therapy group than in the exercise group tended to report improvement (RR 0.63 , 95% CI 0.36 to 1.10 ; 1 study, 50 participants; [Analysis 2.16](#)).

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 exercise group had a higher mean number of specialist medical care contacts (MD 0.60 , 95% CI 0.05 to 1.15 ; [Analysis 2.17](#)) and higher mean numbers of inpatient days (MD 0.80 , 95% CI 0.41 to 1.19 ; [Analysis 2.17](#)) than participants in the CBT group. However, these group differences were not seen when we analysed data at a dichotomous level ([Analysis 2.18](#)).

2.10 Dropout

There is low-certainty evidence suggesting that dropout rates may be higher for CBT than for exercise therapy (RR 0.59 , 95% CI 0.28 to 1.25 ; 1 study, 321 participants; [Analysis 2.19](#)). Only [White 2011](#) provided data for this comparison.

[Wearden 2010](#) reported that more participants discontinued graded exercise therapy than supportive listening, with 12 of 92 participants dropping out of graded exercise therapy and 7 of 91 participants dropping out of supportive listening (RR 1.70 , 95% CI 0.70 to 4.11 ; [Analysis 2.19](#)).

Comparison 3. Exercise therapy versus adaptive pacing therapy

Only one study with 319 participants contributed data for this comparison ([White 2011](#)).

3.1 Fatigue

There is low-certainty evidence showing that exercise therapy may be slightly more effective than adaptive pacing in reducing fatigue at end of treatment (MD -2.00, 95% CI -3.57 to -0.43; 1 study, 305 participants) and at follow-up after 52 weeks (MD -2.50, 95% CI -4.16 to -0.84; 1 study, 307 participants). The only available study (White 2011), assessed fatigue by a 33-point Fatigue Scale (Chalder 1993), as shown in Analysis 3.1.

3.2 Adverse effects

One study (White 2011), reported the rate of serious adverse reactions between exercise therapy and adaptive pacing (RR 0.99, 95% CI 0.14 to 6.97; 1 study, 319 participants; Analysis 3.2). White 2011 defined serious adverse reactions according to European Union Clinical Trials Directive 2001. They observed two serious adverse reactions (deterioration in mobility and self-care, and worse CFS symptoms and function) among 160 participants in the exercise group, and two (one incidence of suicidal thoughts, and one episode of worsened depression) among the 159 participants in the adaptive-pacing group. The confidence interval remains wide due to few events in all intervention groups, and therefore the effect of exercise therapy on serious adverse reactions remains uncertain (very low-certainty evidence).

3.3 Pain

We did not find any studies that investigated pain as an outcome.

3.4 Physical functioning

Low-certainty evidence suggests that exercise therapy may be more effective in improving physical functioning than adaptive pacing at end of treatment (MD -12.20, 95% CI -17.23 to -7.17, 305 participants; Analysis 3.3) and at follow-up after 52 weeks (MD -11.80, 95% CI -17.55 to -6.05; 307 participants; Analysis 3.3). All results were based on one study (White 2011), that measured physical functioning using the SF-36 physical functioning subscale (Ware 1992).

3.5 Quality of life

No data were reported for this outcome.

3.6.1 Depression

No data were reported for this outcome at end of treatment. Low-certainty evidence from one study (White 2011), suggests that exercise therapy may be slightly more effective in reducing depression at 52 weeks' follow-up (MD -1.10, 95% CI -2.09 to -0.11; 1 study, 293 participants; Analysis 3.4). White 2011 assessed depression using the HADS depression subscale (0 to 21 points; Zigmond 1983).

3.6.2 Anxiety

No data were reported for this outcome at end of treatment. Low-certainty evidence from one study (White 2011), suggests that there may be little or no difference in anxiety between exercise therapy and adaptive pacing at 52 weeks' follow-up (MD -0.40, 95% CI -1.40 to 0.60; 1 study, 293 participants; Analysis 3.5). White 2011 assessed anxiety using the HADS anxiety subscale (Zigmond 1983).

3.7 Sleep

No data were reported for this outcome at end of treatment. Low-certainty evidence from one study (White 2011), suggests that

exercise therapy may be slightly more effective in improving sleep at 52 weeks' follow-up (MD -1.60, 95% CI -2.70 to -0.50; 1 study, 294 participants; Analysis 3.6). White 2011 assessed depression using the 20-point Jenkins Sleep Scale (Jenkins 1988).

3.8 Self-perceived changes in overall health

Low-certainty evidence from one study (White 2011), indicates that more participants may report improvement following exercise therapy than following adaptive pacing at end of treatment (RR 1.45, 95% CI 1.02 to 2.07; 1 study, 319 participants; Analysis 3.7). 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).

Only very low-certainty evidence was available at 52 weeks' follow-up, and we are therefore uncertain whether exercise therapy or adaptive pacing have an impact on self-perceived changes in overall health. Briefly, only one study (White 2011), compared the rate of participants who reported some degree of improvement in the two groups (RR 1.31, 95% CI 0.96 to 1.79; 1 study, 319 participants; Analysis 3.7).

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 exercise group had a higher mean number of contacts with complementary healthcare resources (MD 3.80, 95% CI 1.42 to 6.18; Analysis 3.8), higher mean numbers of specialised medical care contacts (MD 0.70, 95% CI 0.14 to 1.26; Analysis 3.8), higher mean numbers of accidents and emergencies (MD 0.50, 95% CI 0.31 to 0.69; Analysis 3.8), and lower mean numbers of inpatient days (MD -1.00, 95% CI -1.54 to -0.46; Analysis 3.8) than participants in the pacing group. However, we did not see these group differences when we analysed data at a dichotomous level (Analysis 3.9).

3.10 Dropout

Only very low-certainty evidence was available for this outcome, implying that we are uncertain whether exercise therapy or adaptive pacing have an impact on self-perceived changes in overall health. The only study reporting on this outcome (White 2011), showed that 10 of the 160 participants in the graded exercise therapy group and 11 of the 160 participants in the adaptive pacing group withdrew (RR 0.91, 95% CI 0.40 to 2.08; Analysis 3.10).

Comparison 4. Exercise therapy versus antidepressants

Only one study (Wearden 1998), contributed data for this comparison, with a total of 69 participants. In this study, investigators compared graded exercise therapy with placebo (n = 34) versus the antidepressant fluoxetine with exercise placebo (n = 35).

4.1 Fatigue

Only very low-certainty evidence was available for this outcome, so we are uncertain whether there are differences in how exercise therapy and antidepressants affect fatigue. In the only available study, with 48 participants, investigators assessed fatigue on a 42-point Fatigue Scale (Chalder 1993), at end of treatment (MD -1.99, 95% CI -8.28 to 4.30; Wearden 1998; Analysis 4.1).

4.2 Adverse effects

No data were reported for this outcome.

4.3 Pain

No data were reported for this outcome.

4.4 Physical functioning

No data were reported for this outcome.

4.5 Quality of life

No data were reported for this outcome.

4.6.1 Depression

Only very low-certainty evidence was available for this outcome, so we are uncertain whether there are differences in how exercise therapy and antidepressants affect depression. [Wearden 1998](#) used the HADS depression subscale (0 to 21 points; [Zigmond 1983](#)), to assess depression among 48 participants at end of treatment (MD 0.15, 95% CI -2.11 to 2.41; [Analysis 4.2](#)).

4.6.2 Anxiety

No data were reported for this outcome.

4.7 Sleep

No data were reported for this outcome.

4.8 Self-perceived changes in overall health

No data were reported for this outcome.

4.9 Health service resources

No data were reported for this outcome.

4.10 Dropout

Only very low-certainty evidence was available for this outcome, so we are uncertain whether there are differences in how exercise therapy and antidepressants affect dropout rate. In the only available study, [Wearden 1998](#) reported similar dropout rates in both groups, with 11 dropouts reported among the 34 participants in the exercise group and 10 dropouts among the 35 participants in the antidepressant group (RR 1.13, 95% CI 0.55 to 2.31; [Analysis 4.3](#)).

Exercise therapy adjunctive to other treatment versus the other treatment alone

Comparison 5. Exercise therapy plus antidepressants versus antidepressants alone

One study contributed data to this comparison ([Wearden 1998](#)). In this study, investigators compared graded exercise therapy used alongside the antidepressant fluoxetine (n = 33) versus graded exercise therapy used alongside an antidepressant placebo (n = 35).

5.1 Fatigue

Only very low-certainty evidence was available for this outcome, so we are uncertain whether there are differences in how exercise therapy or exercise plus antidepressants affects fatigue. In the only available study, researchers assessed fatigue on a 42-point Fatigue Scale ([Chalder 1993](#)), at end of treatment, but the results were inconclusive (MD -3.66, 95% CI -10.41 to 3.09; 1 study, 43 participants; [Analysis 5.1](#)).

5.2 Adverse effects

No data were reported for this outcome.

5.3 Pain

No data were reported for this outcome.

5.4 Physical functioning

No data were reported for this outcome.

5.5 Quality of life

No data were reported for this outcome.

5.6.1 Depression

Only very low-certainty evidence was available for this outcome, so we are uncertain whether there are differences in how exercise therapy or exercise plus antidepressants affect depression. [Wearden 1998](#) used the HADS depression subscale (0 to 21 points; [Zigmond 1983](#)), to assess depression at end of treatment (MD -0.27, 95% CI -2.68 to 2.14; 1 study, 43 participants; [Analysis 5.2](#)).

5.6.2 Anxiety

No data were reported for this outcome.

5.7 Sleep

No data were reported for this outcome.

5.8 Self-perceived changes in overall health

No data were reported for this outcome.

5.9 Health service resources

No data were reported for this outcome.

5.10 Dropout

Only very low-certainty evidence was available for this outcome, so we are uncertain whether there are differences in how exercise therapy or exercise plus antidepressants affect dropout rates. In the only available study, [Wearden 1998](#) observed similar dropout rates in both groups (RR 1.48, 95% CI 0.77 to 2.87; [Analysis 5.3](#)).

DISCUSSION

Summary of main results

Eight studies, with a total of 1518 participants, satisfied inclusion criteria and are included in this review. Investigators compared exercise therapy with 'passive' control in all eight studies, and the results show that exercise therapy probably reduces fatigue at end of treatment ([Analysis 1.1](#); [Summary of findings 1](#)), and that the effect at long-term follow-up is uncertain ([Analysis 1.2](#); [Summary of findings 1](#)). The impact of exercise therapy on serious adverse reactions is uncertain (RR 0.99, 95% CI 0.14 to 6.97; 1 study, 319 participants, very low-certainty evidence; [Analysis 1.3](#); [Summary of findings 1](#)). None of the studies looked at pain at end of treatment, and the long-term effect is uncertain because the certainty of this evidence is very low ([Analysis 1.4](#); [Summary of findings 1](#)). Exercise therapy may moderately improve physical functioning at end of treatment ([Analysis 1.5](#); [Summary of findings 1](#)), whereas the long-term effect is uncertain ([Analysis 1.6](#); [Summary of findings 1](#)). None of the studies looked at quality of life at end of treatment, and the long-term effect is uncertain ([Analysis 1.7](#); [Summary of findings 1](#)).

The effect of exercise therapy on depression is uncertain at end of treatment and at follow-up ([Analysis 1.8](#); [Analysis 1.9](#); [Summary of findings 1](#)). Exercise therapy may slightly improve sleep at end of treatment ([Analysis 1.12](#); [Summary of findings 1](#)) and after 52 to 70 weeks ([Analysis 1.13](#); [Summary of findings 1](#)).

Two studies with 351 participants compared exercise therapy with CBT, suggesting little or no difference in fatigue at end of treatment ([Analysis 2.1](#); [Summary of findings 2](#)) and after 52 weeks ([Analysis 2.2](#); [Summary of findings 2](#)). The impact of exercise therapy on serious adverse reactions is uncertain ([Analysis 2.3](#); [Summary of findings 2](#)). For secondary outcomes there may be little or no difference between exercise therapy and CBT in physical functioning, depression and sleep (low-certainty evidence; [Summary of findings 2](#)). The effect of exercise therapy compared to CBT in quality of life or pain is uncertain (very low-certainty evidence; [Summary of findings 2](#)).

Investigators compared exercise therapy versus adaptive pacing in one study with 305 participants, reporting that exercise therapy may slightly reduce fatigue at end of treatment (low-certainty evidence; [Analysis 3.1](#); [Summary of findings 3](#)) and after 52 weeks (low-certainty evidence). The impact of exercise therapy on serious adverse reactions is uncertain (RR 0.99, 95% CI 0.14 to 6.97; 1 study, 319 participants, low-certainty evidence; [Analysis 3.2](#); [Summary of findings 3](#)). Regarding secondary outcomes, the available evidence suggests exercise may slightly improve physical functioning, depression and sleep compared to adaptive pacing at end of treatment and after 52 weeks (low-certainty evidence; [Summary of findings 3](#)). No studies looked at quality of life or pain.

Comparisons of exercise therapy with or without antidepressants versus antidepressants alone were only reported in one small study. The evidence was rated to very-low quality, implying that the available results are very uncertain.

Overall completeness and applicability of evidence

This evidence was collected from outpatients diagnosed with 1994 CDC criteria or the Oxford criteria. Our comparison of the two studies using 1994 CDC criteria ([Moss-Morris 2005](#); [Wallman 2004](#)), versus the five studies using the Oxford criteria ([Fulcher 1997](#); [Powell 2001](#); [Wearden 1998](#); [Wearden 2010](#); [White 2011](#)), did not reveal any subgroup differences ($I^2 = 0\%$, $P = 0.76$ (SMD -0.73 , 95% CI -1.17 to -0.28 versus SMD -0.63 , 95% CI -1.07 to -0.19)), but participants diagnosed using other criteria may experience different effects. All studies were conducted in high-income countries (Australia, New Zealand, USA and the UK), and the evidence base was limited to participants able to participate in exercise therapy as it was offered. Settings varied from primary to tertiary care, which suggests easy generalisation. Most studies used aerobic exercise, but it would be preferable if we had found studies that offered different types of exercise therapy, as this would reflect clinical practice.

Quality of the evidence

Formal blinding of participants and clinicians is not inherently possible in studies of exercise therapy, due to the nature of the intervention. This increases risk of performance and detection bias in particular because outcomes were measured subjectively (e.g. questionnaires, visual analogue scales). However, many groups representing the interests of those with CFS are opposed to exercise

therapy, and this may in contrast reduce the outcome estimate. Six of the seven studies reported that investigators used intention-to-treat analysis, but this was done in different ways and may have influenced the effect estimate. One study ([Jason 2007](#)), reported large baseline differences across groups, 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. We observed a large between-study variation with regard to type of exercise, intensity of exercise and incremental procedures used ([Table 2](#)). We acknowledge that the effect of exercise therapy is likely to depend on how training is conducted, and that inclusion of studies using different exercise regimens is likely to introduce some heterogeneity. Further, the treatment provided to participants in the control group was also not uniform across the included studies. Whereas the difference between waiting list, relaxation and treatment as usual may seem obvious, it is important to recognise that the actual ingredients of 'treatment as usual' differed widely among the included studies. This may have contributed to variation in the reported effect estimates. Regarding participants and their health status, we noted substantial differences in baseline illness severity, as illustrated by the wide range in baseline physical functioning, depression, co-morbidity and illness duration ([Table 1](#)). Some studies applied narrow participant eligibility criteria, whereas others included more heterogeneous samples, and these differences may have caused variation in the reported effect estimate. Our finding of similar outcomes with different definitions of CFS mitigates this risk.

All the potential sources of heterogeneity mentioned above could have contributed to variation in results derived from the aggregate analysis presented in this 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 studies is low; we were therefore 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 eligible studies, such as studies 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 we have used data from the 2004 review ([Edmonds 2004](#)), and from published versions of included articles.

The authors of this review had to decide 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 continue.

One of the included studies ([Powell 2001](#)), is an outlier as it reports very positive results in favour of exercise therapy. We have reviewed the study thoroughly and discussed it with clinical experts, but we have not identified good reasons to exclude it. Nevertheless, we decided to perform post hoc sensitivity analyses to explore how [Powell 2001](#) affects the overall estimates. The inclusion of [Powell 2001](#) in meta-analysis was rarely associated with large distortions of the overall pooled estimate, and the most important impact of [Powell 2001](#) was the introduction of extensive heterogeneity into many meta-analyses ([Summary of findings 1](#)).

The review authors noted potential bias regarding how the comparators in this review were categorised and pooled. We decided to report diverse comparators such as CBT, cognitive therapy and supportive therapy together as a single comparator called 'psychological treatments', although, because of clinical and contextual heterogeneity, we decided not to pool the results in meta-analyses. These different psychological treatments do have elements in common. For example, both CBT and cognitive therapy use cognitive approaches and goal setting, but they differ in certain respects. CBT aims to change unhelpful thoughts, whereas cognitive therapy, as described and implemented by [Jason 2007](#), aims to accept them.

Meta-analysis of individual patient data (IPD) constitutes an alternative approach to meta-analysis of aggregate data. Analysis based on IPD in general will enable us to use a wider range of statistical and analytical approaches ([Stewart 2011](#)). By utilising IPD, it is possible to explore the relative importance of the various heterogeneity factors mentioned above, and to ensure that missing data and baseline differences are dealt with in standardised ways. IPD also allows the possibility of performing subgroup analyses that have not been previously undertaken. A project aimed at undertaking IPD analyses of the studies included in this review has started, and should shed new light on the aggregate level analyses presented here.

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. In line with recent updates provided in the *Cochrane Handbook for Systematic Reviews of Interventions* ([Higgins 2011c](#)), we have implemented several methodological improvements, including a thorough risk of bias assessment for all included studies ([Higgins 2011a](#)). Also, the updated search for literature led to the inclusion of three new studies 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 increase in study numbers and total study participants has important implications. First, statistical power has been increased by the inclusion of new data. Second, the most recent studies offered longer follow-up times, and 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 studies 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.

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 graded exercise therapy ([Castell 2011](#)), suggests that the two treatments are equally efficacious, especially for people with comorbid anxiety or depressive symptoms.

A randomised study 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 study did not meet our a priori inclusion criteria and we excluded it 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. Two RCTs identified as ongoing in our search in May 2014 have been published and report positive effects of physical activity. One study, with 91 participants, compared a self-regulation-based physical activity programme with standard medical care, and found that the programme had, "... a significant effect on fatigue, fatigue severity, leisure time physical activity, personal activity goal progress and health related quality of life. No significant effect was found on daily number of steps and somatic and psychological distress" ([Marques 2015](#)). The other RCT, with 211 participants, compared guided graded exercise self-help plus specialist medical care versus specialist medical care alone, and found significant improvements in fatigue and physical function, and did not record any serious adverse reactions ([Clarke 2017](#)). The conclusions presented in our review correspond well with those of other relevant studies and reviews.

AUTHORS' CONCLUSIONS

Implications for practice

Low- to moderate-certainty evidence suggests that exercise therapy may contribute to alleviation of some of the symptoms of chronic fatigue syndrome (CFS), especially fatigue. Long-term effects are in general more uncertain than short-term effects mainly because studies did not always have long-term follow-up. The impact of exercise therapy on serious adverse reactions is uncertain. Due to few studies with a small number of participants it is difficult to draw conclusions about the comparative effectiveness of cognitive behavioural therapy (CBT), adaptive pacing or other interventions. This evidence is collected from outpatients diagnosed with 1994 criteria of the Centers for Disease Control and Prevention (CDC) or the Oxford criteria, or both, and people diagnosed using other criteria may experience different effects.

Implications for research

Further randomised controlled trials are needed to clarify the most effective type, intensity and duration of exercise therapy. These studies should carefully report the characteristics of the

exercise therapy provided, and meet the requirements of the TIDieR checklist (Hoffmann 2014). It is important that these studies measure health service resource use, alongside the primary outcomes of fatigue and adverse effects, and other relevant secondary outcomes. Researchers should take care to describe how they operationalised the diagnostic process. Further work to identify which subgroups of patients that are most likely to benefit from treatment would be valuable.

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 studies in which they were involved, as well as the Cochrane Common Mental Disorders' 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. In: 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-Morris 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 RO, Nye FJ, Edwards RH. Patient education to encourage graded exercise in chronic fatigue syndrome: 2-year follow-up of randomised controlled trial. *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 K. Confirmation of ages (means and SDs) of groups in trial [personal communication]. Email to: L Larun 2 November 2009.

* 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.

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.

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). In: 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 AJ. Raw data to facilitate calculations for meta-analysis [personal communication. Email to: L Larun March 2009.

Wearden 2010 {published and unpublished data}74156610

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, 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](https://doi.org/10.1136/bmj.c1777)]

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.

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. www.controlled-trials.com/ISRCTN74156610/ISRCTN74156610 (accessed 2 September 2014).

White 2011 {published data only}54285094

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](https://doi.org/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;**2**:1067-74. [DOI: [10.1016/S2215-0366\(15\)00317-X](https://doi.org/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. *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;**72**(6):509.

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. *Lancet* 2011;**377**:832-6. [DOI: [10.1016/S0140-6736\(11\)60096-2](https://doi.org/10.1016/S0140-6736(11)60096-2)]

* 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. *Lancet* 2011;**377**:611-90.

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](https://doi.org/10.1186/1471-2377-7-6)]

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. www.controlled-trials.com/ISRCTN54285094 (accessed 1 September 2014).

References to studies excluded from this review

Broadbent 2012 {unpublished data only}

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).

Broadbent S, Coutts R. The protocol for a randomised controlled trial comparing intermittent and graded exercise to usual care for chronic fatigue syndrome patients. *BMC Sports Science, Medicine & Rehabilitation* 2013;**5**(1):1-6.

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. Optimization of cognitive behavioral therapy (CBT) for CFS patients in rehabilitation by means of ambulatory activity-based feedback (ABF). trialregister.nl/trialreg/admin/rctview.asp?TC=1513 (accessed 7 May 2013).

Gordon 2010 {published data only}

* Gordon BA, Knapman LM, Lubitz L. Graded 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](https://doi.org/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.

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* 1998:0.

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).

Liu 2010 {published data only}

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

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}

* Ridsdale 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} [10.1017/S0033291712000256](https://doi.org/10.1017/S0033291712000256)

* 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](https://doi.org/10.1017/S0033291712000256)]

Sabes-Figuera R, McCrone P, Hurley M, King M, Donaldson AN, Ridsdale 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 RJ, Atkinson CM, Champion PD. Group rehabilitation for adult chronic fatigue syndrome. Unpublished article (received 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} **0.1017/S0033291712000232**

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](https://doi.org/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](https://doi.org/10.1136/adc.2003.035154)]

Vos-Vromans 2008 {unpublished data only} **77567702**

* 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]. www.controlled-trials.com/isrctn/pf/77567702 (accessed 7 May 2013). [ISRCTN77567702]

Vos-Vromans DC, Smeets RJ, Rijnders LJ, Gorrissen RR, Pont M, Köke AJ, et al. Cognitive behavioural therapy versus multidisciplinary rehabilitation treatment for patients with chronic fatigue syndrome: study protocol for a randomized controlled trial (FatiGo). *Trials* 2012;**13**:71.

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](https://doi.org/10.1136/adc.2003.046649)]

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 studies awaiting assessment

Marques 2012 {unpublished data only} **70763996**

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](https://doi.org/10.1186/1471-2458-12-202)]

White 2012 {published data only} **22975026**

White PD. Therapy guided self-help treatment (GETSET) for patients with chronic fatigue syndrome/myalgic encephalomyelitis: a randomised controlled trial in secondary care. 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.

Alderson 2004

Alderson P, Green S, Higgins JP, editors. Cochrane Reviewers' Handbook 4.2.2 [updated March 2004]. In: The Cochrane Library, Issue 1, 2004. 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* 2002;**11**(3):284-8.

Beck 1996

Beck AT, Steer RA, Brown GK. 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.

Buyse 1989

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

Carruthers 2011

Carruthers BM, Van 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.

Clarke 2017

Clarke LV, Pesola F, Thomas JM, Vergara-Williamson M, Beynon M, White PG. Guided graded exercise self-help plus specialist medical care versus specialist medical care alone for chronic fatigue syndrome (GETSET): a pragmatic randomised controlled trial. *Lancet* 2017;**390**(10092):363-73. [DOI: [10.1016/S0140-6736\(16\)32589-2](https://doi.org/10.1016/S0140-6736(16)32589-2)]

Cleeland 1994

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

Crawley 2013

Crawley E, Collin SM, White PD, Rime K, Sterne JA, May MT, and CFS/ME National Outcomes Database. Treatment outcome in adults with chronic fatigue syndrome: a prospective study in England based on the CFS/ME National Outcomes Database. *Quarterly Journal of Medicine* 2013;**106**:555-65.

Deeks 2011

Deeks JJ, Higgins JP, Altman DG (editors). Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

Egger 1997

Egger M, Davey-Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 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 European Parliament and the Council of the European Union. Directive 2001/20/EC European Parliament and the Council of the European Union of 4 April 2001. Official Journal of the European Communities 2001;**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.

Goligher 2008

Goligher EC, Pouchot J, Brant R, Kherani RB, Avina-Zubieta JA, Lacaillie D, et al. Minimal clinically important difference for 7 measures of fatigue in patients with systemic lupus erythematosus. *Journal of Rheumatology* 2008;**35**:635-42.

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*. Rockville, MD: National Institute of Mental Health, 1976.

Hewitt 1993

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

Higgins 2003

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

Higgins 2011a

Higgins JP, Altman DG, Sterne JA (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March

2011). The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

Higgins 2011b

Higgins JP, Deeks JJ (editors). Chapter 7: Selecting studies and collecting data. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

Higgins 2011c

Higgins JP, Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

Higgins 2011d

Higgins JP, Deeks JJ, Altman DG (editors). Chapter 16: Special topics in statistics. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

Hoffmann 2014

Hoffmann T, Glasziou P, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ* 2014;**348**:1687.

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 utmattelsessyndrom]. *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. Art. No: CD011040. [DOI: [10.1002/14651858.CD011040](https://doi.org/10.1002/14651858.CD011040)]

Marques 2015

Marques M, De Gucht V, Leal I, Maes S. Effects of a self-regulation based physical activity program (the "4-STEPS") for

unexplained chronic fatigue: a randomized controlled trial. *International Journal of Behavioral Medicine* 2015;**2**:187-96. [DOI: [10.1007/s12529-014-9432-4](https://doi.org/10.1007/s12529-014-9432-4)]

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. guidance.nice.org.uk/CG53/guidance/pdf/English (last accessed November 2009).

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](https://doi.org/10.1111/j.1365-2362.2011.02575.x)]

Oxford English Dictionary

OED Online December 2014 Oxford University Press. "therapy, n.". 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. Art. No: CD001027. [DOI: [10.1002/14651858.CD001027.pub2](https://doi.org/10.1002/14651858.CD001027.pub2)]

Prins 2006

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

Puhan 2008

Puhan MA, Frey M, Büchi S, Schünemann HJ. The minimal important difference of the hospital anxiety and depression scale in patients with chronic obstructive pulmonary disease. *Health and Quality of Life Outcomes* 2008;**6**(43):1-6. [DOI: [10.1186/1477-7525-6-46](https://doi.org/10.1186/1477-7525-6-46)]

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.

Review Manager 2014 [Computer program]

Review Manager 5 (RevMan 5). Version 5.3. Copenhagen: Nordic Cochrane Centre, The Cochrane Collaboration, 2014.

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.

Schünemann 2011

Schünemann HJ, Oxman AD, Vist GE, Higgins JP, Deeks JJ, Glasziou P, et al. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JPT, Green S (editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

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.

Smid 2017

Smid DE, Franssen FM, Houben-Wilke S, Vanfleteren LE, Janssen DJ, Wouters EF, et al. Responsiveness and MCID estimates for CAT, CCQ, and HADS in patients with COPD undergoing pulmonary rehabilitation: a prospective analysis. *Journal of the American Medical Directors Association* 2017;**18**:53-8.

Stewart 2011

Stewart LA, Tierney JF, Clarke M. Chapter 19: Reviews of individual patient data. In: Higgins JPT, Green S

(editors), *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from www.handbook.cochrane.org.

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). Bristol: Westcare 1994.

Ward 2014

Ward MM, Guthrie LC, Alba MI. Clinically important changes in short form 36 health survey scales for use in rheumatoid arthritis clinical trials: the impact of low responsiveness. *Arthritis Care & Research* 2014;**66**:1783-9.

Ware 1992

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

Wyrwich 2007

Wyrwich KW, Metz SM, Kroenke K, Tierney WM, Babu AN, Wolinsky FD. Triangulating patient and clinician perspectives on clinically important differences in health-related quality of life among patients with heart disease. *Health Services Research* 2007;**42**:2257-74.

Zigmond 1983

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

References to other published versions of this review

Edmonds 2004

Edmonds M, McGuire H, Price J. Exercise therapy for chronic fatigue syndrome. *Cochrane Database of Systematic Reviews* 2004, Issue 3. Art. No: CD003200. [DOI: [10.1002/14651858.CD003200.pub2](https://doi.org/10.1002/14651858.CD003200.pub2)]

* Indicates the major publication for the study

CHARACTERISTICS OF STUDIES

Characteristics of included studies [ordered by study ID]

Fulcher 1997

Study characteristics

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

Exercise therapy for chronic fatigue syndrome (Review)

Fulcher 1997 (Continued)

Earlier treatment: NS

Co-morbidity: 20 (30%) possible cases of depression (HADS). 30 (45%) on full-dose AD (n = 20) or low-dose tricyclic ADs as hypnotics (n = 10)

Average illness duration: 2.7 (0.6-19) years

Work and employment status: 26 (395) working or studying at least part time

Setting: secondary care (CF clinic in a general hospital of psychiatry)

Country: UK

Interventions	Group 1: ET (12 sessions) with 1 weekly supervised session and 5 home sessions a week, initially lasting between 5 and 15 min (n = 33) Group 2: flexibility and relaxation (12 sessions) with 5 home sessions prescribed per week (n = 33)
Outcomes	<ul style="list-style-type: none"> • Changes in overall health (Global Impression Scale, score between 1 and 7, where 1 = very much better, 4 = no change) • Anxiety and depression (HADS) • Fatigue (FS; 14-item questionnaire) • Sleep (PSQI) • Physical functioning (SF-36) • Physiological assessments (maximal voluntary contraction of quadriceps, peak oxygen consumption, lactate, HR) • Perceived exertion (Borg Scale) <p>Outcomes were assessed at end of treatment (12 weeks)</p>
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

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 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"

Fulcher 1997 (Continued)

Selective reporting (reporting bias)	Unclear risk	All primary outcomes stated under Methods were reported; however, as the study 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
Study characteristics

Methods	RCT, 4 parallel arms
Participants	<p>Diagnostic criteria: CDC 1994</p> <p>Number of participants: n = 114</p> <p>Gender: 95 (83.3%) female</p> <p>Age: 43.8 years</p> <p>Earlier treatment: NS</p> <p>Co-morbidity: 44 (39%) with a current Axis I disorder (depression and anxiety most common). Use of AD not stated</p> <p>Illness duration: > 5 years</p> <p>Work and employment status: 52 (46%) working or studying at least part time, 24% unemployed, 6% retired, 25% on disability</p> <p>Setting: secondary care, but recruitment from different sources</p> <p>Country: USA</p>
Interventions	<p>13 sessions every 2 weeks lasting 45 min</p> <p>Group 1: CBT aimed at showing participants that activity could be done without exacerbating symptoms (n = 29)</p> <p>Group 2: ACT focused on developing individualised and pleasurable activities accompanied by reinforcement of progress (n = 29)</p> <p>Group 3: COG focused on developing strategies to better tolerance, reduce stress and symptoms and lessen self-criticism (n = 28)</p> <p>Group 4: relaxation treatment, introducing several types of relaxation techniques along with expectations of skill practice (n = 28)</p>
Outcomes	<p>Several outcomes are reported (~25), among others</p> <ul style="list-style-type: none"> • Physical functioning (SF-36) • Fatigue (FSS) • Depression (BDI-II) • Anxiety (BAI) • Self-efficacy (self-efficacy questionnaire) • Stress (PSS) • Pain (BPI) • QoL (QOLS) • 6-MWT

Jason 2007 (Continued)

- Changes in overall health (Clinical Global Impression - Improvement Scale (CGI-I), score between 1 and 7, where 1 = very much better, 4 = no change)

Outcomes assessed at 12 months' follow-up

Notes Fidelity ratings and dropout reported across study arms

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	NS
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 study 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 relaxation group 53.77 (26.66))

Moss-Morris 2005
Study characteristics

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 AD NS

Exercise therapy for chronic fatigue syndrome (Review)

Moss-Morris 2005 (Continued)

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: GET (12 weeks), met weekly, final goal 30 min for 5 days/week, 70% of VO ₂ max (n = 25) Group 2: standard medical care provided by a CFS specialist physician (n = 24)
Outcomes	<ul style="list-style-type: none"> • 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 functioning subscale score) • Fatigue (FS) • Activity levels • Cognitive function • Physiological assessments (e.g. maximum aerobic capacity, HR) • Acceptability <p>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</p>
Notes	The exact components involved in 'treatment as usual' are not explained

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"
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 dropout: 1 had to return to the USA, 1 had an injured calf and 1 was not reached at follow-up. 3 of 24 participants (12.5%) in control group did not return follow-up questionnaire at 12 weeks. To determine whether dropout affected the calculated treatment effect, study authors completed ITT analysis
Selective reporting (reporting bias)	Unclear risk	All primary outcomes stated under Methods were reported; however, as the study 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
Study characteristics

Methods	RCT, 4 parallel arms
Participants	<p>Diagnostic criteria: Oxford</p> <p>Number of participants: n = 148</p> <p>Gender: 116 (78%) female</p> <p>Age, mean: 33 years</p> <p>Earlier treatment: NS</p> <p>Co-morbidity: 58 (39%) possible cases of depression (HADS), 27 (18%) used ADs</p> <p>Illness duration: 4.3 years</p> <p>Work and employment status: 50 (34%) working, 64 (43%) on disability</p> <p>Setting: secondary/tertiary care</p> <p>Country: UK</p>
Interventions	<p>Group 1: treatment as usual (n = 34)</p> <p>Group 2: ET + 2 sessions (total 3 h, n = 37)</p> <p>Group 3: ET + 7 telephone sessions (total 3.5 h, n = 39)</p> <p>Group 4: ET + 7 sessions (total 7 h, n = 38)</p> <p>Sessions, whether telephone or face-to-face, were used to reiterate the treatment rationale and to discuss problems associated with GET</p>
Outcomes	<ul style="list-style-type: none"> 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, 10-30) Fatigue (FS; 11 items; scores > 3 indicate excessive fatigue) Anxiety and depression, HADS; score range from 0-21 worst) Sleep (Jenkins Sleep Scale, 4 items; lower scores indicate better outcomes; score range 0-20 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) <p>Outcomes assessed at 3 (end treatment), 6 and 12 months</p>
Notes	<p>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.</p> <p>SF-36 physical functioning subscale is reported on a 10-30 scale. We transformed scores from the 10-30 scale to the more common 0-100 scale by using the following formula: $mean_{new} = (mean_{old} - 10) * 5$ and $SD_{new} = 5 * SD_{old}$</p>

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"

Powell 2001 (Continued)

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 study 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

Wallman 2004
Study characteristics

Methods	RCT, 2 parallel arms
Participants	<p>Diagnostic criteria: CDC 1994</p> <p>Number of participants: n = 61</p> <p>Gender: 47 (77%) female</p> <p>Age: 16-74 years (average 43.3 (12.7) in the exercise group and 45.7 (12.5) in the control group)</p> <p>Earlier treatment: NS</p> <p>Co-morbidity: HADS depression score at baseline was 6.8 points, 16 (26%) used ADs</p> <p>Illness duration: no detectable initial difference between groups</p> <p>Work and employment status: NS</p> <p>Setting: primary care</p> <p>Country: Australia</p>
Interventions	<p>Group 1: prescribed ET, 12 weeks (n = 32)</p> <p>Group 2: flexibility and relaxation, 12 weeks (n = 29)</p>
Outcomes	<ul style="list-style-type: none"> Physiological assessments (HR, blood pressure at rest and during exercise, lactate and oxygen consumption) Perceived exertion (Borg Scale, RPE) Energy expenditure (Older Adult Exercise Status Inventory) Fatigue (FS; 11 items) Anxiety and depression (HADS)

Wallman 2004 (Continued)

- 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	We obtained supplementary HADS data from study authors for first version of this review
-------	---

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
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 from participating in the exercise test"
Selective reporting (reporting bias)	Unclear risk	All primary outcomes stated under Methods were reported; however, as the study 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
Study characteristics

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 AD NS

Wearden 1998 (Continued)

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: GET + fluoxetine (n = 33) Group 2: GET + drug placebo, 26 weeks, preferred aerobic exercise 20 min \geq 3 times/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 they felt they needed to (n = 34)
Outcomes	<ul style="list-style-type: none"> Fatigue (FS; 14 items; \geq 4 were used as cutoff to designate caseness) General health status (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 (\leq 50), social function (\leq 40), pain (\leq 50), health perception (\leq 70) and mental health (\leq 67) Anxiety or depression, HADS; cutoff of \geq 11 designated cases) Psychiatric diagnoses (Clinical Interview Schedule + supplementary questions by psychologist) Physiological assessments (grip strength and functional work capacity) <p>Outcomes assessed at weeks 12 and 26 (end of treatment)</p>
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. We obtained supplementary HADS data from study authors for the first version of this review

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 dropout rates in all intervention groups

Wearden 1998 (Continued)

Selective reporting (reporting bias)	High risk	<p>It is clear (p 488) that investigators collected data for all 6 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 of the MOS, with the exception of health perceptions, which is significant and favours the intervention group.</p> <p>NB: Data for forced work capacity were collected by investigators but are not reported in this review</p>
Other bias	Low risk	We do not suspect other bias

Wearden 2010
Study characteristics

Methods	RCT, 3 parallel arms
Participants	<p>Diagnostic criteria: Oxford (31% fulfilled London ME criteria)</p> <p>Number of participants: n = 296</p> <p>Gender: 230 (78%) female</p> <p>Age, mean (SD): 44.6 (11.4) years</p> <p>Earlier treatment: 264 (89%) reported medication during the past 6 months with AD (n = 160) or analgesic (n = 79)</p> <p>Co-morbidity: 53 (18%) had a depression diagnosis, 160 (54%) were prescribed ADs the last 6 months</p> <p>Illness duration (mean): 7 (range from 0.5-51.7) years</p> <p>Work and employment status: NS</p> <p>Setting: primary care</p> <p>Country: UK</p>
Interventions	<p>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)</p> <p>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 participants can freely discuss their prioritised concerns (n = 101)</p> <p>Group 3: GP 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)</p>
Outcomes	<ul style="list-style-type: none"> Physical functioning (SF-36 physical functioning subscale, percentage score in which higher scores indicate better outcomes) Fatigue (FS; 11 items; each item was scored dichotomously on a 4-point scale (0, 0, 1 or 1)); total scores of ≥ 4 designated significant levels of fatigue. Lower scores indicated better outcomes) Anxiety and depression (HADS, depression and anxiety scale; lower scores indicate better outcomes) Sleep (Jenkins Sleep Scale; 4 items; lower scores indicate better outcomes)

Wearden 2010 (Continued)

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

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	<p>Number of dropouts (did not complete treatment): 18/95 (group 1), 17/101 (group 2). Reasons for dropout: 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)</p> <p>Loss to follow-up at 20 weeks: 10/95 (group 1), 4/101 (group 2), 8/100 (group 3)</p> <p>Loss to follow-up at 70 weeks: 14/95 (group 1), 11/101 (group 2), 14/100 (group 3)</p>
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
Study characteristics

Methods	RCT, multicentre, 4 parallel arms
Participants	<p>Diagnostic criteria: Oxford (56% satisfied London ME criteria)</p> <p>Number of participants: n = 641</p> <p>Gender: 495 (77%) female</p> <p>Age, mean (SD): 38 (12) years</p> <p>Earlier treatment: NS</p>

Exercise therapy for chronic fatigue syndrome (Review)

White 2011 (Continued)

Co-morbidity: 219 (34%) with any depressive disorder, 260 (41%) used ADs

Illness duration: median 32 (IQR 16-68) months (GET 35 (18-67); SMC 25 (15-57) 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, 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 CFS, 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, 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 > 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, 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, GET: done on the basis of deconditioning and exercise intolerance theories of CFS. 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 HR ranges were set when necessary to avoid overexertion, which eventually aimed at 30 min of light exercise 5 times/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 (FS; Likert scoring 0, 1, 2, 3; range 0-33; lowest score is least fatigue)
- Physical functioning (SF-36 physical functioning subscale version 2; range 0-100; highest score is best functioning)
- Safety outcomes (non-serious adverse events, serious adverse events, serious adverse reactions to study 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-MWT (distance in meters walked)
- Sleep (Jenkins Sleep Scale score for disturbed sleep)

White 2011 (Continued)

- Anxiety and depression (HADS)
- Number of CFS 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

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"
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	The protocol and the statistical analysis plan were not formally published prior to recruitment of participants, and some readers therefore claim the study should be viewed as being a post hoc study. The study authors oppose this, and have published a minute from a Trial Steering Committee (TSC) meeting stating that any changes made to the analysis since the original protocol was agreed by TSC and signed off before the analysis commenced.
Other bias	Low risk	We do not suspect other types of bias

ACT: anaerobic activity therapy; **AD:** antidepressant; **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; **CF:** chronic fatigue; **CFS:** chronic fatigue syndrome; **COG:** cognitive therapy; **DSM-III:** Diagnostic and Statistical Manual of Mental Disorders from the American Psychiatric Association, 3rd edition (Revised); **ET:** exercise therapy; **FS:** Fatigue Scale; **FSS:** Fatigue Severity Scale; **GET:** graded exercise therapy; **GP:** general practitioner; **HADS:** Hospital Anxiety and Depression Scale; **HR:** heart rate; **IQR:** interquartile range; **ITT:** intention-to-treat; **ME:** myalgic encephalitis; **MOS:** Medical Outcome Survey; **NS:** not stated; **PSQI:** Pittsburgh Sleep Quality Index; **PSS:** Perceived Stress Scale; **QoL:** quality of life; **QOLS:** Quality of Life Scale; **RCT:** randomised controlled trial; **RPE:** rating of perceived exertion; **SD:** standard deviation; **SF-36:** Short Form 36; **SMC:** specialist medical care; **VO₂:** oxygen consumption; **6MWT:** six-minute walking test

Characteristics of excluded studies [ordered by study ID]

Study	Reason for exclusion
Broadbent 2012	Compared 2 exercise interventions: intermittent and graded, outside scope of review
Evering 2008	Intervention was feedback on physical activity
Gordon 2010	Compared 2 different types of exercise therapy, outside scope of review
Guarino 2001	Ineligible study population: "Gulf War veterans"
Hatcher 1998	The study author stated in an email of 5 February 2018 that the data were not available
Kos 2012	Study intervention was an Activity Pacing Self-Management programme where exercise was only one part. Quote: "Participants conferred with the therapists to set relevant and achievable activity and exercise goals, ..."
Liu 2010	Ineligible outcomes ("... therapeutic effects and the changes of malondialdehyde (MDA) content and the activity of serum superoxide dismutases (SOD) and serum glutathione peroxidase (GSH-Px)")
Nunez 2011	Combination treatment of which exercise therapy was a minor part
Ridsdale 2004	No clinical diagnosis of CFS and intervention did not include exercise, so did not meet inclusion criteria
Ridsdale 2012	Ineligible population: "people presenting with chronic fatigue in primary care"
Russel 2001	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	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	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	Study used a "cross-sectional design"
Thomas 2008	Study used "between-group comparisons were used." This was a controlled trial, but participants were not randomly assigned
Tummers 2012	Interventions included variations of CBT not exercise: "additional CBT (stepped care) or regular CBT (care as usual)"
Viner 2004	Ineligible population: "young people (aged 9–17 years) with CFS/ME"
Vos-Vromans 2008	Study compared 2 different types of exercise therapy
Wright 2005	Ineligible population, included young people 0-19 years of age
Zhuo 2007	No randomisation procedure was described

CBT: cognitive-behavioural therapy; **CFS:** chronic fatigue syndrome; **ME:** myalgic encephalitis;

Characteristics of studies awaiting classification [ordered by study ID]

Marques 2012

Methods	Multicentre, RCT
Participants	<p>Fulfilling operationalised criteria for ICF and for CFS</p> <p>Patients visiting their physician with a main complaint of unexplained fatigue of at least 6 months' duration are recruited for the study</p> <p>Inclusion criteria: meeting the operationalised criteria for ICF or CFS (CDC criteria); aged 18-65 years; fluent in spoken Portuguese; capacity to provide informed consent</p> <p>Exclusion criteria: presence of a concurrent somatic condition that can explain the fatigue symptoms; severe psychiatric disorders</p>
Interventions	<p>SC or SC plus a self-regulation-based physical activity programme (4-STEPS)</p> <p>In addition to SC, participants in the intervention group received the 4-STEPS programme consisting of the following.</p> <ul style="list-style-type: none"> • 2 face-to-face individual 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 h. Details on topics addressed during the MI sessions are presented in Table 1 • 2 brief telephone counselling sessions: sessions take about 20 min and are provided 2 weeks and 6 weeks after the last MI session. • Self-regulation booklets: 2 booklets were designed to help participants 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. • 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 • Daily activities record: 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 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 receive a leaflet for their partner or significant other to increase social support
Outcomes	<p>Primary outcome: reduction in perceived fatigue severity, assessed 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</p>
Notes	<p>ISRCTN: ISRCTN70763996</p> <p>Copied from the published protocol: www.biomedcentral.com/1471-2458/12/202</p>

White 2012

Methods	Randomised interventional trial
Participants	<p>Inclusion</p> <ul style="list-style-type: none"> • 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 • 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 <p>Exclusion</p> <ul style="list-style-type: none"> • Not receiving a diagnosis of CFS/ME • Co-morbid condition that requires that exercise be performed only in the presence of a doctor • < age 18 • Active suicidal thoughts
Interventions	<p>Guided support, a copy of the GETSET booklet, a 30-min consultation face-to-face by Skype or by telephone, 3 further Skype telephone contacts</p> <p>Intervention over 9 weeks: follow-up length: 3 months; study entry: single randomisation only</p>
Outcomes	<p>Primary: SF-36 physical function subscale, measured 12 weeks from randomisation</p> <p>Secondary: Clinical Global Impression Change Scale, score measured 12 weeks from baseline</p>
Notes	www.controlled-trials.com/ISRCTN22975026/GETSET

CDC: Centers for Disease Control and Prevention; **CFS:** chronic fatigue syndrome; **ICF:** idiopathic chronic fatigue; **MI:** motivational interviewing; **RCT:** randomised controlled trial; **SC:** standard care; **SF-36:** Short Form 36

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
1.1 Fatigue (end of treatment)	7	840	Std. Mean Difference (IV, Random, 95% CI)	-0.66 [-1.01, -0.31]
1.2 Fatigue (follow-up)	4	670	Std. Mean Difference (IV, Random, 95% CI)	-0.62 [-1.32, 0.07]
1.3 Participants with serious adverse reactions	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
1.4 Pain (follow-up)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected

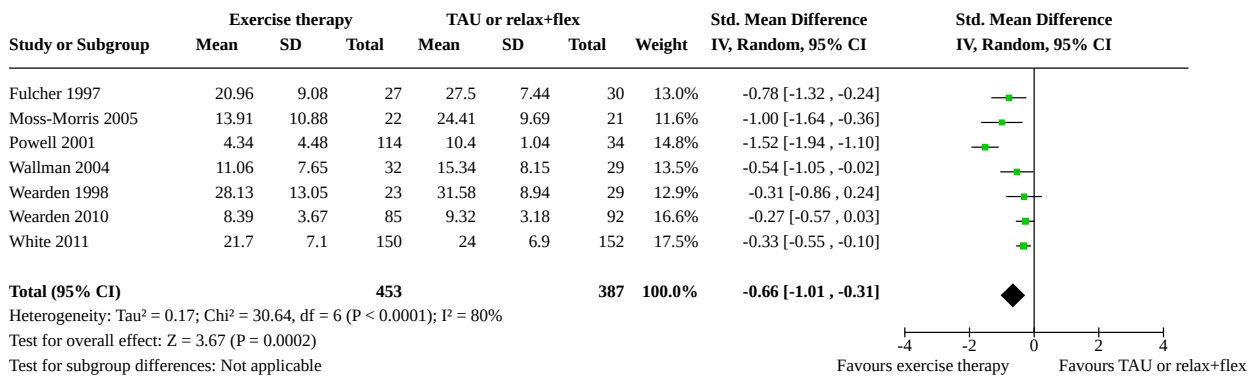
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.4.1 Brief Pain Inventory, pain severity subscale (0 to 10 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.4.2 Brief Pain Inventory, pain interference subscale (0 to 10 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.5 Physical functioning (end of treatment)	5		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.5.1 SF-36, physical functioning subscale (0 to 100 points)	5	725	Mean Difference (IV, Random, 95% CI)	-13.10 [-24.22, -1.98]
1.6 Physical functioning (follow-up)	3		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.6.1 SF-36, physical functioning subscale (0 to 100 points)	3	621	Mean Difference (IV, Random, 95% CI)	-16.33 [-36.74, 4.08]
1.7 Quality of life (follow-up)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.7.1 Quality of Life Scale (16 to 112 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.8 Depression (end of treatment)	5		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.8.1 HADS, depression score (7 items/21 points)	5	504	Mean Difference (IV, Random, 95% CI)	-1.63 [-3.50, 0.23]
1.9 Depression (follow-up)	4	654	Std. Mean Difference (IV, Random, 95% CI)	-0.35 [-0.93, 0.23]
1.10 Anxiety (end of treatment)	3		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.10.1 HADS, anxiety score (0 to 21 points)	3	387	Mean Difference (IV, Random, 95% CI)	-1.48 [-3.58, 0.61]
1.11 Anxiety (follow-up)	4	652	Std. Mean Difference (IV, Random, 95% CI)	-0.17 [-0.50, 0.15]
1.12 Sleep (end of treatment)	2		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.12.1 Jenkins Sleep Scale (0 to 20 points)	2	323	Mean Difference (IV, Random, 95% CI)	-1.49 [-2.95, -0.02]
1.13 Sleep (follow-up)	3		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.13.1 Jenkins Sleep Scale (0 to 20 points)	3	610	Mean Difference (IV, Random, 95% CI)	-2.04 [-3.84, -0.23]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.14 Self-perceived changes in overall health (end of treatment)	4	489	Risk Ratio (M-H, Random, 95% CI)	1.83 [1.39, 2.40]
1.15 Self-perceived changes in overall health (follow-up)	3	518	Risk Ratio (M-H, Random, 95% CI)	1.88 [0.76, 4.64]
1.16 Health resource use (follow-up) (Mean no. of contacts)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.16.1 Primary care	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.16.2 Other doctor	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.16.3 Healthcare professional	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.16.4 Inpatient	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.16.5 Accident and emergency	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.16.6 Other health/social services	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.16.7 Complementary health care	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.16.8 Standardised medical care	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
1.17 Health resource use (follow-up) (No. of users)	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
1.17.1 Primary care	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
1.17.2 Other doctor	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
1.17.3 Healthcare professional	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
1.17.4 Inpatient	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
1.17.5 Accident and emergency	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
1.17.6 Medication	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected

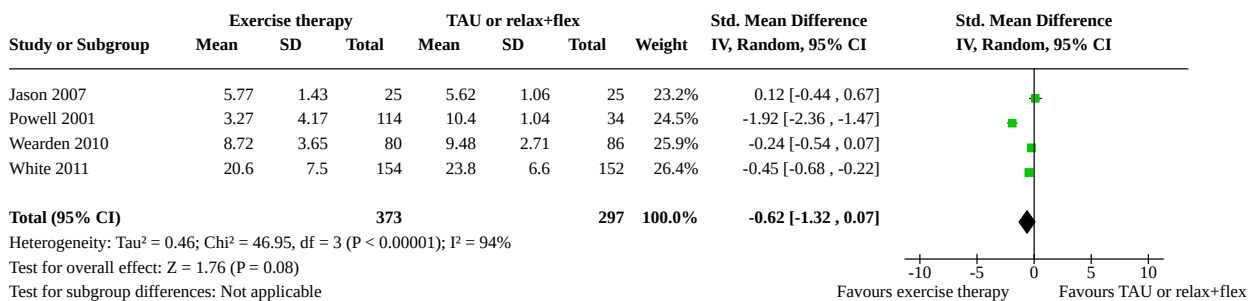
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.17.7 Complementary health care	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
1.17.8 Other health/social services	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
1.17.9 Standardised medical care	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
1.18 Dropout	6	843	Risk Ratio (M-H, Random, 95% CI)	1.63 [0.77, 3.43]
1.19 Sensitivity analysis for fatigue (end of treatment)	7		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.19.1 Fatigue Scale, FS (11 items/0 to 11 points)	2	325	Mean Difference (IV, Random, 95% CI)	-3.50 [-8.53, 1.53]
1.19.2 Fatigue Scale, FS (11 items/0 to 33 points)	2	363	Mean Difference (IV, Random, 95% CI)	-2.57 [-4.04, -1.10]
1.19.3 Fatigue Scale, FS (14 items/0 to 42 points)	3	152	Mean Difference (IV, Random, 95% CI)	-6.80 [-10.31, -3.28]
1.20 Sensitivity analysis for fatigue (follow-up)	4		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.20.1 Fatigue Scale, FS (11 items/0 to 11 points)	1	148	Mean Difference (IV, Random, 95% CI)	-7.13 [-7.97, -6.29]
1.20.2 Fatigue Scale, FS (11 items/0 to 33 points)	2	472	Mean Difference (IV, Random, 95% CI)	-2.87 [-4.18, -1.55]
1.20.3 Fatigue Severity Scale, FSS (9 items/1 to 7 points)	1	50	Mean Difference (IV, Random, 95% CI)	0.15 [-0.55, 0.85]
1.21 Subgroup analysis for fatigue (end of treatment)	7	840	Std. Mean Difference (IV, Random, 95% CI)	-0.66 [-1.01, -0.31]
1.21.1 Graded exercise therapy	6	779	Std. Mean Difference (IV, Random, 95% CI)	-0.68 [-1.08, -0.28]
1.21.2 Exercise with self-pacing	1	61	Std. Mean Difference (IV, Random, 95% CI)	-0.54 [-1.05, -0.02]
1.22 Subgroup analysis for fatigue (follow-up)	4	670	Std. Mean Difference (IV, Random, 95% CI)	-0.62 [-1.32, 0.07]
1.22.1 Graded exercise therapy	3	620	Std. Mean Difference (IV, Random, 95% CI)	-0.85 [-1.67, -0.03]
1.22.2 Anaerobic exercise	1	50	Std. Mean Difference (IV, Random, 95% CI)	0.12 [-0.44, 0.67]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
1.23 Sensitivity analysis for depression (follow-up)	4		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.23.1 Beck Depression Inventory (0 to 63 points)	1	45	Mean Difference (IV, Random, 95% CI)	3.44 [-3.00, 9.88]
1.23.2 HADS, depression subscale (0 to 21 points)	3	609	Mean Difference (IV, Random, 95% CI)	-2.26 [-5.09, 0.56]
1.24 Sensitivity analysis for anxiety (follow-up)	4		Mean Difference (IV, Random, 95% CI)	Subtotals only
1.24.1 Beck Anxiety Inventory (0 to 63 points)	1	45	Mean Difference (IV, Random, 95% CI)	0.70 [-4.52, 5.92]
1.24.2 HADS, anxiety score (0 to 21 points)	3	607	Mean Difference (IV, Random, 95% CI)	-1.01 [-2.75, 0.74]

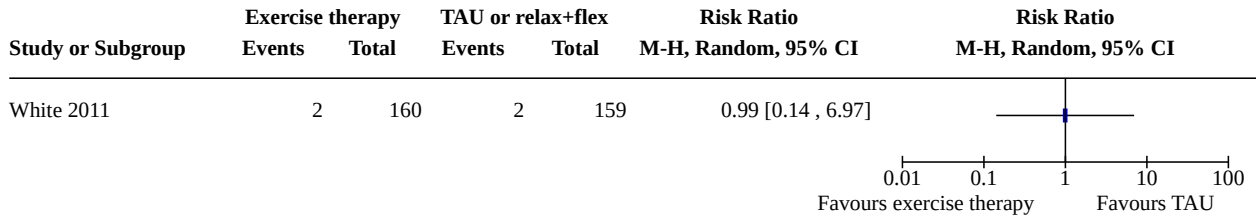
Analysis 1.1. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 1: Fatigue (end of treatment)



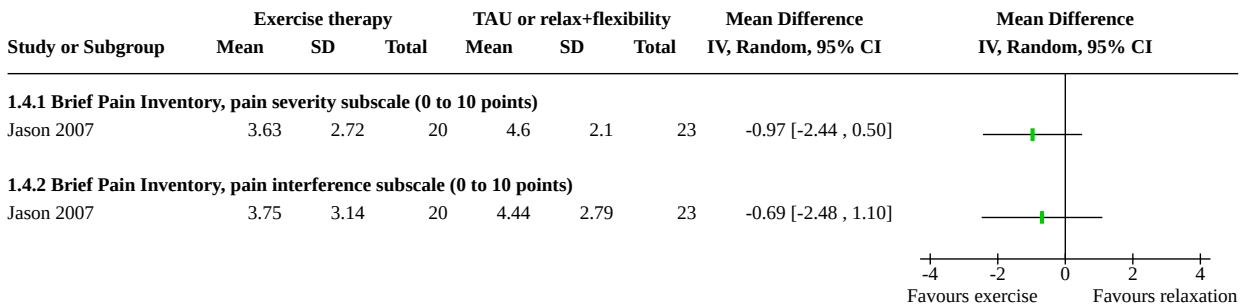
Analysis 1.2. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 2: Fatigue (follow-up)



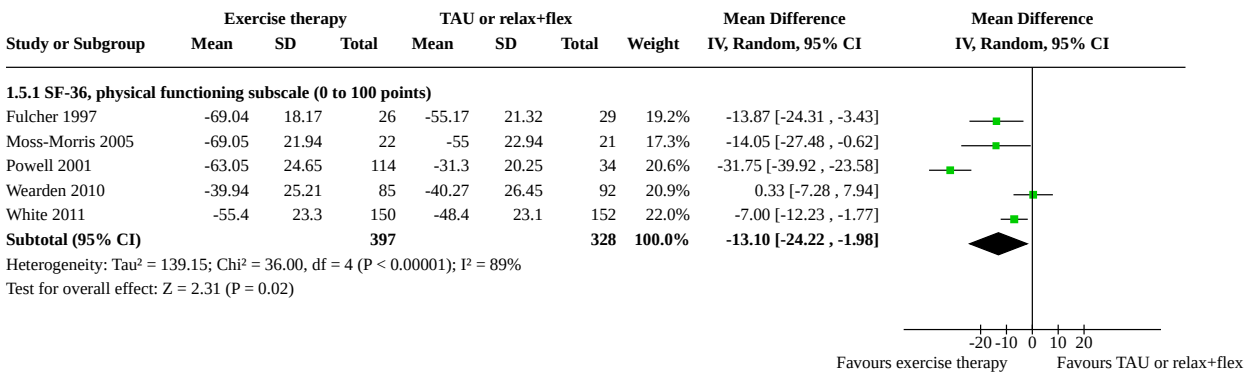
Analysis 1.3. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 3: Participants with serious adverse reactions



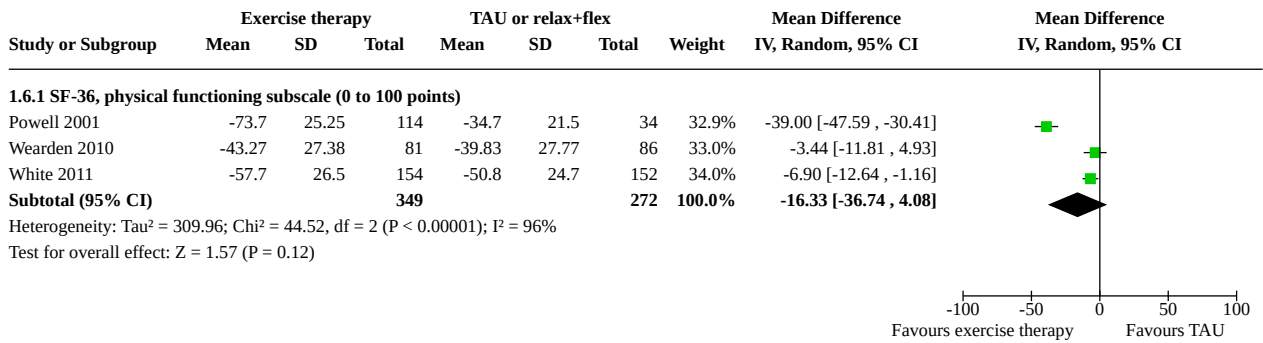
Analysis 1.4. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 4: Pain (follow-up)



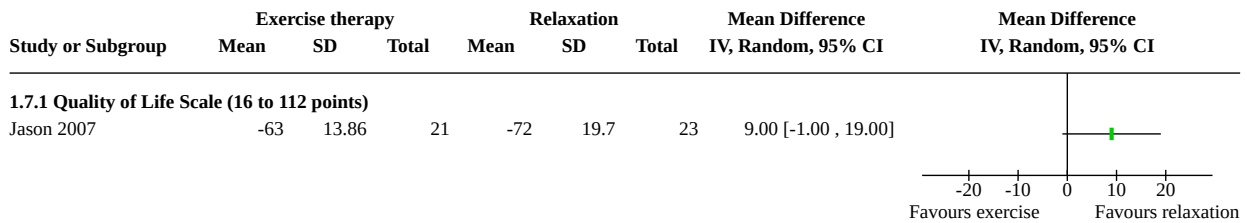
Analysis 1.5. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 5: Physical functioning (end of treatment)



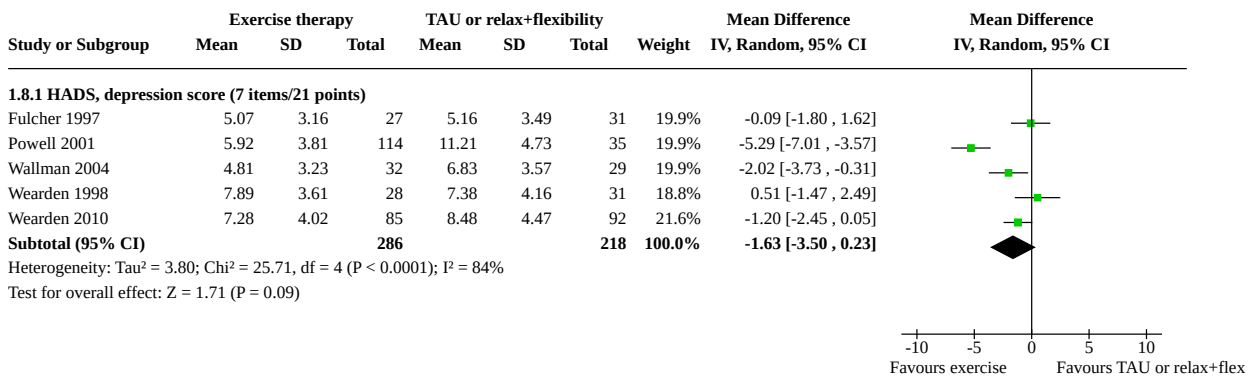
Analysis 1.6. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 6: Physical functioning (follow-up)



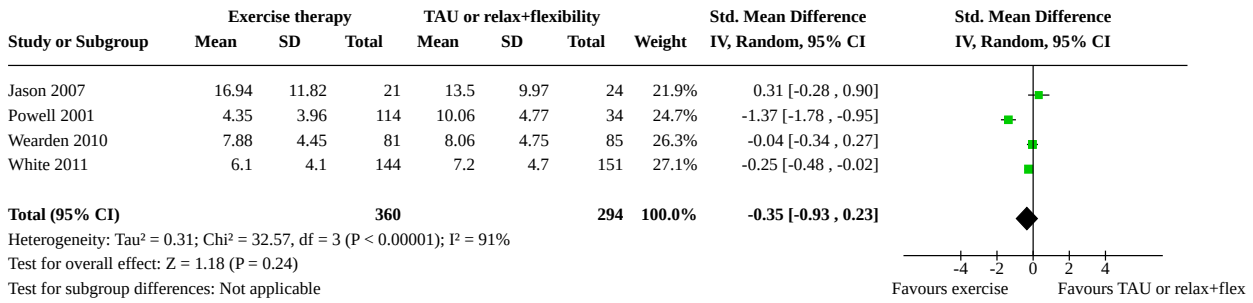
Analysis 1.7. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 7: Quality of life (follow-up)



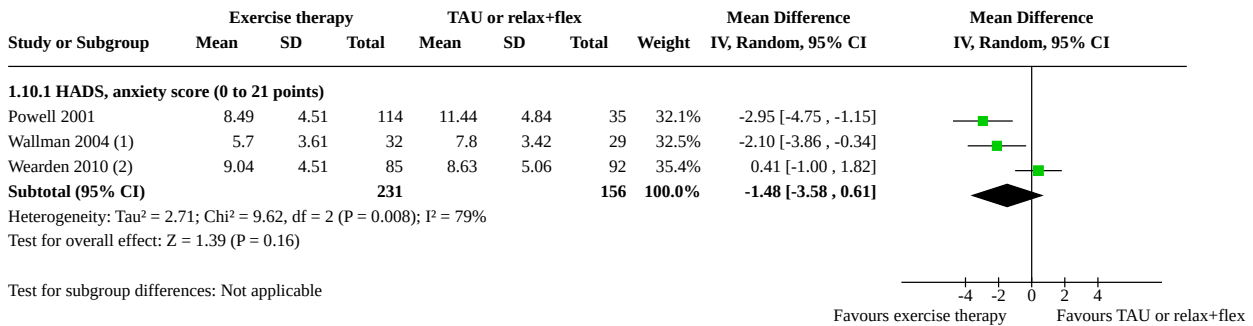
Analysis 1.8. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 8: Depression (end of treatment)



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



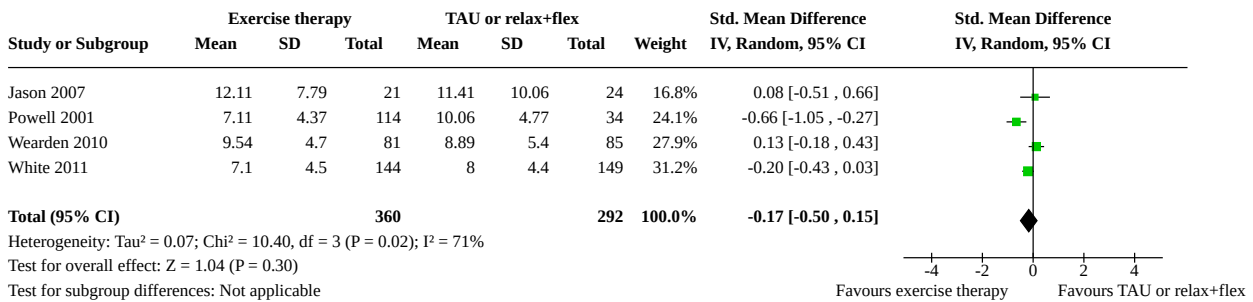
Analysis 1.10. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 10: Anxiety (end of treatment)



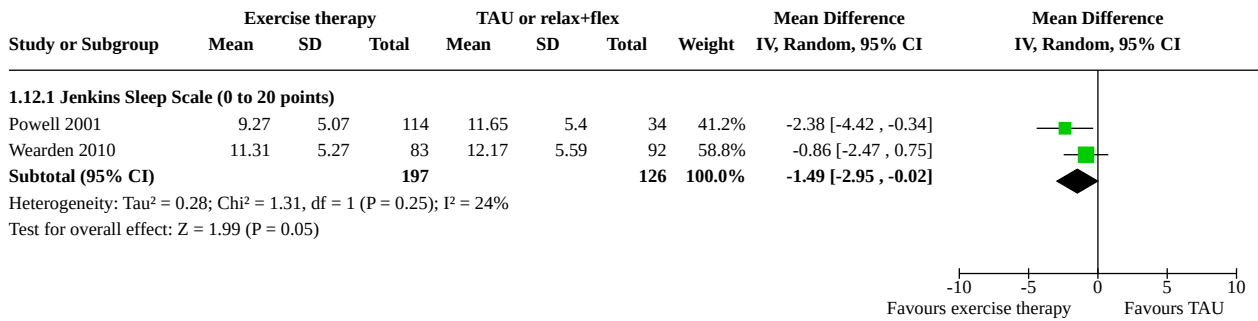
Footnotes

- (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)

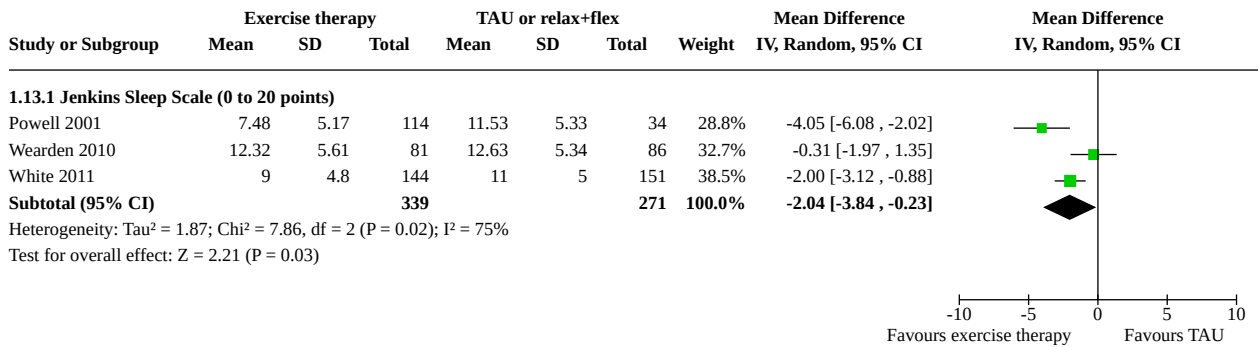
Analysis 1.11. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 11: Anxiety (follow-up)



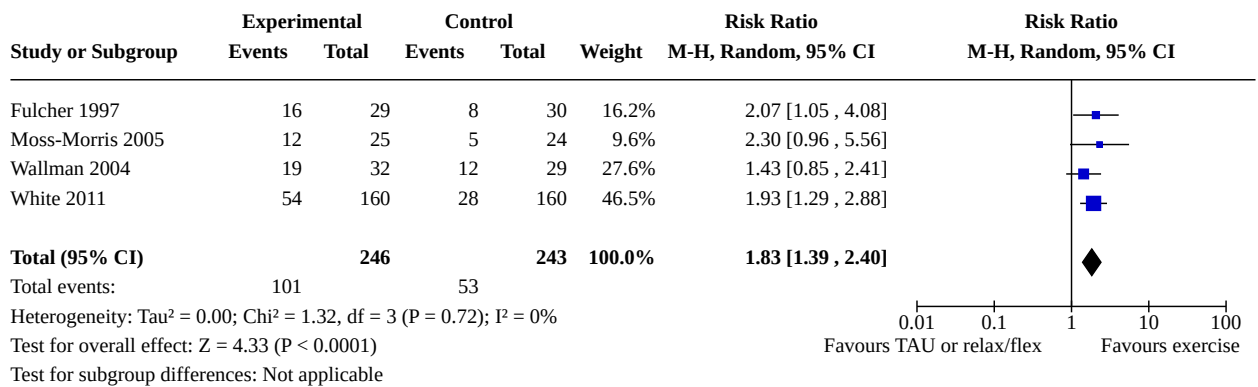
Analysis 1.12. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 12: Sleep (end of treatment)



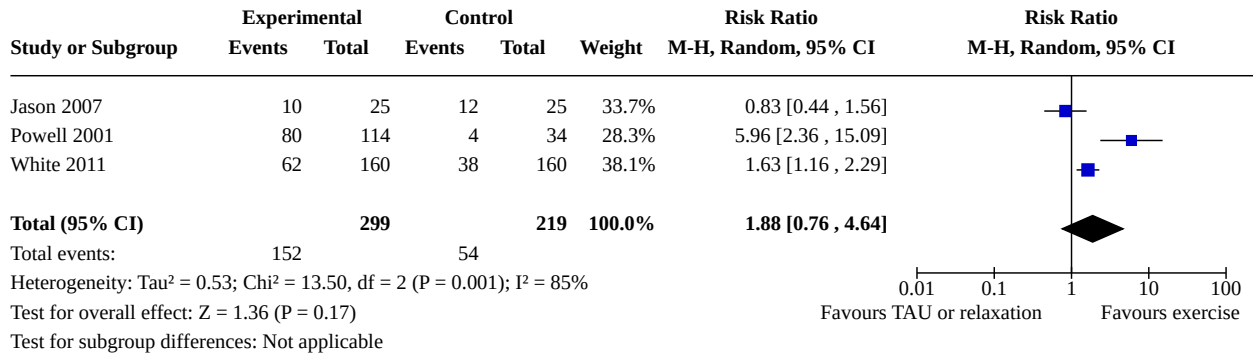
Analysis 1.13. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 13: Sleep (follow-up)



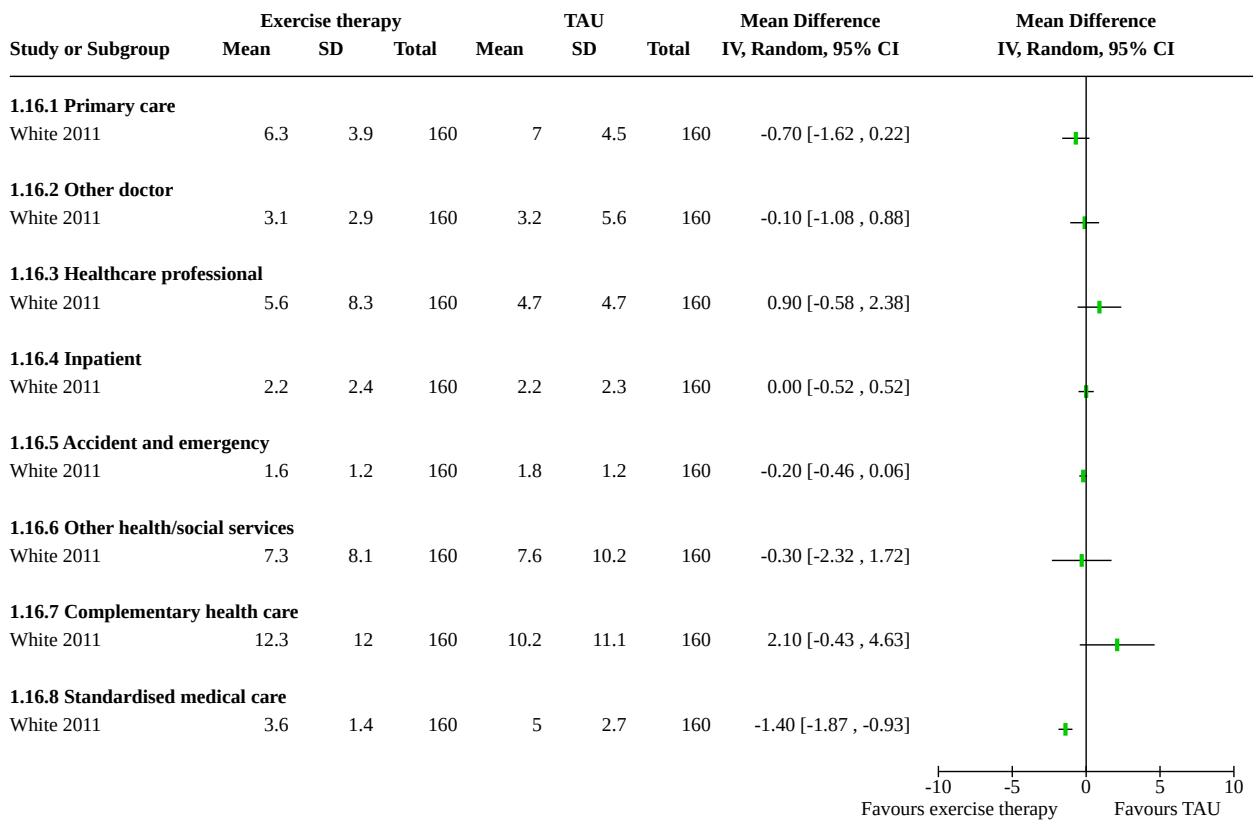
Analysis 1.14. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 14: Self-perceived changes in overall health (end of treatment)



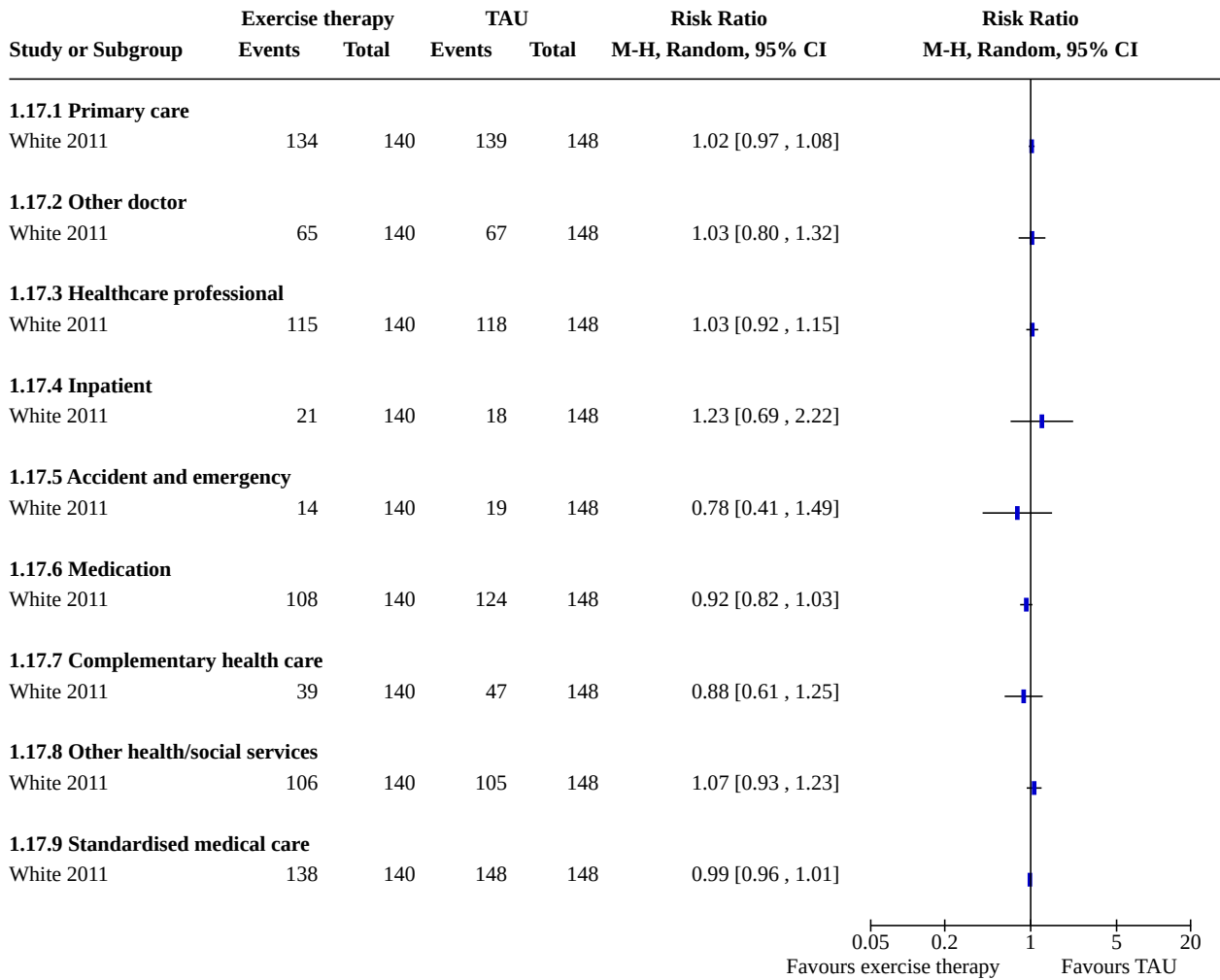
Analysis 1.15. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 15: Self-perceived changes in overall health (follow-up)



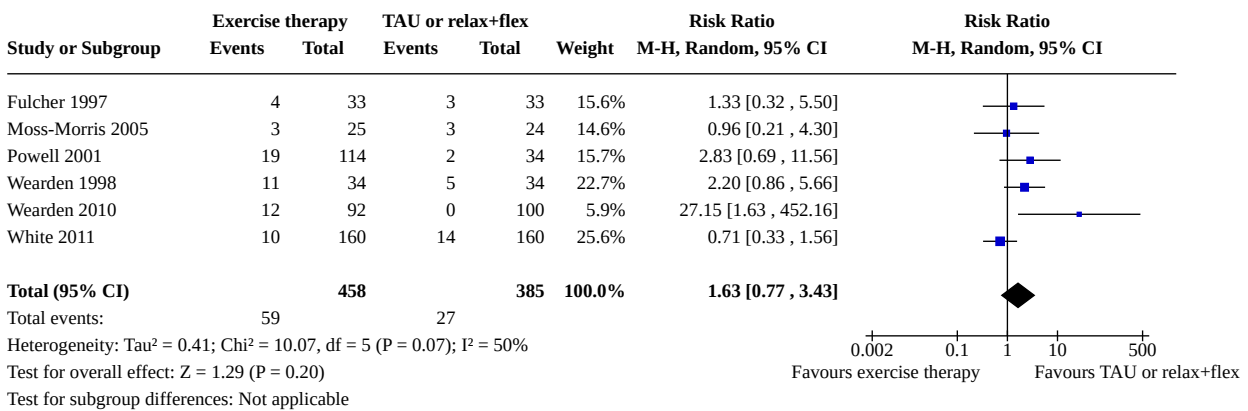
Analysis 1.16. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 16: Health resource use (follow-up) (Mean no. of contacts)



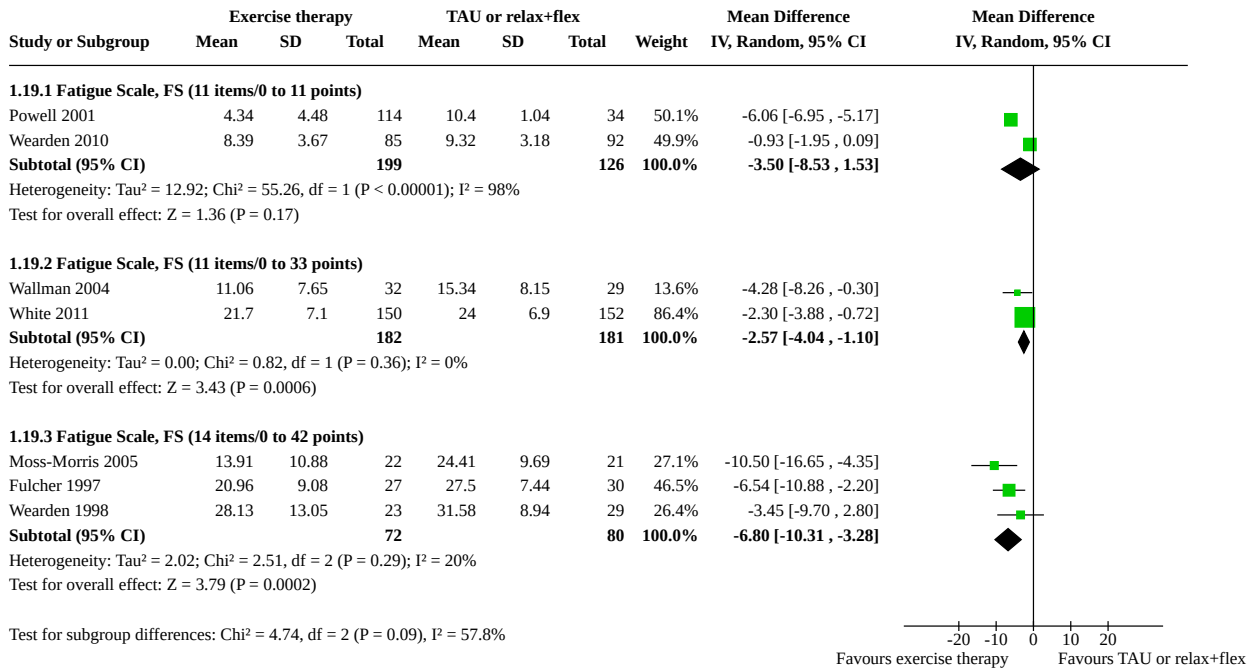
Analysis 1.17. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 17: Health resource use (follow-up) (No. of users)



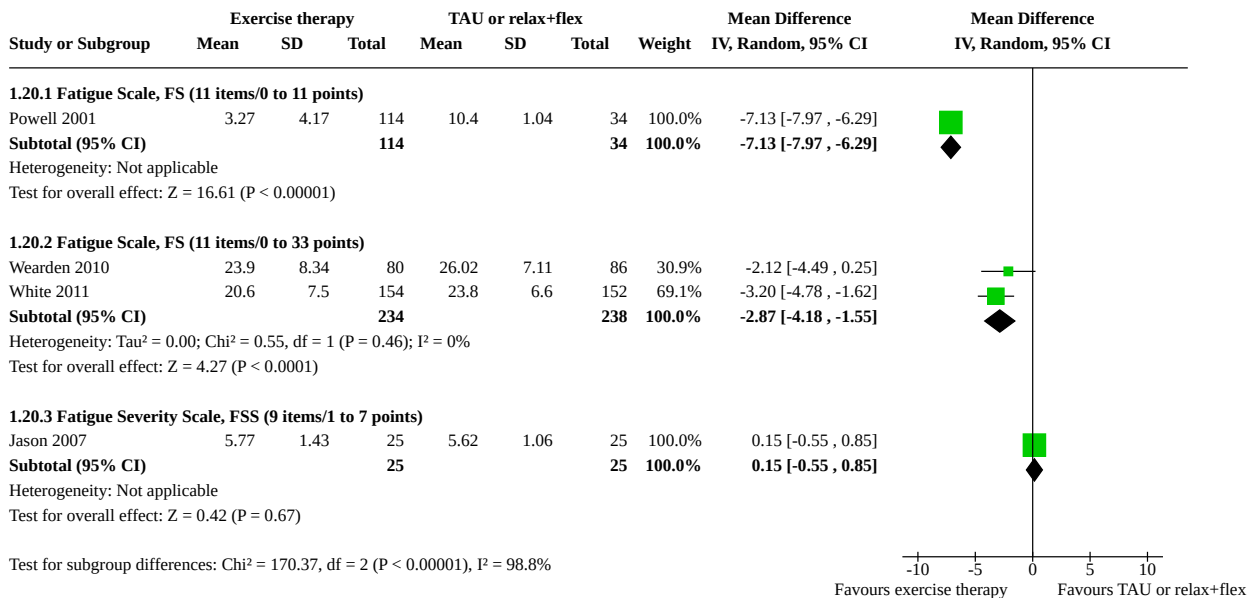
Analysis 1.18. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 18: Dropout



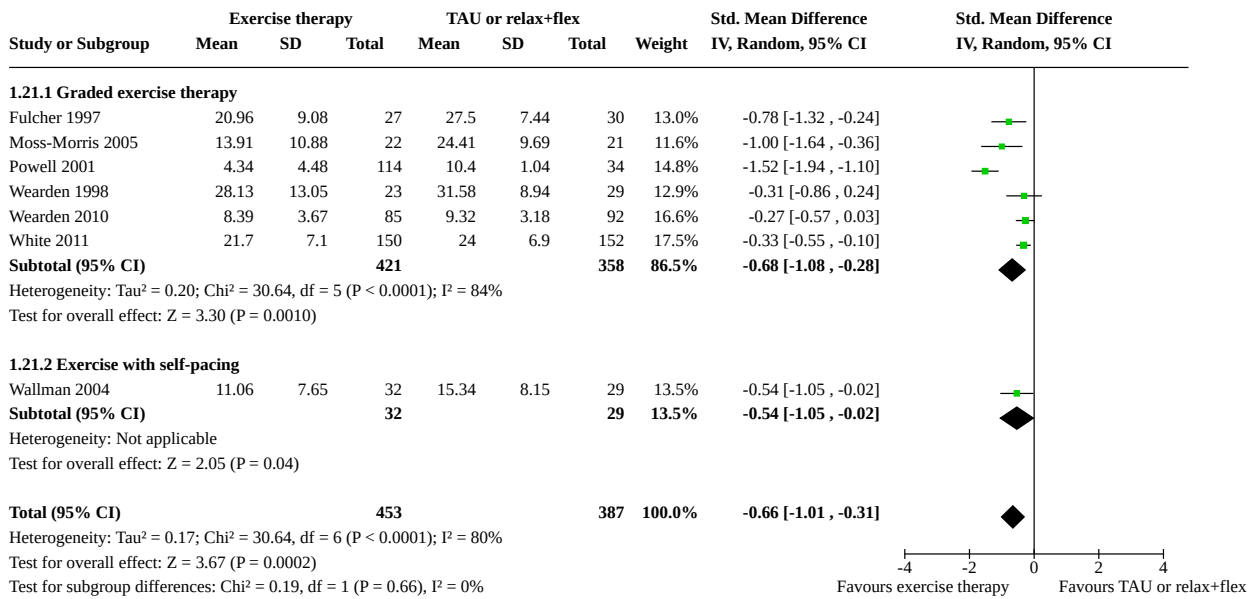
Analysis 1.19. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 19: Sensitivity analysis for fatigue (end of treatment)



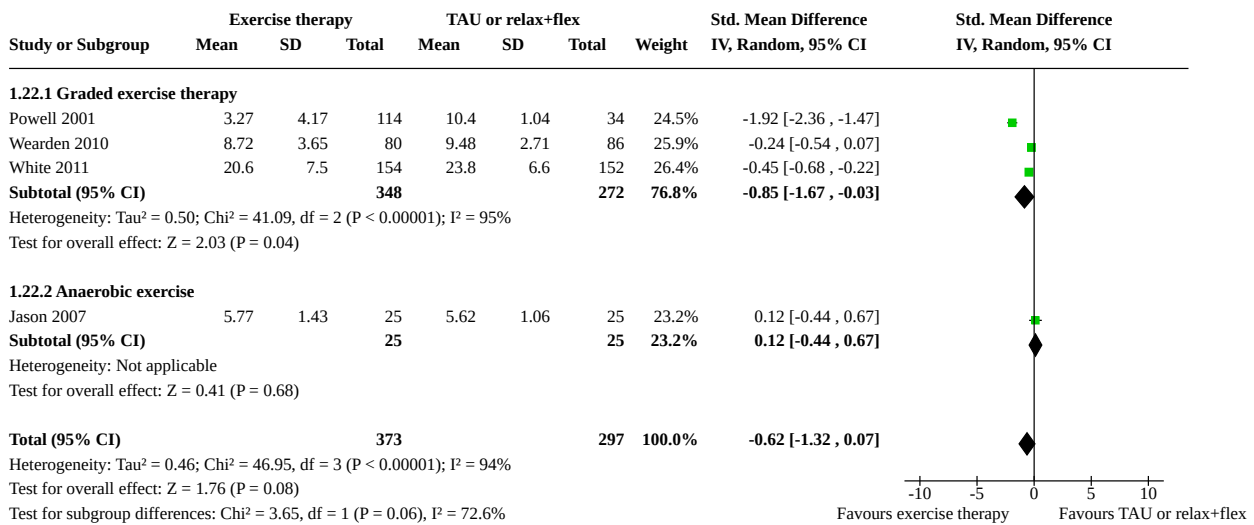
Analysis 1.20. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 20: Sensitivity analysis for fatigue (follow-up)



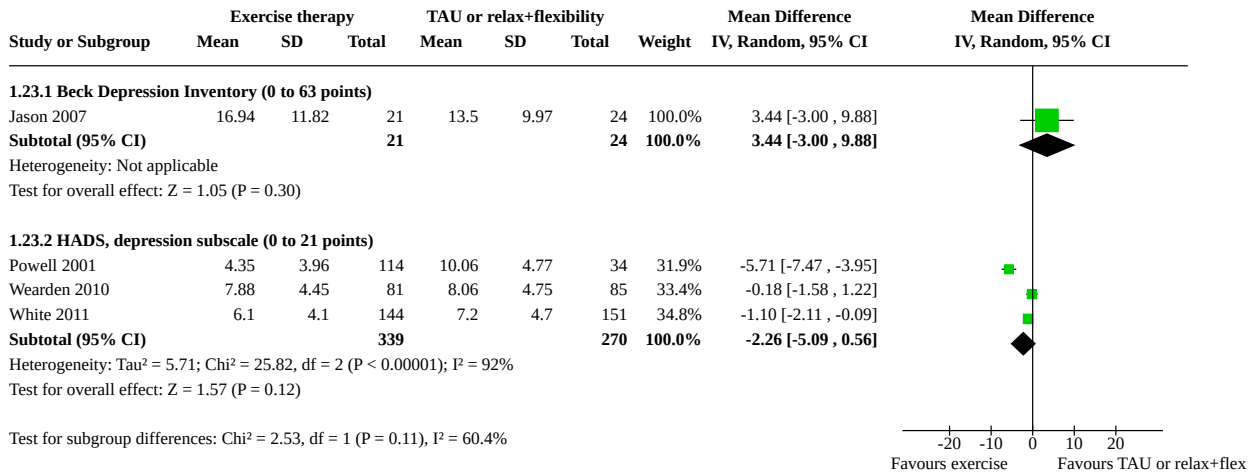
Analysis 1.21. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 21: Subgroup analysis for fatigue (end of treatment)



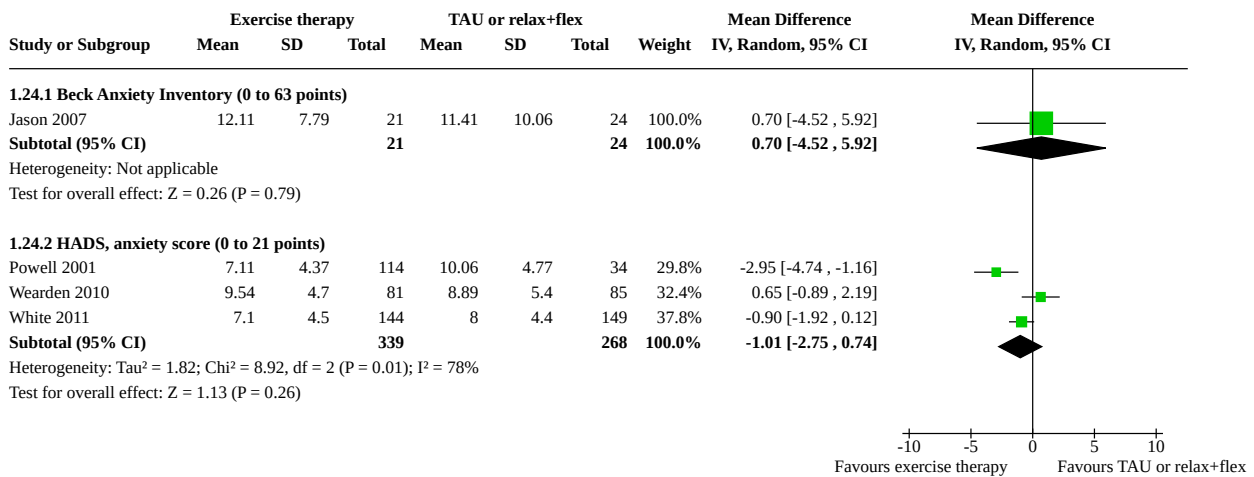
Analysis 1.22. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 22: Subgroup analysis for fatigue (follow-up)



Analysis 1.23. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 23: Sensitivity analysis for depression (follow-up)



Analysis 1.24. Comparison 1: Exercise therapy versus treatment as usual, relaxation or flexibility, Outcome 24: Sensitivity analysis for anxiety (follow-up)



Comparison 2. Exercise therapy versus psychological treatment

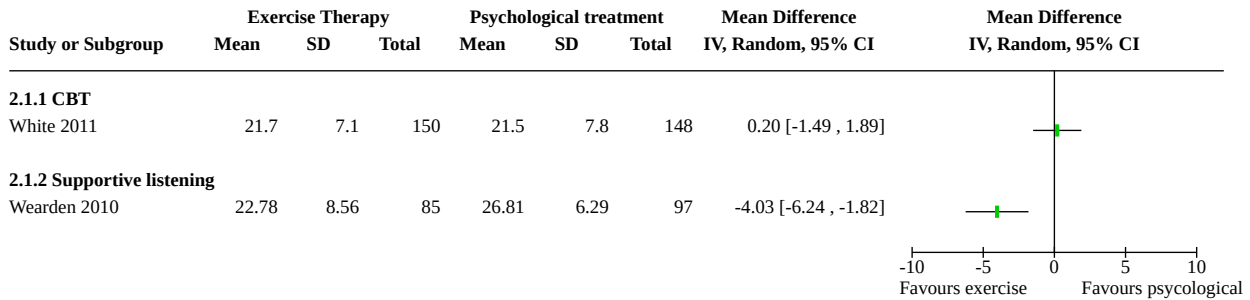
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.1 Fatigue at end of treatment (FS; 11 items/0 to 33 points)	2		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.1.1 CBT	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.1.2 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.2 Fatigue at follow-up (SMD)	2	351	Std. Mean Difference (IV, Random, 95% CI)	0.07 [-0.13, 0.28]

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.3 Participants with serious adverse reactions	2		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.3.1 CBT	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.3.2 Supportive listening	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.4 Pain at follow-up (Brief Pain Inventory, pain severity subscale; 0 to 10 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.4.1 CBT	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.4.2 Cognitive therapy	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.5 Pain at follow-up (Brief Pain Inventory, pain interference subscale; 0 to 10 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.5.1 CBT	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.5.2 Cognitive therapy	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.6 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
2.6.1 CBT	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.6.2 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.7 Physical functioning at follow-up (SF-36, physical functioning subscale; 0 to 100 points)	3		Mean Difference (IV, Random, 95% CI)	Subtotals only
2.7.1 CBT	2	348	Mean Difference (IV, Random, 95% CI)	7.92 [-9.79, 25.63]
2.7.2 Cognitive therapy	1	47	Mean Difference (IV, Random, 95% CI)	21.37 [6.61, 36.13]
2.7.3 Supportive listening	1	171	Mean Difference (IV, Random, 95% CI)	-7.55 [-15.57, 0.47]
2.8 Quality of life (follow-up)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected

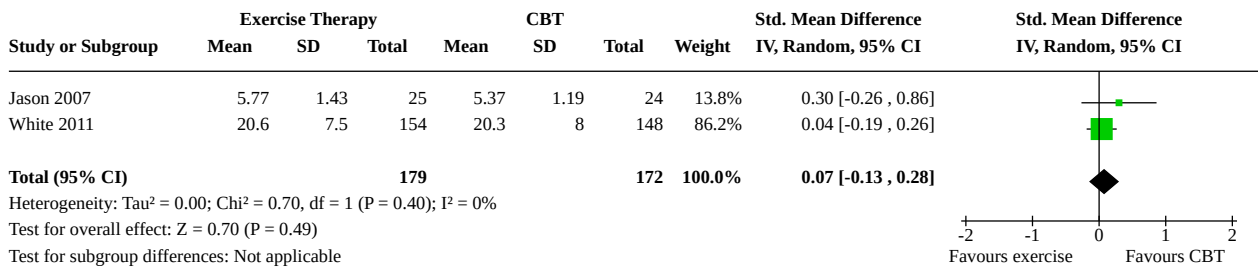
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.8.1 Quality of Life Scale (16 to 112 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.9 Depression at end of treatment (HADS depression score; 7 items/21 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.9.1 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.10 Depression at follow-up (SMD)	2	331	Std. Mean Difference (IV, Random, 95% CI)	0.01 [-0.21, 0.22]
2.11 Anxiety at end of treatment (HADS anxiety; 7 items/21 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.11.1 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.12 Anxiety at follow-up (SMD)	2	331	Std. Mean Difference (IV, Random, 95% CI)	0.07 [-0.15, 0.28]
2.13 Sleep at end of treatment (Jenkins Sleep Scale; 0 to 20 points)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.13.1 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.14 Sleep at follow-up (Jenkins Sleep Scale; 0 to 20 points)	2		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.14.1 CBT	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.14.2 Supportive listening	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.15 Self-perceived changes in overall health at end of treatment	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.15.1 CBT	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.16 Self-perceived changes in overall health at follow-up	2		Risk Ratio (M-H, Random, 95% CI)	Subtotals only
2.16.1 Cognitive therapy	1	50	Risk Ratio (M-H, Random, 95% CI)	0.62 [0.36, 1.10]
2.16.2 CBT	2	368	Risk Ratio (M-H, Random, 95% CI)	0.71 [0.33, 1.54]
2.17 Health resource use (follow-up) (Mean no. of contacts)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
2.17.1 Primary care	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.17.2 Other doctor	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.17.3 Healthcare professional	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.17.4 Inpatient	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.17.5 Accident and emergency	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.17.6 Other health/social services	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.17.7 Complementary health care	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.17.8 Standardised medical care	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
2.18 Health resource use (follow-up) (No. of users)	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.18.1 Primary care	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.18.2 Other doctor	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.18.3 Healthcare professional	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.18.4 Inpatient	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.18.5 Accident and emergency	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.18.6 Medication	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.18.7 Complementary health care	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.18.8 Other health/social services	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.18.9 Standardised medical care	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.19 Dropout	2		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.19.1 CBT	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
2.19.2 Supportive listening	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected

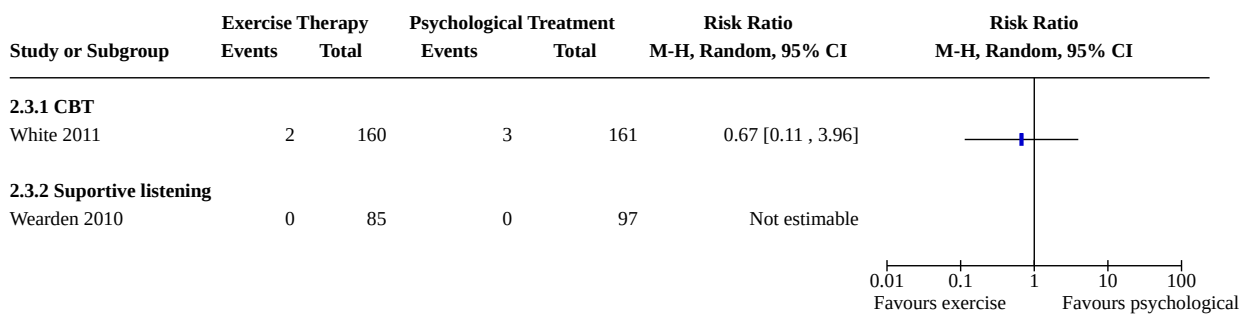
Analysis 2.1. Comparison 2: Exercise therapy versus psychological treatment, Outcome 1: Fatigue at end of treatment (FS; 11 items/0 to 33 points)



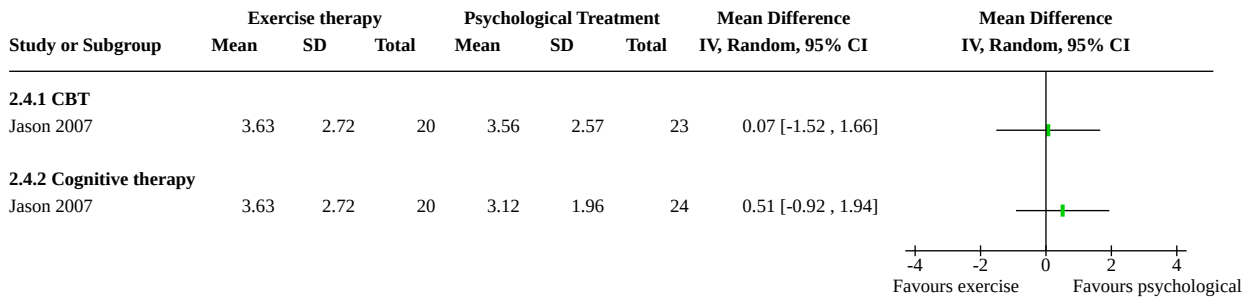
Analysis 2.2. Comparison 2: Exercise therapy versus psychological treatment, Outcome 2: Fatigue at follow-up (SMD)



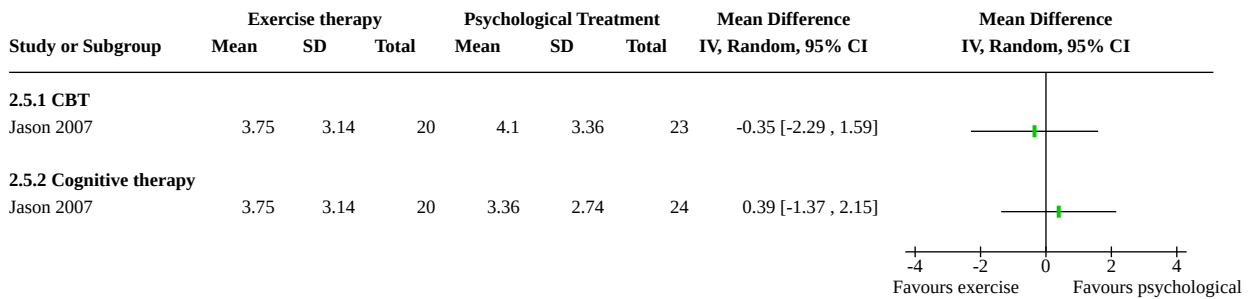
Analysis 2.3. Comparison 2: Exercise therapy versus psychological treatment, Outcome 3: Participants with serious adverse reactions



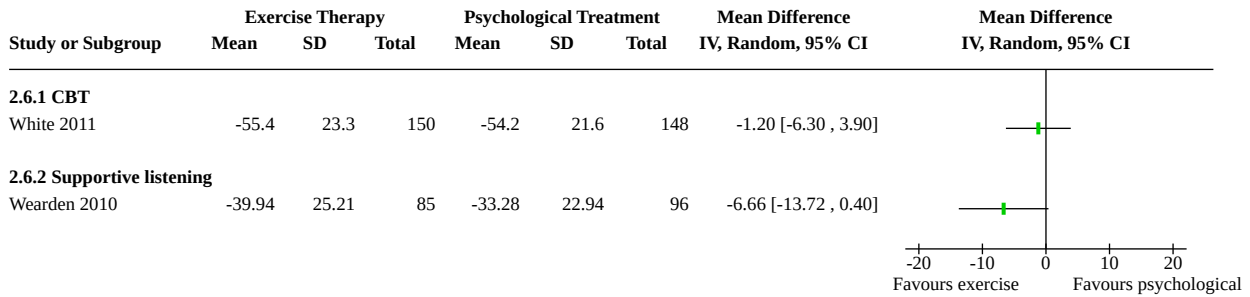
Analysis 2.4. Comparison 2: Exercise therapy versus psychological treatment, Outcome 4: Pain at follow-up (Brief Pain Inventory, pain severity subscale; 0 to 10 points)



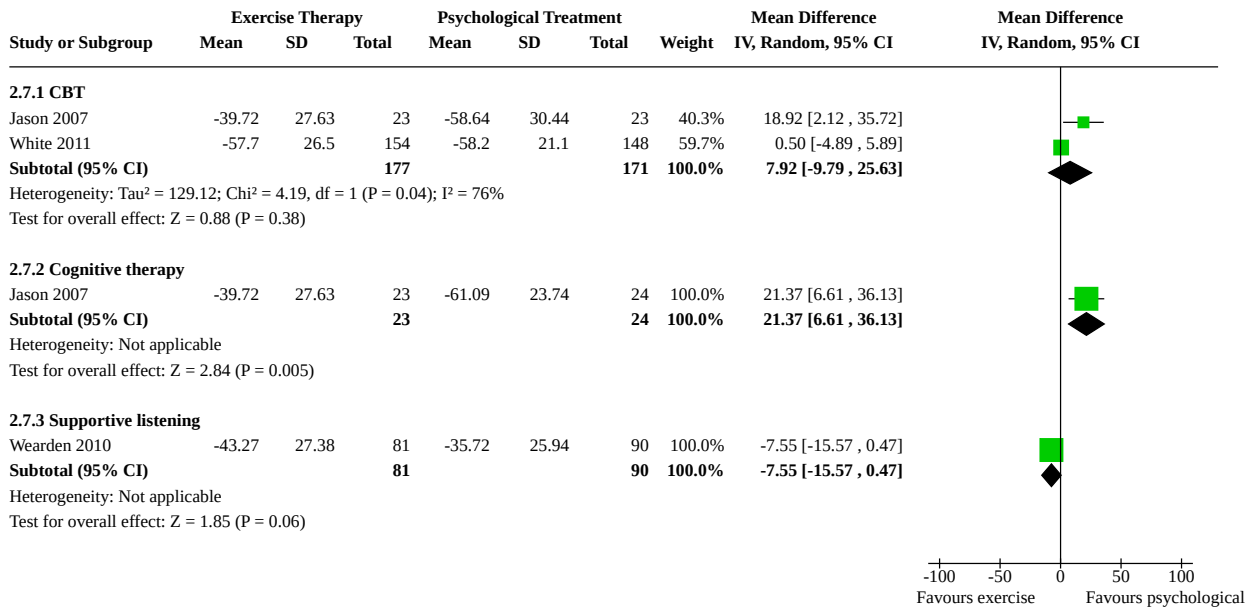
Analysis 2.5. Comparison 2: Exercise therapy versus psychological treatment, Outcome 5: Pain at follow-up (Brief Pain Inventory, pain interference subscale; 0 to 10 points)



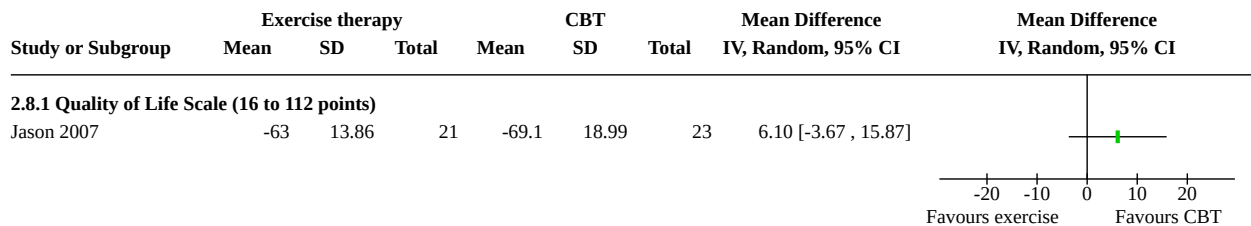
Analysis 2.6. Comparison 2: Exercise therapy versus psychological treatment, Outcome 6: Physical functioning at end of treatment (SF-36, physical functioning subscale; 0 to 100 points)



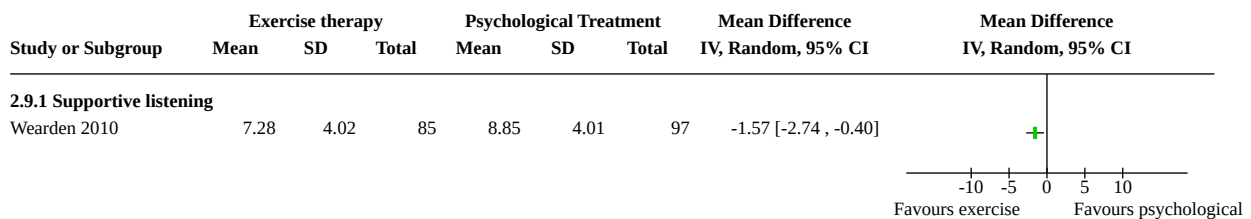
Analysis 2.7. Comparison 2: Exercise therapy versus psychological treatment, Outcome 7: Physical functioning at follow-up (SF-36, physical functioning subscale; 0 to 100 points)



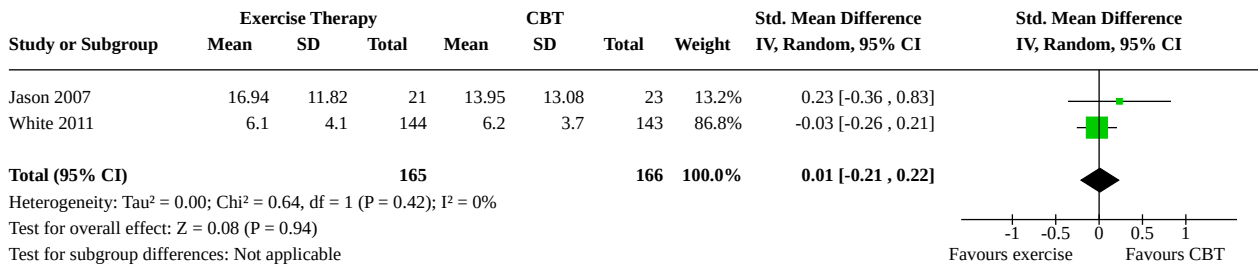
Analysis 2.8. Comparison 2: Exercise therapy versus psychological treatment, Outcome 8: Quality of life (follow-up)



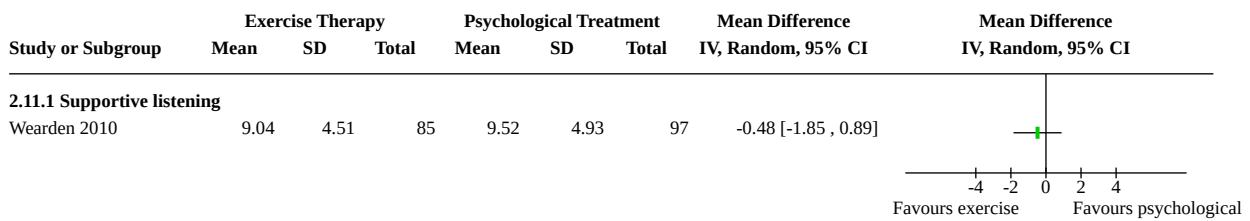
Analysis 2.9. Comparison 2: Exercise therapy versus psychological treatment, Outcome 9: Depression at end of treatment (HADS depression score; 7 items/21 points)



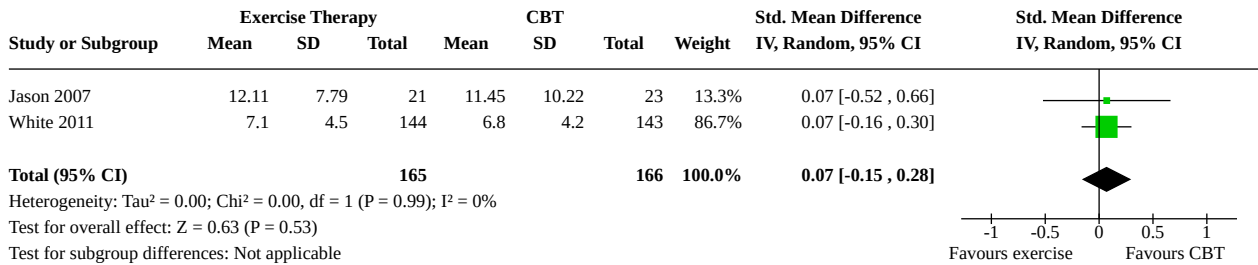
Analysis 2.10. Comparison 2: Exercise therapy versus psychological treatment, Outcome 10: Depression at follow-up (SMD)



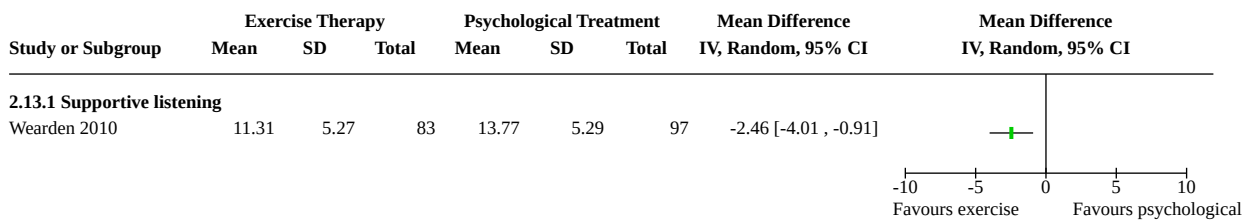
Analysis 2.11. Comparison 2: Exercise therapy versus psychological treatment, Outcome 11: Anxiety at end of treatment (HADS anxiety; 7 items/21 points)



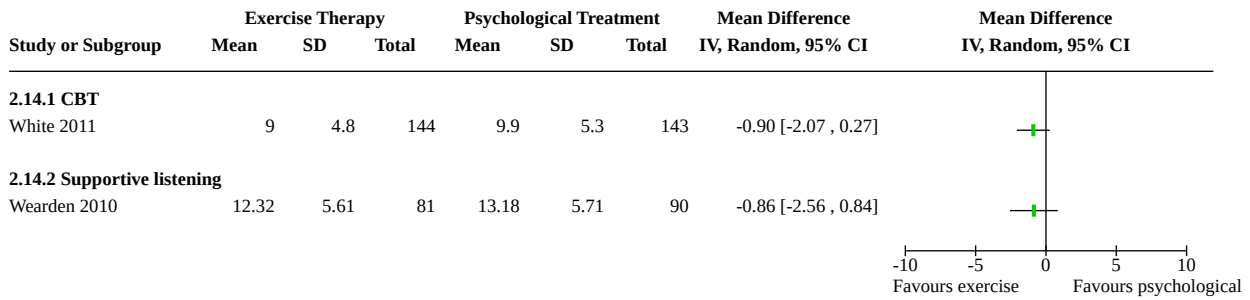
Analysis 2.12. Comparison 2: Exercise therapy versus psychological treatment, Outcome 12: Anxiety at follow-up (SMD)



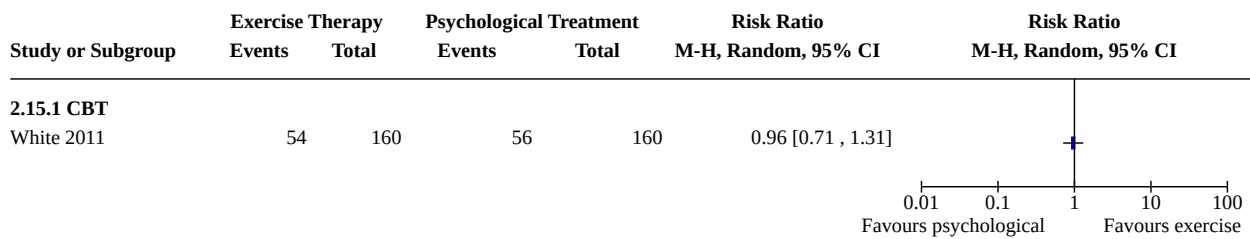
Analysis 2.13. Comparison 2: Exercise therapy versus psychological treatment, Outcome 13: Sleep at end of treatment (Jenkins Sleep Scale; 0 to 20 points)



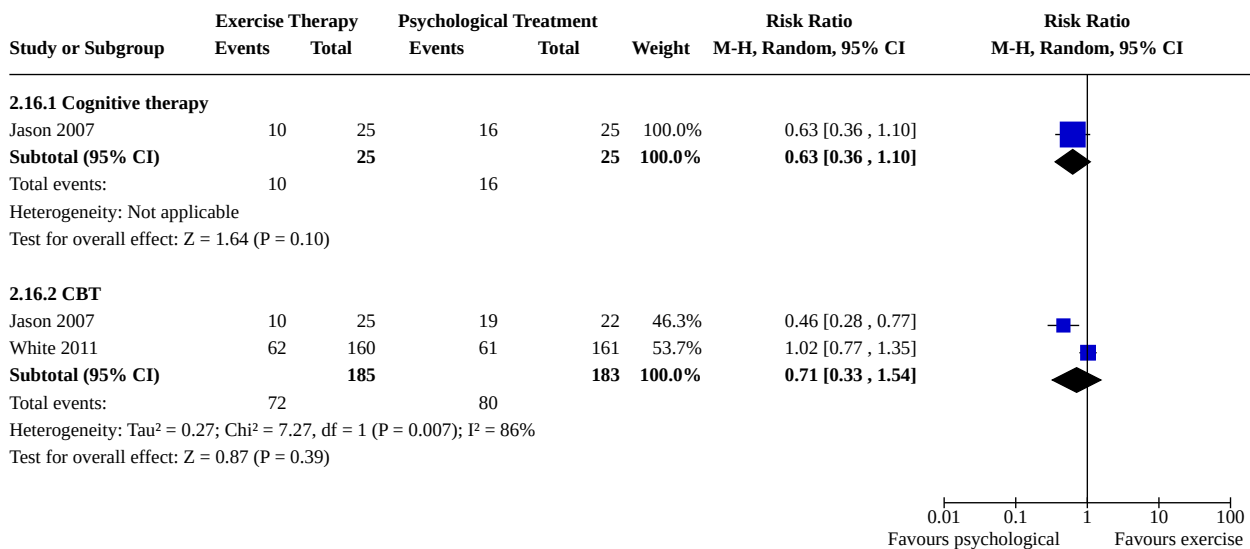
Analysis 2.14. Comparison 2: Exercise therapy versus psychological treatment, Outcome 14: Sleep at follow-up (Jenkins Sleep Scale; 0 to 20 points)



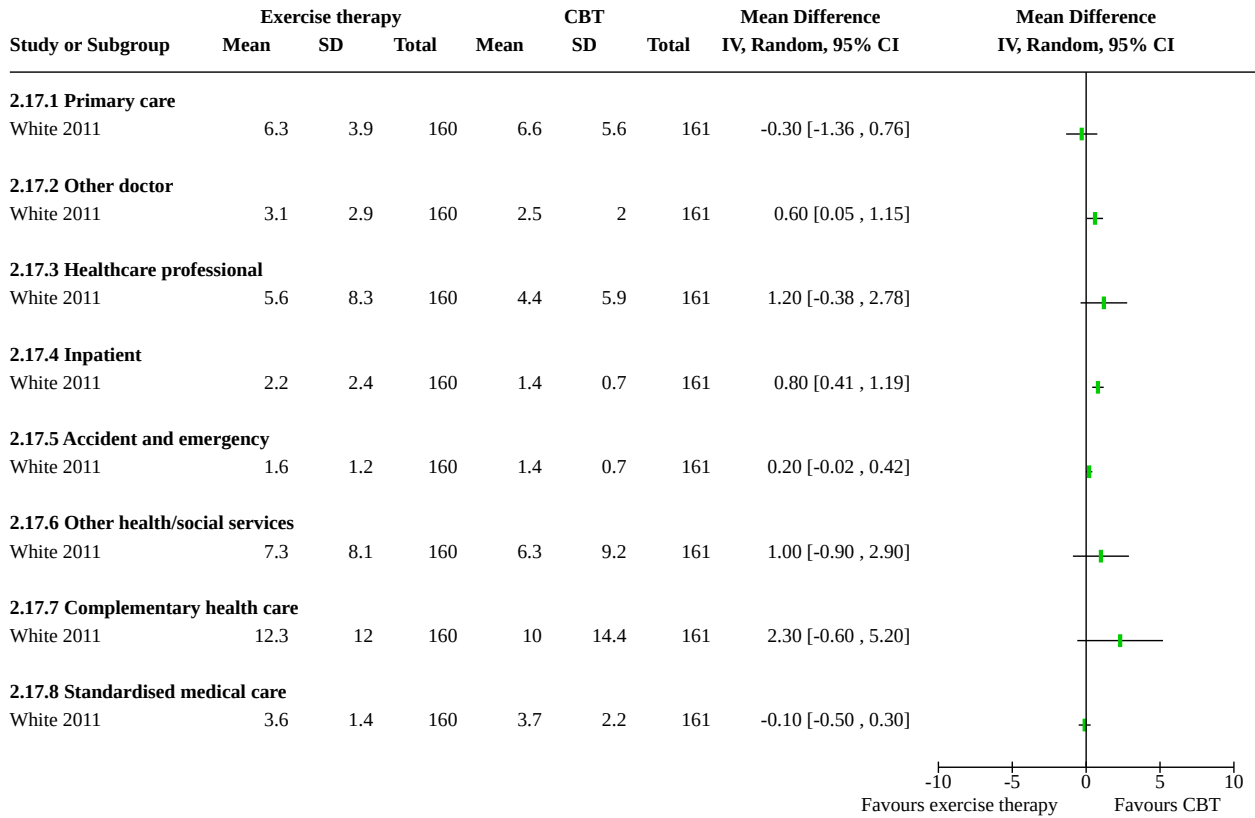
Analysis 2.15. Comparison 2: Exercise therapy versus psychological treatment, Outcome 15: Self-perceived changes in overall health at end of treatment



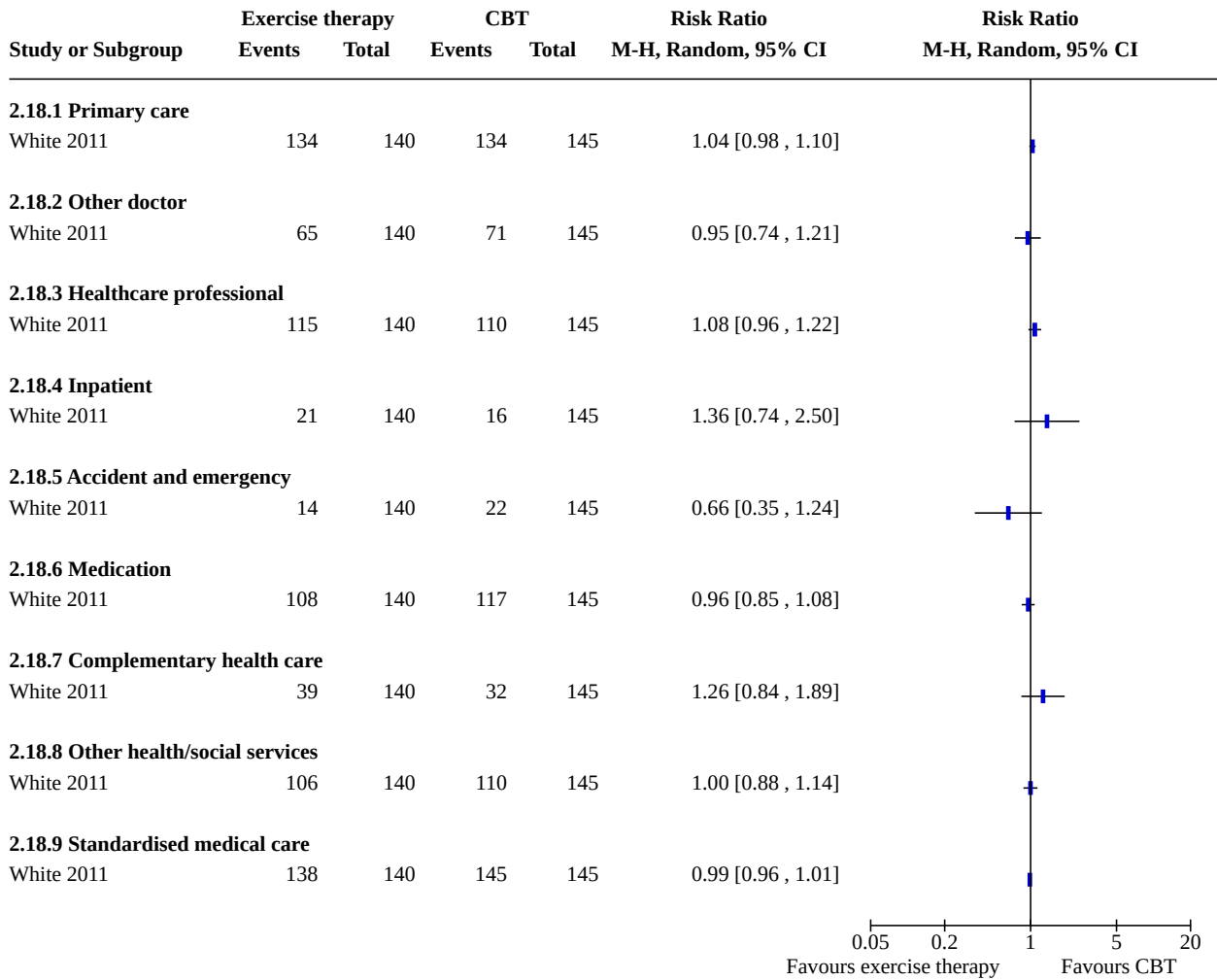
Analysis 2.16. Comparison 2: Exercise therapy versus psychological treatment, Outcome 16: Self-perceived changes in overall health at follow-up



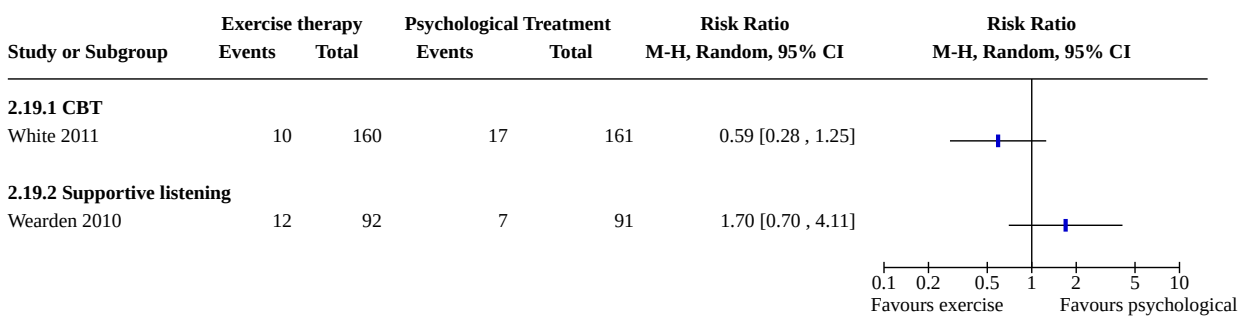
Analysis 2.17. Comparison 2: Exercise therapy versus psychological treatment, Outcome 17: Health resource use (follow-up) (Mean no. of contacts)



Analysis 2.18. Comparison 2: Exercise therapy versus psychological treatment, Outcome 18: Health resource use (follow-up) (No. of users)



Analysis 2.19. Comparison 2: Exercise therapy versus psychological treatment, Outcome 19: Dropout

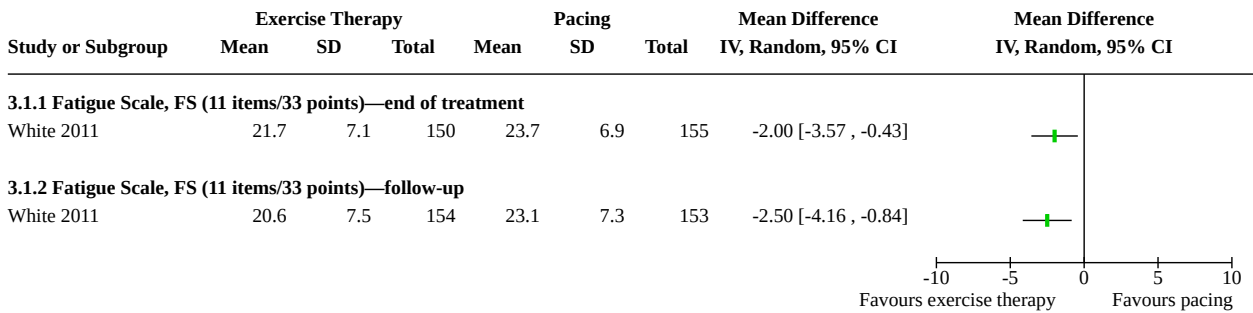


Comparison 3. Exercise therapy versus adaptive pacing

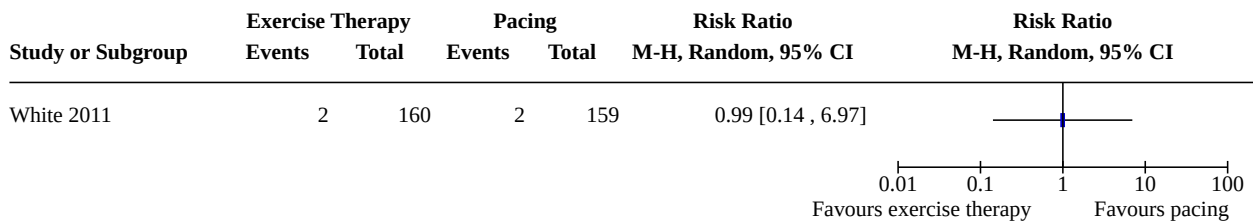
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.1 Fatigue	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.1.1 Fatigue Scale, FS (11 items/33 points)—end of treatment	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.1.2 Fatigue Scale, FS (11 items/33 points)—follow-up	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.2 Participants with serious adverse reactions	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.3 Physical functioning	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.3.1 SF-36, physical functioning subscale (0 to 100)—end of treatment	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.3.2 SF-36, physical functioning subscale (0 to 100)—follow-up	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.4 Depression	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.4.1 HADS, depression score (7 items/21 points)—follow-up	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.5 Anxiety	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.5.1 HADS, anxiety score (0 to 21 points)—follow-up	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.6 Sleep	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.6.1 Jenkins Sleep Scale (0 to 20 points)—follow-up	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.7 Self-perceived changes in overall health	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.7.1 End of treatment	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.7.2 Follow-up	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.8 Health resource use (follow-up) (Mean no. of contacts)	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.8.1 Primary care	1		Mean Difference (IV, Random, 95% CI)	Totals not selected

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
3.8.2 Other doctor	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.8.3 Healthcare professional	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.8.4 Inpatient	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.8.5 Accident and emergency	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.8.6 Other health/social services	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.8.7 Complementary health care	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.8.8 Standardised medical care	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
3.9 Health resource use (follow-up) (No. of users)	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.9.1 Primary care	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.9.2 Other doctor	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.9.3 Healthcare professional	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.9.4 Inpatient	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.9.5 Accident and emergency	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.9.6 Medication	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.9.7 Complementary health care	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.9.8 Other health/social services	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.9.9 Standardised medical care	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected
3.10 Dropout	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected

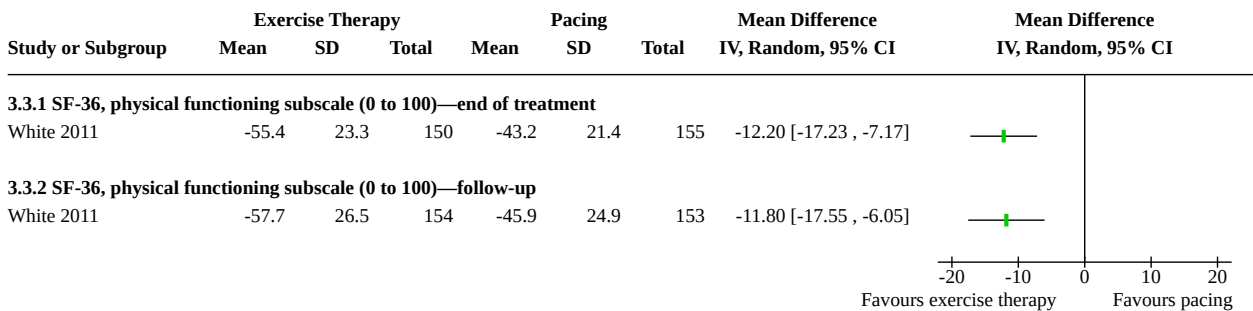
Analysis 3.1. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 1: Fatigue



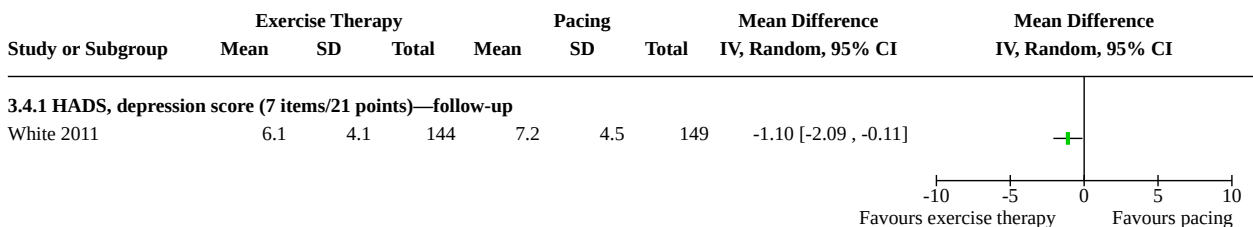
Analysis 3.2. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 2: Participants with serious adverse reactions



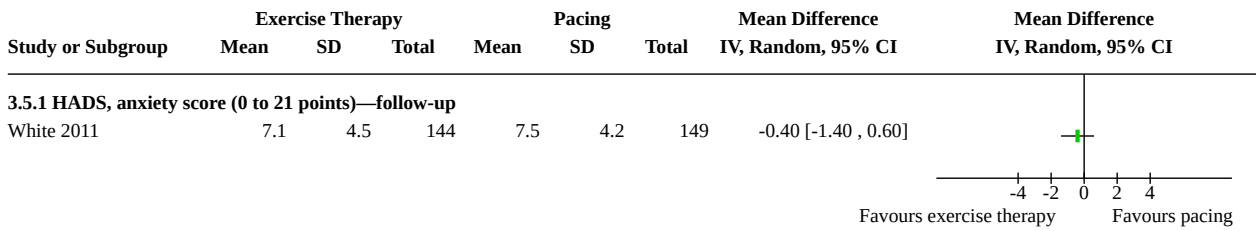
Analysis 3.3. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 3: Physical functioning



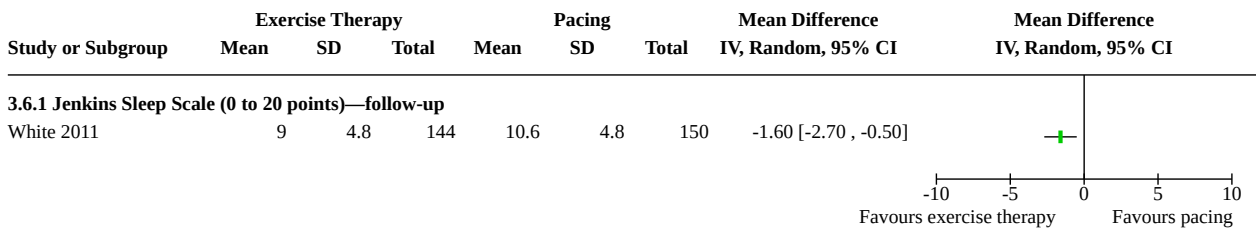
Analysis 3.4. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 4: Depression



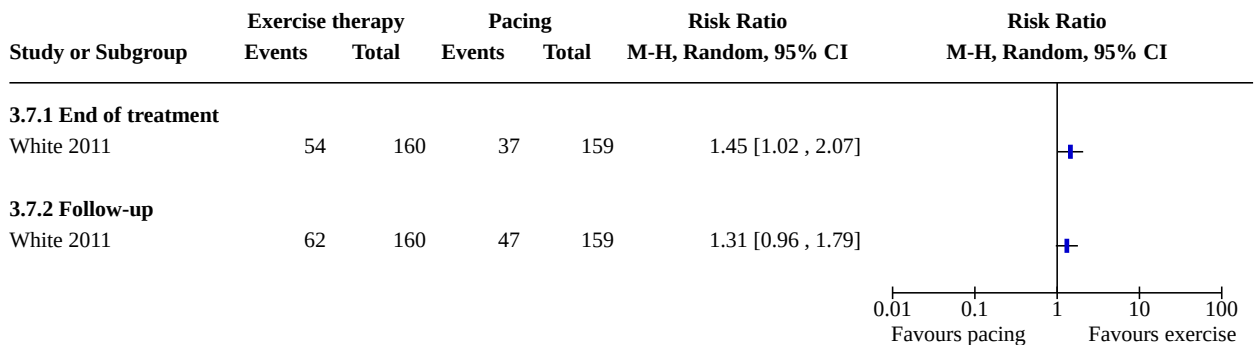
Analysis 3.5. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 5: Anxiety



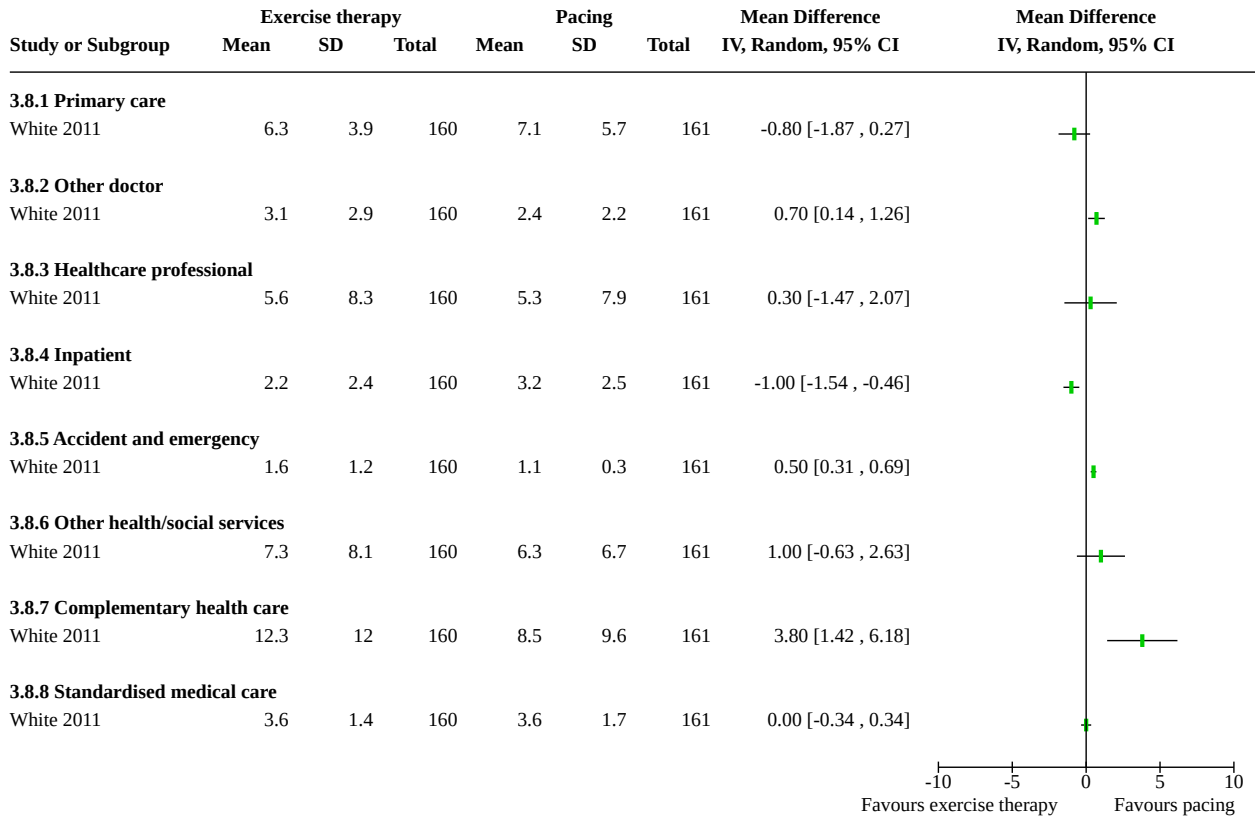
Analysis 3.6. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 6: Sleep



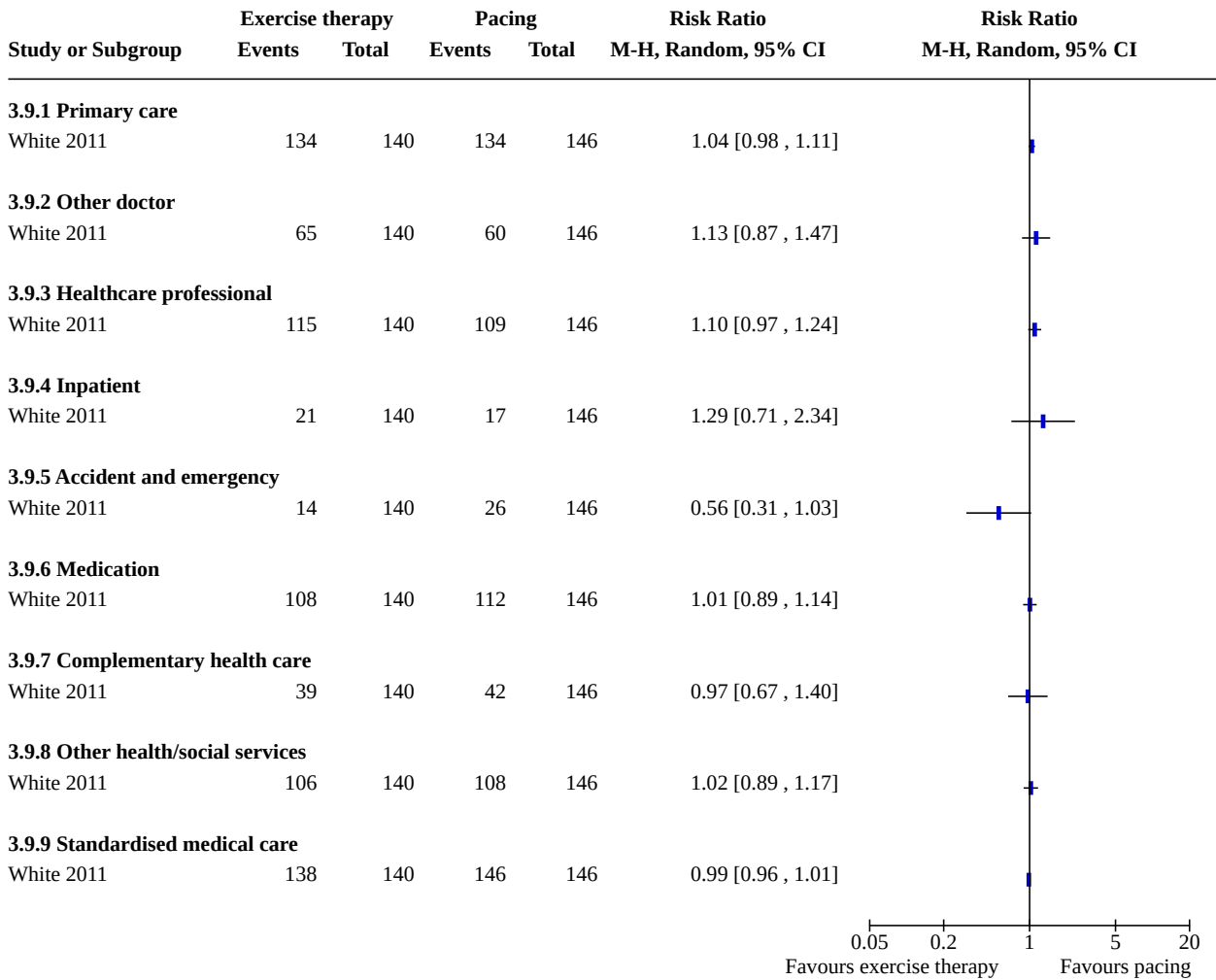
Analysis 3.7. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 7: Self-perceived changes in overall health



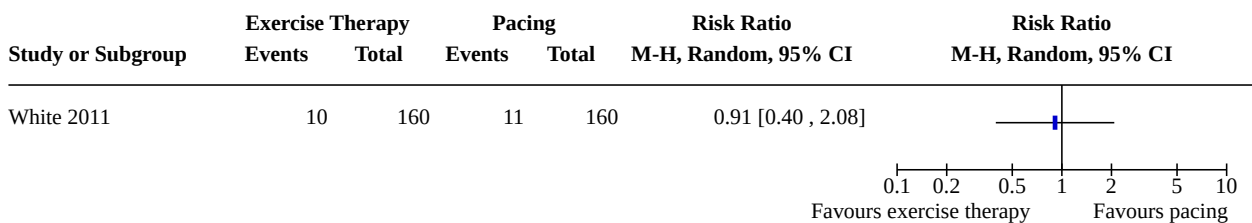
Analysis 3.8. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 8: Health resource use (follow-up) (Mean no. of contacts)



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



Analysis 3.10. Comparison 3: Exercise therapy versus adaptive pacing, Outcome 10: Dropout

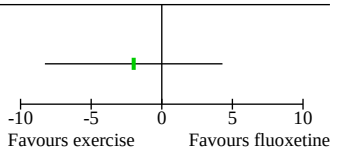


Comparison 4. Exercise therapy versus antidepressant

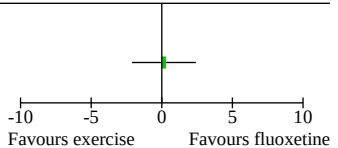
Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4.1 Fatigue	1		Mean Difference (IV, Random, 95% CI)	Totals not selected

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
4.1.1 Fatigue Scale, FS (14 items/0 to 42 points), end of treatment	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
4.2 Depression	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
4.2.1 HADS, depression score (7 items/21 points), end of treatment	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
4.3 Dropout	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected

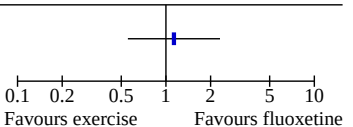
Analysis 4.1. Comparison 4: Exercise therapy versus antidepressant, Outcome 1: Fatigue

Study or Subgroup	Exercise therapy			Antidepressant (fluoxetine)			Mean Difference IV, Random, 95% CI	Mean Difference IV, Random, 95% CI
	Mean	SD	Total	Mean	SD	Total		
4.1.1 Fatigue Scale, FS (14 items/0 to 42 points), end of treatment								
Wearden 1998	28.13	13.05	23	30.12	8.49	25	-1.99 [-8.28, 4.30]	

Analysis 4.2. Comparison 4: Exercise therapy versus antidepressant, Outcome 2: Depression

Study or Subgroup	Exercise therapy			Antidepressant (fluoxetine)			Mean Difference IV, Random, 95% CI	Mean Difference IV, Random, 95% CI
	Mean	SD	Total	Mean	SD	Total		
4.2.1 HADS, depression score (7 items/21 points), end of treatment								
Wearden 1998	7.47	4.05	23	7.32	3.93	25	0.15 [-2.11, 2.41]	

Analysis 4.3. Comparison 4: Exercise therapy versus antidepressant, Outcome 3: Dropout

Study or Subgroup	Exercise therapy		Antidepressant (fluoxetine)		Risk Ratio M-H, Random, 95% CI	Risk Ratio M-H, Random, 95% CI
	Events	Total	Events	Total		
Wearden 1998	11	34	10	35	1.13 [0.55, 2.31]	

Comparison 5. Exercise therapy + antidepressant versus antidepressant

Outcome or subgroup title	No. of studies	No. of participants	Statistical method	Effect size
5.1 Fatigue	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
5.1.1 Fatigue Scale, FS (14 items/0 to 42 points), end of treatment	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
5.2 Depression	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
5.2.1 HADS, depression score (7 items/21 points), end of treatment	1		Mean Difference (IV, Random, 95% CI)	Totals not selected
5.3 Dropout	1		Risk Ratio (M-H, Random, 95% CI)	Totals not selected

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

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

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

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

Analysis 5.3. Comparison 5: Exercise therapy + antidepressant versus antidepressant, Outcome 3: Dropout

Study or Subgroup	Exercise + antidepressant		Antidepressant		Risk Ratio M-H, Random, 95% CI	Risk Ratio M-H, Random, 95% CI
	Events	Total	Events	Total		
Wearden 1998	14	33	10	35	1.48 [0.77, 2.87]	

ADDITIONAL TABLES
Table 1. Study demographics

Study ID	N	Gender	Duration of illness	Depression comorbidity	Use of anti-depressants (ADs)	Work and employment status
Fulcher 1997	66	49 F/17 M 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	95 F/19 M 83% female	> 5.0 years	44 (39%) with a current Axis I disorder (depression and anxiety most common)	NS	52 (46%) working or studying at least part time, 24% unemployed, 6% retired, 25% on disability
Moss-Morris 2005	49	34 F/15 M 69% female	3.1 years	14 (29%) possible or probable cases of depression (HADS)	NS	11 (22%) were unemployed and were unable to work because of disability
Powell 2001	148	116 F/32 M 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
Wallman 2004	61	47 F/14 M 77% female	NS	Mean HADS depression score at baseline was 6.8 points	16 (26%) used AD	No detectable initial difference between the groups (numbers not reported)
Wearden 1998	136	97 F/39 M 71% female	2.3 years	46 (34%) with depressive disorder according to DSM-III-R criteria	NS	114 (84%) had recently changed occupation
Wearden 2010	296	230 F/66 M 78% female	7.0 years	53 (18%) had a depression diagnosis	160 (54%) were prescribed AD in the past 6 months	NS
White 2011	641	495 F/146 M 77% female	2.7 years	219 (34%) with any depressive disorder	260 (41%) used AD	NS

AD: antidepressant; **DSM-III-R:** Diagnostic and Statistical Manual of Mental Disorders from the American Psychiatric Association, 3rd edition (Revised) **F:** female; **HADS:** Hospital Anxiety and Depression Scale; **M:** male; **NS:** 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 h), talking only	5 d/week	5-15 min increasing to 30 min/d	5-15 min at 40% of peak O ₂ consumption (target HR of resting + 50% of HRR)	Duration increased 1-2 min/week up to 30 min then intensity increased	Ambulatory HR monitors	If increased fatigue, continue at the same level for an extra week
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 min), 13 sessions	3/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
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	4-5 d/wk	Set collaboratively approx 5-15 min	HR at 40% of VO ₂ max	Duration 3-5 min/week Intensity increased after 6 weeks 5 bpm/week	Ambulatory HR monitors	If increased fatigue, continue at the same level for an extra week
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 h 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 min twice/d	Duration of exercise	Discouraged, but restart at lower level and rapidly re-increase



Table 2. Characteristics of exercise interventions (Continued)

Wallman 2004	Single physical therapist	Small laminated Borg Scale and HR monitor	Walking/jogging, swimming or cycling	Phone contact every 2 weeks	Every second day	From 5-15 min, increasing to 30 min	Initial exercise duration was between 5 and 15 min, and intensity was based on the mean HR value achieved midpoint during submaximal exercise tests	Duration increased by 2-5 min/2 wk	HR monitoring, Borg Exertion Scale	Keep Borg within 11-14. Adjust every 2 weeks. Average peak HR when exercising comfortably at a typical day represents participant's target HR (± 3 bpm) for future sessions
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 d/week	20 min	75% of VO ₂ max from bike test	Intensity increased	Borg Exertion Scale chart, before and after HR	Increase if: 10 bpm drop post-exercise and 2-point drop in Borg Scale score
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/d	First 90 min, then alternating 60 and 30 min	Determined collaboratively with the participant	"Increased very gradually," examples show 50% increase/d	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 (8-10 d 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 x 4, then fortnightly; total of 15 sessions	5-6 d/week	Negotiated, goal to get to 30 min/session	Test of fitness (step test and 6MWT), perceived physical exertion, actigraph data	"20% increases" per fortnight; increase duration to 30 min, then increase intensity	Exercise diary + Borg scale + "Use non-symptoms to monitor" and HR monitor	Do not increase if global increase in symptoms

(for in-
tensity in-
creases)

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

bpm: beats per minute; **GET:** graded exercise therapy; **HR:** heart rate; **HRR:** heart rate reserve; **VO₂:** oxygen consumption; **6MWT:** six-minute walking test

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

APPENDICES

Appendix 1. Search strategy—CCMDCTR-References

CCMDCTR-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

Cochrane Central Register of Controlled Trials (CENTRAL) (2014, Issue 4)

#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
- #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)

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

International trials registers

World Health Organization International Clinical Trials Portal available at 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 submitted, 2 December 2018

Summary

Recently I have published a reanalysis of this Cochrane review. Unfortunately there are many problems with the review and the trials in it. For example, P-Hacking, extensive endpoint changes, overlap in entry/recovery criteria, selecting patients who don't have the disease, ignoring null effects, relying on subjective outcomes in unblinded trials and ignoring the absence of objective improvement. The reanalyses which looked at the objective outcomes showed that graded exercise therapy is not an effective treatment for ME/CFS. The studies in the review do not provide any evidence that graded exercise therapy is safe, on the other hand, patient evidence and the literature show that it is not safe.

The open access reanalysis can be read here: <https://journals.sagepub.com/doi/full/10.1177/2055102918805187>

Reply

Many thanks for your feedback on this review. Cochrane recognises the importance of the review and is committed to providing a high quality review that reflects the best current evidence to inform decisions. The Editor-in-Chief is currently holding discussions with colleagues and the author team to determine a series of steps that will lead to a full update of this review. Your feedback will be considered as part of this process so that it can inform future versions of the review. These discussions will be concluded as soon as possible.

Contributors

Feedback submitted by: Mark Vink

Response: Jessica Hendon (Managing Editor of the Cochrane Common Mental Disorders Review Group)

Feedback submitted, 5 November 2018

Summary

A few questions about where the disease ME/CFS will be placed by Cochrane in the future. If Cochrane moves myalgic encephalomyelitis/chronic fatigue syndrome (ME/CFS) into the Long term conditions and Aging Network, will it be moved into the Metabolic and Endocrine Disorders review group within this network? I'm making this assumption, based on the metabolic abnormalities, found in people with ME/CFS, when objective metabolic exercise tests, are carried out, as per "Cardiopulmonary Exercise Test Methodology for Assessing Exertion Intolerance in Myalgic Encephalomyelitis/Chronic Fatigue Syndrome" - <https://www.ncbi.nlm.nih.gov/pubmed/30234078>? If, in the alternative, Cochrane decides that ME/CFS is to remain in the Brain, Nerves and Mind (BNM) Network, will it be moved into a separate group of its own? As whilst the disease fits within the BNM Network, it doesn't fit into any of the listed Cochrane Review Groups. The closest fit is probably the Multiple Sclerosis and rare diseases of the CNS? However, given that ME/CFS is thought to be more prevalent than multiple sclerosis and is not rare, it doesn't really fit into this group. Will a new Cochrane Review Group, be made for ME/CFS, that is in line with the published biomedical and physiological findings?

Reply

Many thanks for your follow-on comments related to Cochrane's decision to consider repositioning its chronic fatigue syndrome (CFS)/myalgic encephalomyelitis (or encephalopathy) (ME) reviews. The repositioning of the editorial oversight of CFS/ME reviews is ongoing. Your feedback has been forwarded to the Cochrane Editor-in-Chief so that it can be considered as part of this process.

Contributors

Feedback submitted by: Adrienne Wooding

Response: Jessica Hendon (Managing Editor of the Cochrane Common Mental Disorders Review Group)

Feedback submitted, 18 October 2018**Summary**

It has been raised by others that the Cochrane review erroneously places ME/CFS in its mental health category. The response provided by Cochrane when this issue was raised does not inspire confidence in their knowledge/advice on the disease. ME/CFS was the subject of a comprehensive literature review carried out by the USA National Academy, published in 2015 this report categorically determined that the disease ME/CFS is not a mental health disorder. The large volume of biomedical research findings of a wide range of organic abnormalities is also at odds with a mental health disorder. Further more the World Health Organisation has categorised ME/CFS as a neurological disease. The National Centre for Neuroimmune and Emerging Diseases has patented a blood test for the disease and is in the early stages of validating it. It would be much appreciated if Cochrane would categorise ME/CFS in the appropriate group i.e. along with other neurological diseases such as Parkinson's, Huntington's, multiple sclerosis etc. Delighted to see that the latest review has been suspended and look forward to its replacement with a review that bases its findings on OBJECTIVE outcome data.

Reply

Many thanks for your comment and for noting recent categorisations of Chronic fatigue syndrome (CFS)/myalgic encephalomyelitis (or encephalopathy) (ME). Feedback on reviews is normally dealt with by the relevant review author, but in this case as your query relates more to an organisational management issue, we are responding on behalf of the Cochrane Common Mental Disorders (CMD) Review Group and the Cochrane Editor-in-Chief.

We value your observations about the placement of CFS/ME reviews in The Cochrane Library. We want our evidence to properly support those with lived experience of CFS/ME and to ensure that the CFS/ME community have confidence in our portfolio of reviews. We are also aware that the hosting of this topic by the Cochrane CMD Review Group has been antagonistic to some in the CFS/ME community.

Cochrane has recently created eight new Networks of Cochrane Review Groups (CRGs). The formation of these networks provides a timely opportunity to review the scope of all CRGs and to consider changes where appropriate. In response to concerns raised by members of the CFS/ME community, Cochrane has been considering repositioning the editorial oversight of CFS/ME reviews. The Cochrane CMD Review Group currently sits within the Brain, Nerves and Mind (BNM) Network. In the future, reviews on this topic might sit with another Cochrane Review Group within the BNM Network, or they might transfer to another Network altogether, such as the Long Term Conditions and Ageing 2 Network. Please be reassured that this is currently under consideration and a decision is anticipated before the end of 2018.

We would also like to refer you to the recent published note for the latest information about the status of this particular review 'This review is subject to an ongoing process of review and revision following the submission of a formal complaint to the Editor-in-Chief. Cochrane considers all feedback and complaints carefully, and revises or updates reviews when it is appropriate. The review author team have advised us that a resubmission of this review is imminent. A decision on the status of this review will be made once this resubmission has been through editorial process, which we anticipate will be towards the end of November 2018'.

Contributors

Feedback submitted by: Adrienne Wooding

Response: Peter Coventry and Jessica Hendon

Peter Coventry is the Feedback Editor of the CMD Review Group and Jessica Hendon is the Managing Editor of the CMD Review Group. No other conflicts of interest declared.

Feedback submitted, 16 June 2017**Summary**

Comment: I'm concerned regarding your conclusion that no evidence suggests that exercise therapy may worsen outcome, as you have stated that no conclusions were possible for the drop-out rate.

Whilst I appreciate that you are unable to draw conclusions about drop-out rates due to insufficient data, is it perhaps potentially misleading or ambiguous to summarise that in general patients may benefit from GET with there being no evidence for symptoms worsening, when there are a researchers that support the claim that CBT/GET is detrimental to the long term prognosis of patients with ME/CFS. Without assessment of data concerning those whom have dropped out (those most likely to experience worsening symptoms) the conclusions you have stated could prove harmful if taken as encouragement for GPs to place their patients on GET regimes.

I do not question your analysis of the data, but rather I am concerned with the way in which you have expressed your findings.

Reply

Thank you for your interest in the review and your comment.

In our systematic review, we aim to summarise the effect estimates associated with the use of exercise therapy for patients diagnosed with chronic fatigue syndrome (CFS)/ME. We decided to rely on data from randomised controlled trials (RCT), as RCTs provide much more robust data than for example anecdotal evidence. We held serious adverse reactions (SAR) and serious adverse events (SAE) as our primary outcome, whereas the drop-out rate was added as a secondary outcome.

Systematic reviews based on aggregated data dependent on the data reported in the included trials. One trial reported that SARs and SAEs were rare in both groups, suggesting that the difference between the groups is small when measured in absolute terms. Analysis of drop-out rates did not reveal statistical differences between the groups, and we cannot conclude that exercise is associated with higher drop-out rates. Even if we had seen differences between the groups, however, drop-out rates must be interpreted with caution. It is important to be aware that drop-out is not a direct measure of harm. There might be several reasons patients drop out, and some of these reasons are not expected to distribute equally between the groups. Harm is one possible reason for drop-out, but patients may also withdraw because they are unhappy with the randomisation (preconceptions), because they feel better or because they don't experience the expected level of improvement etc.

Systematic reviews aim to bring the best evidence to the clinical encounter, but shared decision making includes patient preferences and clinical expertise when a treatment plan is decided upon.

Contributors

Feedback submitted by: Richard Gardner

Response submitted by: Lillebeth Larun

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 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 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 assignment 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.

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 than 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

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 ($I^2 = 78\%$, $P < 0.0001$; 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 ($I^2 = 13\%$, $P = 0.33$)."

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 ($I^2 = 93\%$, $P < 0.00001$)."

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, Schmalong 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.
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 than 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. *Psychological 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

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

"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

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].

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 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

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 reevaluation 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 [...]"

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 reevaluation 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]

"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 cephalopathy. *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

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 0-3 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 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:

- Larun L, Brurberg KG, Odgaard Jensen J, Price JR. Exercise therapy for chronic fatigue syndrome. Cochrane Database Syst Rev. 2016; CD003200.
- 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 not affect the conclusion of the review.

Regards,

Lillebeth Larun

Contributors

Feedback submitted by: Robert Courtney

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 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 diagnostic criteria 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.pacetrials.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
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

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 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, analyses 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.
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 Behavioral 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

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

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.

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.

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?

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.

WHAT'S NEW

Date	Event	Description
18 March 2021	Amended	Editorial Note amended to correct an error in the website address.

HISTORY

Protocol first published: Issue 3, 2001

Review first published: Issue 3, 2004

Date	Event	Description
1 March 2021	Amended	A note on the status of the review has been moved from the Abstract to an Editorial note.
15 June 2020	Amended	The comments received about the review have been reordered to present the most recent comments first.
21 May 2020	Amended	Addition of the following text to the beginning of the Abstract, 'A statement from the Editor in Chief about this review and its planned update is available here: www.cochrane.org/news/pub-

Date	Event	Description
		lication-cochrane-review-exercise-therapy-chronic-fatigue-syn-drome.
12 March 2020	Amended	Note added from the editorial team at Cochrane Editorial and Methods Department on 12 March 2019, 'A webpage providing information and regular updates on the progress of the planned update of this Cochrane Review is available here: community.cochrane.org/organizational-info/people/central-executive-team/editorial-methods/projects/stakeholder-engagement-high-profile-reviews-pilot '.
13 February 2020	Amended	Added text to clarify the date and nature of the changes made to the review version published on 2 October 2019 (doi.org/10.1002/14651858.CD003200.pub8). Specifically, text has been added to the events below dated 8 August 2019 and the published note section of the review.
6 February 2020	Amended	Addition of new published note from the editorial team at Cochrane Editorial and Methods Department, 'A statement from the Editor in Chief about this review and its planned update is available here: www.cochrane.org/news/publication-cochrane-review-exercise-therapy-chronic-fatigue-syndrome '.
8 August 2019	New citation required and conclusions have changed	The following changes were made as part of the review version published on 2 October 2019 (doi.org/10.1002/14651858.CD003200.pub8). Interpretation of the evidence now reflects the following changes to the review: <ul style="list-style-type: none"> • restructuring the analysis of fatigue to combine data with standardised mean differences; • inclusion of long-term effects on fatigue as a 'Summary of findings' table outcome; • changed GRADE rating for adverse reactions; and • acknowledgement of the criteria used to define chronic fatigue syndrome by study investigators. This event, generated on 8 August 2019, relates to the review version published on 2 October 2019 (doi.org/10.1002/14651858.CD003200.pub8).
8 August 2019	Amended	Review amended in response to a formal complaint and editorial review. See new published note for details. This event generated on 8 August 2019, relates to the review version published on 2 October 2019 (doi.org/10.1002/14651858.CD003200.pub8).
17 June 2019	Amended	Addition of new published note, 'Cochrane's Editor in Chief has received the revised version of the review from the author team with changes made in response to the complaint by Robert Courtney. The process has taken longer than hoped; the amended review is being finalised and it will be published during the next 2 months.'
8 March 2019	Amended	Addition of new published note 'Cochrane's editors and the review author team have jointly agreed that there will be a further

Date	Event	Description
		period up to the end of May 2019, in which time the author team will amend the review to address changes aimed at improving the quality of reporting of the review and ensuring that the conclusions are fully defensible and valid to inform health care decision making. The changes will also address concerns raised in feedback since the Robert Courtney complaint. The amendment will not include a full update, but a decision about this will be made subsequently.'
5 December 2018	Feedback has been incorporated	Feedback has been added, along with a response from the Cochrane Common Mental Disorders (CMD) Review Group
30 November 2018	Amended	<p>Addition of new published note 'The author team has re-submitted a revised version of this review following the complaint by Robert Courtney. The Editor in Chief and colleagues recognise that the author team has sought to address the criticisms made by Mr Courtney but judge that further work is needed to ensure that the review meets the quality standards required, and as a result have not approved publication of the re-submission. The review is also substantially out of date and in need of updating.</p> <p>Cochrane recognises the importance of this review and is committed to providing a high quality review that reflects the best current evidence to inform decisions.</p> <p>The Editor in Chief is currently holding discussions with colleagues and the author team to determine a series of steps that will lead to a full update of this review. These discussions will be concluded as soon as possible'.</p>
9 November 2018	Feedback has been incorporated	Feedback has been added, along with a response from the Cochrane Common Mental Disorders (CMD) Review Group
2 November 2018	Feedback has been incorporated	Feedback has been added, along with a response from the Cochrane Common Mental Disorders (CMD) Review Group
25 October 2018	Amended	<p>Addition of new published note 'This review is subject to an ongoing process of review and revision following the submission of a formal complaint to the Editor in Chief. Cochrane considers all feedback and complaints carefully, and revises or updates reviews when it is appropriate. The review author team have advised us that a resubmission of this review is imminent. A decision on the status of this review will be made once this resubmission has been through editorial process, which we anticipate will be towards the end of November 2018'.</p>
5 October 2017	Feedback has been incorporated	Feedback has been added, along with the authors' response
5 May 2017	Feedback has been incorporated	Feedback has been added, along with the authors' response.
21 June 2016	Feedback has been incorporated	Feedback has been added, along with the authors' response.
1 February 2016	Feedback has been incorporated	Feedback has been added along with the authors' response.
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.

Date	Event	Description
2 October 2014	New search has been performed	This review has been updated with newer methodology, and new studies have been incorporated.
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 studies 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

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 2004 version of the review

We changed the Objectives 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 for adults with CFS compared with any other intervention or control' in this update.

We changed comparisons 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' 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 in a particular routine.
- Psychological therapies: CBT/cognitive therapy/supportive therapy/behavioural therapies/psychodynamic therapies
- Adaptive pacing therapy
- Pharmacological therapy

We have revised and reordered the list of secondary outcomes for clarity and have added self-perceived 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* (Higgins 2011c). For the first version of this review (Edmonds 2004), the review authors conducted assessment of methodological quality according to contemporary criteria of the handbook of The Cochrane Collaboration (Alderson 2004). Review authors rated the adequacy of allocation concealment as adequate (A), unclear (B) or inadequate (C) or as not used (D), and applied the Cochrane Collaboration Depression, Anxiety and Neurosis (CCDAN) Quality Rating System (Moncrieff 2001). For this update, we re-extracted data on risk of bias to comply with current recommendations, and we used concealment of allocation as the main quality criterion for included studies (Higgins 2011a).

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.

Re-expressing standardised mean differences and defining minimal important differences

If available studies measured the same outcome using different scales or different versions of the same scale, we presented the pooled effect estimates in terms of standardised mean differences (SMD). SMD units may be hard to interpret, and based on feedback from the Cochrane Editorial and Methods Department we have also calculated/re-expressed SMD using more familiar instruments. Chapter 12 in the *Cochrane Handbook for Systematic Reviews of Interventions* (Schünemann 2011), recommends using standard deviation from a representative observational study as a basis for the recalculation, and we therefore use standard deviations reported in Crawley 2013 for this purpose.

Post hoc, we have also been encouraged to define minimal important differences (MID) for commonly used outcome measures. MID thresholds and relevant research literature are reported under [Measures of treatment effect](#) in the Methods section.

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 intra-class 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).

We did not identify any cluster-randomised controlled trials (RCTs) 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 intra-class correlation coefficients 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 intra-class correlation coefficients 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 chronic fatigue syndrome, randomised cross-over trials 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 studies 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, we will combine means and standard deviations using the methods described in Chapter 7 (Section 7.7.3.8) of the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011b).

Multiple medications

Some studies may combine several interventions with one comparison group. Should we identify studies 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; [Higgins 2011d](#)).

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).

NOTES

Note added from the editorial team at Cochrane Editorial and Methods Department on 12 March 2019: A webpage providing information and regular updates on the progress of the planned update of this Cochrane Review is available here: community.cochrane.org/organizational-info/people/central-executive-team/editorial-methods/projects/stakeholder-engagement-high-profile-reviews-pilot.

Previously published note

February 2020

Note added from the editorial team at Cochrane Editorial and Methods Department on 6 February 2020: A statement from the Editor in Chief about this review and its planned update is available here: www.cochrane.org/news/publication-cochrane-review-exercise-therapy-chronic-fatigue-syndrome.

August 2019

This published note, generated in August 2019, relates to the review version published on 2 October 2019 (doi.org/10.1002/14651858.CD003200.pub8).

In 2018, following receipt of a formal complaint about the Cochrane Review, 'Exercise therapy for chronic fatigue syndrome', the then Editor-in-Chief of the Cochrane Library, Dr David Tovey, commissioned an appraisal of the published review by his team. The findings of this assessment were shared with the authors in September 2018. It was judged that the authors could have an opportunity to address the complaint by amending the published review, instead of withdrawing it.

The authors submitted an amended version of the review, which was assessed further by independent editors in November 2018. Following their assessment, in December 2018 the authors were asked to make additional changes. In early 2019, the Editor-in-Chief and the authors jointly agreed to an extension until the end of May 2019 to address all the comments. This amended review has now been accepted for publication by Dr Karla Soares-Weiser, who took over as Editor-in-Chief in June 2019.

June 2019

Cochrane's Editor-in-Chief has received the revised version of the review from the author team with changes made in response to the complaint by Robert Courtney. The process has taken longer than hoped; the amended review is being finalised and it will be published during the next two months.

March 2019

Cochrane's editors and the review author team have jointly agreed that there will be a further period up to the end of May 2019, in which time the author team will amend the review to address changes aimed at improving the quality of reporting of the review and ensuring that the conclusions are fully defensible and valid to inform health care decision making. The changes will also address concerns raised in feedback since the Robert Courtney complaint. The amendment will not include a full update, but a decision about this will be made subsequently.

November 2018

The author team has re-submitted a revised version of this review following the complaint by Robert Courtney. The Editor-in-Chief and colleagues recognise that the author team has sought to address the criticisms made by Mr Courtney but judge that further work is needed to ensure that the review meets the quality standards required, and as a result have not approved publication of the re-submission. The review is also substantially out of date and in need of updating.

Cochrane recognises the importance of this review and is committed to providing a high quality review that reflects the best current evidence to inform decisions.

The Editor-in-Chief is currently holding discussions with colleagues and the author team to determine a series of steps that will lead to a full update of this review. These discussions will be concluded as soon as possible.

October 2018

This review is subject to an ongoing process of review and revision following the submission of a formal complaint to the Editor-in-Chief. Cochrane considers all feedback and complaints carefully, and revises or updates reviews when it is appropriate. The review author team have advised us that a resubmission of this review is imminent. A decision on the status of this review will be made once this resubmission has been through editorial process, which we anticipate will be towards the end of November 2018.

February 2015

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 Behavioral Therapy; Depression [therapy]; Exercise; Exercise Therapy [adverse effects] [*methods]; Fatigue Syndrome, Chronic [psychology] [*therapy]; Health Status; Patient Dropouts [statistics & numerical data]; Quality of Life; Randomized Controlled Trials as Topic; Sleep Wake Disorders [therapy]

MeSH check words

Adult; Humans