eman ta zabal zazu

Universidad del País Vasco

Euskal Herriko Unibertsitatea

MEDIKUNTZA ETA ERIZAINTZA FAKULTATEA

FACULTAD DE MEDICINA Y ENFERMERÍA

MÁSTER EN NEUROCIENCIAS

2019-2020

Trabajo de Fin de Máster

FIBROMYALGIA

Effectiveness of multidisciplinary treatments for fibromyalgia: A systematic review.

Author: Varenka Oca Navarro

Tutor: Francisco Javier Ballesteros Rodríguez

INDEX

ABSTRACT

Introduction The American College of Rheumatology (ACR) defines fibromyalgia (FM) as a neurologic chronic disorder. Widespread musculoskeletal pain is accompanied by a wide range of symptoms and co-morbid health problems. Studies suggest that prevalence in the general population vary between 2% and 4% with a peak incidence in middle-aged females. FM is considered a public health problem, due to its high incidence among adult population, the lack of knowledge concerning etiology in addition to the non-curative available treatments. There is none specific drug for FM and as medical management is usually partially successful, non-pharmacological treatments are accomplished for a bio-psycho-social approach, comprising the biological, psychological and social factors that encompass the syndrome. The combination of approaches such as education, exercise and psychotherapeutic interventions seem to create synergy together with pharmacological therapies so multidisciplinary treatment programs have been recommended for FM syndrome.

Objective The main objective of this study is to determine the efficacy of multidisciplinary treatments for FM.

Methodology 119 articles were obtained from both MEDLINE and Cochrane databases and 19 clinical trials were extracted from two systematic reviews obtained from an overview of treatments for FM. Finally, 17 suitable RCTs were included in the present systematic review. Improvement on pain, fatigue, quality of sleep, physical function, depression and/or anxiety were measured. GRADEpro GDT was used to assess the quality of the evidence.

Results Low evidence was found for physical function and anxiety while very low evidence was found for pain, fatigue, quality of sleep and depression. No clinically meaningful results were found for any of the assessed outcomes.

Conclusions The evidence is very uncertain about the effectiveness of multidisciplinary treatments in FM. A great heterogeneity was found with regard to multidisiciplinary treatments and the evaluation criteria.

KEY ABBREVIATIONS

ACR	American College of Rheumatology
BDI	Beck Depression Inventory
CBT	Cognitive Behavioral Therapy
CNS	Central Nervous System
FIQ	Fibromyalgia Impact Questionnaire
FM	Fibromyalgia
HADS	The Hospital Anxiety and Depression Scale
MOSS-S	Medical Outcome Study Sleep Scale
PGWB	Psychological General Well-Being Index

ABSTRACT

Introduction El Colegio Americano de Reumatología (ACR) define fibromialgia (FM) como un trastorno neurológico crónico. Dolor musculo-esquelético generalizado se acompaña de una gran variedad de síntomas y problemas de salud co-mórbidos. La literatura revela que la prevalencia en la población general varía entre 2-4%, presentando un pico de incidencia en mujeres de mediana edad. La FM es considerada un problema de salud pública debida a su alta incidencia entre la población adulta, la falta de conocimiento sobre su etiología además de los tratamientos no-curativos disponibles para paliar los síntomas de la enfermedad. No existe un fármaco específico para ésta afección, y ya que las terapias farmacológicas suelen ser parcialmente exitosas, se llevan a cabo tratamientos no farmacológicos para un abordaje bio-psico-social, que abordan los factores biológicos, psicológicos y sociales que constituyen el síndrome. La combinación de abordajes parece crear sinergia junto con terapias farmacológicas así que programas de tratamiento multidisciplinares han sido recomendados para el síndrome de FM.

Objective El objetivo principal de este estudio es determinar la eficacia de tratamientos multidisciplinares para FM.

Methodology Se obtuvieron 119 artículos de las bases de datos MEDLINE y Cochrane y se extrajeron 19 ensayos clínicos de dos revisiones sistemáticas obtenidas de una revisión de mapeo realizada anteriormente. Finalmente, se incluyeron 17 ECAs en la presente revisión sistemática. Se midió la mejoría en dolor, fatiga, calidad del sueño, función física depresión y/o ansiedad. Para evaluar la calidad de la evidencia se utilizó GRADEpro GDT.

Results Se encontró baja evidencia para función física y ansiedad mientras que la calidad fue muy baja para dolor, fatiga, calidad de sueño y depresión. No se encontraron resultados clínicamente significativos para ninguno de los dominios.

Conclusions La evidencia sobre la efectividad de los tratamientos multidisciplinares para FM es muy incierta. Se encontró una gran heterogeneidad en relación con los tratamientos multidisciplinares y el criterio de evaluación.

KEY ABBREVIATIONS

ACR	American College of Rheumatology
BDI	Beck Depression Inventory
CBT	Cognitive Behavioral Therapy
CNS	Central Nervous System
FIQ	Fibromyalgia Impact Questionnaire
FM	Fibromyalgia
HADS	The Hospital Anxiety and Depression Scale
MOSS-S	Medical Outcome Study Sleep Scale
PGWB	Psychological General Well-Being Index

1. INTRODUCTION

The American College of Rheumatology (ACR) defines fibromyalgia (FM) as a neurologic chronic disorder. Widespread musculoskeletal pain is accompanied by a wide range of symptoms, such as, tenderness to touch in muscles, joints and skin – characterized by a painful hypersensitivity to pressure, cold and hot –, severe fatigue, non restorative sleep and cognitive impairment. Some patients develop co-morbid health problems, among which stand out depressive or anxiety disorders, migraine, digestive problems – irritable bowel syndrome (IBS) or gastroesophageal reflux disease (GERD) –, pelvic pain or temporomandibular disorders (TMJ) (1). Some of these affections are gathered under the name of central sensitization syndromes (CSS) and share a multifactorial and not completely established underlying mechanisms, as well as FM (2).

Genetic bases have been described for FM and several chronic pain conditions. The genetic risk of developing FM is set by many polymorphisms, these related to changes in the metabolism of neurotransmitters which are involved in pain modulation (3). Nevertheless, the processing of pain in central nervous system (CNS) is determined by many genetic markers that interact during lifetime events, such as **infections** (4)**, physical trauma** (5) or **psychological stressors** (6), so the genetic set point for sensory regulation – including pain – can be modulated in the course of life.

Researchers have found **alterations at muscular levels** in FM patients (7), like atrophy (8), ischemia (9), metabolic alterations (10), qualitative and quantitative abnormalities in mitochondria (7), lack of organization of Z lines in sarcomeres (10), decreased collagen concentrations (11) or disorders in the motor recruitment (8), yet they have not been able to relate muscular pain and fatigue with these phenomena (12). **Structural changes in the CNS** have also been described (12). These changes involve brain blood-perfusion abnormalities (13), changes related with the volume and density of white and grey matter (14–16), metabolic changes regarding neurotransmitter imbalances (12) – eg. dopamine (15) and glutamate (17) –, altered functional connectivity (3) and changes in the hypothalamic–pituitary–adrenal (HPA) axis, which trigger the autonomic system dysfunction (18). **An alteration at the perception and processing of pain** constitutes the base of the phenomenon of pain centralization. This theory has emerged as a prominent hypothesis for the pathogenesis of FM, in which has been described to be an imbalance between pain inhibitory and pain facilitator systems and an amplification of pain related to hypersensitivity (eg. allodynia and hyperalgesia) due to hypervigilance conditions (3) **(Figure 1)**. Enhanced pain responses, recruitment of low-threshold sensory inputs or reduced descending pain modulation pathways have been studied by neuroimaging (3,12). These studies have shown that people suffering from FM need lower pressure stimulus to activate central somatosentory cortical areas, and limbic regions appear to be more activated when catastrophizing or depression are present (19). Described altered mechanisms in patients with FM are not unique to this affection and may reflect susceptibility factors. It is required a better understanding of the mechanisms that contribute to the onset and course of this chronic condition.

Regarding to the prevalence of this disorder, literature reveals that values in the general population vary between 2% and 4% (3). In Spain, this percentage is set on a 2.4% – diagnosis based on the 1990 ACR diagnostic criteria **(Table 1)** – with a peak incidence in middle-aged females (3,12,20). FM affects a

ACR 1990 and 2010 diagnostic criteria (1,3) **Table 1**

1990	Presence of 11 out of 18 standardized tender points (TP) and presence of chronic widespread pain (for ≥3 months).
2010	Does not include TP examination. Addresses the examination of important symptoms that characterize FM.

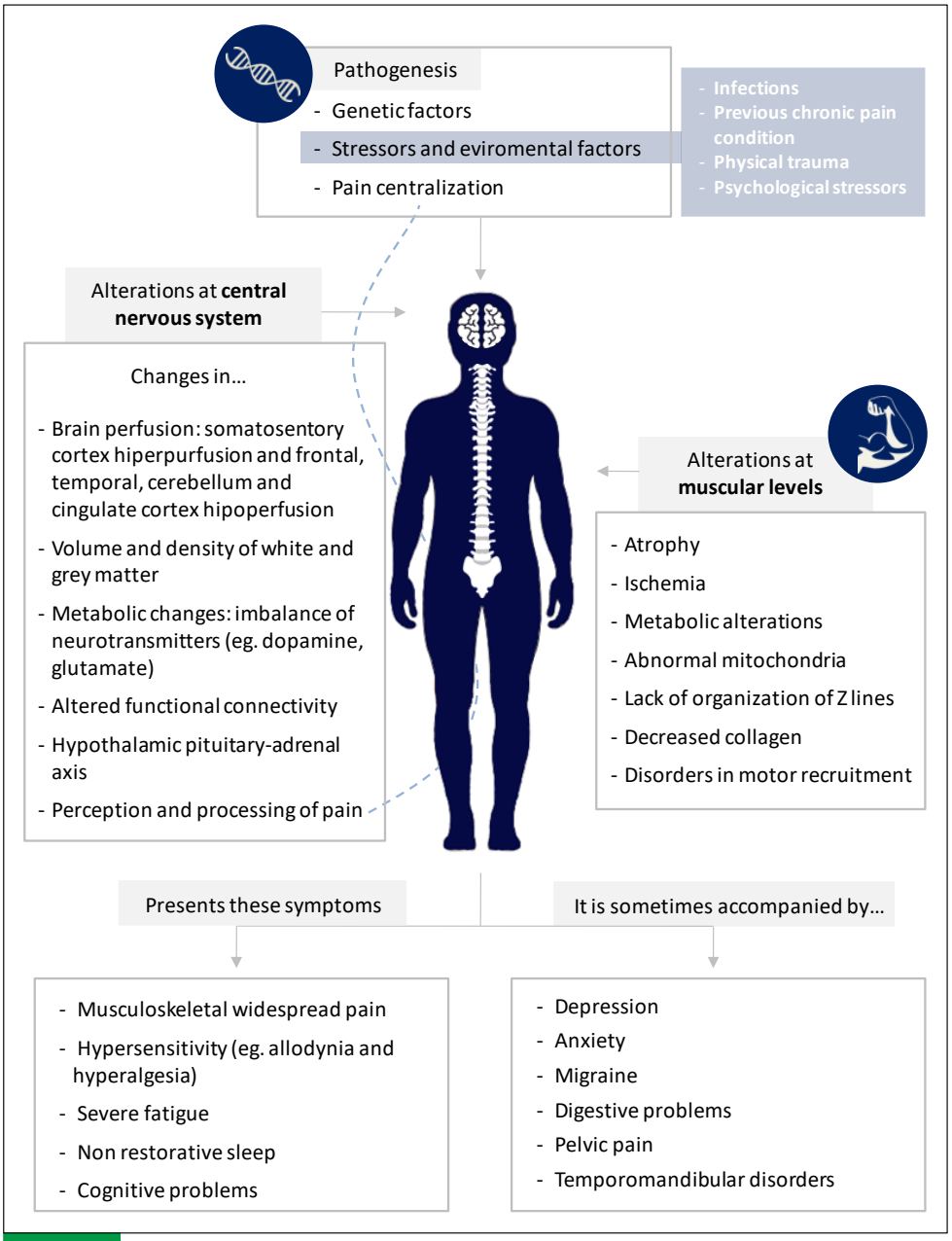


Figure 1 **Pathogenesis, physiopatology and symptomatology and comorbidity of FM (1, 3, 12). Source: Own elaboration.**

heterogeneous group of patients who can differ with regard to their symptoms and psycho-physical features. Many researchers have studied the role of personality factors in FM, and some have found that patients show higher levels of alexithymia – difficulty in identifying and describing emotions and feelings – compared with healthy controls, which plays an important role affective aspects of pain (21). As mentioned before, pain catastrophizing behaviors can also enhance pain severity in FM patients. This cognitive and emotional process is related to pessimist thoughts in which rumination on pain, constant complaints and disability assumption are present and contribute to the maintenance of the condition (22).

FM is considered a public health problem, due to its high incidence among adult population, the lack of knowledge concerning etiology and the components that cause the affection in addition to the non-curative available treatments (23). Both pharmacological and non-pharmacological treatments are used. These last ones play an important role, as the effect sizes for non-pharmacological approaches have demonstrated to be larger than those for drug therapies (24).

There is not a specific drug for FM, so available drug therapies are drugs which have demonstrated to be effective in pathologies with analogous characteristics (23). They have a supportive role in symptom management and they attempt to improve the quality of life of patients by reducing suffering and improving functionality, yet there is not a consolidated criterion in their use and a benefit-side effect balance is hard to obtain. As explained before, it is possible that an imbalance of neurotransmitters at CNS level cause an alteration at the modulation of pain, responsible for the clinical profile. For this reason, drugs of choice in FM have their targets at CNS, such as, antidepressants, anticonvulsant drugs or opioid analgesics among others.

Antidepressants are used as maintenance treatment. Among tricyclic agents, **amitriptyline** has showed potential at treating pain and the sleeping disorders (25) by inhibiting the reuptake and increasing the synaptic concentrations of both, noradrenalin and serotonin in the CNS, which leads to a reduction in pain signaling. This drug has shown strong evidence for efficacy (26). **Cyclobenzaprine** has also demonstrated to be beneficial and to have strong evidence for efficacy (26); even though it is classified as a tryciclic remains unknown its antidepressant effect so it is frequently used as a muscle relaxant (27). Serotonin and noradrenaline reuptake inhibitors (SNRIs) use a similar mechanism of action and have demonstrated modest evidence for efficacy (26). **Duloxetine** (28), reports the best efficacy evidence while **milnacipran** (29,30), seems to be effective in cognitive difficulties. Both of them exhibit a better tolerability and side effect reaction profile than tricyclics (31). **Mirtazapine** keeps serotonin and noradrenalin levels up, being beneficial for the quality of sleep, fatigue, mood and pain (23).

Pain facilitatory neurotransmitter concentrations – glutamate (32) and P substance (23) – seem to be heightened in CNS in FM. **Pregabalin**, a gabapentinoid used as anticonvulsant that interferes in pain transmission by inhibiting the release of excitatory neurotransmitters, has shown improvements in pain, sleep and overall well-being. Its analgesic effect seems to be more effective than the one for **gabapentin**, another commonly used anticonvulsant (23,24,31). These anticonvulsants have shown modest evidence for efficacy (26).

Even though descending inhibitory pathways are one of the key targets for an effective pharmacological treatment in FM, **opioids** have not shown remarkable effectiveness and their effects in the modulation of central pain still remain unknown (33,34), thus there is no evidence for efficacy (26). **Cannabinoids** have shown a little reduction of pain and depressive and anxiety symptoms. Regarding **non-steroidal anti-inflammatory drugs**, these are used in low doses and for short periods as they show low indication in FM syndrome, as well as **common analgesics** (35).

As medical management is usually partially successful, non-pharmacological treatments are accomplished for a biopsychosocial approach, comprising the biological, psychological and social factors that encompass the syndrome. Several non-pharmacological treatments have been described: therapies based on physical exercise, educational therapies, cognitive behavioral approaches, mindfulness, relaxation techniques, biofeedback treatments, acupuncture, stimulation techniques, nutritional interventions, chiropractic or even hypnosis. Exercise and psychoeducational approaches are the non-pharmacological therapies which have demonstrated to be most effective (35). **Aerobic exercise** – at least for 12 weeks long – shows moderate evidence of efficacy on overall well being, physical condition and on pain and hyperalgesia (23). Among **psychotherapeutic** approaches, therapies that focus the attention on mind and body connection (e.g. relaxation techniques or biofeedback), have revealed improvement in physical functioning, pain and state of mind (36). Learning and memory processes appear to play a significant role in chronic pain, so **cognitive-behavioral treatments** (CBT) address these levels.

Painful perception emerges from the interaction between sensory input and previous experiences; therefore, CBT alters brain connections between pain signals, emotions and cognition. This is achieved by modifying pain behaviors and cognitions to reduce negative feelings and lack of control over the pain (31). **Acupuncture** is used in many chronic conditions – including FM – to reduce pain intensity by decreasing inflammation, releasing endogenous opioids and decreasing anxiety, and it has shown a little improvement in pain and fatigue (37). Other approaches, such as **transcranial magnetic stimulation**, have been effective in reducing pain intensity too (38).

FM appears to be a complex syndrome which incorporates a wide range of symptoms and functional alterations that might not have a definite cause, so dealing with it remains a challenging task not only for patients but also for healthcare professionals. Due to the complex manifestations and development of the disorder, leading the treatment as a unique and isolated approach may not comprise the complexity that characterizes FM; therefore there are required simultaneous approaches for multiple aspects of the condition. **Multidisciplinary treatment programs** have been recommended for FM syndrome as they have demonstrated to have beneficial effects compared to mono disciplinary treatment programs (39,40). The combination of approaches such as education, exercise and psychotherapeutic interventions seem to create synergy together with pharmacological therapies so physical and mental symptoms are embraced (40). By the use of multidisciplinary treatments patients gain knowledge on the disease, are made aware of a range of possibilities regarding treatments for FM and acquire self-efficacy – perceived ability to manage symptoms that provides sense of control –, which has been described to show a positive relationship with beneficial health behaviors and health status (41).

2. OBJECTIVE

The main objective of this study is to determine the efficacy of multidisciplinary treatments for FM. PICO model was used to define the clinical question **(Table 2)**. A systematic review on the different multidisciplinary approaches was performed and improvements on key symptoms were evaluated.

Table 2 **PICO clinical question framework for the search strategy. Population, Intervention, Comparison, Outcomes and Study design are gathered.**

P	Adults women with ACR based fibromyalgia diagnosis
I	Multidisciplinary treatments
C	Usual care or/and control group
O	Effectiveness of the approach based on the improvement of significant symptoms/co-morbid health problems in this condition
S	Randomized Controlled Trials (RCTs)

3. METHODOLOGY

First of all, a protocol was developed using the intervention review methodological framework proposed by Cochrane and structured in Review Manager 5.3. The final version of the protocol is available upon request from the corresponding author.

Prior to the present study, a mapping review on pharmacological and non-pharmacological interventions for FM was accomplished. A comprehensive literature search was conducted in MEDLINE (PubMed), Cochrane and Epistemonikos databases. The search was conducted the 21st of November of 2019, with no restriction on publication date. **140 studies were included** in this overview and two of the included studies reviewed the effectiveness of multidisciplinary and multicomponent therapies in fibromyalgia syndrome.

Table 3 Search strategy: combination of terms, booleans and truncations used in Pubmed and Cochrane databases.

Data bases	Key words			Article type	Query translation
PubMed	Fibromyalgia [MeSH Terms]	AND	Multidisciplinary treatment	Clinical Trial	("fibromyalgia"[MeSH Terms] OR "fibromyalgia"[All Fields]) AND "multidisciplinary treatment"[All Fields] OR "multicomponent treatment"[All Fields] AND Clinical Trial[ptyp]
	OR		OR		
	Fibromyalgia [All fields]		Multicomponent treatment		
Cochrane library	Fibromyalgia (in all text)	AND	Multidisciplinary treatment	Trials	
			OR		
			Multicomponent treatment		

Note: The election of key words was done based on terms that were found in both systematic reviews retrieved from the overview.

Two data bases were consulted to accomplish the comprehensive research for the current systematic review: Cochrane and MEDLINE (PubMed). The search was conducted the 11th of February of 2020. The search query consisted of terms extracted from the systematic reviews retrieved from the overview. The combination of key words, booleans and truncations used for the search is been gathered in **Table 3**.

In this systematic review, RCTs were exclusively included. The studies had to be performed in adult patients (≥18 years old) with diagnosis of fibromyalgia based on 1990-2010 ACR criteria (42,43) or similar. At least one study group had to be treated in a multidisciplinary approach. To be considered multidisciplinary treatment, separately addressed concomitant therapeutic interventions had to be accomplished. At least two out of all the interventions found in the previous mapping review of FM treatments needed to be part of the program. At least one of the following domains had to be measured: pain, fatigue, quality of sleep, physical function, depression and/or anxiety.

3.1. Measured outcomes

A study conducted in 252 rheumatologists and 86 patients via Delphi method, revealed a high percentage of doctors as well as of patients who emphasized the importance of symptoms like pain, fatigue and sleep disturbances when evaluating FM syndrome. Among patients and in concordance with clinicians, the multidimensional function – which refers to the physical function and the impact of this in daily activities –, depression and anxiety appeared to be also remarkable domains (44).

This criterion was used to establish which outcomes to be measured in the present study **(Table 4)**. Pain was chosen as the **primary outcome**. Nowadays, pain is defined by IASP (The International Association for the Study of Pain) as an unpleasant sensory and emotional experience associated with actual or potential tissue damage (45). This definition avoids tying pain exclusively to the stimulus and brings a biopsicosocial perspective of pain in which the affective side is also included.

TP examination was used as diagnostic criteria (ACR, 1990 (42)) and it is usually used to measure pain improvement in many trials. 18 tender points are described for this syndrome, which appear to be positive when, by applying 4kg/cm^2 with an algometer, unpleasant responses of pain are achieved (43). The latest 2010 ACR diagnostic criteria, measures pain through a body schema questionnaire, in which 19 painful areas are represented, so the extension of painful regions can be assessed. Afterwards, relevant somatic and cognitive symptoms are measured. The Widespread Pain Index (WPI) and Symptom Severity

Table 4 Classification of domains measured in the meta-analysis.

Primary outcome	1. Pain
Secondary outcomes	2. Fatigue
	3. Quality of sleep
	4. Physical function
	5. Depression
	6. Anxiety

Table 5 In this table are gathered the measurement items that were used in the clinical trials to measure the outcomes.

		OUTCOMES					
		Pain	Fatigue	Quality of sleep	Physical function	Depression	Anxiety
QUESTIONNAIRES	Burckhardt et al.	FIQ	FIQ	-	FIQ, 6MW	FIQ	FIQ
	Casanueva-Fernandez et al.	MPQ, FIQ	FSS, FIQ	PSQI	6MW, SF-36	BDI, ZDS, FIQ	BAI, HADS, FIQ
	Castel et al.	NRS, FIQ	-	MOSS -Scale	-	HADS	HADS
	Cedraschi et al.	FIQ, RPS	FIQ	-	FIQ, SF-36	PGWB, FIQ	PGWB, FIQ
	Gowans et al.	FIQ	FIQ	-	6MW	FIQ	FIQ
	Hammond et al.	FIQ	FIQ	-	FIQ	FIQ	FIQ
	King et al.	FIQ, SE Scale	-	-	6MW	-	-
	Lemstra et al.	VAS, PDI	-	-	-	BDI	-
	Lera et al.	FIQ	-	-	MOS SF-36	-	-
	Luciano et al.	FIQ	FIQ	-	FIQ	FIQ	FIQ
	Racine et al.	BPI	BFI	MOSS-Scale	PCS	HADS	HADS
	Rooks et al.	FIQ, VAS	FIQ	-	Health Survey SF-36, 6MW	BDI, FIQ	FIQ
	van Eijk-Hustings et al.	FIQ	FIQ	FIQ	-	FIQ	FIQ
	van Koulil et al.	-	-	-	Walking test, cycling test	-	-
	Vlaeyen et al.	MPQ	-	-	BAT	BDI	-
	Wigers et al.	VAS	VAS	-	-	VAS	-
	Zijlstra et al.	FIQ, MPQ	FIQ	VAS	FIQ, RAND-36 (PCS)	BDI, FIQ	FIQ

(SS) scale are used for these tasks (43) being TP examination excluded for pain assessment. For this reason, TP examination was not used in this systematic review for the measurement of pain. Instead, 3 questionnaires were chosen for the measurement of this outcome.

The **Fibromyalgia Impact Questionnaire** (FIQ) is a multidimensional self-reported questionnaire and one of the most frequently used assessment tools in the evaluation of FM (46). 3 domains are measured: function, overall impact and symptoms. Pain is measured by a visual analogue scale (VAS), from “no pain” to “very severe pain”. These items are scored in numerical increments from 0 (no impairment) to 10 (maximum impairment).

The **Brief Pain Inventory** (BPI) is also a self reported questionnaire designed to measure several aspects of pain and validated for central pain conditions (46). The brief version is composed of a total of 15 items in which pain is measured by a dichotomous “yes/no” question, a body map in which painful areas must be shaded and marked with a “X” the most painful one, 11 point-rating intensity scales from 0 (no pain) to 10 (pain as bad as you can imagine), a pain medications list, a percentage item of pain relief from 0% (no relief) to 100% (complete relief) and a 11-point numeric rating scale for pain interference from 0 (does not interfere) to 10 (completely interferes).

The **McGill Pain Questionnaire** (MPQ) is a multidimensional pain questionnaire designed to measure different dimensions of pain in adults with chronic pain, including pain due to rheumatic conditions (47). The questionnaire is composed by 78 pain descriptors categorized into 20 subclasses, each containing 2-6 words. The descriptors fall into four major groups: sensory, 1-10; affective, 11-15; evaluative, 16; and miscellaneous, 17-20. There is also a pain intensity scale ranging from 0 (no pain) to 5 (excruciating).

Secondary measured outcomes are fatigue, quality of sleep, physical function, depression and anxiety. Neither every symptom is measured nor are the same questionnaires and/or tests used in all the clinical trials about fibromyalgia. Due to this fact, symptoms chosen as secondary outcomes which are measured in the clinical trials are gathered in **Table 5**. At the same time, measurement tools of each outcome are summarized.

3.2. Measurement periods

There is not a standardized criterion for measurement periods so in the present study 3 measurement terms

have been established regarding the follow up measurement dates of the outcomes. **Short-term follow up** refers to assessments from post-treatment to <3 months follow up; **middle-term follow up**, from ≥3 to 6 months follow up and **long-term follow up** from 6 to 12 months follow up or ≥12 months follow up assessments.

3.3. Methodological quality assessment

To assess the methodological aspects of the included studies, risk of bias was evaluated. Selection bias (random sequence generation), attrition bias (incomplete outcome data) and other bias were assessed. Review Manager 5.3 was used to perform this task. Reporting bias was not evaluated due to no protocols were found for the included clinical trials. Regarding performance bias, masking of the participants and professionals was not measured. The aim of the included studies was to evaluate the effectiveness of multidisciplinary approaches – non-pharmacological ones – so healthcare professional cannot be blinded to the treatment they are providing and participants cannot be blinded to the treatment they are receiving.

3.4. Analysis of features of the included studies

IBM SPSS Statistics 25 was used for the statistical analysis of parameters like means and percentages when examining the features of the included studies. This same software was used for the assessment of average of scores at baseline measurements.

3.5. Statistical analysis and interpretation of results

First of all, extraction of data from the included studies was carried out. Data was classified based on 2 aspects: the measured outcome and the instrument used to measure each outcome. For pain assessment data from studies that measured this outcome with FIQ, BPI, MPQ was collected. For the rest of outcomes, all data was collected independent of the questionnaire used for the assessment of the domain. All the assessment periods in both the experimental group and control group were gathered (see **Supplementary material 1)**.

Once all data was collected, both researchers, independently, extracted effect sizes (number of participants, mean and standard deviation for each outcome scale) with a 100% agreement. Preferences to analyze – if number permitted – were studies reporting outcomes which used the same measure. It allowed analyzing mean differences (MD) – more interpretable than standardized mean differences (SMD) –, imputing missing values (SDs) and searching for minimal important differences (MID) associated to specific measures and reported in the literature. For compatibility and interpretability, we included only studies reporting raw data (means, SDs and Ns). We did not include studies reporting transformed data (change from baseline or others). Review Manager 5.3 was used to run the analysis assuming a random effects model. Variables were categorized as continuous ones. When the same measurement item was used for the assessment of efficacy of multidisciplinary treatments on a certain symptom, mean values were used in the statistical analysis, so mean difference was achieved. In contrast, when different questionnaires were combined in the statistical analysis, standard deviations were used so SD mean difference was represented.

After introducing the values in Review Manager 5.3, graphic representations were achieved in which mean or standard deviation differences were shown and the overall effect value was represented – the width represented the general confidence interval –. The confidence interval (CI) was set in 95%, which allowed understanding the precision of the estimation within an established margin of error. The 95% CI represents a 95% certainty that the association studied is not given by chance. In the graphic representations, risk of bias was also included so it could be taken into account the contribution of the risk of each study

depending on the weight based on the N and the confidence interval of each study.

Finally, Grading of Recommendations Assessment, Development and Evaluation - Guide Development Tool (GRADEpro GDT) was used for the interpretation of the results. The following dimensions were analyzed for the certainty assessment: study design, risk of bias, indirectness, imprecision and other considerations.

Risk of bias was assessed so it could be determined if this was a not serious, serious or very serious factor. The classification obtained in Review Manager 5.3 was used for the evaluation. High risk of bias was considered when the proportion of information from studies with high risk of bias was sufficient to affect the interpretation of the results.

Inconsistency refers to the unexplained heterogeneity of the results. To analyze heterogeneity between the studies I^2 was take into consideration. In the present study, when this was >60% and there was no way to explain heterogeneity, it was considered estimations showed significant heterogeneity among them. In this way, inconsistency was considered to be a very serious factor. In contrast, when heterogeneity of the estimation was <60%, inconsistency was considered not to be a serious factor.

Indirectness was assessed to measure if the evidence was sufficiently direct, ranging from no indirectness to very serious indirectness. The questionnaire provided by the GRADEpro GDT was followed for the assessment. Population, intervention, comparator, directness of comparison and the measured outcome were the domains taken into consideration to determine the indirectness.

Imprecision was based on the number of subjects the result was based on. The results are considered imprecise when the studies include relatively few patients and few events. For quantitative outcomes, the number of subjects is set in N>300 (OIS, optimal information size) so it can be considered imprecision is not a serious factor.

Other consideration taken into account was publication bias. This was analyzed when ≥10 studies were included in the combination of data and for the analysis the funnel plot was examined. If less than ten studies were used in the analysis, publication bias was classified as undetected.

After the assessment of the items mentioned above, certainty of the evidence was achieved. In the present study two kinds of grades of evidence were reached. Very low certainty, when very little confidence in the effect was estimated (the true effect is likely to be substantially different from the estimate of effect) and low certainty, when the confidence in the effect estimate is limited (the true effect may be substantially different from the estimate of the effect).

4. RESULTS

4.1. Included studies in the meta-analysis

The flowchart, based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) flow diagram, is shown in **Figure 2**.

119 articles were obtained from both data bases mentioned before and 19 clinical trials were extracted from the two systematic reviews obtained from the overview of treatments for FM conducted before (represented as “Other sources” in **Figure 2**). In this way, a total of 138 records were achieved. After removing duplicates, via Zotero, 108 articles were obtained and assessed for eligibility.

Two researchers (J.B. and V.O.) included and excluded – on the basis of the title and summary and guided by the eligibility criteria established before **(Table 6)** – the potentially suitable articles for the systematic review. Both researchers performed the choice in an independent way and blind on to the decision of the other one, so each researcher chose freely without

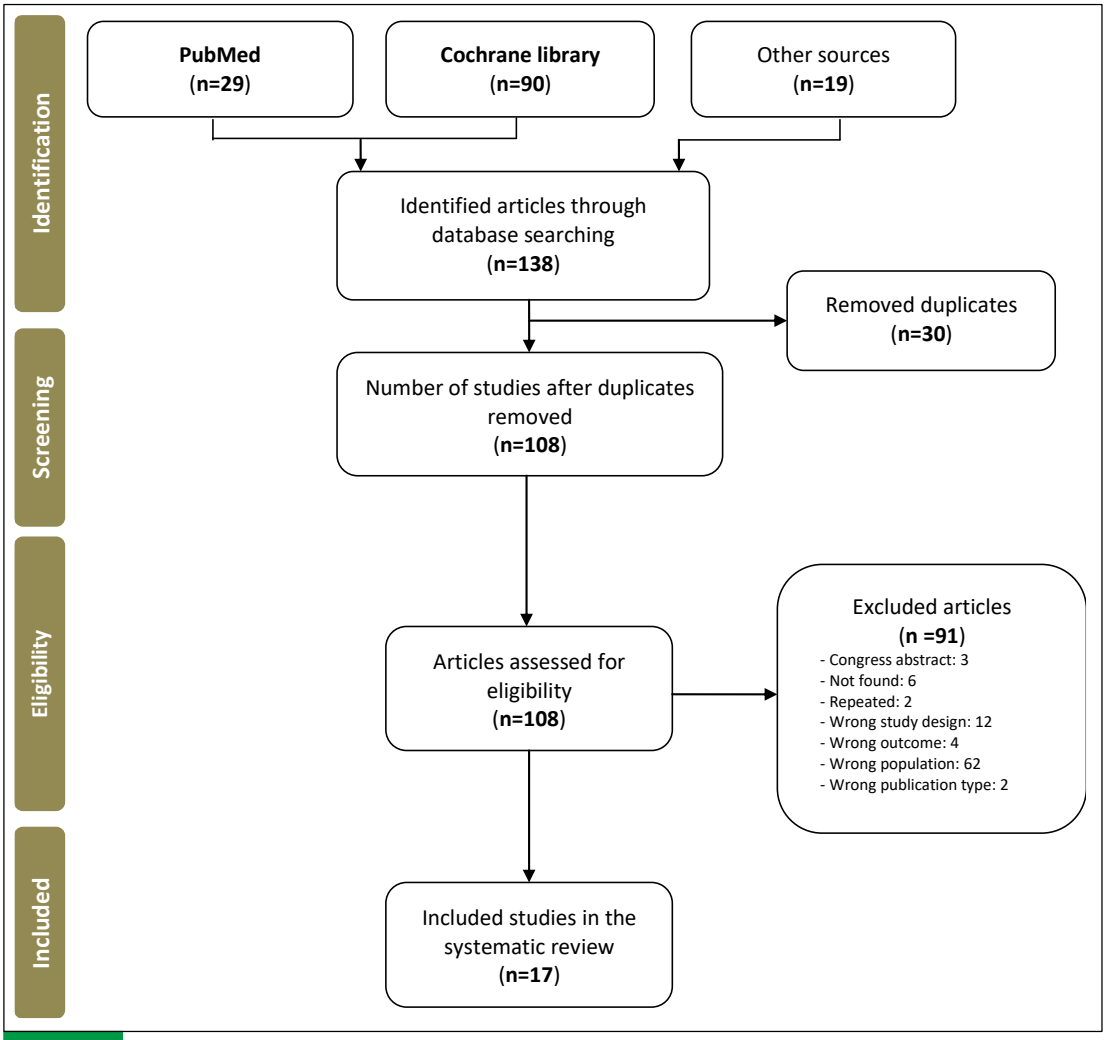


Figure 2 **PRISMA diagram. Summary of the election and exclusion of records.**

having influence over the others election. Rayyan QCRI program was used to perform this task. Subsequently, both of the researchers analyzed the other researcher's classification of the records and discussed discrepancies in the election of suitable articles. After unanimity, a total of **17 records** were assessed full text and included in the systematic review.

7 out of the 89 (48–54) records retrieved from the research in both databases were included. Reasons for exclusion were: 60 were accomplished in a wrong population, 12 were a wrong study design, 3 measured another outcome, 3 were congress abstracts, 2 were not found and 2 were another kind of publication type. An ongoing suitable article was found which was not completed due to budget limitations. After contacting the author via email, he

Table 6 **The inclusion and exclusion criteria used for the selection.**

Publication type	For more reliability, selected studies for the systematic review are based on Randomized Controlled Trials (RCTs).
Publication date	All publications to date.
Population	Only patients with fibromyalgia diagnosis, this based on the American College of Rheumatology (ACR) criteria or similar. There will not be included individuals with co-morbilities, except insomnia, depression, anxiety, migraine, IBS, GERD or TMJ which are described to appear together with fibromyalgia by ACR criteria. With regard to the sex of the subjects, women as well as men will be added, these being >18 years old.
Interventions	There will only be added articles which have reviewed the efficacy of multidisciplinary treatments for fibromyalgia.
Language	Only articles published in English or Spanish will be added.

sent us a paper based on a smaller number of participants that they had at the time the study was terminated. This trial was included in the systematic review (52).

10 out of the 19 (55–64) clinical trials extracted from the two systematic reviews from the overview about treatments for FM were included in this study. Reasons for exclusion were: 4 were not found, 2 were done in a wrong population, 2 were duplicated and 1 measured another outcome.

Features of included studies in the present meta-analysis are gathered in **Table 7**.

4.2. Features of participants from the experimental group

Several aspects of the participants of the experimental groups were assessed. All the studies were conducted in adult patients. With regard to the mean age ± SD of the intervention group participants this is set in 47.37 ± 3.6. The 94.1% (48–51,53–64) of the studies were taken into account due to 1 of the RCTs did not provide the mean age of the participants on the experimental condition (52).

Concerning the gender of the intervention group patients, most of them were females. The smallest female sample constituted of 14 patients (57), while the biggest group of women was formed by 105 (51) participants . A total of 884 women were assessed whereas the total number of male participants was of only 52 patients. Only 16 RCTs (48–51,53–64) were taken into consideration as the remaining one (52) did not provide data about sex distinction. The total cluster in the experimental group, based on the 17 included studies, consisted of 1043 participants.

All the patients were diagnosed from FM based on ACR 1990 criteria and a unique study used both the 1990 and 2010 ACR diagnosis criteria (52).

4.3. Characteristics of multidisciplinary treatments and assessment periods

Multidisciplinary treatments are constituted by more than an approach and some of these treatments share similar approaches: pharmacotherapy, physical exercise, education and/or psychotherapy among others. A graphic representation of the components of the multidisciplinary treatments is shown in **Supplementary fig. 3**. Physical exercise was the most included approach among the multidisciplinary treatments, present in the 94.1% of the treatments (48–50,52–64). Education was found in 70.6% of the included treatments (50,51,53,55–62,64), becoming the second most utilized approach. Instructive sessions focused their attention on diverse topics: information about FM (symptoms, co-morbid medical conditions, treatments…), psychological factors of pain, the role of stress, self-management techniques, social behaviors and strategies, barriers to behavior change, problem solving/coping strategies, physical education, goal setting and/or nutritional information.

Psychotherapy formed part of the 47.1% of the multidisciplinary treatments (49–54,62,63), being CBT the most used approach. 23.5% of the treatments included pharmacotherapy (48–51). Percentages are graphically represented in **Figure 3**.

Figure 3 **Components of the multidisciplinary treatments. Percentages based on the 17 included studies.**

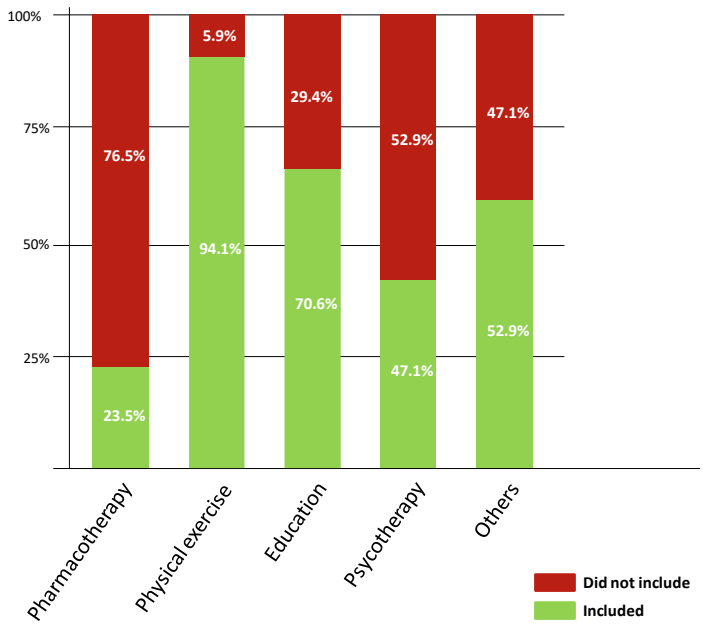


Table 7. Details of studies included in the meta-analysis

Author, year, country	Mean age, years	Women, (%)	Diagnostic criteria	Measurement		Treatment group			Control group	
				Periods	Items	N	Multidisciplinary treatment, approaches	Duration	N	Treatment, duration
Burckhardt et *al.*, 1994, Sweden (55)	46.5	33 (100)	ACR, 1990	T_0, T_1, 3 month- and long term FU	FIQ, FAI, QOLS-S, SELF, physical fitness (6MW, flexibility, the chair test), TP Index, BDI	33	Education and physical therapy (stretching, pool therapy)	6 sessions (1h each) and an additional training hour	35	Usual care, delayed treatment control group
									31	Education, 6 sessions
Casanueva-Fernandez et *al.*, 2011, Spain (48)	47.8	16 (94.1)	ACR, 1990	T_0, T_1 and 1 month FU	TP Index, Myalgic Score, Pressure Pain Threshold, Grip Strength Test, VAS, MPQ, FSS, HAS, BAI, ZDS, BDI, PSQI, FIQ, HAQ, MOS SF-36	17	Pharmacotherapy and massage therapy, ischemic pressure on the 18 tender points, aerobic exercise and thermal therapy	8 weeks (1h sessions)	17	Pharmacotherapy and education, 4 sessions
Castel et *al.*, 2013, Spain (49)	49	81 (100)	ACR, 1990	T_0, T_1, 3, 6 and 12 month FU	NRS, HADS, Coping Strategies Questionnaire, FIQ, Darmouth COOP/WONCA Functional Health Assessment Charts, MOSS-Scale	81	Pharmacotherapy, CBT and physical therapy (aerobic exercise, strengthening, flexibility, stretching, hydro-kinesiotherapy and kinesiotherapy)	24 sessions (1h each)	74	Conventional pharmacotherapy
Cedraschi et *al.*, 2004, Switzerland (56)	48.9	78 (92.8)	ACR, 1990	T_0 and 6 month FU	TP Index, PGWB, SF-36, FIQ, RPS, Potts and Silverman	84	Swimming pool sessions, relaxation exercises, low impact land based exercises, activities of daily living, education-discussion sessions	12 sessions (90 min each)	80	Waiting list intervention
Gowans et *al.*, 1998, Canada (57)	44.3	14 (70)	ACR, 1990	T_0, T_1 and 3 month FU	6MW, Borg rating scale, ASES, FIQ	23	Exercise classes (stretching, therapeutic pool) and multidisciplinary educational sessions	24 sessions (30 min -1h each)	22	Waiting list intervention
Hammond et *al,* 2006, London (58)	48.4	63 (88.7)	ACR, 1990	T_0, 4 month and 8 month FU	FIQ, ASES, RAI, SPAQ and number of self-reported visits to a doctor	97	Education, exercise (stretching, strengthening, Tai-Chi) and other interventions (activity pacing, sleep hygiene, problem solving, negative automatic thoughts…)	10 weeks (2h each)	86	Relaxation techniques, for 10 weeks (1h each)
King et *al.,* 2002, Canada (59)	47.4	37 (100)	ACR, 1990	T_0, T_1 and 3 month FU	Chronic pain-self efficacy scale, Lorig's Arthritis Self-Efficacy Scale, FIQ, 6MW, TP Index and total survey site score	37	Education and exercise combination (aerobic exercises, mild stretches)	12 weeks (exercise twice per week and education once per week)	39	Instructions for stretches and general coping strategies, 12 weeks (contacted 1-2 times)
									46	Aerobic exercise, 12 weeks (from 10 to 40 min each session)
									48	Education, 12 sessions (75 min-2h each)
Lemstra et *al.*, 2005, Canada (60)	49.7	37 (86)	ACR, 1990	T_0, T_1 and 15 month FU	VAS, PDI, BDI and the state of change	43	Exercise therapy (stretching, aerobic exercise, strengthening), pain/stress management lectures, education lectures, dietary lecture and massage therapy	6 weeks	36	Waiting list intervention, usual care
Lera et *al.*, 2008, Spain (50)	51.1	66 (100)	ACR, 1990	T_0, T_1 and 6 month FU	TP Index, FIQ, MOS SF-36, SCL-90-R	33	Pharmacotherapy, physical education, physical exercise (aerobic exercise, stretching), CBT	15 CBT sessions (90 min each) before MT sessions	31	Pharmacotherapy, physical education, physical exercise, 14 multidisciplinary (MT) sessions (1h per week)

Luciano et *al,* 2011, Spain (51)	55.2	105 (97.2)	ACR, 1990	T_0 and T_1	FIQ, Chronic Medical Conditions Checklist, STAI, MCSDS	108	Usual care and psychoeducational program (education, autogenic training)	9 sessions (2h each)	108	Usual care
Racine et *al.*, 2018, Canada (52)	-	-	ACR, 1990-2010	T_0, T_1 and 3 month FU	BPI, BFI, MOS SF-36, PCS, MCS, MOSS- Scale, HADS, POAM-P, PGIC	54	OL treatment. Exercise/sports, chores, social/leisure activities, cognitive tasks and work/volunteering/housework activities	10 weeks (2h each)	36	Operant learning (OL) delayed
						53	EC treatment. Exercise/sports, chores, social/leisure activities, cognitive tasks and work/volunteering/housework activities	10 weeks (2h each)	35	Energy conservation (EC) delayed
Rooks et *al.*, 2007, United States (61)	50.3	156 (100)	ACR, 1990	T_0 and 6 month FU	FIQ, Health Survey SF-36, 6MW, BDI, ASES	55	Combination group (ST+FSHC)	16 weeks	51	Aerobic exercise and flexibility exercises, 32 sessions (1h each)
									50	Aerobic exercise, strength training, flexibility exercises (ST) , 32 sessions (1h each)
									55	Education (FSHC), 7 sessions (2h each)
van Eijk-Hustings et *al.,* 2012, Netherlands (53)	41.6	63 (94)	ACR, 1990	T_0, T_1, 3, 21 and 24 month FU	HR-Qol (EQ-5D), FIQ	108	Sociotherapy (education), physiotherapy (aerobic exercise, strengthening), psychotherapy and creative arts. Strategies to preserve changes	1 year (12 week therapy sessions and 9 month after-care program)	48	Usual care
									47	Aerobic exercise, stretching, resistance training, 24 sessions
van Koulil et *al.*, 2010, Netherlands (54)	41.7	65 (95.5)	ACR, 1990	T_0 and 3 month FU	SWT, the cycling test (Borg scale ranging)	29	PA treatment, CBT, exercise training, hydrotherapy and relaxation therapy	32 sessions (2h each)	45	Pain-avoidance (PA) waiting list condition
						39	PP treatment, CBT, exercise training, hydrotherapy and relaxation therapy	32 sessions (2h each)	45	Pain-persistence (PP) waiting list condition
Vlaeyen et *al.*, 1996, Netherlands (62)	44.6	80 (90.9)	ACR, 1990	T_0, T_1, 6 and 12 month FU	PCL, CSQ, BAT, UAB, CHIP, MPLC, MPQ, Dutch Hyperventilation Questionnaire, FSS-III-R, BDI, MOCI, Social Desirability	39	Educational program, physical exercise (swimming, bicycling…) and discussion group (EDI)	12 sessions (2h each)	43	Waiting list intervention
						49	Educational program and cognitive treatment (imagery and EMG biofeedback) (ECO)	12 sessions (2h each)		
Wigers et *al.*, 1996, Norway (63)	43.5	36 (90)	ACR, 1990	T_0, T_1 and 4 years after completion	Wallace's "rule of nine", VAS, TP Index, cycling test, VRS	20	Stress management treatment. Cognitive behavioral stress management, relaxation training, thermal biofeedback measure	20 sessions (90 min each)	20	Usual care
									20	Aerobic exercise and stretching, 40 sessions (45 min each)
Zijlstra et *al.*, 2005, Netherlands (64)	48	55 (94.8)	ACR, 1990	T_0, T_1 1, 3, 6 and 12 month FU	RAND-36 (PCS/MCS), VAS, FIQ, BDI, MPQ, CIS, TP Index (GTPS), modified 6MW	58	Spa treatment (thalassotherapy), exercise (stretching, low impact aerobics), education program.	2.5 weeks	76	Usual care

T_0= Baseline measurement; T_1= Post-treatment measurement; FU: Follow-up
Meaning of the abbreviations of measurement items are gathered in Supplementary fig. 2

More than a half of the multidisciplinary treatments (52.9%) included other approaches (48,52–54,56,58,60,63,64): massage therapy (48,60), ischemic pressure therapy (48), thermal therapy (48), spa treatment (thalassotherapy) (64), hydrotherapy (54), relaxation exercises (54,56), operant learning and energy conservation activity pacing approaches (52), sleep hygiene (58), activity pacing (58), problem solving strategies (58), pain/stress management (60,63), dietary lectures (60) and creative arts (53) among others.

Duration of the experimental treatments also differed between studies **(Figure 4)**. The majority of the included treatments lasted 6 weeks (29.4%) (55–57,60,62). Longest treatments lasted 16 weeks (11.8% of the studies) (54,61), while the shortest one lasted 2 and a half weeks (64). Other lengths of multidisciplinary treatments were: 8 weeks (11.8%) (48,51), 10 weeks (11.8%) (52,58), 12 weeks (17.6%) (49,53,59) and 14 weeks (11.8%) (50,63). Mean duration of the intervention approaches was between 9 and 10 weeks long.

Regarding follow-up assessments, short-term and middle-term follow-up measures were accomplished in more than three quarters of the studies, 70.6% (48–52,55,57,59,60,62–64) and 76.5% (49,50,52–59,61,62,64) respectively. Long-term follow-up assessments were carried out in 47.1% (49,53,55,58,60,62–64) of the studies (summarized in **Supplementary fig. 3)**.

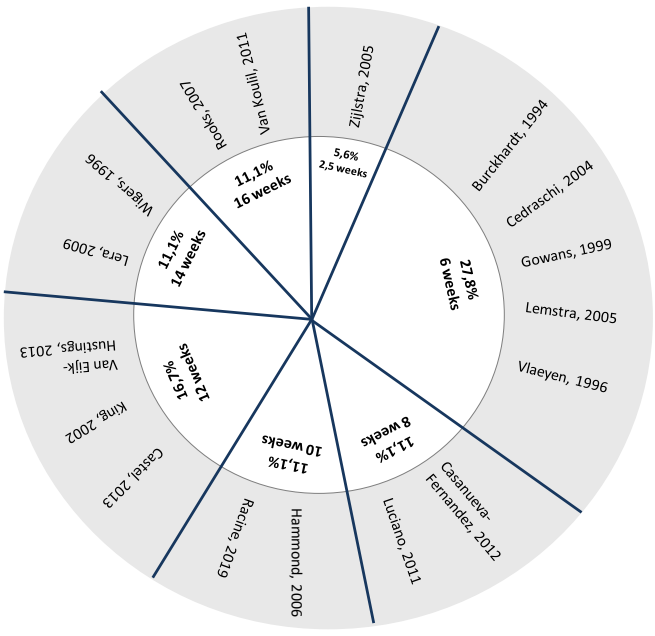


Figure 4 **Duration of the multidisciplinary treatments. Percentages based on the 17 included studies.**

4.4. Measurement items

As mentioned before in methodology, to analyze pain (primary outcome), 3 questionnaires – FIQ, BPI, MPQ – were chosen. When analyzing the items that were used on the included RCTs, we realized that most of the studies used FIQ as the questionnaire of election to measure the effect of multidisciplinary treatments on this outcome. 13/16 (48–51,53,55–59,61,63,64) studies used FIQ (VAS) as an item for pain measure and 10/12 (49,51,53,55–59,61,63) provided enough data to include in the statistical analysis. Only 2/16 (48,64) studies used MPQ to measure pain and an unique study (60) utilized the BPI questionnaire. The data provided was insufficient so only FIQ-Pain was used in the present study for the combination of data.

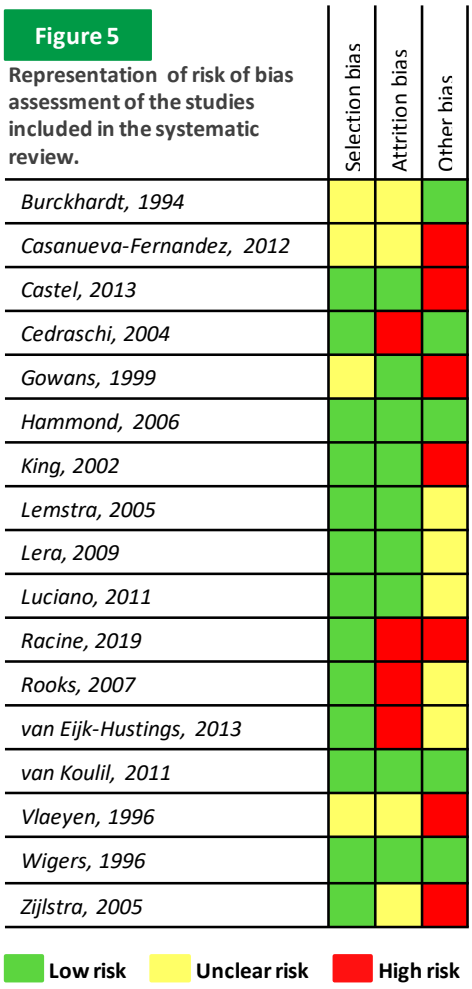
Regarding secondary outcomes, fatigue, physical function and anxiety were also measured by the scores achieved from FIQ. Quality of sleep and depression were measured by the combination of several questionnaires. With regard to quality of sleep, MOSS-S and FIQ scales were chosen. For the measure of depressive symptoms, BDI, HADS and PGWB questionnaires were selected.

Studies that were not included in the statistical analysis despite measuring the outcomes with the mentioned instruments were due to the fact that no mean, standard deviation or standard errors were provided.

4.5. Risk of bias

Results of methodological features of the included studies are gathered in **Figure 5**. Regarding selection bias, 76.5% of the studies showed low risk, 23.5% unclear risk and none high risk. With regard to

Figure 5
Representation of risk of bias assessment of the studies included in the systematic review.

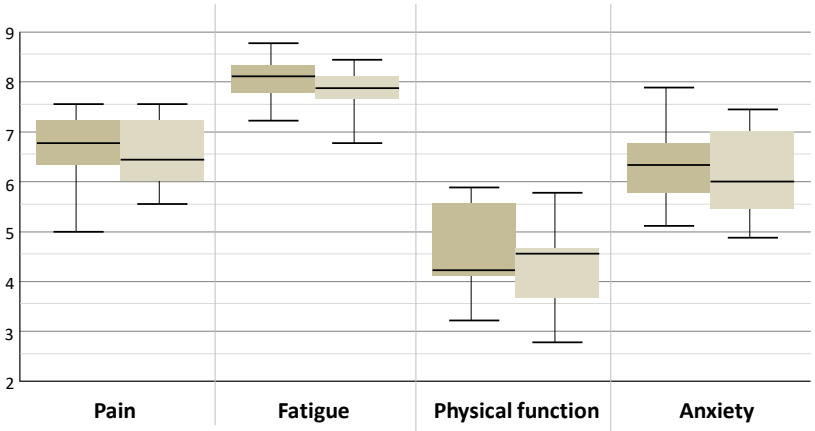


attrition bias, 52.9% of the studies presented low risk and 23.5% presented unclear risk and as well as high risk. Reasons for these considerations are gathered in **Supplementary material 4** (see for more detailed data).

Finally, at the time of evaluating other bias, 29.4% showed low risk and unclear risk while 41.2% of the studies showed high risk. Reasons for considering there was unclear risk in other bias were: a small sample size (48,49), Hawthorne effect (60), the absence of follow-up assessments (51) and a worst condition at inflow in treatment group despite randomization (53). It was considered high risk when contamination of the intervention-control group was not taken into account (59), when no control group was included in the study (61), when the sample size was very limited (50), when despite Zelen design treatments were so dissimilar that it was unlikely to have not influenced outcomes beyond randomization (64) and the fact that multiple outcomes were assessed without control for multiple comparisons (57,62).

4.6. Assessment of baseline mean scores

Baseline data was analyzed. In this way we ensured no clinically meaningful differences were found between experimental groups and control groups for the different outcomes at baseline assessments. For pain, fatigue, physical function and anxiety mean scores were used for the analysis. Mean scores of FIQ questionnaire for each sub-scale were utilized. In the case of quality of sleep (MOSS-S, FIQ) and depression



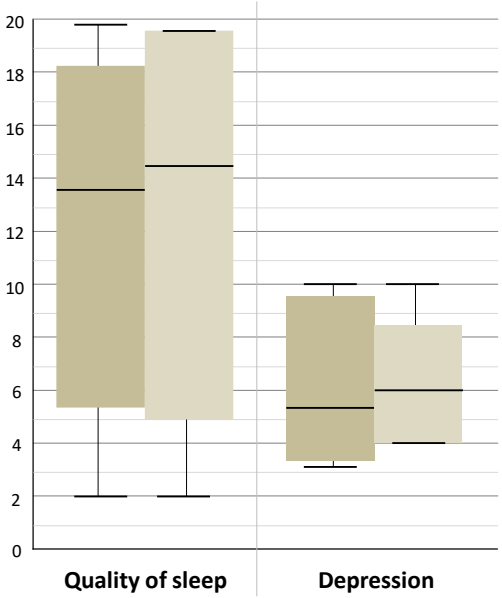


Figure 6 **Box diagram. The following figures show baseline data from the experimental group vs. control group. Mean scores from FIQ pain, fatigue, physical function and anxiety and SD scores from questionnaires used to asses quality of sleep and depression have been used for the analysis. Darker boxes represent experimental groups while lighter ones control groups. Minimal, maximum, mean scores and SD are analyzed.**

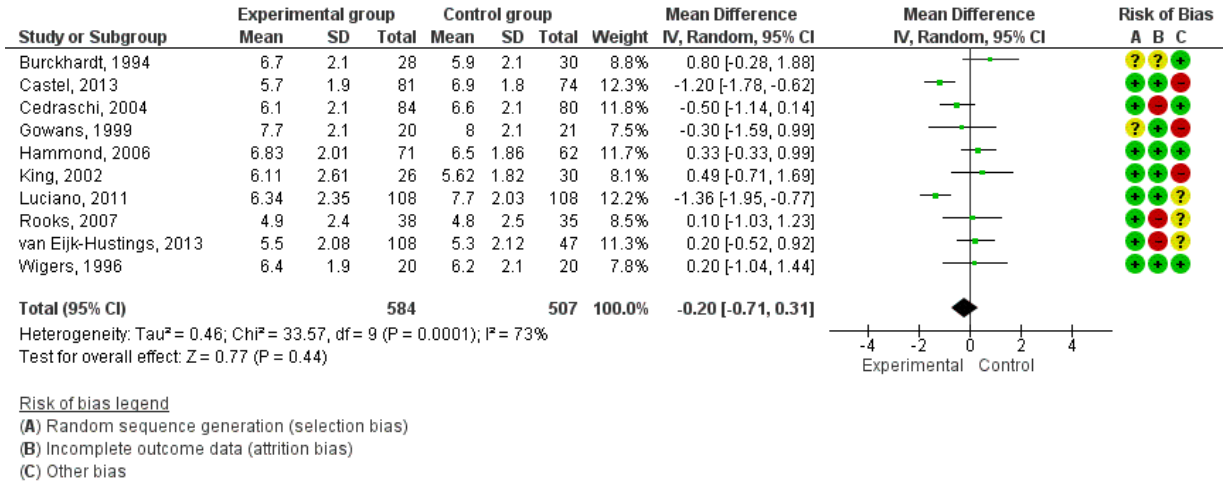


Figure 7 **Effect size of multidisciplinary treatment on pain at post-treatment.**

(BDI, HADS, PGWB) due to different questionnaires were used for the analysis SD scores were examined.

In **Figure 6** is represented by a box diagram the minimal and maximum scores, the average score and the standard deviation from each outcome. No randomization bias was found (For detailed data see **Supplementary fig. 5**).

4.7. Efficacy of multidisciplinary treatments on pain

The effect size on pain is shown in **Figure 7**. 10 studies were included to assess the effectiveness of multidisciplinary treatments on pain. Mean scores of the questionnaire FIQ-Pain were used for the statistical analysis. 584 patients constituted the N of the multidisciplinary treatment, while the sample of the control group was formed of 507 patients. The total sample size was of 1091 patients with diagnosis of FM. Due to the fact that the total N was higher than 300 and the confidence interval (95% CI) was -0.2 [-0.71, 0.31] in a scale from 0 to 10, not serious imprecision was found for this outcome.

Risk of bias entailed a very serious factor. 6 out of the 10 studies (60%) included in the statistical analysis in FIQ-Pain showed high risk in at least one of the analyzed bias. 3 showed high risk in attrition bias and the remaining 3 showed high risk in other bias (reasons specified in **Supplementary fig. 6**). Studies appeared to show significant heterogeneity between them (I^2=73%). No publication bias were found (funnel plot shown in **Supplementary fig. 7**).

Very low certainty of the evidence was found for this outcome. There is very little confidence in the effect estimate so the true effect is likely to be substantially different from the estimate of effect. For this reason, the evidence is very uncertain about the effect of multidisciplinary treatments on pain.

4.8. Efficacy of multidisciplinary treatments on fatigue

The effect size on fatigue is shown in **Figure 8**. The improvement on fatigue in FM condition after receiving a multidisciplinary treatment compared to the improvement after receiving usual care was measured based on 8 RCTs. The total sample size constituted of 878 patients, 477 in the experimental group and 401 in the control group. Not serious imprecision was found as more than 300 patients with FM diagnosis were used in the statistical analysis and the confidence interval (95% CI) was -0.18 [-0.72, 0.36].

Risk of bias was considered a very serious factor. 4 out of the 8 studies (50%) included in the statistical analysis in FIQ-Fatigue showed high risk in at least one of the analyzed bias. 3 showed high risk in attrition

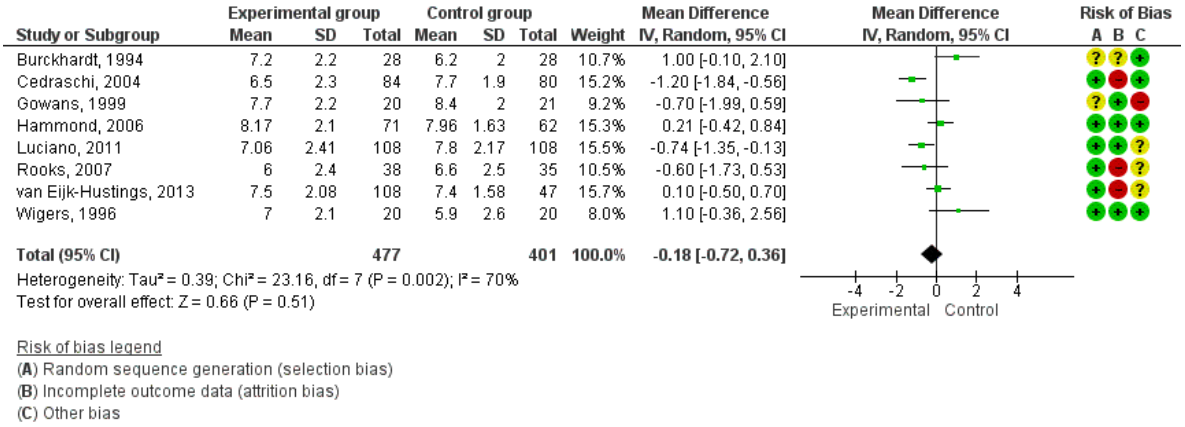


Figure 8 **Effect size of multidisciplinary treatment on fatigue at post-treatment.**

bias and the remaining one showed high risk in other bias (reasons specified in **Supplementary fig. 8**).

Studies appeared to show significant differences between them as a high heterogeneity on the estimations was found (I^2=70%). No publication bias were detected as only 8 studies were analyzed for this outcome (<10).

Very low evidence was found for this outcome. There is very little confidence in the effect estimate so the true effect is likely to be substantially different from the estimate of effect. Therefore, multidisciplinary treatments may have little to no effect on fatigue but the evidence is very uncertain.

4.9. Efficacy of multidisciplinary treatments on quality of sleep

The effect size on quality of sleep is shown in **Figure 9**. 3 RCTs were included and 4 comparisons were accomplished for the statistical analysis of quality of sleep. Several questionnaires were used for the assessment of this outcome (MOSS-S, FIQ) so standard deviation was used to calculate the mean difference.

230 patients were allocated in the experimental group and 207 patients in the control group. The total sample size was of 437 participants. The total confidence interval (95% CI) was 0.37 [0.16, 0.57]. Not serious imprecision was found for this outcome.

Risk of bias was considered a very serious factor. All the studies (100%) included in the statistical analysis in the assessment of quality of sleep showed high risk in at least one of the analyzed bias. 2 showed high risk in attrition bias and another 2 studies showed high risk in other bias (reasons specified in **Supplementary fig. 9**).

A huge heterogeneity was found between the included studies (I^2=95%). No publication bias was detected as only 4 studies were analyzed for this outcome (<10).

Very low evidence was found for this outcome. There is very little confidence in the effect estimate so the

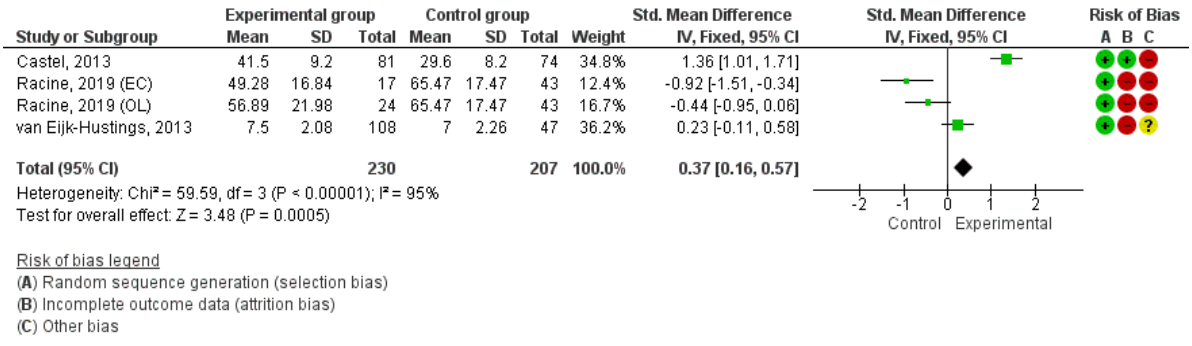


Figure 9 **Effect size of multidisciplinary treatment on quality of sleep at post-treatment.**

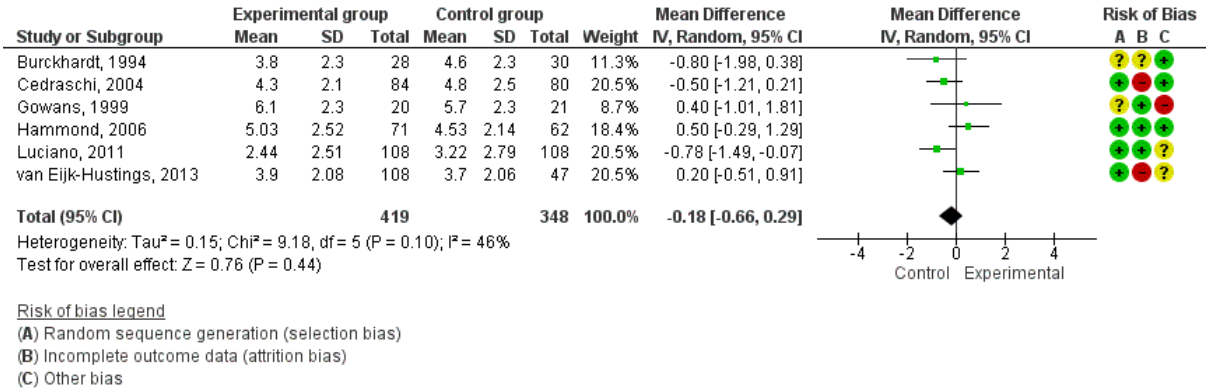


Figure 10 **Effect size of multidisciplinary treatment on physical function at post-treatment.**

true effect is likely to be substantially different from the estimate of effect. The evidence is very uncertain about the effect of multidisciplinary treatments on quality of sleep.

4.10. Efficacy of multidisciplinary treatments on physical function

The effect size on physical function is shown in **Figure 10**. Mean scores of FIQ-Physical function from 6 RCTs were used in the statistical analysis for the assessment of this outcome. The total sample size was of 767 participants, 419 patients in the experimental group and 348 in the control group. The confidence interval (95% CI) was -0.18 [-0.66, 0.29] in a scale from 0 to 10. Not serious imprecision was found for this outcome.

Very serious risk of bias was found in the included RCTs. 3 out of the 6 studies (50%) included in the statistical analysis in FIQ-Physical showed high risk in at least one of the analyzed bias. 2 showed high risk in attrition bias and the remaining one showed high risk in other bias (reasons specified in **Supplementary fig. 10**).

Heterogeneity between the studies was not considered to be a serious factor (I^2=46%). No publication bias was detected as only 6 studies were analyzed for this outcome (<10).

Low certainty was achieved for this outcome. The confidence in the effect estimate is limited as the true effect may be substantially different from the estimate of the effect. Multidisciplinary treatments may result in little to no difference in physical function.

4.11. Efficacy of multidisciplinary treatments on depression

The effect size on depression is shown in **Figure 11**. Many questionnaires were used for the assessment of the present outcome (BDI, HADS, PGWB) so standard deviations were used for the statistical analysis. 4 RCTs were included yet 6 comparisons were accomplished. 445 participants formed the total cluster, divided in 248 patients in the experimental group and 197 in the control group. The confidence interval (95% CI) was -0.30 [-0.74, 0.14]. Not serious imprecision was found.

All the studies (100%) included in the statistical analysis in the assessment of depression showed high risk in at least one of the analyzed bias. 3 showed high risk in attrition bias and 2 studies showed high risk in other bias. For this reason, risk of bias was considered a very serious factor (reasons specified in **Supplementary fig. 11**).

Studies appeared to show significant differences heterogeneity (I^2=79%). Nonetheless, no publication bias was detected as only 4 studies were analyzed for this outcome (<10).

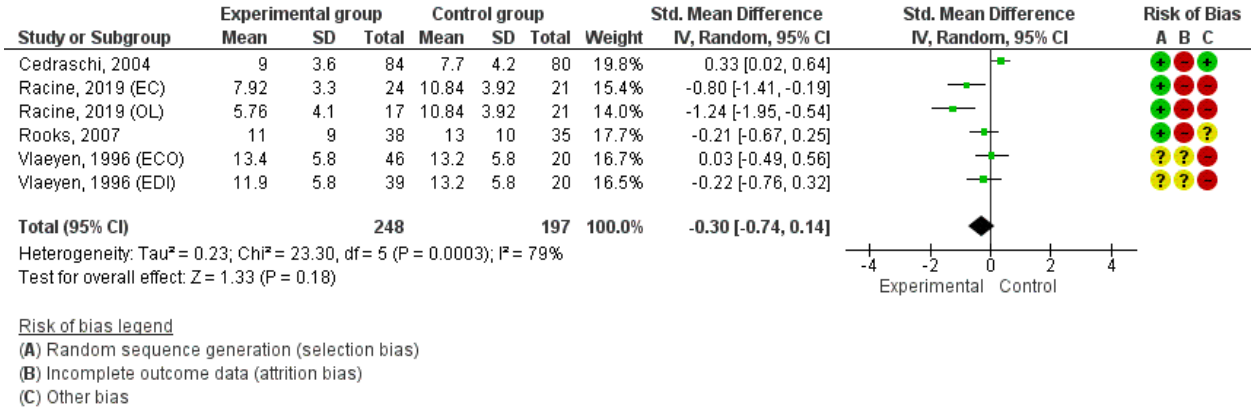


Figure 11 **Effect size of multidisciplinary treatment on depressive symptoms at post-treatment.**

Very low certainty was determined for this outcome. There is very little confidence in the effect estimate as the true effect is likely to be substantially different from the estimate of effect. Multidisciplinary treatments may reduce depressive symptoms in FM but the evidence is very uncertain.

4.12. Efficacy of multidisciplinary treatments on anxiety

The effect size on anxiety is shown in **Figure 12**. 7 studies that used FIQ-Anxiety as measurement item for this outcome were used in the statistical analysis. 457 participants took part in the experimental treatment while 383 patients were allocated in the control group. 840 patients with FM diagnosis formed the total cluster used to assess this outcome. The confidence interval (95% CI) was -0.57 [-1.17, 0.02]. Not serious imprecision was found for this outcome.

Once again, risk of bias was considered a very serious factor as 4 out of the 7 studies (≈ 60%) included in the statistical analysis in FIQ-Anxiety showed high risk in at least one of the analyzed bias. 3 showed high risk in attrition bias and the remaining one showed high risk in other bias (reasons specified in **Supplementary fig. 12**).

Neither significant heterogeneity was found between the included studies (I^2=54%) nor any publication bias were detected as 7 studies were analyzed for this outcome (<10).

Low certainty of the evidence was found for this outcome. The confidence in the effect estimate is limited as the true effect may be substantially different from the estimate of the effect. Due to this fact, multidisciplinary treatments may result in little to no difference in anxiety.

5. DISCUSSION

The main objective of this meta-analysis was to determine the efficacy of multidisciplinary treatments

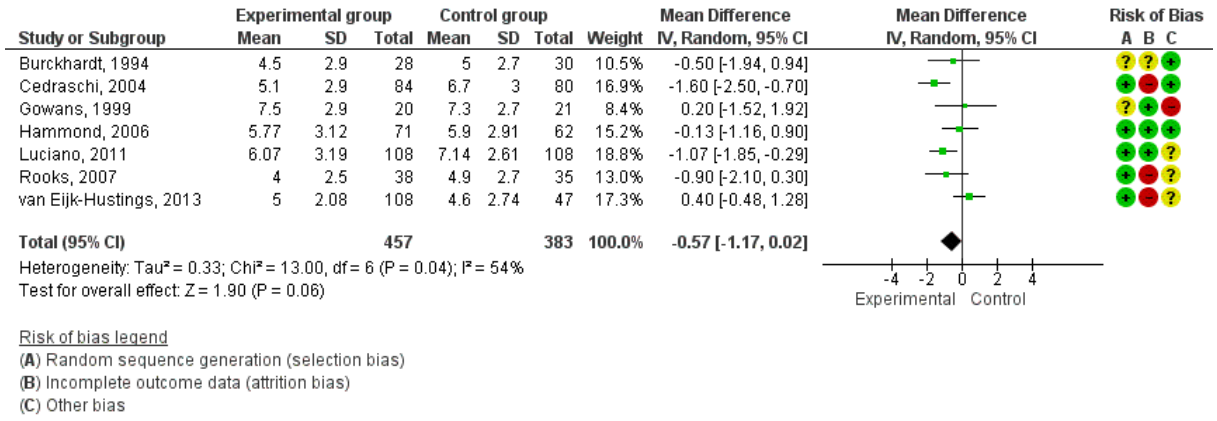


Figure 12 **Effect size of multidisciplinary treatment on anxiety at post-treatment.**

in FM. There is not strong evidence of the efficacy of multidisciplinary treatments to reduce some key symptoms in FM, such as pain, fatigue, sleeping disturbances, physical impairment, anxiety and depressive symptoms at post-treatment assessments. Low evidence was found for physical function and anxiety while very low evidence was found for pain, fatigue, quality of sleep and depression. No clinically meaningful results were found for any of the assessed outcomes. The provided data could be applied to FM patients sharing similar characteristics: middle aged FM diagnosed females. No statements on the efficacy of multidisciplinary treatments in infants, adolescents, elderly, males and people suffering from FM together with other co-morbidities are possible.

In the present meta-analysis suitable RCTs from two previously performed systematic reviews on this topic were included. Häuser et al (65) performed a meta-analysis to review the short- and long-term efficacy of multicomponent treatments for FM. The results obtained differ from our findings. Strong evidence of efficacy was found for reduction of pain, fatigue and depressive symptoms as well as for the improvement of self-efficacy and physical function at post-treatment assessment. Nonetheless, only improvements on self-efficacy and physical function endured over follow up assessments. These discrepancies could be due to the criteria used for the assessment of the methodological quality. Van Tulder score was used for the analysis. 5/7 of the studies (58–61,64) included in the meta-analysis were classified as moderate quality and 2/7 (56,57) as low quality ones. Strong evidence was considered when at least 2 moderate-quality RCTs constituted the findings and moderate evidence was considered when at least 2 low-quality RCTs and/or 1 moderate quality RCT supported the results. In our study, the methodological assessment of the trial was performed following the Review Manager framework for risk of bias evaluation and classified for many reasons as a very serious factor in GRADEpro GDT. So the proportion of information from studies with high risk of bias was sufficient to affect the interpretation of the results. Regarding heterogeneity, no inconsistency (I^2) was found for pain and fatigue in Häusers study while in the present study a very serious inconsistency was found for both of the outcomes. The data provided was based on a smaller N compared to the one in our meta-analysis.

Karjalainen et al (66) analyzed the effectiveness of multidisciplinary rehabilitation for adults suffering from FM and widespread musculoskeletal pain. The results presented were similar to the results obtained in our study. Due to the low quality of the methodological frame of the included studies, the evidence obtained was limited and inconsistent so multidisciplinary treatments were graded as ineffective.

There are several limitations in the present study. First of all, the approaches that formed the multidisciplinary treatments included in the meta-analysis differed from one another. As there is not a unified criterion on which components should multidisciplinary treatments gather, a great heterogeneity was found. Regarding the duration of the therapies, the same limitation was found, concerning the duration – on weeks long – and the sessions that were provided. Many reviews have focused their attention on defining components for multidisciplinary approaches aimed to improve symptoms of FM. Multidisciplinary treatments have been defined by entailing a physical approach and at least another element from psychological, social and occupational dimensions (67). These include exercise interventions, pain management strategies, cognitive-behavioral therapies, coping skills training, educational programs, mind-body therapies, complementary and alternative medicine and/or pharmacotherapy (68). Despite the heterogeneity on the combination of approaches between studies, the goal of multidisciplinary treatments should be focused on providing patients information about the condition

and coping skills so they are able to manage their symptoms on their daily life and they know how to self-manage their symptoms leading to positive changes (41).

94.1% (16/17) of the multidisciplinary treatments included in the present systematic review were formed by at least an approach aimed at improving physical function. Bidonde et al performed an umbrella systematic review to identify and evaluate approaches based on exercise for adults with FM. Results showed that there is enough evidence to determine that different modalities of exercise training improve pain, multidimensional and physical function for individuals with FM diagnosis. These findings support the effectiveness of therapeutic exercise in the management of the syndrome. Nonetheless, it remains unclear what are the most appropriate parameters to carry out exercise interventions (69). A study that evaluated the therapeutic validity of exercise interventions included in the mentioned umbrella systematic review determined that the validity found was low. The findings showed that most of the exercise approaches did not follow the recommendations published by the American College of Sports Medicine (ACSM) regarding frequency, time, intensity, sets, repetitions and type of exercise (70). Standardization in exercise variables could help determining the effectiveness of this approach and would help in clinical practice at the time of prescribing and recommending exercises for individuals with FM (71).

Efficacy of non-pharmacological interventions has been studied not only at clinical levels but on biomarker expression too (72). Among pro-inflammatory cytokines, serum levels of interleukin-8 (IL-8) were observed to be decreased after providing a multidisciplinary pain intervention at the 6 month post-treatment assessment. The multidisciplinary treatment lasted for 3 weeks (120h) and integrated physical exercise, ergonomic training, psychotherapy, education, behavioral therapy and work-placed interventions. 6 months after providing the therapy IL-8 was reduced nearly to normal range and correlated positively with pain intensity (73). IL-8 levels were also reduced 4 months after a pool based aquatic exercise intervention (72). The mechanisms of central sensitization and hyperalgesia involve the interaction between neurons and activated glia cells. Activated glia cells release pro-inflammatory cytokines – substances with the potential for pain amplification – such as IL-8. This mechanism could be an important factor for the development and maintenance of chronic pain conditions (74). The endocannabinoid system also plays an important role in the modulation of pain. Anandamide levels were reported to be significantly increased after a 15 week resistance exercise program which came along with decreased scores of pain intensity, depressive symptoms and an increment on muscle strength (75). Consequently exercise approaches could ameliorate the inflammatory status of FM patients by reducing values of inflammatory factors, so these are closer to the baseline levels of healthy individuals (72), and increase levels of endocannabinoids which are related with anti-nociceptive effects (76).

Pharmacotherapy was not controlled in most studies. Therefore, remains uncertain whether the reported effects are only due to the applied multidisciplinary treatments or even to treatment condition effect. Chen et al (77) studied the effect of placebo treatment in FM. A wide range of different treatments for FM were compared to whether placebo or no treatment condition. Participants receiving placebo treatment obtained significant improvements in all the main outcome measures such as pain, fatigue, sleep quality, function and overall well being. Effects seem to be superior in placebo treatment compared to no treatment ones. So people with FM treated with placebo can show significant improvement in pain and other outcomes which support the clinically significant analgesic effect of placebo treatments. Further studies are needed in this field.

It would be interesting to have a default consensus on the outcomes that should be measured and the most appropriate items for the measurement of each outcome. In the included primary studies questionnaires that were used to assess each outcome varied. The lack of standard outcome measures makes interpretation, combination and comparison of the results of clinical trials a challenging task. The Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT) are responsible for reaching consensus on what should be measured and how these should be measured in clinical trials (78). Pain, fatigue, multidimensional function and sleep disturbances are included as core domains. There are discrepancies about including depression as a core domain as it is still uncertain the best way to assess it in FM. Anxiety is not considered to be an essential measured outcome (79). Regarding questionnaires of election for each outcome measurement many have been established. For pain assessment are employed McGill Pain Questionnaire (MPQ) (80) and the Brief Pain Inventory (BPI) (46). Tenderness is separately assessed from patient reported pain and measured by TP examination and dolorimetry (78). Fatigue should ideally be measured by The Multidimensional Assessment of Fatigue index, the Functional Assessment of Chronic Illness Therapy (81), the Fatigue Severity Scale (82) or even the FIQ subscale for fatigue. MOS sleep scale (83) and FIQ are indicated for sleep disturbances. Co-morbid psychiatric disorders such as depression and anxiety are measured by Beck Depression Inventory and the Beck Anxiety Inventory (84), the Hospital Anxiety and Depression Scale (85) and the Hamilton Depression Rating Scale (86). Some of the mentioned scales have been the questionnaires of election in this meta-analysis.

Due to the limitation of time, in the present systematic review no grey literature was included. Nonetheless, to ensure the majority of the literature available on multidisciplinary treatments for FM was introduced, prior to this study an overview on pharmacological and non-pharmacological treatments for FM was accomplished. The aim of this quick scoping review was to create a wide perspective of the available treatments for FM and help classifying all the different types of approaches described to date. In this way, indirectly, reviews of multidisciplinary treatments were achieved and the primary studies were extracted to be assessed and included in the current study. In addition, a comprehensive literature research was carried out complementary to the study extraction from the overview.

Our study also shows some significant strengths. A strict inclusion and exclusion criterion was applied, and as a result, the quality of studies was satisfactory. Only high quality RCTs were included so potential biases were minimized by excluding non-randomized and/or uncontrolled trials. Due to the variability in the clinical frame that characterizes FM specific diagnostic criteria was required for study inclusion (ACR 1990-2010 diagnostic criteria). In this way, we ensured that multidisciplinary treatments were exclusively tested on patients with this condition. Intention-to-treat data was analyzed when this was available (53).

6. CONLUSIONS

The evidence is very uncertain about the effectiveness of multidisciplinary treatments in fibromyalgia. If multidisciplinary treatments are already characterized for incorporating a wide variety of therapeutic approaches, the lack of standardized therapies, even standardized approaches within the therapies (eg. activities aimed at improving physical activity), result in a great heterogeneity that is difficult to manage at the time of comparing data and evaluating its effectiveness. This heterogeneity is also contemplated between authors of studies on this topic when concluding in the interpretation of the results. In our study, and in line with a previous meta-analysis also based on the structure proposed by Cochrane, it has been determined that the evidence is not strong

enough to declare that multidisciplinary treatments have the ability to improve core symptoms in this syndrome. Probably, if future RCTs followed a pattern both in the therapies – approaches and duration – and in the way of measuring results – items and measurement periods – results obtained in meta-analysis could acquire a more conclusive character. It would be interesting if future researches in this area were aimed at creating more precise guidelines on the characteristics that this type of treatments should meet as well as more detailed evaluation criteria.

7. BIBLIOGRAPHY

[1] Fibromyalgia [Internet]. Rheumatology.org. 2020 [cited 31 March 2020]. Available from: https://www.rheumatology.org/I-Am-A/Patient-Caregiver/Diseases-Conditions/Fibromyalgia.

[2] Central sensitivity syndromes: a new paradigm and group nosology for fibromyalgia and overlapping conditions, and the related issue of disease versus illness. Seminars in arthritis and rheumatism: Elsevier; 2008.

[3] Häuser W, Ablin J, Fitzcharles M, Littlejohn G, Luciano JV, Usui C, et al. Fibromyalgia. Nature reviews Disease primers 2015;1(1):1-16.

[4] Buskila D, Atzeni F, Sarzi-Puttini P. Etiology of fibromyalgia: the possible role of infection and vaccination. Autoimmunity reviews 2008;8(1):41-43.

[5] Buskila D, Neumann L, Vaisberg G, Alkalay D, Wolfe F. Increased rates of fibromyalgia following cervical spine injury. A controlled study of 161 cases of traumatic injury. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology 1997;40(3):446-452.

[6] Häuser W, Kosseva M, Üceyler N, Klose P, Sommer C. Emotional, physical, and sexual abuse in fibromyalgia syndrome: A systematic review with metaanalysis. Arthritis care & research 2011;63(6):808-820.

[7] Le Goff P. Is fibromyalgia a muscle disorder? Joint Bone Spine 2006;73(3):239-242.

[8] Bazzichi L, Dini M, Rossi A, Corbianco S, De Feo F, Giacomelli C, et al. Muscle modifications in fibromyalgic patients revealed by surface electromyography (SEMG) analysis. BMC Musculoskeletal disorders 2009;10(1):36.

[9] Elvin A, Siösteen A, Nilsson A, Kosek E. Decreased muscle blood flow in fibromyalgia patients during standardised muscle exercise: a contrast media enhanced colour Doppler study. European journal of pain 2006;10(2):137-144.

[10] Sprott H, Salemi S, Gay RE, Bradley LA, Alarcon GS, Oh SJ, et al. Increased DNA fragmentation and ultrastructural changes in fibromyalgic muscle fibres. Ann Rheum Dis 2004 Mar;63(3):245-251.

[11] Gronemann S, Ribel-Madsen S, Bartels E, Danneskiold-Samsøe B, Bliddal H. Collagen and muscle pathology in fibromyalgia patients. Rheumatology 2004;43(1):27-31.

[12] Escolar Martín JM, Durán Barbosa R. Fisiopatología de la fibromialgia: Alteraciones a nivel cerebral y muscular. Fisioterapia. 2011;33(4):173–82.

[13] Guedj E, Cammilleri S, Niboyet J, Dupont P, Vidal E, Dropinski JP, et al. Clinical correlate of brain SPECT perfusion abnormalities in fibromyalgia. J Nucl Med 2008 Nov;49(11):1798-1803.

[14] Kuchinad A, Schweinhardt P, Seminowicz DA, Wood PB, Chizh BA, Bushnell MC. Accelerated brain gray matter loss in fibromyalgia patients: premature aging of the brain? J Neurosci 2007 Apr 11;27(15):4004-4007.

[15] Wood PB, Glabus MF, Simpson R, Patterson 2nd JC. Changes in gray matter density in fibromyalgia: correlation with dopamine metabolism. The Journal of Pain 2009;10(6):609-618.

[16] Lutz J, Jäger L, de Quervain D, Krauseneck T, Padberg F, Wichnalek M, et al. White and gray matter abnormalities in the brain of patients with fibromyalgia: A diffusion -tensor and volumetric imaging study. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology 2008;58(12):3960-3969.

[17] Harris RE, Sundgren PC, Craig A, Kirshenbaum E, Sen A, Napadow V, et al. Elevated insular glutamate in fibromyalgia is associated with experimental pain. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology 2009;60(10):3146-3152.

[18] Autonomic dysfunction in patients with fibromyalgia: application of power spectral analysis of heart rate variability. Seminars in arthritis and rheumatism: Elsevier; 2000.

[19] Cook DB, Lange G, Ciccone DS, Liu WC, Steffener J, Natelson BH. Functional imaging of pain in patients with primary fibromyalgia. J Rheumatol 2004 Feb;31(2):364-378.

[20] Marques AP, Santo A de S do E, Berssaneti AA, Matsutani LA, Yuan SLK. Prevalence of fibromyalgia: literature review update. Rev Bras Reumatol (English Ed. 2017;57(4):356–63.

[21] Conversano C, Marchi L, Rebecca C, Gemignani A, Bazzichi LM, Carmassi C, et al. Personality traits in fibromyalgia (Fm): Does fm personality exists? a systematic review. Clin Pract Epidemiol Ment Heal. 2018;14:223–32.

[22] Ellingson LD, Stegner AJ, Schwabacher IJ, Lindheimer JB, Cook DB. Catastrophizing interferes with cognitive modulation of pain in women with fibromyalgia. Pain Medicine 2018;19(12):2408-2422.

[23] Álvarez Mena MG, Álvarez Mena PR, Montes Reina MJ, Castillo Jumbo EP, Mafla Andrade J. Fibromialgia. Avances en su tratamiento. Revista Cubana de Reumatología 2019;21(2):1-8.

[24] Kwiatek R. Treatment of fibromyalgia. Aust Prescr. 2017;40(5):179–83.

[25] Moore RA, Derry S, Aldington D, Cole P, Wiffen PJ. Amitriptyline for neuropathic pain and fibromyalgia in adults. Cochrane Database of Systematic Reviews 2012(12).

[26] Goldenberg DL. Pharmacological treatment of fibromyalgia and other chronic musculoskeletal pain. Best practice & research Clinical rheumatology 2007;21(3):499-511.

[27] Bennett RM, Gatter RA, Campbell SM, Andrews RP, Clark SR, Scarola JA. A comparison of cyclobenzaprine and placebo in the management of fibrositis. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology 1988;31(12):1535-1542.

[28] Lunn MP, Hughes RA, Wiffen PJ. Duloxetine for treating painful neuropathy, chronic pain or fibromyalgia. Cochrane Database of Systematic Reviews 2014(1).

[29] Clauw DJ, Mease P, Palmer RH, Gendreau RM, Wang Y. Milnacipran for the treatment of fibromyalgia in adults: a 15-week, multicenter, randomized, double-blind, placebo-controlled, multiple-dose clinical trial. Clin Ther 2008;30(11):1988-2004.

[30] Mease PJ, Clauw DJ, Gendreau RM, Rao SG, Kranzler J, Chen W, et al. The efficacy and safety of milnacipran for treatment of fibromyalgia. a randomized, double-blind, placebo-controlled trial. J Rheumatol 2009 Feb;36(2):398-409.

[31] de Oliveira Júnior, José Oswaldo, De Almeida MB. The current treatment of fibromyalgia. Br J Pain.São Paulo 2018;1(3):255-262.

[32] Fink K, Dooley DJ, Meder WP, Suman-Chauhan N, Duffy S, Clusmann H, et al. Inhibition of neuronal Ca2

influx by gabapentin and pregabalin in the human neocortex. Neuropharmacology 2002;42(2):229-236.

[33] Baraniuk JN, Whalen G, Cunningham J, Clauw DJ. Cerebrospinal fluid levels of opioid peptides in fibromyalgia and chronic low back pain. BMC musculoskeletal disorders 2004;5(1):48.

[34] Harris RE, Clauw DJ, Scott DJ, McLean SA, Gracely RH, Zubieta JK. Decreased central mu-opioid receptor availability in fibromyalgia. J Neurosci 2007 Sep 12;27(37):10000-10006.

[35] Macfarlane GJ, Kronisch C, Dean LE, Atzeni F, Häuser W, Fluß E, et al. EULAR revised recommendations for the management of fibromyalgia. Ann Rheum Dis 2017;76(2):318-328.

[36] Theadom A, Cropley M, Smith HE, Feigin VL, McPherson K. Mind and body therapy for fibromyalgia. Cochrane Database of Systematic Reviews 2015(4).

[37] Deare JC, Zheng Z, Xue CC, Liu JP, Shang J, Scott SW, et al. Acupuncture for treating fibromyalgia. Cochrane Database of Systematic Reviews 2013(5).

[38] Hou W, Wang T, Kang J. The effects of add-on non-invasive brain stimulation in fibromyalgia: a meta-analysis and meta-regression of randomized controlled trials. Rheumatology 2016;55(8):1507-1517.

[39] Scascighini L, Toma V, Dober-Spielmann S, Sprott H. Multidisciplinary treatment for chronic pain: A systematic review of interventions and outcomes. Rheumatology. 2008;47(5):670–8.

[40] Goldenberg DL. Multidisciplinary modalities in the treatment of fibromyalgia. J Clin Psychiatry 2008;69(Suppl 2):30-34.

[41] Bennett RM. Multidisciplinary group programs to treat fibromyalgia patients. Rheum Dis Clin North Am 1996 May;22(2):351-367.

[42] Wolfe F, Smythe HA, Yunus MB, Bennett RM, Bombardier C, Goldenberg DL, et al. The American College of Rheumatology 1990 criteria for the classification of fibromyalgia. Arthritis & Rheumatism: Official Journal of the American College of Rheumatology 1990;33(2):160-172.

[43] Gur M, Gulkesen A, Akgol G. Comparison of ACR 1990 and ACR 2010 classification criteria in fibromyalgia syndrome. Med Sci | Int Med J. 2019;8(3):1.

[44] Salaffi F, Ciapetti A, Puttini PS, Atzeni F, Iannuccelli C, Di Franco M, et al. Preliminary identification of key clinical domains for outcome evaluation in fibromyalgia using the Delphi method: the Italian experience. Reumatismo 2012:27-34.

[45] H. Merskey and N. Bogduk, IASP Press S. IASP Terminology - IASP [Internet]. Classification of Chronic Pain, Second Edition, IASP Task Force on Taxonomy. 1994. p. 209–14. Available from: https://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698%0Ahttps://www.iasp-pain.org/terminology?navItemNumber=576

[46] Williams DA, Arnold LM. Measures of fibromyalgia: Fibromyalgia Impact Questionnaire (FIQ), Brief Pain Inventory (BPI), Multidimensional Fatigue Inventory (MFI-20), Medical Outcomes Study (MOS) Sleep Scale, and Multiple Ability Self-Report Questionnaire (MASQ). Arthritis Care Res (Hoboken) 2011 Nov;63 Suppl 11:S86-97.

[47] Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF. Arthritis Care Res. 2011;63(SUPPL. 11):240–52.

[48] Casanueva-Fernández B, Llorca J, i Rubió JB, Rodero-Fernández B, González-Gay MA. Efficacy of a

multidisciplinary treatment program in patients with severe fibromyalgia. Rheumatol Int 2012;32(8):2497-2502.

[49] Castel A, Fontova R, Montull S, Periñán R, Poveda MJ, Miralles I, et al. Efficacy of a multidisciplinary fibromyalgia treatment adapted for women with low educational levels: a randomized controlled trial. Arthritis care & research 2013;65(3):421-431.

[50] Lera S, Gelman SM, López MJ, Abenoza M, Zorrilla JG, Castro-Fornieles J, et al. Multidisciplinary treatment of fibromyalgia: does cognitive behavior therapy increase the response to treatment? J Psychosom Res 2009;67(5):433-441.

[51] Luciano JV, Martínez N, Peñarrubia-María MT, Fernandez-Vergel R, García-Campayo J, Verduras C, et al. Effectiveness of a psychoeducational treatment program implemented in general practice for fibromyalgia patients: a randomized controlled trial. Clin J Pain 2011;27(5):383-391.

[52] Racine M, Jensen MP, Harth M, Morley-Forster P, Nielson WR. Operant learning versus energy conservation activity pacing treatments in a sample of patients with fibromyalgia syndrome: A pilot randomized controlled trial. The Journal of Pain 2019;20(4):420-439.

[53] van Eijk-Hustings Y, Kroese M, Tan F, Boonen A, Bessems-Beks M, Landewé R. Challenges in demonstrating the effectiveness of multidisciplinary treatment on quality of life, participation and health care utilisation in patients with fibromyalgia: a randomised controlled trial. Clin Rheumatol 2013;32(2):199-209.

[54] van Koulil S, van Lankveld W, Kraaimaat FW, van Helmond T, Vedder A, van Hoorn H, et al. Tailored cognitive-behavioural therapy and exercise training improves the physical fitness of patients with fibromyalgia. Ann Rheum Dis 2011 Dec;70(12):2131-2133.

[55] Burckhardt CS, Mannerkorpi K, Hedenberg L, Bjelle A. A randomized, controlled clinical trial of education and physical training for women with fibromyalgia. J Rheumatol 1994 Apr;21(4):714-720.

[56] Cedraschi C, Desmeules J, Rapiti E, Baumgartner E, Cohen P, Finckh A, et al. Fibromyalgia: a randomised, controlled trial of a treatment programme based on self management. Ann Rheum Dis 2004 Mar;63(3):290-296.

[57] Gowans S, DeHueck A, Voss S, Richardson M. A randomized, controlled trial of exercise and education for individuals with fibromyalgia. Arthritis Care & Research 1999;12(2):120-128.

[58] Hammond A, Freeman K. Community patient education and exercise for people with fibromyalgia: a parallel group randomized controlled trial. Clin Rehabil 2006;20(10):835-846.

[59] King SJ, Wessel J, Bhambhani Y, Sholter D, Maksymowych W. The effects of exercise and education, individually or combined, in women with fibromyalgia. J Rheumatol 2002 Dec;29(12):2620-2627.

[60] Lemstra M, Olszynski W. The effectiveness of multidisciplinary rehabilitation in the treatment of fibromyalgia: a randomized controlled trial. Clin J Pain 2005;21(2):166-174.

[61] Rooks DS, Gautam S, Romeling M, Cross ML, Stratigakis D, Evans B, et al. Group exercise, education, and combination self-management in women with fibromyalgia: a randomized trial. Arch Intern Med 2007;167(20):2192-2200.

[62] Vlaeyen JW, Teeken-Gruben NJ, Goossens ME, Rutten-van Mölken M, Pelt R, Eek Hv, et al. Cognitive-educational treatment of fibromyalgia: a randomized clinical trial. I. Clinical effects. 1996.

[63] Wigers SHr, Stiles T, Vogel P. Effects of aerobic exercise versus stress management treatment in fibromyalgia. Scand J Rheumatol 1996;25(2):77-86.

[64] Zijlstra T, van de Laar, Mart AFJ, Bernelot Moens H, Taal E, Zakraoui L, Rasker JJ. Spa treatment for primary fibromyalgia syndrome: a combination of thalassotherapy, exercise and patient education improves symptoms and quality of life. Rheumatology 2005;44(4):539-546.

[65] Häuser W, Bernardy K, Arnold B, Offenbächer M, Schiltenwolf M. Efficacy of multicomponent treatment in fibromyalgia syndrome: a meta-analysis of randomized controlled clinical trials. Arthritis Care & Research 2009;61(2):216-224.

[66] Karjalainen KA, Malmivaara A, van Tulder MW, Roine R, Jauhiainen M, Hurri H, et al. Multidisciplinary rehabilitation for fibromyalgia and musculoskeletal pain in working age adults. Cochrane Database of Systematic Reviews 1999(3).

[67] Banerjee S, Argaez C. Multidisciplinary Treatment Programs for Patients with Chronic Non-Malignant Pain : A Review of Clinical Effectiveness , Cost-Effectiveness , and Guidelines. CADTH Rapid Response Rep. 2017.

[68] Burckhardt C. Multidisciplinary approaches for management of fibromyalgia. Curr Pharm Des 2006;12(1):59-66.

[69] Bidonde J, Jean Busch A, Bath B, Milosavljevic S. Exercise for adults with fibromyalgia: an umbrella systematic review with synthesis of best evidence. Current rheumatology reviews 2014;10(1):45-79.

[70] Pescatello, Linda S.; Arena, Ross; Riebe, Deborah; Thompson PD. ACSM's Guidelines for Exercise Testing and Prescription. Ninth Edit. (Pescatello LS, ed.) Wolters Kluwer Health; 2013.

[71] Álvarez-Gallardo IC, Bidonde J, Busch A, Westby M, Kenny GP, Delgado-Fernández M, et al. Therapeutic validity of exercise interventions in the management of fibromyalgia. J Sports Med Phys Fitness. 2019;59(5):828–38.

[72] Sanada K, Díez MA, Valero MS, Pérez-Yus MC, Demarzo MM, García-Toro M, et al. Effects of non-pharmacological interventions on inflammatory biomarker expression in patients with fibromyalgia: a systematic review. Arthritis research & therapy 2015;17(1):272.

[73] Wang H, Buchner M, Moser MT, Daniel V, Schiltenwolf M. The role of IL-8 in patients with fibromyalgia: A prospective longitudinal study of 6 months. Clin J Pain. 2009;25(1):1–4.

[74] Kosek E, Altawil R, Kadetoff D, Finn A, Westman M, Le Maître E, et al. Evidence of different mediators of central inflammation in dysfunctional and inflammatory pain-interleukin-8 in fibromyalgia and interleukin-1 β in rheumatoid arthritis. J Neuroimmunol 2015;280:49-55.

[75] Stensson N, Gerdle B, Ernberg M, Mannerkorpi K, Kosek E, Ghafouri B. Increased Anandamide and Decreased Pain and Depression after Exercise in Fibromyalgia. Med Sci Sports Exerc 2020 Jul;52(7):1617-1628.

[76] Cravatt BF, Lichtman AH. The endogenous cannabinoid system and its role in nociceptive behavior. J Neurobiol. 2004;61(1):149–60.

[77] Chen X, Zou K, Abdullah N, Whiteside N, Sarmanova A, Doherty M, et al. The placebo effect and its determinants in fibromyalgia: meta-analysis of randomised controlled trials. Clin Rheumatol. 2017;36(7):1623–30.

[78] Choy EH. Clinical domains of fibromyalgia syndrome: Determination through the OMERACT process. J Musculoskelet Pain. 2010;18(4):380–6.

[79] Mease P, Arnold LM, Choy EH, Clauw DJ, Crofford LJ, Glass JM, et al. Fibromyalgia syndrome module at

OMERACT 9: Domain construct. J Rheumatol. 2009;36(10):2318–29.

[80] Melzack R. The short-form McGill pain questionnaire. Pain. 1987;30(2):191–7.

[81] Webster K, Cella D, Yost K. The Functional Assessment of Chronic Illness Therapy (FACIT) Measurement System: Properties, applications, and interpretation. Health Qual Life Outcomes. 2003;1:1–7.

[82] Krupp LB, Larocca NG, Muir Nash J, Steinberg AD. The fatigue severity scale: Application to patients with multiple sclerosis and systemic lupus erythematosus. Arch Neurol. 1989;46(10):1121–3.

[83] Stewart AL, Ware JE. Measuring functioning and well-being: the medical outcomes study approach. : duke university Press; 1992.

[84] Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol 1988;56(6):893.

[85] Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983;67(6):361-370.

[86] Hamilton M. The Hamilton rating scale for depression. Assessment of depression: Springer; 1986. p. 143-152.

Supplementary material 1 **The following tables gather all the data collected from the included studies in the systematic review.**

		Treatment group	Control group		
Measured outcome: Pain		Mean (SD)	Mean (SD)		
FIQ-Pain (VAS)					
Burckhardt					
		n=28	**n=30**	n=28	
	Baseline	7.0	6.5	5.8	
	Post treatment	6.7	5.9	5.6	
	3 month FU	6.0	ND	6.0	
	Long term FU	6.1	5.7	5.4	
Casanueva-Fernandez					
		n=17	**n=17**		
	Baseline	7.71	6.82		
	Post treatment	6.79 (% of improve)	ND		
	1 month FU	4.53 (% of improve)	13.75 (% of improve)		
Castel					
	(NRS, 0-10)	**n=81**	**n=74**		
	Baseline	6.8 (1.4)	7.1 (1.6)		
	Post treatment	5.7 (1.9)	6.9 (1.8)		
	3 week FU	6.4 (1.9)	6.8 (1.8)		
	6 month FU	6.4 (1.9)	7.0 (1.9)		
	12 month FU	6.7 (1.6)	7.1 (1.8)		
Cedraschi					
		n=84	**n=80**		
	Baseline	6.3 (1.9)	6.0 (2.1)		
	6 month FU	6.1 (2.1)	6.6 (2.1)		
Gowans					
		n=20	**n=21**		
	Baseline	7.5	7.5		
	Post treatment	7.7	8.0		
	FU	6.9	ND		
Hammond					
		n=71	**n=62**		
	Baseline	7.01 (2.08)	6.37 (2.07)		
	4 month FU	6.83 (2.01)	6.50 (1.86)		
	8 month FU	6.86 (2.27)	6.24 (2.02)		
King					
		n=26	**n=30**	**n=21**	**n=18**
	Baseline	5.01 (1.93)	5.49 (1.97)	5.04 (1.86)	4.19 (1.88)
	Post treatment	6.11 (2.61)	5.62 (1.82)	5.06 (2.08)	4.88 (1.95)
	FU	6.01 (2.15)	5.99 (1.75)	5.48 (2.04)	4.67 (2.16)
Lemstra					
		n=43	**n=36**		
	Change in VAS	1.02 (0.25)	0.22 (0.20)		
	15 month FU (95%CI)	−0.21 (−0.80–0.38)			
Luciano					
		n=108	**n=108**		
	Baseline	7.37 (1.86)	7.37 (2.10)		
	Post treatment	6.34 (2.35)	7.70 (2.03)		
Rooks					
		n=38	**n=35**	**n=35**	**n=27**
	Baseline	6.6 (2.1)	6.0 (2.1)	5.6 (2.1)	6.0 (2.1)
	Post treatment	4.9 (2.4)	4.8 (2.5)	5.2 (2.0)	5.9 (2.2)

van Eijk-Hustings				
		n=108	**n=47**	**n=48**
	Baseline	6.3 (0.2)	6.2 (0.26)	5.5 (0.3)
	Post treatment	5.5 (0.2)	5.3 (0.31)	5.7 (0.3)
	FU	5.3 (0.2)	5.2 (0.37)	5.3 (0.3)
Wigers				
		n=20	**n=20**	**n=20**
	Baseline	7.2 (1.8)	7.2 (1.9)	6.5 (1.7)
	Post treatment	6.4 (1.9)	6.2 (2.1)	7.2 (2.4)
	4 year FU	7.0 (1.8)	6.8 (2.4)	6.9 (2.4)
Zijlstra				
		n=58	**n=76**	
	Baseline (T_0)	5.9 (1.8)	5.8 (1.7)	
	1 month FU	-1.6 (2.3) (change from T_0)	ND	
	3 month FU	-0.7 (1.9) (change from T_0)	0.0 (1.5)	
	6 month FU	-0.1 (2.3) (change from T_0)	0.1 (1.7)	
	12 month FU	-0.1 (1.7) (change from T_0)	-0.3 (1.9)	
Range (0-10)				

		Treatment group		**Control group**
Measured outcome: Pain		Mean (SD)		Mean (SD)
MPQ				
Casanueva-Fernandez				
		n=17		**n=17**
	Baseline	40.5		37.15
	Post treatment	28.81 (% of improve)		ND
	1 month FU	26.6 (% of improve)		ND
Vlaeyen				
		ECO n=46	**EDI n=39**	**n=40**
	Baseline	0.1 (2.4)	-0.1 (1.8)	0.0 (1.6)
	Post treatment	1.0 (1.8)	0.4(1.8)	0.4 (1.8)
	FU1	1.1 (1.7)	0.1 (1.7)	
	FU2	1.0 (1.9)	0.8 (1.9)	
Zijlstra				
		n=58		76
	Baseline (T_0)	21.1 (8.4)		21.0 (8.8)
	1 month FU	-7.3 (8.3) (change from T_0)		ND
	3 month FU	-4.0 (7.3) (change from T_0)		-2.5 (8.6) (change from T_0)
	6 month FU	-2.8 (8.5) (change from T_0)		-2.0 (8.3) (change from T_0)
	12 month FU	-1.4 (6.9) (change from T_0)		-1.5 (7.9) (change from T_0)
Range (0-45)				

		Treatment group	**Control group**
Measured outcome: Pain		Mean (SD)	Mean (SD)
PDI			
Lemstra			
		n=43	**n=36**
	Change in PDI	8.70 (1.51)	1.97 (1.56)
	15 month FU (95%CI)	-6.51 (-11.33- -1.69)	
Range (0-70)			

		Treatment group		Control group
Measured outcome: Pain		Mean (SD)		Mean (SD)
BPI				
Racine				
		OL n=17	**EC n=24**	**n=43**
	Baseline	5.53 (2.12)	6.75 (1.36)	6.19 (2.05)
	Post treatment	6.06 (1.98)	6.38 (1.66)	6.44 (1.82)
Range (0-10)				

		Treatment group	Control group		
Measured outcome: Fatigue		Mean (SD)	Mean (SD)		
FIQ-Fatigue (VAS)					
Burckhardt					
		n=28	**n=30**	**n=28**	
	Baseline	7.9	6.8	8.1	
	Post treatment	7.2	6.2	7.9	
	3 month FU	7.3	6.9	ND	
	Long term FU	7.1	6.5	7.2	
Casanueva-Fernandez					
		n=17	**n=17**		
	Baseline	8.31	7.19		
	Post treatment	19.07 (% of improve)	ND		
	1 month FU	15.92 (% of improve)	7.59 (% of improve)		
Cedraschi					
		n=84	**n=80**		
	Baseline	7.5 (1.7)	7.4 (2.4)		
	6 month FU	6.5 (2.3)	7.7 (1.9)		
Gowans					
		n=20	**n=21**		
	Baseline	8.1	8.4		
	Post treatment	7.7	8.4		
	3 month FU	7.4	ND		
Hammond					
		n=71	**n=62**		
	Baseline	8.76 (1.54)	7.84 (1.81)		
	4 month FU	8.17 (2.10)	7.96 (1.63)		
	8 month FU	8.18 (2.26)	7.66 (1.73)		
Luciano					
		n=108	**n=108**		
	Baseline	8.18 (1.83)	8.13 (1.89)		
	Post treatment	7.06 (2.41)	7.80 (2.17)		
Rooks					
		n=38	**n=35**	**n=35**	**n=27**
	Baseline	7.2 (2.0)	7.7 (1.6)	7.0 (2.3)	6.9 (1.6)
	Post treatment	6.0 (2.4)	6.6 (2.5)	6.6 (2.2)	7.2 (1.7)
van Eijk-Hustings					
		n=108	**n=47**	**n=48**	
	Baseline	8.3 (0.2)	8.0 (0.2)	7.4 (0.3)	
	Post treatment	7.5 (0.2)	7.4 (0.23)	7.2 (0.3)	
	FU	7.2 (0.3)	7.0 (0.4)	7.5 (0.4)	
Wigers					
		n=20	**n=20**	**n=20**	
	Baseline	8.0 (2.2)	8.0 (2.0)	6.6 (2.7)	
	Post treatment	7.0 (2.1)	5.9 (2.6)	6.3 (3.3)	
	4 year FU	6.8 (2.0)	6.6 (2.7)	6.1 (3.0)	

Zijlstra			
		n=58	**n=76**
	Baseline (T_0)	6.5 (2.0)	6.3 (1.9)
	1 month FU	-1.6 (2.8) (change from T_0)	ND
	3 month FU	-1.0 (2.0) (change from T_0)	-0.1 (1.7) (change from T_0)
	6 month FU	-0.8 (2.5) (change from T_0)	0.1 (1.6) (change from T_0)
	12 month FU	-0.3 (2.1) (change from T_0)	-0.3 (2.0) (change from T_0)
Range (0-10)			

		Treatment group	**Control group**
Measured outcome: Fatigue		Mean (SD)	Mean (SD)
FSS			
Casanueva-Fernandez			
		n=17	**n=17**
	Baseline	6.34	6.24
	Post treatment	5.71 (% of improve)	ND
	1 month FU	13.47 (% of improve)	0.87 (% of improve)
Range (1-7)			

		Treatment group		**Control group**
Measured outcome: Fatigue		Mean (SD)		Mean (SD)
BFI				
Racine				
		OL n=17	**EC n=24**	**n=43**
	Baseline	6.41 (1.94)	6.83 (1.81)	6.74 (1.81)
	Post treatment	5.82 (2.19)	6.88 (1.90)	6.88 (1.94)
Range (0-10)				

		Treatment group	**Control group**
Measured outcome: Quality of sleep		Mean (SD)	Mean (SD)
PSQI			
Casanueva-Fernandez			
		n=17	**n=17**
	Baseline	13.65	12.53
	Post treatment	27.43 (% of improve)	5.52
	1 month FU	21.72 (% of improve)	6.76 (% of improve)
Range (0-21)			

		Treatment group		**Control group**
Measured outcome: Quality of sleep		Mean (SD)		Mean (SD)
MOS-SS				
Castel				
		n=81		**n=74**
	Baseline	29.0 (8.9)		27.9 (8.1)
	Post treatment	41.5 (9.2)		29.6 (8.2)
	3 month FU	40.5 (10.4)		31.2 (9.4)
	6 month FU	38.7 (10.5)		29.0 (8.9)
	12 month FU	36.3 (11.0)		28.8 (8.6)
Racine				
		OL n=17	**EC n=24**	**n=43**
	Baseline	54.80 (17.30)	62.82 (19.65)	66.23 (19.32)
	Post treatment	49.28 (16.84)	56.89 (21.98)	65.47 (17.47)

Measured outcome: Quality of sleep		Treatment group Mean (SD)	Control group Mean (SD)
VAS			
Zijlstra			
		n=58	**n=76**
	Baseline (T_0)	5.5 (2.1)	5.9 (2.2)
	1 month FU	-1.5 (2.8) (change from T_0)	ND
	3 month FU	-0.1 (2.2) (change from T_0)	-0.3 (2.3) (change from T_0)
	6 month FU	0.1 (2.5) (change from T_0)	-0.1 (2.3) (change from T_0)
	12 month FU	0.4 (2.5) (change from T_0)	-0.4 (2.5) (change from T_0)
Range (0-10)			

Measured outcome: Quality of sleep		Treatment group Mean (SD)	Control group Mean (SD)	
FIQ				
van Eijk-Hustings				
		n=108	**n=47**	**n=48**
	Baseline	8.2 (0.2)	8.1 (0.26)	7.6 (0.3)
	Post treatment	7.5 (0.2)	7.0 (0.33)	7.2 (0.3)
	FU	7.1 (0.3)	7.2 (0.4)	7.6 (0.4)
Range (0-10)				

Measured outcome: Physical function		Treatment group Mean (SD)	Control group Mean (SD)	
FIQ-Physical function (VAS)				
Burckhardt				
		n=28	**n=30**	**n=28**
	Baseline	4.4	4.6	4.4
	Post treatment	3.8	4.6	3.9
	3 month FU	3.8	ND	4.1
	Long term FU	3.4	4.5	3.5
Cedraschi				
		n=84	**n=80**	
	Baseline	4.2 (2.0)	4.5 (2.2)	
	6 month FU	4.3 (2.1)	4.8 (2.5)	
Gowans				
		n=20	**n=21**	
	Baseline	5.8	5.7	
	Post treatment	6.1	5.7	
	3 month FU	5.7	ND	
Hammond				
		n=71	**n=62**	
	Baseline	5.26 (2.50)	4.54 (2.16)	
	4 month FU	5.03 (2.52)	4.53 (2.14)	
	8 month FU	5.14 (2.55)	4.52 (2.39)	
Luciano				
		n=108	**n=108**	
	Baseline	3.31 (2.27)	2.80 (2.40)	
	Post treatment	2.44 (2.51)	3.22 (2.79)	

van Eijk-Hustings				
		n=108	**n=47**	**n=48**
	Baseline	4.2 (0.2)	3.6 (0.2)	3.4 (0.3)
	Post treatment	3.9 (0.2)	3.7 (0.3)	4.0 (0.3)
	FU	3.6 (0.2)	3.6 (0.4)	3.9 (0.3)
Zijlstra				
		n=58	**n=76**	
	Baseline (T_0)	4.5 (1.7)	4.5 (1.6)	
	1 month FU	-0.9 (1.6) (change from T_0)	ND	
	3 month FU	-0.6 (1.4) (change from T_0)	-0.0 (1.4) (change from T_0)	
	6 month FU	-0.3 (1.6) (change from T_0)	-0.1 (1.5) (change from T_0)	
	12 month FU	-0.1 (1.8) (change from T_0)	-0.2 (1.5) (change from T_0)	
Range (0-10)				

		Treatment group	**Control group**		
Measured outcome: Physical function		Mean (SD)	Mean (SD)		
6MW					
Burckhardt					
		n=28	**n=30**	**n=28**	
	Baseline	488.6	479.9	494.5	
	Post treatment	493.5	466.8	499.2	
Casanueva-Fernandez					
		n=17	**n=17**		
	Baseline	177.14	220.2		
	Post treatment	6.13 (% of improve)	ND		
	1 month FU	10.47 (% of improve)	ND		
Gowans					
		n=20	**n=21**		
	Baseline	330.7	350.6		
	Post treatment	402.7	372.6		
	3 month FU	389.2	ND		
King					
		n=26	**n=30**	**n=21**	**n=26**
	Baseline	452.0 (73.5)	495.4 (74.3)	495.4 (74.3)	494.6 (93.6)
	Post treatment	501.1 (81.9)	494.3 (96.2)	494.3 (96.2)	498.7 (125.6)
	FU	465.2 (107.4)	520.9 (80.9)	476.6 (109.9)	479.4 (112.3)
Rooks					
		n=38	**n=35**	**n=35**	**n=27**
	Baseline	457 (76)	488 (79)	462 (80)	467 (86)
	Post treatment	485 (73)	515 (68)	496 (74)	442 (123)
Score: meters/6 minutes					

		Treatment group	**Control group**
Measured outcome: Physical function		Mean (SD)	Mean (SD)
SF-36			
Casanueva-Fernandez			
		n=17	**n=17**
	Baseline	44.27	50.17
	Post treatment	11.4 (% of improve)	ND
	1 month FU	14.31 (% of improve)	0.98 (% of improve)
Cedraschi			
		n=84	**n=80**
	Baseline	41.8 (18.1)	46.8 (19.4)
	6 month FU	42.2 (19.8)	43.9 (19.6)

Lera					
		n=35	**n=31**		
	Baseline	38.6 (22.1)	32.3 (17.6)		
	Post treatment	39.5 (20.4)	30.7 (14.4)		
Rooks					
		n=38	**n=35**	**n=35**	**n=27**
	Baseline	42.5 (18.9)	42.9 (19.1)	47.8 (23.0)	46.3(23.9)
	Post treatment	59.1 (19.1)	58.9 (20.3)	56.8 (19.6)	49.3 (23.9)
Range (0-100)					

		Treatment group		**Control group**
Measured outcome: Physical function		Mean (SD)		Mean (SD)
BAT		**ECO n=46**	**EDI n=39**	**n=40**
Vlaeyen				
	Baseline	0.2 (2.6)	-0.1 (2.4)	-0.1 (2.9)
	Post treatment	0.3 (1.7)	-0.5 (1.7)	0.2 (1.7)
	FU1	0.1 (2.3)	0.2 (2.3)	
	FU2	1.0 (1.9)	0.8 (1.9)	
Range				

		Treatment group		**Control group**	
Measured outcome: Physical function		Mean (SD)		Mean (SD)	
Walking test		**PA n=29**	**PP n=39**	**PA WLC n=45**	**PP WLC n=45**
Van Koulil					
Walked meters					
	Baseline	259.6 (156.7) (n=28)	305.7 (122.8) (n=37)	245.5 (133.4) (n=31)	339.2 (133.7) (n=24)
	FU	438.7 (128.9) (n=23)	496.9 (149.9) (n=36)	250.3 (136.9) (n=30)	381.7 (150.8) (n=24)
Perceived exertion					
	Baseline	3.8 (1.5) (n=28)	3.9 (1.8) (n=37)	4.8 (1.5) (n=30)	4.6 (1.9) (n=24)
	FU	2.9 (1.3) (n=23)	2.8 (1.4) (n=36)	5.3 (1.9) (n=29)	2.8 (1.4) (n=24)
Range (0-10)					

		Treatment group		**Control group**	
Measured outcome: Physical function		Mean (SD)		Mean (SD)	
Cycling test		**PA n=29**	**PP n=39**	**PA WLC n=45**	**PP WLC n=45**
Van Koulil					
Walked meters					
	Baseline	7.8 (4.6) (n=28)	8.9 (3.8) (n=37)	6.9 (3.7) (n=30)	12.3 (5.0) (n=24)
	FU	12.5 (4.5) (n=23)	12.4 (3.8) (n=36)	7.1 (3.9) (n=30)	12.0 (4.8) (n=24)
Perceived exertion					
	Baseline	4.9 (2.1) (n=27)	4.8 (1.5) (n=36)	5.2 (2.0) (n=30)	5.0 (1.6) (n=24)
	FU	3.8 (1.5) (n=23)	4.0 (1.7) (n=36)	5.6 (1.7) (n=28)	4.9 (1.5) (n=22)
Range (0-10)					

Measured outcome: Physical function		Treatment group Mean (SD)		Control group Mean (SD)
PCS				
Racine				
		OL n=17	**EC n=24**	**n=43**
	Baseline	31.58 (9.22)	29.09 (5.73)	29.96 (4.88)
	Postreatment	32.86 (8.32)	30.67 (4.80)	30.26 (6.45)
Zijlstra				
		n=58		**n=76**
	Baseline (T_0)	28.6 (8.0)		27.8 (7.4)
	1 month FU	6.3 (8.2) (change from T_0)		ND
	3 month FU	3.6 (8.8) (change from T_0)		0.8 (6.7) (change from T_0)
	6 month FU	1.3 (9.6) (change from T_0)		0.5 (5.8) (change from T_0)
	12 month FU	2.6 (7.4) (change from T_0)		1.6 (7.8) (change from T_0)
Range (10-50)				

Measured outcome: Depression		Treatment group Mean (SD)		Control group Mean (SD)	
FIQ-Depression (VAS)					
Burckhardt					
		n=28		**n=30**	**n=28**
	Baseline	4.3		4.2	3.4
	Post treatment	3.8		3.8	3.0
	3 month FU	3.2		ND	3.6
	Long term FU	2.5		4.8	4.1
Casanueva-Fernandez					
		n=17		**n=17**	
	Baseline	5.41		5.06	
	Post treatment	9.49 (% of improve)		ND	
	1 month FU	7.05 (% of improve)		9.46 (% of improve)	
Cedraschi					
		n=84		**n=21**	
	Baseline	5.5 (3.1)		5.9 (3.5)	
	6 month FU	4.6 (3.1)		6.1 (3.4)	
Gowans					
		n=20		**n=21**	
	Baseline	6.6		6.2	
	Post treatment	5.9		7.1	
	FU	6.3		ND	
Hammond					
		n=71		**n=62**	
	Baseline	4.98 (3.02)		5.05 (2.98)	
	4 month FU	4.83 (3.12)		5.11 (3.26)	
	8 month FU	5.44 (3.38)		5.08 (3.20)	
Luciano					
		n=108		**n=108**	
	Baseline	7.42 (3.02)		6.82 (3.11)	
	Post treatment	5.24 (3.54)		6.45 (3.09)	
Rooks					
		n=38		**n=35**	**n=35** **n=27**
	Baseline	5.1 (2.7)		4.9 (2.9)	4.2 (2.9) 5.0 (2.4)
	Post treatment	3.3 (2.6)		4.3 (3.0)	3.0 (2.5) 4.2 (2.8)

van Eijk-Hustings				
		n=108	**n=47**	**n=48**
	Baseline	5.2 (0.3)	4.8 (0.3)	4.2 (0.4)
	Post treatment	4.1 (0.3)	4.6 (0.4)	4.5 (0.4)
	FU	3.9 (0.3)	5.0 (0.5)	4.2 (0.4)
Wigers				
		n=20	**n=20**	**n=20**
	Baseline	4.4 (3.2)	3.4 (2.9)	4.0 (3.7)
	Post treatment	2.4 (2.2)	3.1 (3.2)	3.6 (3.5)
	4 year FU	4.0 (2.8)	3.2 (3.4)	3.0 (3.1)
Zijlstra				
		n=58	**n=76**	
	Baseline (T_0)	2.6 (2.4)	2.8 (2.3)	
	1 month FU	-0.7 (2.1) (change from T_0)	ND	
	3 month FU	-0.2 (2.4) (change from T_0)	-0.1 (2.1) (change from T_0)	
	6 month FU	0.1 (2.4) (change from T_0)	-0.1 (2.2) (change from T_0)	
	12 month FU	0.6 (2.8) (change from T_0)	-0.3 (2.4) (change from T_0)	
Range (0-10)				

		Treatment group		**Control group**	
Measured outcome: Depression		Mean (SD)		Mean (SD)	
BDI					
Casanueva-Fernandez					
		n=17		**n=17**	
	Baseline	24.59		19.35	
	Post treatment	7.09 (% of improve)		ND	
	1 month FU	11.48 (% of improve)		8.07 (% of improve)	
Lemstra					
		n=43		**n=36**	
	Change in BDI	7.74 (1.17)		0.97 (0.75)	
	15 month FU(95% CI)	2.77 (-0.85-6.39)			
Rooks					
		n=38		**n=35**	**n=35** **n=27**
	Baseline	18.0 (10)		17.0 (10)	13.0 (9) 14.0 (10)
	Post treatment	11.0 (9)		13.0 (10)	9.0 (8) 14.0 (12)
Vlaeyen					
		ECO n=46	**EDI n=39**	**n=40**	
	Baseline	12.8 (6.6)	12.5 (9.4)	15.5 (8.3)	
	Post treatment	13.4 (5.8)	11.9 (5.8)	13.2 (5.8)	
	6 month FU	12.8 (6.8)	12.9 (6.8)	ND	
	12 month FU	14.5 (10.6)	13.0 (10.6)	ND	
Zijlstra					
		n=58		**n=76**	
	Baseline T_0	13.2 (6.8)		13.0 (6.7)	
	1 month FU	-2.9 (3.8) (change from T_0)		ND	
	3 month FU	-1.7 (5.5) (change from T_0)		-1.2 (5.3)	
	6 month FU	-2.0 (4.5) (change from T_0)		-0.8 (5.4)	
	12 month FU	-0.3 (5.5) (change from T_0)		-0.8 (4.7)	
Range (0-63)					

		Treatment group	Control group
Measured outcome: Depression		Mean (SD)	Mean (SD)
ZDS			
Casanueva-Fernandez			
		n=17	**n=17**
	Baseline	68.75	64.15
	Post treatment	3.79 (% of improve)	ND
	1 month FU	5.7 (% of improve)	2.26(% of improve)
Range (25-100)			

		Treatment group		Control group
Measured outcome: Depression		Mean (SD)		Mean (SD)
HADS				
Racine				
		OL n=17	**EC n=24**	**n=43**
	Baseline	7.47 (3.22)	7.75 (3.84)	10.14 (3.79)
	Post treatment	5.76 (4.10)	7.92 (3.30)	10.84 (3.92)
Range (0-21)				

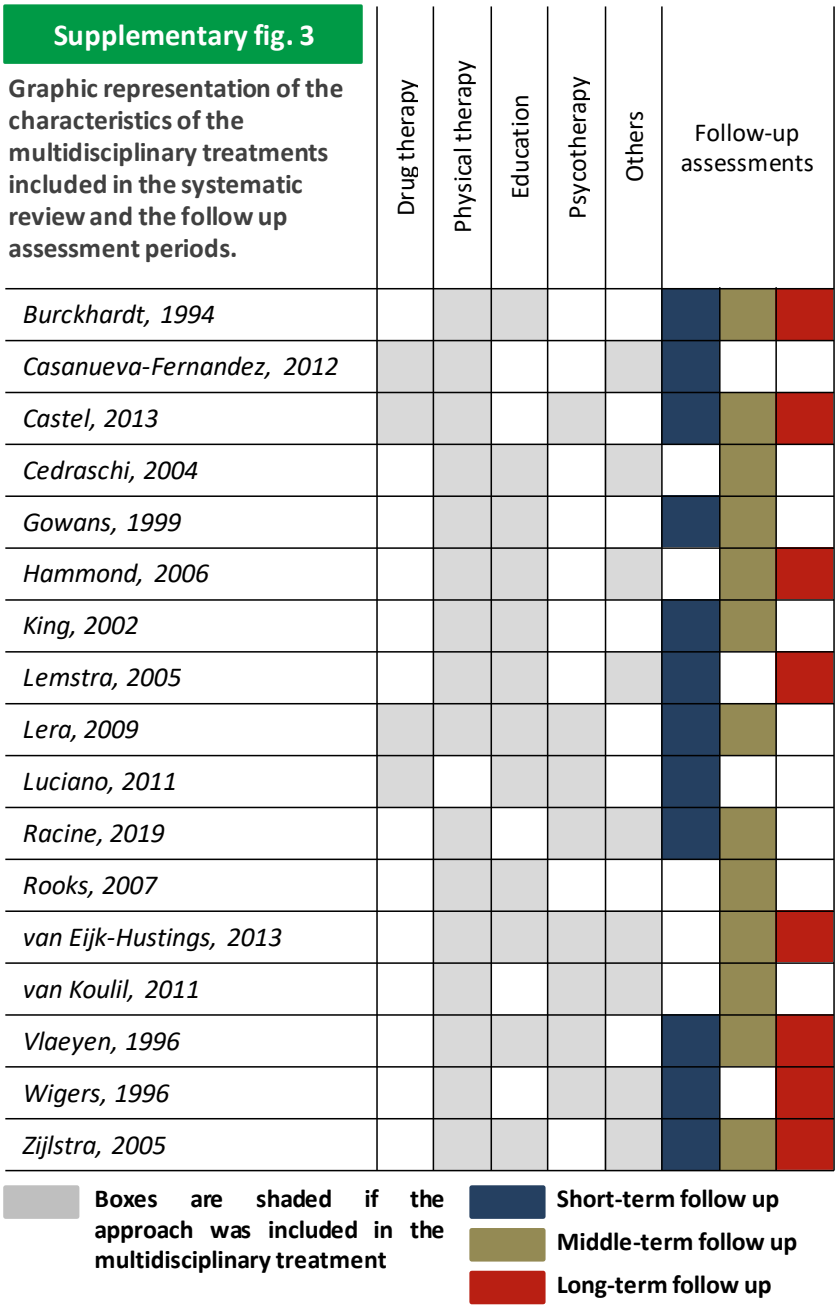
		Treatment group	Control group
Measured outcome: Depression		Mean (SD)	Mean (SD)
PGWB			
Cedraschi			
		n=84	**n=80**
	Baseline	8.3 (3.4)	7.6 (4.0)
	6 month FU	9.0 (3.6)	7.7 (4.2)
Range (0-110)			

Supplementary fig. 2 **Meaning of the abbreviations of the questionnaires gathered in the table of the details of the studies included in the meta-analysis.**

Abbreviation	Questionnaire	Abbreviation	Questionnaire
6MW	6-Minute-Walk	**MOSS-Scale**	Medical Outcome Study Sleep Scale
ASES	Arthritis Self-Efficacy Scale	**MPLC**	Multidimensional Pain Locus of Control Scale
BAI	Beck Anxiety Inventory	**MPQ**	McGill Pain Questionnaire
BAT	Behavioral Approach Test	**NRS**	Numerical Rating Scale
BDI	Beck Depression Inventory	**PCL**	Pain Cognition List
BFI	Brief Fatigue Inventory	**PCS**	Physical Component Summary
BPI	Brief Pain Inventory	**PDI**	Pain Disability Index
CHIP	Checklis for Interpersonal Pain Behavior	**PGIC**	Patients' Global Impression of Change
CIS	Checklist Individual Strenght	**PGWB**	Psychological General Well-Being Index
CSQ	Coping Strategies Questionnaire	**POAM-P**	Patterns of Activity Measure-Pain
EQ-5D	Five-dimensional EuroQol	**PSQI**	Pittsburgh Sleep Quality Index
FAI	Fibromyalgia Attitudes Index	**QOLS-S**	Quality of Life Scale
FHAQ	Fibromyalgia Health Assessment Questionnaire	**RAI**	Rheumatology Attitudes Index
FIQ	Fibromyalgia Impact Questionnaire	**RPS**	Regional Pain Score
FSS	Fatigue Severity Scale	**SCL-90-R**	The Symptom Checklist-90-Revised
FSS-III-R	Fear Survey Schedule	**SPAQ**	Scottish Physical Activity Questionnaire
GTPS	Graded Tender Point Score	**STAI**	The State Trait Anxiety Inventory
HADS	The Hospital Anxiety and Depression Scale	**SWT**	Shuttle Walking Test
HAS	Hamilton Anxiety Scale	**TP Index**	Tender Point Index
MCS	Mental Component Summary	**UAB**	Pain Behavior Scale
MCSDS	Marlow-Crowne Social Desirability Scale	**VAS**	Visual Analog Scale
MOCI	Maudsley Obssessive Compulsive Inventory	**VRS**	Verbal Rating Scale
MOS (SF-36)	Medical Outcomes Study, Short Form	**ZDS**	Zung Self-Rating Depression Scale

Supplementary fig. 3

Graphic representation of the characteristics of the multidisciplinary treatments included in the systematic review and the follow up assessment periods.



Supplementary material 4 **Risk of bias assessment. Detailed data about reasons for considering high,unclear or low risk. High risk is represented in red, unclear in yellow and low in green.**

Burckhardt, 1994		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		Randomized Controlled Clinical Trial. 99 patients were randomly assigned to 1 of the 3 groups. Randomization not specified.
Incomplete outcome data (attrition bias)		13 subjects were dropped from the posttest analysis either because they did not return for retesting or because they attended only 1-2 classes if they were in the experimental groups. 5 patients from the education group and 1 from the MT treatment did not complete or dropped out of the study. Reasons not specified.
Other bias		None declared.

Casanueva-Fernandez, 2012		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		Double-blinded randomized and controlled study (RCT). Unspecified randomization.
Incomplete outcome data (attrition bias)		5 patients from the education group and 1 from the MT group did not complete de study. Reasons not specified. There is no report of the trial flow.
Other bias		Small number of patients included. 16 registered outcomes for a total of 34 participants without control for multiple comparisons.

Castel, 2013		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		Patients were randomly assigned in a 1:1 ratio in blocks of 32 according to a computer-generated random number table.
Incomplete outcome data (attrition bias)		The sample size consisted of 155 participants. 142 completed treatment. Reasons not specified. Low dropout rate (13/155 [8.4%]).
Other bias		Effect of contact was not taken into account, so results could have been due to different degrees of time and attention that were dedicated to each of the groups. Study sample was limited to low educational women, so generalizing results to other population groups is partially limited.

Cedraschi, 2004		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		RCT. After baseline medical evaluation, participants were randomly allocated to a treatment group or a control group. The assignment was performed in blocks of 20, split into treatment program (n=10) or control (n=10). Randomization was made by means of an electronic numbers generator (SPSS). An independent person who was not responsible for determining the participant's eligibility provided sequentially numbered, sealed, and opaque envelopes.
Incomplete outcome data (attrition bias)		164 participants were randomly allocated to the treatment (n=84) or the control group (n=80). 61 patients (73%) in the treatment group and 68 (85%) in the control group completed the 6 month follow up (included in the final analyses). 2 participants in the treatment group explicitly cited an increase in pain as the reason for dropping out. Differential dropout rates (23/84 [27.3%] experimental, 12/80 [15%] control).
Other bias		None declared.

Gowans, 1999		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		In phase I (randomized, controlled trial), subjects were randomly assigned to a 6-week exercise and educational program (intervention subjects) or served as waiting list controls for 6 weeks. Randomization not specified.
Incomplete outcome data (attrition bias)		**Phase I:** 45 subjects (23 intervention subjects, 22 waiting list control subjects) were randomized and enrolled in cohorts of 14 to 18 individuals in the spring (2 cohorts) or fall (1 cohort). Three intervention subjects were subsequently excluded for attending less than 50% of the program (attendance: 0%, 0%, 8%, respectively).One control subject did not return for testing at 6 weeks. Therefore, phase I analyses were based on 41 subjects (20 intervention subjects, 21 waiting list control subjects).

		Phase II: 15 of the 21 original waiting list control subjects satisfied study inclusion criteria (attended greater than 50% of the 6-week program and returned for testing immediately post-program). Six of the original 21 subjects were excluded: 2 subjects attended the educational sessions but refused hydrotherapy, 3 subjects attended less than 50% of the program (attendance: 45%, 36%, 0%, respectively), and 1 subject did not return for testing immediately post-program. Thus, data analyses for phase II are based on 35 subjects (20 intervention subjects and 15 previous waiting list control subjects). **Phase III (followup):** 30 of the 35 subjects (19 intervention subjects + 11 previous control subjects) who completed phase II returned for testing 3 (n 5 23) or 6 (n 5 7) months post-program. Subjects who completed phase III did not differ on demographic characteristics from the 11 subjects who completed only phase I.
Other bias		Multiple outcomes assessed without control for multiple comparisons.

Hammond, 2006		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		Participants were randomized to the patient education-exercise group or the relaxation group using computer-generated random numbers in pre-prepared sealed numbered envelopes. No stratification was used. They were telephoned to arrange program attendance, with up to three opportunities given to attend.
Incomplete outcome data (attrition bias)		From the relaxation group (n=86) 24 did not receive the treatment: withdrew (n=9), postponed due to illness (n=2), unable to attend (n=13). From the educational-exercise program (n=97), 26 did not receive the treatment: withdrew (n=11), postponed due to illness (n=2), unable to attend (n=13). At the final assessment 79% (n=49) of patients in the relaxation group and 73% (n=52) of patients in the education-exercise program were assessed. Although there was initially a surprisingly high level of consent, about a quarter dropped out when asked to attend and a quarter stopped attending early in treatment. There was a higher dropout rate in the relaxation (attention control) group, although almost half considered it beneficial. High dropout rates but for similar reasons (26/97 [26.8%] versus 24/86 [27.9%]).
Other bias		None declared.

King, 2002		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		Randomized controlled trial with repeated measures design. Random assignment of subjects to groups was done in blocks of 4 to 16 subjects. A list was prepared prior to start of study using a table of random numbers and subject ID number (order of admission to study). The investigator with the list who assigned subjects to groups was unaware of their baseline test results. Both assessors were blinded to the subject's group randomization on subsequent visits.
Incomplete outcome data (attrition bias)		196 women attended the pre-test session and were randomized into one of 4 groups. After randomization and before the first session, 26 subjects decided not to participate. The number of non-participants for each group was: exercise-only 3; education-only 11; exercise and education 5; and control 7. Reasons for not participating included lack of time (n = 6), sessions conflicted with previous commitments (n = 5), distance to travel (n = 2), moved out of country (n = 1), and unknown (n = 12). The reasons that subjects dropped out after attending at least one session were: lack of time (n = 11), sessions conflicted with previous commitments (n = 1), family health/personal problems (n = 8), felt program would not help (n = 5), or they could not be reached or refused to return for testing (n = 9). Study presents reasons for dropout and ITT analysis.
Other bias		Group contamination.

Lemstra, 2005		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		79 men and women were randomly assigned to 1 of 2 groups each with 6-week duration. The unit of randomization was individual, computer-generated, and envelope concealed. This process, as well as intervention allocation, was under the supervision of a data manager.
Incomplete outcome data		43 subjects began the intervention group, and 36 subjects began the control group. 7

(attrition bias)		subjects quit the intervention protocol prior to completion (16.3%). 6 subjects quit within the first 2 weeks citing lack of time as a reason. One subject quit the intervention due to complications with a medical condition not related to the study. In total,36 subjects completed the 6-week intervention, and all 36 subjects completed the control. 35 intervention subjects out of 36 completed the 15-month follow-up on the intervention group (1 moved to unknown new location).
Other bias		Hawthorne effect (different response to treatment in intervention group as result of special attention and interest that they received). Study subjects were volunteers. Long term exercise adherence was based on self reporting.

Lera, 2009		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		RCT. Flip of a coin for randomization.
Incomplete outcome data (attrition bias)		Of the initial 83 participants, 66 completed their program and post treatment evaluations (MT n=35 and MT+CBT n=31). The dropout rate was similar for both groups. Reasons not specified.
Other bias		Limited sample size.

Luciano, 2011		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		RCT. Patients were randomly assigned to the intervention group or to the control group using a computer-generated randomization list drawn up by one of the investigators.
Incomplete outcome data (attrition bias)		The dropout rate was 6.5% (n=7) in the intervention group and 13% (n=14) in the control group. The reasons for dropout were: not interested in the study (n=16), family burden (n=2), not able to comply with the treatment schedule (n=2) and relocation (n=1).
Other bias		Results in the follow-up assessments have not been reported, so permanent improvement cannot be determined. Psychiatric disorders were not assessed, so the distribution of these disorders may be different in both groups, influencing the results.

Racine, 2019		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		The randomization plan included 12 participants per block and was generated using online software (http://www.randomization.com, using the seed 3034). A total of 15 blocks were generated. The first 144 participants (first 12 blocks) were assigned to each of the four treatment conditions (i.e., 3 participants per 4 treatment conditions). However, in order to maximize the number of study participants as the study came to an end, we began enrolling participants into only the immediate treatment conditions starting with the 13th block, as there was not enough time left at this point to enroll participants into either of the two delayed conditions and have them complete participation. Thus, the last 34 participants (two 12-person blocks and one 10person block) were randomized to either the OL or EC immediate treatment conditions, resulting in more participants in the immediate treatment groups than the delayed treatment groups.
Incomplete outcome data (attrition bias)		A total of 517 patients with a diagnosis of FMS were referred from a variety of sources to the study. Of these, 66% were excluded because they (1) declined participation (49%), (2) could not be contacted (8%) or (3) did not meet eligibility criteria (8%), leaving us with a final total potential sample of 178 patients (34% of those referred). Of these patients, 34% (27 patients in OL group and 33 in EC group) did not receive the allocated treatment (i.e., they completed the questionnaires, but dropped out before treatment started), another 25% (28 patients in OL group and 16 in EC group) dropped out of the study after they had begun treatment, while an additional 3% were ineligible after treatment allocation (3 patients from the OL group and 2 patients from the EC group). Huge number of dropouts (52/90 [57.8%] experimental, 61/88 [69.3%] control).
Other bias		Many. Participants were allowed to continue receiving any other form of concomitant pain treatments, which could have introduced additional biases.

Rooks, 2007		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		Participants were randomly assigned to 1 of 4 groups: aerobic and flexibility exercise (AE); strength training, aerobic, and flexibility exercise (ST); the Arthritis Foundation's Fibromyalgia Self-Help Course (FSHC); or a combination of ST and FSHC (ST-FSHC). Members of the hospital's Biometrics Center not involved in the study used a computer program that generated single-page listings of random group assignment. Individual pages were placed in opaque envelopes, sealed, numbered sequentially, and stored in a locked cabinet. We stratified randomization by level off functional status indicated by a score of less than 40 or 40 or higher on the Fibromyalgia Impact Questionnaire (FIQ) to reduce the chance of an imbalance in this primary outcome variable.
Incomplete outcome data (attrition bias)		Of the 207 participants randomized, 135 (65%) completed the 16-week intervention period and underwent a follow-up assessment. There was 31% attrition in each of the 3 exercise groups and 46% in the FSHC group. Reasons for dropping out centered on health problems other than fibromyalgia and schedule conflicts with work or family. 1/3 of the overall sample dropped out.
Other bias		No comparison with no treatment (to minimize the bias of interpersonal contact).

Van Eijk-Hustings, 2013		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		Designed as a pragmatic RCT. The randomization was performed using computer-generated random numbers in opaque, sealed envelopes, following the order of consent to participate in the observational study. Only those who were randomized to MD or AE were invited to participate in the intervention without being informed about alternative treatment conditions. Patients in the UC (usual care) group were not informed about any intervention.
Incomplete outcome data (attrition bias)		Reasons for not consenting to participate in interventions: children, work/study, distance no interest, other or unknown (n=41 in MD and n=28 in AE). 100% participation in UC. Pragmatic RCT with huge discrepancies between number randomized and number started interventions because of Zelen-like design.
Other bias		Despite de randomization, the MD group turned out to be the worst condition at inflow and thus had the largest potential for improvement.

Van Koulil, 2011		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		The total sample of 158 patients was randomly assigned in clusters to the treatment condition (TC), patients who received the treatment, or the waiting list control condition (WLC) separate for the PA and PP group. These randomization clusters consisted of 8 patients because of the group size for the treatment, but the size of the clusters varied from 1 to 9 due to reasons such as the inclusion of WLC patients who were offered treatment at the end of their follow-up according to protocol and treatment that was scheduled on fixed months. As a result of randomization, 39 patients (5 clusters) were allocated to pain-persistence TC, 45 patients (6 clusters) to the pain-persistence WLC, 29 patients (6 clusters) to pain-avoidance TC, and 45 patients (6 clusters) to pain-avoidance WLC.
Incomplete outcome data (attrition bias)		Flow chart showing participant selection and loss to posttest (T2) and followup (T3) for the 2 study conditions. TC = treatment condition; WLC = waiting list control condition; T1 = pretreatment: https://onlinelibrary.wiley.com/doi/full/10.1002/acr.20268
Other bias		None declared.

Vlaeyen, 1996		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		RCT. Before starting the pretreatment assessments, patients were randomly assigned to an educational cognitive condition (ECO), an educational discussion condition (EDI) or to a waiting list condition (WLC). Randomization not specified.
Incomplete outcome data (attrition bias)		Of 131 patients, 49 were randomly assigned to the ECO, 39 to the EDI and 43 to the WLC. 6 patients dropped out just after randomization (3 from ECO, 3 from WLC). Of 125 patients who started treatment, results were available from 112 (90%) at post-treatment, 67 (79%) at 6 month follow-up, and 66 (78%) at 12 month follow-up. The attrition rates were 22, 23, and 2.5% for the ECO, EDI, and WLC groups, respectively The small number

		of dropouts from the WLC condition reflects the much shorter follow-up period (8 weeks) and the patients' anticipation of treatment after the 8 week waiting period. Reasons for dropout not specified.
Other bias		A limitation of this study is that only pain related and affective outcome variables have been chosen. No conclusions can be made about improvements in disease process measures such as measures of fatigue, number of fibromyalgia tender points, and sleep ratings. Multiple outcomes were assessed without control for multiple comparisons

Wigers, 1996		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		60 patients were randomized to 14 weeks of treatment by either AE, SMT or TAU. Randomized by drawing lots.
Incomplete outcome data (attrition bias)		48 patients completed the program. Dropout reasons for the 16 patients that either withdrew or failed to complete the trial according to the protocol were: health problems, transport problems, moved to another town, initiated additional treatments, not wish to participate and a dead. Dropouts: AE (n=5), SMT (n=7), TAU (n=4).
Other bias		None declared.

Zijlstra, 2005		
Item	**Risk**	**Description**
Random sequence generation (selection bias)		Pre-randomized controlled trial. Patients eligible for inclusion were randomly allocated to the treatment or control group using a computer-generated randomization list and closed numbered envelopes.
Incomplete outcome data (attrition bias)		A total of 170 patients initially fulfilled inclusion criteria; 84 of them were randomized into the SPA and 86 into the CTL group. After randomization and information, more patients refused participation in the SPA group than in the CTL group. Reasons for refusal in the SPA group were: three job related, six family related, two financial, 12 unknown. In the CTL group these were: one marked improvement of FM symptoms, one disappointment with the study protocol, six unknown. Three SPA and two CTL subjects withdrew due to co-morbidity occurring in the period between inclusion and study start. Finally, 58 patients received spa treatment and 76 patients entered the control protocol. Zelen design impedes its assessment.
Other bias		Despite Zelen design, treatments are so absolutely dissimilar that it is unlikely have not influenced outcome beyond randomization.

Supplementary fig. 5 **Baseline minumum, maximun and mean (SD) scores from the questionnaires used for the post-treatment assessment in the meta-analysis.**

		N	Minimum	Maximum	Mean	SD
Pain	**Exp.**	10	5.01	7.5	6.7	0.72
	Control	10	5.49	7.5	6.57	0.68
Fatigue	**Exp.**	8	7.2	8,76	7.99	0.48
	Control	8	6.8	8.4	7.78	0.49
Physical function	**Exp.**	6	3.31	5.8	4.52	0.88
	Control	6	2.8	5.7	4.29	0.99
Anxiety	**Exp.**	7	5.1	7.94	6.36	0.95
	Control	7	4.9	7.45	6.19	1.02
Depression	**Exp.**	6	3.22	10.0	6.07	3.06
	Control	6	3.79	10.0	6.36	2.81
Quality of sleep	**Exp.**	4	2.08	19.65	11.98	8.05
	Control	4	1.78	19.32	12.13	8.69

Supplementary fig. 6 Summary of findinds (GRADE) for pain.

Multidisciplinary treatment compared to usual care for FM - PAIN

Patient or population: Adults with ACR FM diagnosis
Setting: Adults
Intervention: Multidisciplinary treatment
Comparison: Usual care

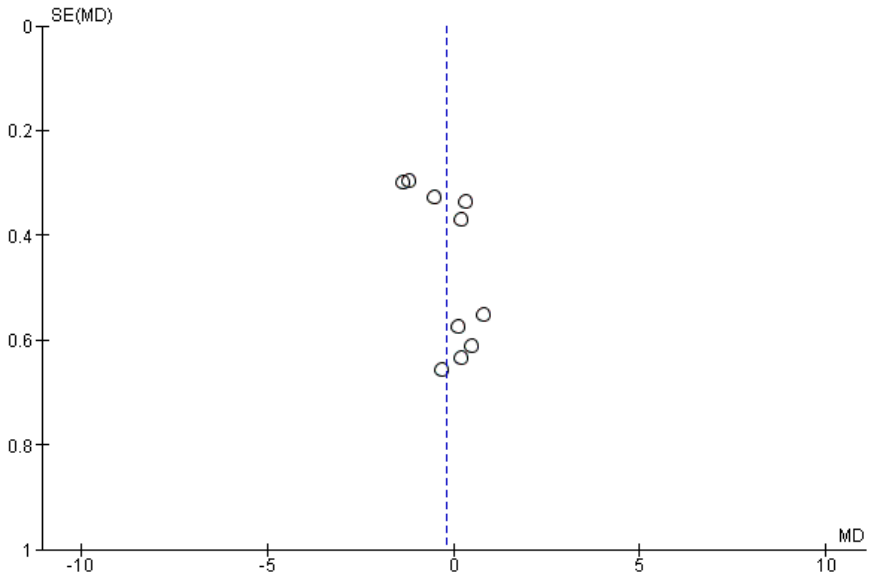
Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with Usual care	Risk with Multidisciplinary treatment				
Pain assessed with: FIQ-Pain Scale from: 0 to 10 follow up: range 3 months to 4 years	The mean pain was **6.35** Scale unit	MD **0.2 Scale unit lower** (0.71 lower to 0.31 higher)	-	1091 (10 RCTs)	⨁◯◯◯ VERY LOW a,b,c	The evidence is very uncertain about the effect of multidisciplinary treatments on pain.

***The risk in the intervention group** (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).

Explanations

a. 6 out of the 10 studies included in the statistical analysis in FIQ-Pain showed high risk in at least one of the analyzed bias. 3 showed high risk in attrition bias (differential dropout rates, high number of dropouts of the overall sample, discrepancies between number of randomized and number of started intervention). The remaining 3, showed high risk in other bias (small number of patients, multiple outcomes assessed without control for multiple comparisons, group contamination).
b. I^2=73%
c. Symmetric funnel plot

Supplementary fig. 7 Funnel plot for pain. Simetric funnel plot indicating undetected risk of publication bias.



Supplementary fig. 8 Summary of findinds (GRADE) for fatigue.

Multidisciplinary treatment compared to usual care for FM - FATIGUE

Patient or population: Adults with ACR FM diagnosis
Setting: Adults
Intervention: Multidisciplinary treatment
Comparison: Usual care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with Usual care	**Risk with Multidisciplinary treatment**				
Fatigue assessed with: FIQ-Fatigue Scale from: 0 to 10 follow up: range 3 months to 4 years	The mean fatigue was **7.24** Scale Unit	MD **0.18 Scale Unit lower** (0.72 lower to 0.36 higher)	-	878 (8 RCTs)	⨁◯◯◯ VERY LOW a,b	Multidisciplinary treatments may have little to no effect on fatigue but the evidence is very uncertain.

***The risk in the intervention group** (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; **MD:** Mean difference

Explanations

a. 4 out of the 8 studies included in the statistical analysis in FIQ-Fatigue showed high risk in at least one of the analyzed bias. 3 showed high risk in attrition bias (differential dropout rates, high number of dropouts of the overall sample, discrepancies between number of randomized and number of started intervention). The remaining one showed high risk in other bias (multiple outcomes assessed without control for multiple comparisons).
b. I^2=70%

Supplementary fig. 9 Summary of findinds (GRADE) for quality of sleep.

Multidisciplinary treatment compared to usual care for FM - QUALITY OF SLEEP

Patient or population: Adults with ACR FM diagnosis
Setting: Adults
Intervention: Multidisciplinary treatment
Comparison: Usual care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with Usual care	Risk with Multidisciplinary treatment				
Quality of sleep assessed with: MOSS-S, FIQ follow up: range 3 months to 12 months	-	SMD **0.37 higher** (0.16 higher to 0.57 higher)	-	437 (4 RCTs)	⨁◯◯◯ VERY LOW a,b	The evidence is very uncertain about the effect of multidisciplinary treatments on quality of sleep.

***The risk in the intervention group** (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; **SMD:** Standardised mean difference

Explanations

a. All the studies included in the statistical analysis in the assessment of quality of sleep showed high risk in at least one of the analyzed bias. 2 showed high risk in attrition bias (huge number of dropouts of the overall sample, discrepancies between number of randomized and number of started intervention). 2 studies showed high risk in other bias (small number of patients, participants were allowed to receive concomitant treatments).
b. I^2=95%

Supplementary fig. 10 Summary of findinds (GRADE) for physical function.

Multidisciplinary treatment compared to usual care for FM - PHYSICAL FUNCTION

Patient or population: Adults with ACR FM diagnosis
Setting: Adults
Intervention: Multidisciplinary treatment
Comparison: Usual care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with Usual care	**Risk with Multidisciplinary treatment**				
Physical function assessed with: FIQ-Physical Scale from: 0 to 10 follow up: range 3 months to 8 months	The mean physical function was **4.42** Scale unit	MD **0.18 Scale unit lower** (0.66 lower to 0.29 higher)	-	767 (6 RCTs)	⨁⨁◯◯ LOW [a]	Multidisciplinary treatments may result in little to no difference in physical function.

***The risk in the intervention group** (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; **MD:** Mean difference

Explanations

a. 3 out of the 6 studies included in the statistical analysis in FIQ-Physical showed high risk in at least one of the analyzed bias. 2 showed high risk in attrition bias (differential dropout rates, discrepancies between number of randomized and number of started intervention). The remaining one showed high risk in other bias (multiple outcomes assessed without control for multiple comparisons). Similar weight of studies presenting at least a high risk bias and studies not showing high risk in any bias. None of the studies showed risk in selection bias.

Supplementary fig. 11 Summary of findinds (GRADE) for depressive symtomps.

Multidisciplinary treatment compared to usual care for FM - DEPRESSION

Patient or population: Adults with ACR FM diagnosis
Setting: Adults
Intervention: Multidisciplinary treatment
Comparison: Usual care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with Usual care	**Risk with Multidisciplinary treatment**				
Depression assessed with: BDI, HADS, PGWB follow up: range 6 months to 12 months	-	SMD **0.3 lower** (0.74 lower to 0.14 higher)	-	445 (4 RCTs)	⨁◯◯◯ VERY LOW a,b	Multidisciplinary treatments may reduce depressive symptoms in FM but the evidence is very uncertain.

***The risk in the intervention group** (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; **SMD:** Standardised mean difference

Explanations

a. All the studies included in the statistical analysis in the assessment of depression showed high risk in at least one of the analyzed bias. 3 showed high risk in attrition bias (differential dropout rates, huge number of dropouts of the overall sample in 2 of the studies). 2 studies showed high risk in other bias (participants were allowed to receive concomitant treatments, multiple outcomes assessed without control for multiple comparisons).
b. I^2=79%

Supplementary fig. 12 **Summary of findinds (GRADE) for anxiety.**

Multidisciplinary treatment compared to usual care for FM - ANXIETY

Patient or population: Adults with ACR FM diagnosis
Setting: Adults
Intervention: Multidisciplinary treatment
Comparison: Usual care

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	Comments
	Risk with Usual care	**Risk with Multidisciplinary treatment**				
Anxiety assessed with: FIQ-Anxiety Scale from: 0 to 10 follow up: range 3 months to 8 months	The mean anxiety was **5.93** Scale unit	MD **0.57 Scale unit lower** (1.17 lower to 0.02 higher)	-	840 (7 RCTs)	⨁⨁◯◯ LOW [a]	Multidisciplinary treatments may result in little to no difference in anxiety.

***The risk in the intervention group** (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; **MD:** Mean difference

Explanations

a. 4 out of the 7 studies included in the statistical analysis in FIQ-Anxiety showed high risk in at least one of the analyzed bias. 3 showed high risk in attrition bias (differential dropout rates, high number of dropouts of the overall sample, discrepancies between number of randomized and number of started intervention). The remaining one showed high risk in other bias (multiple outcomes assessed without control for multiple comparisons).