Effect of Exercise on Depressive Symptoms in Multiple Sclerosis: A Meta-Analysis

[redacted]

This thesis is submitted in partial fulfilment of an Honours degree of

Bachelor of Psychological Science

School of Psychology

The University of Adelaide

September 2020

Word Count: 9131

Table of Contents

Tables and Figures

Tables

Chapter 3: Results

Figures

Chapter 3: Results

Abstract

Background: Depression is debilitating and highly prevalent among adults with multiple sclerosis (MS). Exercise may reduce symptoms of depression; however, the different biases present in uncontrolled studies could overestimate, or underestimate, noted treatment effects. *Aim:* To review the available evidence for the effectiveness of exercise training targeted to MS based on the 'gold' standard randomised controlled trial. *Methods:* Fifteen independent studies, comprising a pooled sample of 544 adults with relapsing-remitting or progressive MS, were identified from the Cochrane Library, Embase, PEDro, PsycINFO, and PubMed databases. The methodological quality of included studies was assessed with the PEDro scale. Hedges' *g*, fail-safe *N*s, heterogeneity, and *p*-values were calculated using random effects modelling. The moderating effects of exercise type (aerobic vs. nonaerobic), exercise volume (i.e., total number of sessions), and baseline depression scores were additionally examined. *Results:* Most studies reported adequate methodological details, although blinding of participants and administering therapists were criteria unfulfilled. Exercise programs resulted in immediate and large improvements in mood (g = .79, 95% CI [.39, 1.19], $p < .01$). However, conclusions could not be drawn for longer-term effectiveness (g = -.18, 95% CI [-.50, .14], p = .28; $N_{studies}$ = 4). Univariate meta-regressions revealed a significant moderating effect for exercise volume. Depressed adults who took part in exercise programs also displayed significantly greater improvements than controls. *Conclusion:* Exercise efficacy for post-MS depression can be maximised by bolstering session volume and duration for those who are most prone to respond. The influence of exercise type on depression outcomes remains unclear.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma in any University, and, to the best of my knowledge, this thesis contains no material previously published except where due reference is made. I give permission for the digital version of this thesis to be made available on the web, via the University of Adelaide's digital thesis repository, the Library Search and through web search engines, unless permission has been granted by the School to restrict access for a period of time.

[redacted]

[redacted]

28 September 2020

Contribution Statement

In writing this thesis, my supervisor and I collaborated to generate a research question of interest. My supervisor guided me on the systematic review and meta-analysis methodology, and together we developed the study eligibility and screening criteria. I created the logic grids for each database, conducted the literature searches and study screening. My supervisor independently assessed 113 records as a measure of screening reliability. I extracted all data, completed the risk of bias assessment, and undertook all statistical analyses. Finally, I wrote up all aspects of the thesis.

Acknowledgements

Firstly, I wish to express my deepest gratitude to my project supervisor, XXXX. I am indebted to you for the utmost support and guidance you have provided me on this journey. Your rapid responses to emails and patience with my endless supply of questions, alongside your relentlessly hard-working nature, dedication, and striking attention to detail are not only greatly appreciated, but also inspiring. I feel very fortunate to have worked with you this year and I cannot ever thank you enough.

Secondly, I thank the researchers who responded to my questions regarding their papers. Specifically, thank you to Dr. Dena Sadeghi Bahmani (The University of Basel) and Raoof Negaresh (Tarbiat Modares University) for their prompt and helpful responses.

Thirdly, to Vikki Langton. Thank you for meeting with me on multiple occasions to share your depth of knowledge and help develop the logic grids which formed the backbone of this project.

Fourthly, to my father, XXXX, and my best friend, XXXX. Despite the physical distance between us, I thank you both for constantly checking in, listening, providing words of encouragement, and making me feel connected during an otherwise lonely year. To my other friends: thank you.

Fifthly, to my boyfriend XXXX. I greatly appreciate your support and laborious listening to the intricacies of this project; I think you know this topic just as well as I do. I appreciate you picking up on the housework, cooking for me when I was glued to my computer, and putting my needs first during this challenging year. I am eternally grateful.

Lastly, I dedicate this thesis to my late mother, XXXX. Losing you has motivated me to chase my dreams, because that is what you always encouraged me to do. My guiding purpose in life is to make you proud and I hope this project has achieved that. I miss, love, and remember you forever.

Chapter 1

Introduction

Multiple Sclerosis: Aetiology, Epidemiology, Symptomology

Multiple sclerosis (MS) is an inflammatory disease that affects the brain and spinal cord. Pathogenically, MS is triggered by an auto-immune response whereby lymphocytic immune cells infiltrate the central nervous system (CNS), via the blood brain barrier, to attack the protective myelin sheath that surrounds the axons of nerve cells (Hemmer et al., 2015; Høglund & Maghazachi, 2014). Myelin damage, or demyelination, inhibits the ability of nerve cells to conduct messages across the nervous system which in turn, impairs motor, sensory, and cognitive functions (Australian Bureau of Statistics, 2012; Høglund & Maghazachi, 2014; McCabe, 2012; Warren & Warren, 2001). Widespread demyelination leads to chronic neurodegeneration and clinically presents as progressive accumulation of disability (Compston & Coles, 2008).

MS is one of the most common forms of disability, affecting more than 2.3 million people worldwide and approximately 25,607 people in Australia (Browne et al., 2014; Campbell et al., 2019). The precise aetiology of MS remains unknown. Rather, a range of environmental and genetic risk factors have been implicated in the disease process, including genetic predisposition (e.g., human leukocyte antigen gene), Epstein-Barr virus contraction, sedentary or unhealthy lifestyle behaviours (e.g., vitamin D deficiency, cigarette smoking, obesity), and geographical location (i.e., residing in higher latitudes; Amato et al., 2017; Simpson et al., 2019). Female gender is also a risk: women are three times more likely to be diagnosed with MS than men (Howard & Younger, 2016; Ribbons et al., 2017).

The burden of MS is exacerbated by the young age of onset, with most people diagnosed between the ages of 20 and 40: during their most active and productive years of life (Amato et al., 2017; Multiple Sclerosis Australia, 2017). A high proportion of this group

subsequently report unemployment, or lower workforce participation due to MS-related impairment or disability (Multiple Sclerosis Research Australia, 2018). Indeed, lost wages and productivity remain one of the largest economic costs in MS care (Multiple Sclerosis Research Australia, 2018).

The clinical symptoms of MS are diverse and depend on the location and severity of myelin damage in the CNS. Somatic and cognitive features include, but are not limited to, paraesthesia (tingling sensation), numbness, visual impairment, gait (walking) disturbance, spasticity (i.e., loss of muscle control), speech difficulty, fatigue, impaired coordination and balance, and reduced concentration and memory (Milo & Miller, 2014). Lhermitte's sign (trigger of an electric shock sensation that moves down the spine or limbs upon neck flexion) and Uhthoff's phenomenon (temporary exacerbation of symptoms when body temperature increases), also present in persons with MS (Compston & Coles, 2008).

Clinical Course

Although the pattern of MS presentation is highly variable, with symptoms fluctuating and even progressing, four disease courses have been identified. *Relapsing-remitting MS (RRMS)* is the most common, affecting 85% of all cases (Klineova & Lublin, 2018). RRMS is characterised by acute relapses of neurologic dysfunction and alternates with periods of remission and clinical stability (Confavreux et al., 2000; Klineova & Lublin, 2018; Lublin & Reingold, 1996). Although relapse frequency varies, persons with MS rarely experience more than two episodes per year (Compston & Coles, 2008). Symptom recovery during remission phases suggests remyelination (Chari, 2007); although, relapse recovery becomes incomplete over time, leading to stepwise accrual of dysfunction (Compston & Coles, 2008; Lublin et al., 2003). When recovery is incomplete and symptoms steadily worsen, the patient has entered the *secondary-progressive* course of MS *(SPMS;* Lublin & Reingold, 1996). SPMS affects approximately 65% of the MS population, occurring an

average of 19 years after RRMS onset (Compston & Coles, 2008; Rovaris et al., 2006). Earlier progression to SPMS has been associated with male gender and older age at diagnosis (Klineova & Lublin, 2018). *Primary-progressive MS (PPMS)* is rare, affecting 10% to 20% of patients (Compston & Coles, 2008). PPMS is characterised by continued disease progression from onset with minor fluctuations (including lack of initial relapse-remitting phase), but no distinct relapses (Lublin & Reingold, 1996). Finally, *clinically isolated syndrome (CIS)* involves an acute episode of demyelination. Many who experience CIS will not develop a second relapse. However, the chance of a second inflammatory attack, and subsequent MS diagnosis, increases from 50% at two years, to 82% at 20 years if CIS is accompanied by white matter lesions (Fisniku et al., 2008).

Diagnosis and Prognosis

MS is diagnosed according to the number of clinical attacks, lesions, and their dissemination across space and time in the CNS (McDonald et al., 2001; Poser et al., 1983; Thompson et al., 2018). Clinical evidence is obtained by magnetic resonance imaging (MRI), laboratory samples (e.g., cerebral spinal fluid to identify an increase in immunoglobulin concentrations), and/or an assessment of medical history. Although advances in medical treatment and care over the last 20 years have improved life expectancy, those with MS can expect to live at least 7 years fewer than others without the disease (Marrie et al., 2015). The prognosis is better for those with relapsing-remitting, rather than progressive forms of the disease, likely due to the availability of disease-modifying medications. Conversely, late onset of disease (i.e., > 40 years of age at onset of symptoms) and a high frequency of relapses predict poor prognosis (Confavreux et al., 1980).

Depression and MS

The variety of disabling physical, sensory, and cognitive symptoms that characterise MS have a profound effect on psychological functioning. Persons with MS have a high

prevalence of psychiatric symptoms and disorders, with major depressive disorder (MDD; Milo & Miller, 2014) being one of the most debilitating comorbidities. Depressive disorder is characterised by the presence of low mood, accompanied by somatic and cognitive changes that significantly impair everyday functioning (American Psychiatric Association, 2013). Rates of depressive disorder among adults living with MS range from 19% to 54% (Goldman Consensus Group, 2005; Marrie et al., 2009; Raskind, 2008): a stark contrast to the lifetime rate of 11% estimated in the general population (Lim et al., 2018). A further 50% report depressive symptoms of concern, such as feelings of worthlessness and guilt, thoughts of suicide, and loss of interest in day-to-day activities (Chwastiak et al., 2002). This subgroup are also approximately 7.5 times more likely to die from suicide (Sadovnick et al., 1996). Regardless of whether depression is a reactive psychological problem or an 'organic' cause of MS (e.g., due to demyelination of key areas of the brain; Haussleiter et al., 2009), the adverse effects of untreated depression on MS disease course and daily functioning are undeniable (Fischer et al., 2015b; Kargiotis et al., 2010). Indeed, depression is considered to be the most important determinant of treatment adherence, functional status, and quality of life (Siegert & Abernethy, 2005).

Depression Assessment

Depression remains underdiagnosed and undertreated in the MS population. Rates of missed diagnoses are reported at approximately 30% and are likely attributable to the somatic and vegetative symptom overlap between the features of MS and psychiatric symptoms (e.g., altered sleeping patterns, appetite, fatigue, concentration), but also symptom denial and false reporting (Cetin et al., 2007; Feinstein et al., 2014; Fischer et al., 2015b; Skokou et al., 2012). Concerningly, up to 65% of adults with MS that meet the criteria for MDD, do not receive pharmacological or psychotherapeutic treatment (Mohr et al., 2006).

Ideally, the 'gold standard' clinician-administered interview should routinely be used in MS care to assess MDD. However, self-report questionnaires have emerged as the predominant format for depression measurement in MS research due to their practicality and ease of administration (Feinstein et al., 2014). Importantly, several depression measures have utility as a screening tool, with optimal cut-off scores for 'caseness' developed from receiver operating characteristic curve analysis (Habibzadeh et al., 2016). Targeted measures, such as the Beck Depression Inventory and its variants (BDI; BDI-II; BDI-FS; Beck et al., 1961, 1996, 2000) or the Center for Epidemiologic Studies Depression Scale (CES-D-10; Kohout et al., 1993), have also demonstrated good to excellent diagnostic validity for persons with MS; that is, the ability to correctly identify someone that meets clinical levels of depression (Amtmann et al., 2014; Benedict et al., 2003; Hind et al., 2016; Strober & Arnett, 2015; Watson et al., 2014). Measures that incorporate depression subscales, such as the Hospital Anxiety and Depression Scales (HADS-D; Zigmond & Snaith, 1983), have also shown good content validity (i.e., correspondence between test items and depression symptoms) and test-retest reliability (i.e., reliability over time; Waston et al., 2014). In sum, available depression measures can reliably be used in applied and research settings as initial screening tools for mental health problems in MS (Hind et al., 2016).

Depression Treatment

In addition to accurate depression assessment for persons with MS, timely treatment is essential to reduce symptom severity and to improve health behaviours, such as treatment adherence (Kalb et al., 2019; Skokou et al., 2012). Despite case control clinical trials demonstrating the benefits of pharmacotherapy (i.e., fluoxetine, sertraline; Feinstein et al., 2014; Koch et al., 2011), a link between cognitive dysfunction—a pervasive MS symptom—and poor treatment response to antidepressants has been shown (Julian & Mohr, 2006). Similarly, whilst psychotherapies (e.g., cognitive behaviour therapy) have demonstrated

robust, immediate benefits to mood (Feinstein et al., 2014; Hind et al., 2010), such therapies need to be carefully adapted to accommodate for the attention and information processing problems typically experienced by persons with MS (Choobforoushzadel et al., 2015; Solaro et al., 2018).

Exercise Training and Depression in MS

Exercise training offers a low-cost and accessible lifestyle treatment for depression in MS. Among the general population, exercise is considered a good complement to both pharmacological and non-pharmacological treatments for depression (e.g., Craft, & Perna, 2004). Defined as "a type of physical activity consisting of planned, structured, and repetitive bodily movement," *exercise* can help to improve and maintain physical fitness, including the ability to perform skill-related (e.g., agility, balance, coordination) or health-related activity (e.g., endurance, strength, flexibility; American College of Sports Medicine, 2014, pp. 2). The exercise rehabilitation literature also makes a distinction between *aerobic* and *nonaerobic exercise*. The former involves an increased and sustained heartrate to develop cardiorespiratory fitness (e.g., walking, running, cycling, swimming, rowing). In comparison, *nonaerobic exercise* typically refers to muscular fitness, flexibility, and neuromotor exercise such as resistance training, yoga, and Pilates (Doyne et al., 1987; Hsu et al., 2019; Stanton & Reaburn, 2014).

There are well-documented recommendations for exercise participation and exercise safety among persons with MS (Motl, 2014; Petajan & White, 1999; Pilutti et al., 2014). These guidelines specify that two weekly supervised sessions of moderate intensity aerobic activity (of ≥ 30 minutes) in addition to strength training of two major muscle groups, is required in order to improve aerobic capacity, muscle strength, and general fitness (Latimer-Cheung et al., 2013). Importantly, systematic reviews have highlighted regular exercise training as a potential solution for not only maximising patient activities and general

functioning, but also reducing disease relapse rate in comparison to no-treatment controls (4.6% for exercise vs. 6.3% for controls; Motl, 2014; Petajan & White, 1999).

Several randomised controlled trials (RCTs) have examined the effect of exercise training on MS-related depression. Notably, studies have typically examined aerobic endurance exercise as an 'add-on' therapy during inpatient MS rehabilitation (e.g., Dettmers et al., 2009). Meta-analytic evidence certainly supports the benefit of exercise on depressed mood and affect when delivered as part of a broader multidisciplinary intervention for chronic disability groups (g = .45, 95% CI [.12, .79], p = .01; Dauwan et al., 2019). The ability to draw conclusions on the reductive effects of exercise based on these findings are, however, limited given their consideration of exercise as a supplement to other evidence-based components (e.g., physiotherapy, occupational therapy, nursing; Dauwan et al., 2019)

The differential effects of unsupervised versus supervised exercise training also needs to be considered. MS studies examining the efficacy of self-managed exercise programs have typically reported little or no improvement in depression symptom severity ratings (e.g., Paul et al., 2014; Romberg et al., 2005). Unsupervised interventions are not only inconsistent with current exercise guidelines for persons with MS (Latimer-Cheung et al., 2013), but may result in a participant being less inclined to initiate or complete sessions due to the self-motivation required to exercise independently. Motivation can further be impaired in individuals with depression. This may explain the non-significant treatment effects experienced by Romberg and colleagues' (2005) exercise group, many of whom exceeded the clinical cut-off score of 15 on the CES-D-10 at baseline (Björgvinsson et al., 2013). Those who are depressed may find it difficult to exercise without the benefits of social exposure and monitoring that an intervention supervised by a qualified health professional affords (Petajan et al., 1996). In saying this, there is also evidence that baseline depression severity does not moderate capacity for treatment change in MS samples (Dalgas et al., 2015; Herring et al., 2017).

The selection of control group is a further consideration. Immediate within-group reductions in depression symptoms have been associated with exercise training; although, improvements in mood have also been reported for participants in an active control condition who received training in abdominal breathing and contraction-relaxation (Grazioli et al., 2019). Similar treatment effects have also been noted when comparing exercise training with yoga or sports climbing (Siengsukon et al., 2016), or even low intensity walking and stretching (Velikonja et al., 2010). Any treatment effects associated with exercise are best examined by using a non-exercise control – in order to highlight the magnitude of any improvements in depression severity associated with exercise, specifically (Temple & Ellenber, 2000).

The type of exercise examined may also contribute to the mixed findings in the MS literature. For example, aerobic exercise has been shown to be highly effective at reducing depression severity (Béland et al., 2020; Morres et al., 2019). The argument is that maintaining an increased rate of cardiovascular activity for sustained periods, such as during running, is associated with increased production of endorphin hormones which promote feelings of wellbeing and reduced pain perception. This phenomenon, referred to as 'runner's high' in the literature (Hicks et al., 2019), is applicable to other forms of aerobic exercise that allow a similar sustained cardiovascular state (e.g., swimming, cycling; Amorosi, 2014; Antunes et al., 2016; Raichlen et al., 2012). Emerging evidence also suggests that general exercise training may promote mood by facilitating psychological improvements in self-esteem and self-efficacy (i.e., belief to achieve a task; Haller et al., 2018; Legrand, 2014; Pickett et al., 2012). Less strenuous forms of nonaerobic exercise, such as yoga may, therefore, not only improve depression due to exercise-mediated psychological improvements, but also appeal more to those experiencing low energy due to a combination of their MS and depression symptoms (Martinsen et al., 1989). To date, however, MS

research has not been able to confirm the particular type of exercise training that is most effective for depression (Ensari et al., 2014).

Exercise volume, characterised by the total number of sessions over the course of a treatment, is an additional feature that may influence treatment efficacy. Meta-regressions have identified exercise volume as a significant ($p < .01$) moderator of MS-related depression: three or more sessions of exercise per week have produced large effects ($g = .81$, 95% CI [.43, 1.20]) compared with the benefits reported by participants who exercised less than twice a week ($g = .36$, 95% CI [.13, .59]; Herring et al., 2017). Although these results indicate an exercise dose-response effect, they are not consistently supported by other MS studies (Dalgas et al., 2015; Ensari et al., 2014). Studies with healthy adults have also been unable to confirm whether a higher session volume has a larger effect on depression symptom severity than sessions that are scheduled on a less frequent basis or smaller volume (Brown et al., 2012; Hamer et al., 2009; Nebiker et al., 2018).

Current Study

Exercise training may help to promote mental health among adults living with MS, by reducing depression symptom severity. However, the efficacy of exercise training remains unclear, with prior reviews in this area including studies using multidisciplinary programs, unsupervised interventions, and/or active controls - all methodological features that limit the ability to draw conclusions on the isolated effect of exercise (Dalgas et al., 2015; Ensari et al., 2014; Herring et al., 2017). There also remains a need to define the most optimal type and dose of exercise. The current study systematically reviews the exercise and MS literature, using meta-analytic techniques to consolidate the available data and help to identify the features and qualities of exercise studies that may impact on treatment effectiveness. A meta-analysis is considered the ‘gold-standard’ of evidence-based healthcare as it increases the

sample size, statistical power, and accuracy of effect size estimates when compared to a single study (Borenstein et al., 2009). The specific aims of this meta-analysis are to:

1. Integrate and analyse quantitative evidence regarding the effectiveness of exercise training, in comparison to waitlist, standard care, or attention controls.
2. Explore potential sources of heterogeneity among included studies, namely the type of exercise training (aerobic vs. nonaerobic), exercise volume (total number of sessions over the course of treatment), and baseline depression scores.
3. Evaluate the strength of evidence through a critical appraisal of the reporting quality of all included studies.

Chapter 2

Method

Literature Search

Five electronic databases (Cochrane Library, Embase, PEDro, PsycINFO, PubMed) were searched from database inception to March 11 2020, with automatic email search alerts enabled up until June 1 2020. Studies examining the effect of exercise interventions in persons with MS were identified using two broad search categories relating to: 1) the population of interest - "multiple sclerosis", and 2) the method of intervention - "exercise". A comprehensive list of index terms, using relevant synonyms for each search category, was developed in consultation with an expert research librarian. Search terms were customised to the vocabulary of each database (i.e., *Emtree* – Embase, *Thesaurus* – PsycINFO, *MeSH* – Pubmed), and combined using truncation and wildcard operators (for full logic grids, see Appendix A). Terms were deliberately kept broad to ensure that all possible studies were identified. The reference lists of included studies and relevant reviews (reviews by: Dauwan et al., 2019; Dalgas et al., 2015; Ensari et al., 2014; Herring et al., 2017) were additionally screened to identify any further studies that may have been missed, with one additional study identified through this process.

Study Eligibility and Screening

In addition to being peer-reviewed articles published in the English language, or with English translation, studies were screened using Population Intervention Comparison Outcome Design (PICO-D) criteria. These criteria helped to facilitate the identification and categorisation of information specific to this review (Agoritsas et al., 2012).

Population

Studies involving adult participants (defined as ≥ 18 years) with a self-reported or physician-confirmed MS diagnosis (i.e., as determined by McDonald or Poser Criteria, MRI,

laboratorial data, medical history; McDonald et al., 2001; Poser et al., 1983; Thompson et al., 2018). There was no restriction on disease severity, with both relapsing and progressive forms included. Studies that examined samples with various neurological or physical disabilities but did not report, or have available, the data for their MS sample, were excluded.

Intervention

All studies examined an exercise training intervention in accordance with the American College of Sports Medicine's (2014) definition of exercise: that is, a planned, structured, repetitive, and purposeful regime to maintain or improve physical fitness. Studies involving single bouts of exercise, or physical activity without a focus on improving fitness specifically, were excluded. Aerobic exercise, which typically involves low to highly strenuous cardiovascular exercise over a continuous and extended period (e.g., walking, running, cycling, aquatics, circuit-style training), as well as nonaerobic exercise, or short bouts of intense exercise efforts (e.g., strength and resistance training, yoga, Pilates), were eligible. The use of standard gymnasium equipment (e.g., jump ropes, treadmills, ergometers, resistance bands, weight training machines) was also acceptable, although the intervention had to involve some form of face-to-face supervised practice provided by a trained health professional (e.g., exercise physiologist, physiotherapist; Latimer-Cheung et al., 2013; Schuch et al., 2016). Exercise interventions that were self-managed, delivered as part of a multidisciplinary rehabilitation program, or involved the use of specialised equipment and techniques (e.g., functional electrical stimulation, vibration machines, transcranial stimulation, neuromodulation, robot assistance), were excluded.

Comparison

To determine whether any observed effect arose directly from the exercise intervention, only studies that used a non-exercise, passive control condition (e.g., waitlist, standard medical, psychosocial monitoring) were eligible (Gharakhanlou et al., 2020).

Outcome

Established self-reported or clinician-administered measures specifically designed to screen for depression symptom severity (e.g., BDI, CES-D-10) or disorder (e.g., MDD) were included. Depression subscales were also eligible, provided that psychometric data (for the MS or the general population) were available (see Appendix B for a list of eligible measures). Measures of general affect or distress, or psychological measures that had been purposely designed for a study, were excluded.

Design

Studies had to randomly allocate participants with MS to an exercise intervention or control group: a design that is considered the 'gold standard' for evaluating treatment effectiveness (Hariton & Locascio, 2018). Studies also had to use a repeated measures design, whereby group depression scores were reported pre and post-intervention.

Study screening was undertaken by the student researcher (X.X.) using Covidence systematic review software (Veritas Health Innovation). Duplicates were first removed and then titles and abstracts of potentially eligible records screened. The full-texts of relevant studies were then accessed and re-screened against the eligibility criteria. To ensure reliability of the screening process, a second researcher (project supervisor, X.X.) independently assessed 113 records. Inter-rater reliability was high (Viera & Garrett, 2005), with reviewers agreeing in 97% of cases (k = .89). The few discrepant articles were resolved through consensus discussion.

Data Extraction, Organisation and Preparation

Corresponding with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines (see Appendix C; Moher et al., 2009), a purposely designed data extraction sheet was used to organise and collate data for each study (see Appendix D). Extracted data included:

a. Study characteristics: Sample size; country; depression measure; control condition; and rate of attrition or dropout from the primary (exercise) intervention. Exercise intervention details were extracted in accordance with the Template for Intervention Description and Replication Checklist (TIDier; Hoffmann et al., 2014; see also http://www.tidierguide.org/) to ensure sufficient description of both the exercise and control conditions examined. This included detail relating to the type of exercise (aerobic or nonaerobic exercise) and, where available, number of sessions (per week), session duration (in minutes), and treatment duration (in weeks).
b. Key sample parameters: Mean age (*SD*) and male:female ratio.
c. MS-specific information: Disease subtype (relapsing-remitting, or progressive forms); mean (*SD*) disease duration (in years) and level of physical disability or function (as defined by the Expanded Disability Status Scale [EDSS]; Kurtzke, 1983).
d. Effect size data: Group means and *SD*s for each depression measure, pre and post-intervention were extracted from each study. Where these data were unavailable, data that could be converted to standardised mean differences (Hedges' g) were obtained. The data for two studies required conversion prior to analysis. Negaresh and colleagues (2019) categorised participants as 'normal weight' and 'overweight.' Here, the means and *SD*s were averaged to provide an overall score for the intervention and control groups. Dalgas (2010) reported 95% confidence intervals (CI) for their depression scores, which were converted to *SD*s (Higgins et al., 2019).

Risk of Bias Assessment

The methodological quality of each study was evaluated using the Physiotherapy Evidence Database (PEDro) Scale: a tool based on the Delphi Checklist for physical therapy interventions (see Appendix E for criteria; Maher et al., 2003; see also Verhagen et al., 1998). Each study was rated against 11 pre-specified criteria relating to measurement, analyses,

reporting transparency (e.g., specifying eligibility criteria), study design (e.g., randomisation, blinding, allocation concealment), and sample characteristics (e.g., comparability of groups at baseline). A score of *1* (criterion met) or *0* (criterion not met) was assigned for each criterion, and each study. Given that not all items equally contribute to a total score for the PEDro Scale (i.e., item relating to eligibility criteria is not scored), a component analysis approach was adopted, whereby only individual items were considered (da Costa et al., 2012). To ensure reliability of the quality assessment, ratings were compared to confirmed scores for each study available on the PEDro database (http://www.pedro.org.au). Inter-rater reliability was high, with 98% accuracy for independent ratings (Viera & Garrett, 2005).

Effect Size Calculations

Effect size data were entered into Comprehensive Meta-Analysis software (CMA Version 3.3.070, November 2014; Borenstein et al., 2013). The effect size estimate selected for this meta-analysis was Hedges' *g*, which corrects for bias in small samples (Lakens, 2013). The formula for *g* involves subtracting the mean (post-pre) change in the control group from the mean change in the intervention group, and then dividing this figure by the pooled and weighted *SD* for each group (Elllis, 2010). A random-effects model was used for all calculations to account for the range of exercise interventions, clinical heterogeneity of MS, and different depression measures used (Borenstein et al., 2009; Cumming, 2012).

Effect size *g* was computed for each depression measure and study. Where a study contributed more than one effect estimate (e.g., multiple depression measures or interventions), individual *g's* were averaged prior to being pooled in a given meta-analysis to ensure no issues of data dependency arose (Cohen, 1988). A pre-post correlation of .70, considered a conservative value for studies with a repeated measures design, was imputed into CMA in order to calculate *g* (Estrada et al., 2019). Individual *g's* were then weighted by their study's inverse variance to account for increased sampling error and variability from

smaller samples ($g_{w;}$ Lipsey & Wilson, 2001). Effect sizes were interpreted using Cohen's (1988) conventions for the social sciences, with values of .20, .50, and .80 corresponding to small, medium, and large effects. The direction of *g* was also standardised so that a positive value indicated greater improvement (i.e., lower depression scores) with exercise training. In addition, *p* values were used to infer statistical significance of individual and pooled *g's* and 95% CIs computed to estimate the precision of each g/g_w. Narrower CIs indicate higher precision (Higgins et al., 2019), whilst CIs which contain the value of zero represent a non-significant estimate (Cumming, 2012).

Between-study heterogeneity was identified and quantified using tau (τ), and I^2 statistics. Tau (τ) represents the *SD* for the overall $g_{w,}$ or degree of dispersion about the mean effects. The I^2 statistic quantifies the total variability across effect estimates as a percentage, whereby values of 25%, 50%, and 100%, indicate low, medium, and substantial heterogeneity, accordingly (Higgins et al., 2003).

Publication bias, also referred to as the 'file drawer' problem, is a key concern in meta-analyses that rely on published data, as studies with significant and large effects are more likely to be published (Borenstein et al., 2009). In the current meta-analysis, this bias was assessed with several statistics. First, a funnel plot analysis was conducted to detect potential publication bias across all included studies, followed by the trim-and-fill method. In a funnel plot, intervention effect estimates from individual studies are plotted against each study's standard error. In the absence of publication bias, a funnel plot should be symmetrical: with larger *N* studies scattered narrowly towards the top of the plot and effect sizes from smaller studies scattered widely at the bottom (Sterne et al., 2011). This result is statistically confirmed by the trim-and-fill method, a non-parametric test which identifies how many hypothetical studies would need to exist to correct for asymmetry in a funnel plot (Sterne et al., 2011). Orwin's fail-safe *N* (N_{fs}; Orwin, 1983) was calculated for each

subsequent meta-analysis. N_{fs} represents the hypothetical number of studies needed to 'nullify' the effects of *g* to a small, meaningless size (i.e., $g = .20$; Orwin, 1983). An N_{fs} value needed to exceed the total number of studies included in this meta-analysis in order to be considered robust.

Sensitivity and Moderator Analyses

Potential outlier effects were identified using a sensitivity analysis. This involved re-running the meta-analysis with all included studies and then removing one study at a time to see whether this changed the overall effect size magnitude or level of significance (Borenstein et al., 2009). A subgroup analysis was then conducted to examine the potential impact of exercise type (i.e., aerobic vs. nonaerobic) on group differences. This involved a *Q*-test based on analysis of variance ($Q_{between}$), with significant results (i.e., $p < .05$) indicating meaningful subgroup differences (Borenstein et al., 2009). Finally, there were sufficient data to undertake two univariate meta-regressions ($N_{studies} > 10$ per independent variable; Higgins et al., 2019) to examine the association between exercise volume (total number of sessions over the course of an exercise program) and baseline depression scores (the average was taken for studies administering multiple interventions) on effect size estimates. This analysis involved a random effects model with inverse variance weighting of effect sizes. Both *Q* model (Q_{model}) statistics, which show variability associated with the regression model, and *Q* residual (Q_{resid}) statistics, which indicate variability unaccounted for by the model, were considered (Borenstein et al., 2009).

Chapter 3

Results

Study Screening

Electronic database searches yielded 10,808 records, of which 6373 potentially relevant titles and abstracts were screened (see Figure 1). A further 113 records were retained for full-text review. Fifteen authors were contacted for additional data, with five responding (Learmonth, 2012; Learmonth et al., 2011; Negaresh et al., 2019; Oken et al., 2004; Sutherland et al., 2001; Tollár et al., 2020). During the screening process, two papers with overlapping samples were identified (Bahmani et al., 2019, 2020): only the study with complete pre post-intervention depression data was retained (Bahmani et al., 2019). The final sample comprised of 15 independent RCTs.

Sample Characteristics

The pooled sample of 544 persons with MS (Exercise = 313; Control = 231) were primarily middle-aged (mean age range = 31.12–54.60; see Table 1). Both groups had been living with chronic (i.e., > 9 years) relapsing (74%) or progressive forms of MS (26%) and mild to moderate levels of disability (i.e., Expanded Disability Status Scale < 5; Meyer-Moock et al., 2014), although MS details were not routinely reported ($N_{studies} \leq 10$). Most (83%) participants were female, consistent with the commonly reported 3:1 ratio of females to males diagnosed (see Appendix F; Ribbons et al., 2017). In the 13 studies that used depression measures with reported cut-off scores, 49% of the pooled sample (i.e., 230 of 472 participants) were classified as ‘depressed’ at baseline (see Appendix B for cut-off scores per measure).

Study Characteristics

Most studies originated in Europe ($N_{studies} = 6$), followed by the Middle East

Figure 1

PRISMA Flowchart of Study Selection Process (Moher et al., 2009)

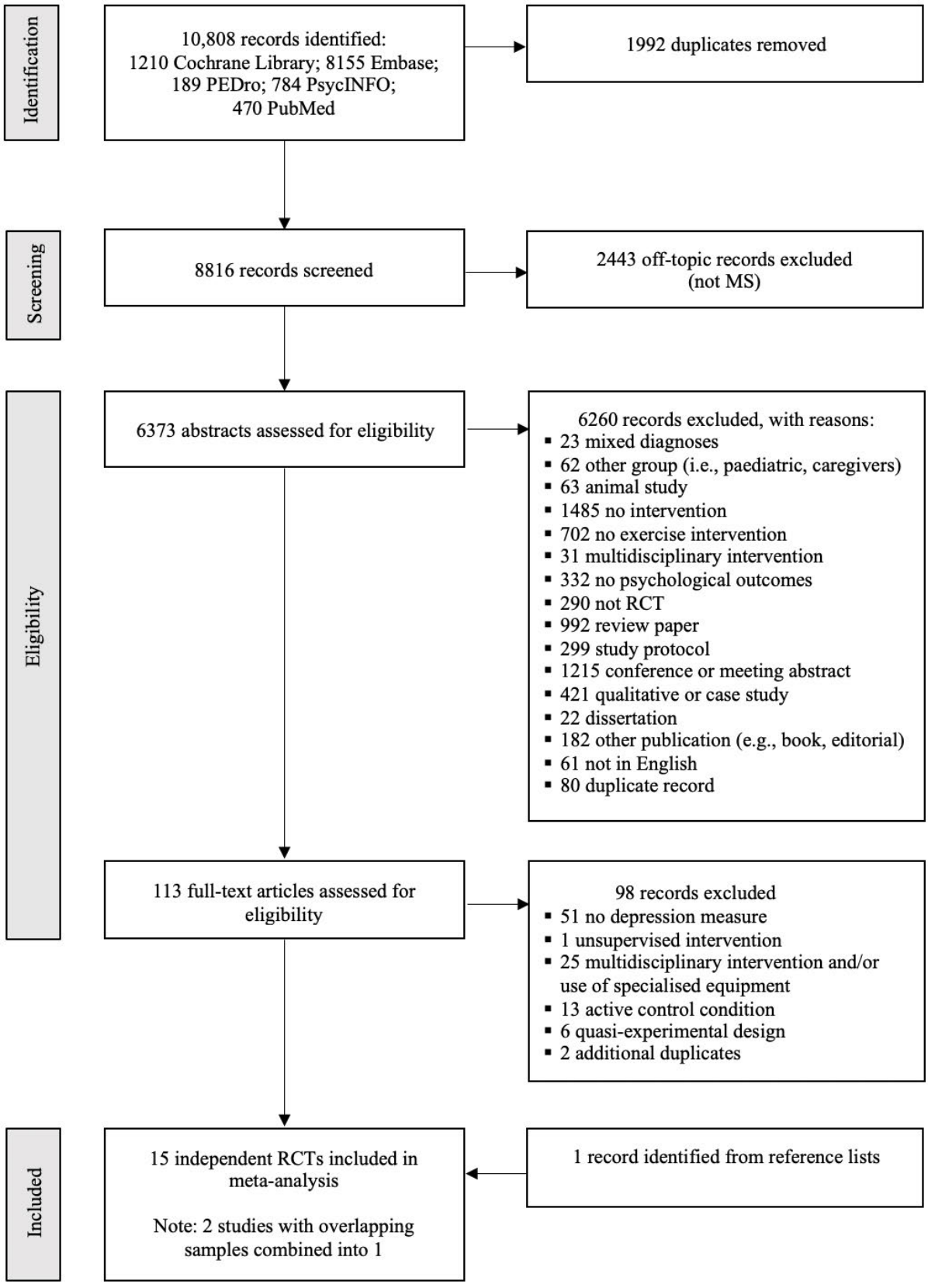


Table 1

Pooled Sample Characteristics

		Exercise				Control			
		$N_{studies}$	$N_{participants}$	*M* (*SD*)	Median	$N_{studies}$	$N_{participants}$	*M* (*SD*)	Median
Sample size		15	313	20.1 (10.1)	15	15	231	15.4 (5.9)	15
Age (in years)		15	313	44.1 (7.8)	47.4	15	231	43.8 (7.6)	45.5
Disease duration (in years)		13	273	9.1 (3.4)	7.3	13	205	9.8 (4.8)	7.4
Disability severity (EDSS)		10	257	3.9 (1.7)	3.7	10	174	3.7 (1.8)	3.2
Gender	Female	12	212			12	153		
	Male	12	39			12	37		
MS sub-type	Relapsing-remitting	6	93			6	75		
	Progressive	6	34			6	26		

Note. $N_{studies}$ = number of studies providing this data; $N_{participants}$ = number of participants providing this data; *M* = mean; *SD* = standard deviation; EDSS = Expanded Disability Status Scale.

($N_{studies}$ = 5), North America ($N_{studies}$ = 3), and Australia ($N_{studies}$ = 1; see Table 2). No individual study contributed more than 12% of the pooled sample size. Sample sizes were typically small, representing pilot or feasibility studies (N_{range} = 9–61).

Depression symptom severity was measured with 10 individual self-assessment tools, more commonly the BDI ($N_{studies}$ = 5), followed equally by the Profile of Mood States – Depression Subscale (POMS-D; $N_{studies}$ = 3; McNair et al., 1971), and the HADS-D ($N_{studies}$ = 3). Measures that align with diagnostic symptom criteria (i.e., Diagnostic and Statistical Manual of Mental Disorders – Fifth edition; DSM-V; American Psychiatric Association, 2013), and considered to be sensitive to treatment change, were also used: the 30-item Inventory of Depressive Disorders ($IDS\text{-}SR_{30}$; Rush et al, 2000) and 16-item Quick Inventory of Depressive Symptomology ($QIDS\text{-}SR_{16}$; Rush et al., 2003).

Intervention Characteristics

Twenty different exercise interventions were evaluated across the 15 studies. Most involved aerobic ($N_{studies}$ = 8), then nonaerobic ($N_{studies}$ = 3) interventions. Four studies delivered multiple respective interventions of either, or both types of exercise. A single study examined three separate aerobic interventions: upper body arm, rowing, and bicycle ergometry (Briken et al., 2014). Other common aerobic interventions included treadmill training (walking/jogging), circuit training, and aquatics. Nonaerobic exercise involved yoga, Pilates, or resistance training.

Exercise sessions were typically delivered in a group setting ($N_{studies}$ = 14) and supervised by various disciplines including certified instructors, neurologists, physical therapists, physiotherapists, or physiatrists - who also provided technique advice. Exercise programs were catered to a group of individuals with predominantly sedentary activity levels. In particular, six studies required participants to have refrained from exercise for periods ranging from 4 to 52-weeks prior to commencement of the study, while three studies

Table 2

Study and Exercise Intervention Characteristics

Lead author (date)	Country	Sample			Exercise Intervention						Control condition
		Total [a] [M:F]	EX	C	Mode	Specific exercise/s	No. sessions (per week)	Session time (mins)	No. weeks	Dropout [b]	
Ahmadi (2013)	Iran	31 [0:31]	1) 10 2) 11	10	1) Aerobic 2) Nonaerobic	1) Treadmill exercise 2) Yoga	3	60-70	8	- -	Waitlist
Bahmani (2019)	Iran	47 [0:47]	26	21	Aerobic	Endurance treadmill & bicycle ergometry	3	30-45	8	19.2%	Attention
Briken (2014)	Germany	42 [-]	1) 10 2) 11 3) 11	10	Aerobic	1) Upper body arm ergometry 2) Rowing ergometry 3) Bicycle ergometry	2-3	15-45	8-10	1) 16.7% 2) 8.3% 3) 8.3%	Waitlist
Çakit (2010)	Turkey	30 [-]	15	15	Aerobic	Progressive resistance bicycle ergometry & balance exercise	2	68-85	8	6.7%	Standard care
Dalgas (2010)	Denmark	31 [-]	15	16	Nonaerobic	Progressive resistance training	2	-	12	15.8%	Standard care
Fleming (2019)	Ireland	9 [0:9]	3	6	Nonaerobic	Pilates	2	60	8	40%	Waitlist
Hebert (2011)	US	26 [4:22]	13	13	Aerobic	Endurance bicycle ergometry & stretching	2	55	6	0%	Waitlist
Learmonth (2011)	Scotland	32 [9:23]	20	12	Aerobic	Endurance circuit training	2	45-60	12	25%	Standard care
Miller (2011)	Scotland	30 [11:19]	15	15	Nonaerobic	Physiotherapy strength & flexibility training	2	60	8	0%	Standard care
Negaresh (2019)	Iran	61 [21:40]	34	27	Aerobic	Interval bicycle ergometry	3	42-66	8	5.6%	Standard care
Oken (2004)	US	57 [4:53]	1) 15 2) 22	20	1) Aerobic 2) Nonaerobic	1) Bicycle ergometry 2) Yoga	1	- 90	26	1) 28.6% 2) 15.4%	Waitlist
Petajan (1996)	US	46 [15:31]	21	25	Aerobic	Arm and leg ergometry	3	40	15	-	Standard care
Razazian (2015)	Iran	54 [0:54]	1) 18 2) 18	18	1) Aerobic 2) Nonaerobic	1) Aquatic training 2) Yoga	3	60	8	- -	Attention
Sutherland (2001)	Australia	22 [10:12]	11	11	Aerobic	Aquatic training & land-based training	3	45	10	-	Standard care
Tollár (2020)	Hungary	26 [2:24]	14	12	Aerobic	High-intensity bicycle ergometry	5	60	5	0%	Waitlist

Note. [a] number of participants allocated to groups at baseline; M = male; F = female; EX = exercise intervention; C = control group; [b] number who did not complete EX; (-) = data not provided.

excluded participants with recent (3–6 months) experience in the particular exercise method used. Sessions mostly occurred twice or thrice weekly over a period of 5 to 26-weeks (M = 10.00, SD = 5.06), to build general strength or endurance. Individual sessions ranged from 15 to 90 minutes (M = 56.11 minutes, SD = 15.53).

Exercise intensity was generally unreported, with only six studies formally evaluating this (e.g., Borg Rating of Perceived Exertion [RPE] Scale; Borg, 1998). Studies did, however, target programs to participant differences in fitness levels - by measuring age-predicted peak heartrate (using heart rate monitors), aerobic capacity (i.e., maximal oxygen uptake or milliliters of oxygen used in one minute per kilogram of body weight; Lundby et al., 2017), or maximum effort on a resistance movement (i.e., maximum amount of weight in kilograms that a person can lift for one repetition). Four studies required self-regulation of exercise intensity, such as choosing when to increase the speed on a treadmill or how long to hold a yoga pose (Ahmadi et al., 2013; Fleming et al., 2019; Learmonth et al., 2011; Oken et al., 2004).

Dropout or attrition rates, where reported, were generally low (M = 13.5%, SD = 11.8; $N_{studies}$ = 11). Three interventions retained all participants until completion, indicating high levels of engagement (Hebert et al., 2011; Miller et al., 2011; Tollar et al., 2020). Conversely, Fleming (2019) reported that 40% of their intervention group did not complete the allocated Pilates intervention. However, the sample size for the intervention condition was small (N = 5) and the reasons for withdrawal were unrelated to the exercise intervention (e.g., side effects from a medication, injury obtained at work). Importantly, no adverse events or injuries due to exercise training were reported across all studies.

Most studies used standard care controls, by maintaining usual treatments aside from any exercise training ($N_{studies}$ = 7). Waitlist controls, whereby exercise treatment was offered upon study completion, were also common ($N_{studies}$ = 6). Two studies used an attention

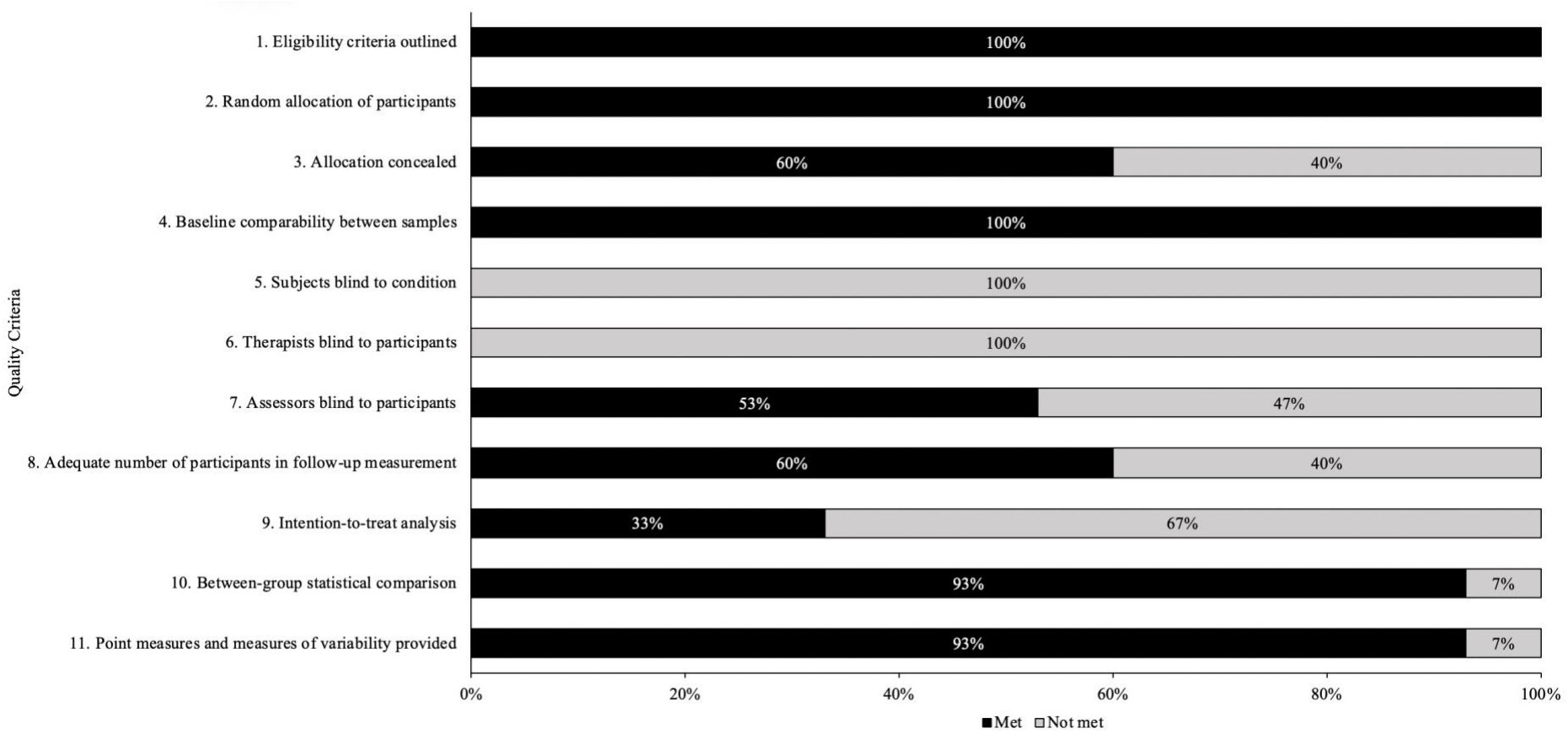
control condition, which involved participants meeting with hospital or research personnel for the same frequency and duration as the exercise group and, thereby, controlling for potential confounding effects of social exposure from exercise training (Bahmani et al., 2019; Razazian et al., 2015).

Risk of Bias Assessment

The proportion of studies meeting each criterion of the PEDro Scale is presented in Figure 2 (see Appendix G for individual study ratings). Eligibility criteria were explicitly stated (Criterion 1: 100% fulfilled) and, as per the requirements of this review, all participants were randomly allocated to their groups (Criterion 2: 100% fulfilled). Allocation was, however, only concealed in some studies (e.g., use of sealed opaque envelopes; Criterion 3: 60% fulfilled). Randomisation methods were, however, effective – as suggested by the comparability of the exercise and control groups on key sample characteristics (i.e., baseline depression scores, MS duration and severity; Criterion 4: 100% fulfilled). Given the physical nature of the intervention under examination, blinding (of subject and therapist) was an issue across all studies (Criteria 5-6: 0% fulfilled). Assessor blinding was also difficult to achieve (Criterion 7: 53% fulfilled). Attrition was minimal, with at least 85% of participants completing both pre and post-intervention assessments (Criterion 8: 60% fulfilled). However, intent-to-treat analysis (i.e., where all participants are included in the analysis according to original group assignment, regardless of whether the treatment was received, completed, or not; McCoy, 2017) was not common, despite being a recommended statistical method for RCTs (Criterion 9: 33% fulfilled). Between-group statistical comparisons (e.g., one-way ANOVA, *p*-values; Criterion 10: 93% fulfilled) and measures of variance (e.g., *SD*s, 95% CIs; Criterion 11: 93% fulfilled) were reported. In sum, most studies ($N_{studies}$ = 11) provided sufficient methodological detail consistent with level one evidence (Burns et al., 2011; Maher et al., 2003).

Figure 2

Proportion of Included Studies Meeting Each Criterion on the PEDro Scale (Maher et al., 2003)



Effectiveness of Exercise

Short-Term Effects

Across all 15 studies, the pooled and weighted *g* was large and significant: exercise training was associated with large and immediate reductions in depression symptom severity scores compared to peers who accessed standard care, psychosocial monitoring, or were waitlisted (Table 3). This finding was, however, susceptible to publication bias. As seen in Figure 3, a funnel plot analysis indicated an underestimation of effect with an additional four studies 'filled' due to lack of symmetrical dispersion within the plot (Duval & Tweedle, 2000). These results were also characterised by substantial between-study heterogeneity, with small to very large effects represented ($I^2 = 81.46$, $\tau = .70$). This included statistically significant group differences ($p < .01$) in favour of exercise reported by seven studies, but also small to large imprecise effects (i.e., wide CIs) as found in eight studies.

Figure 3

Funnel Plot of Standard Error by Hedges' g

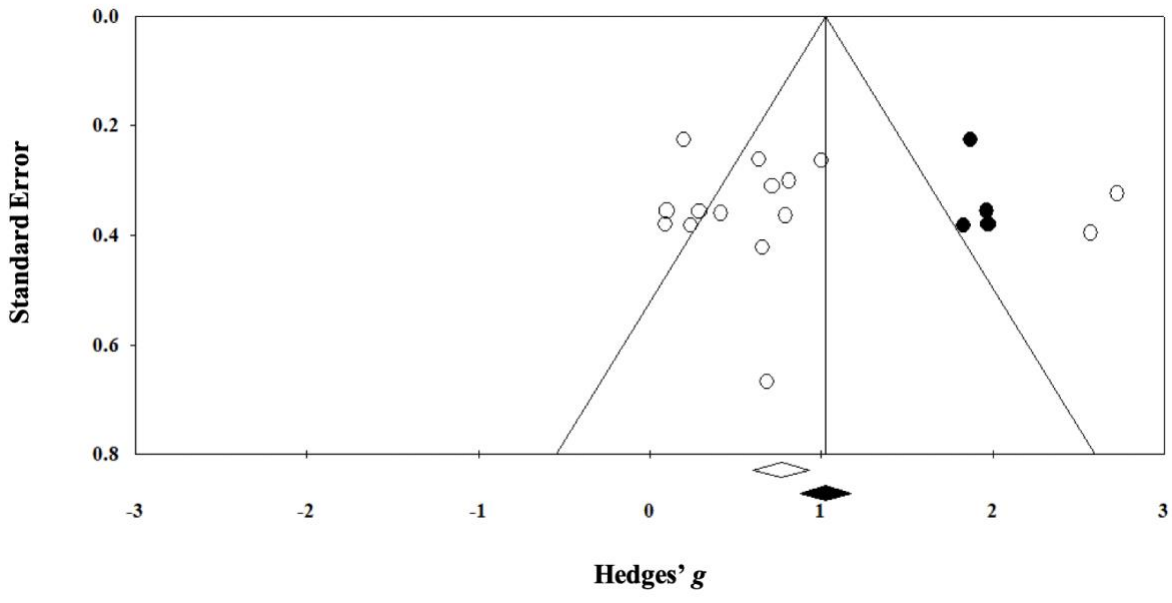
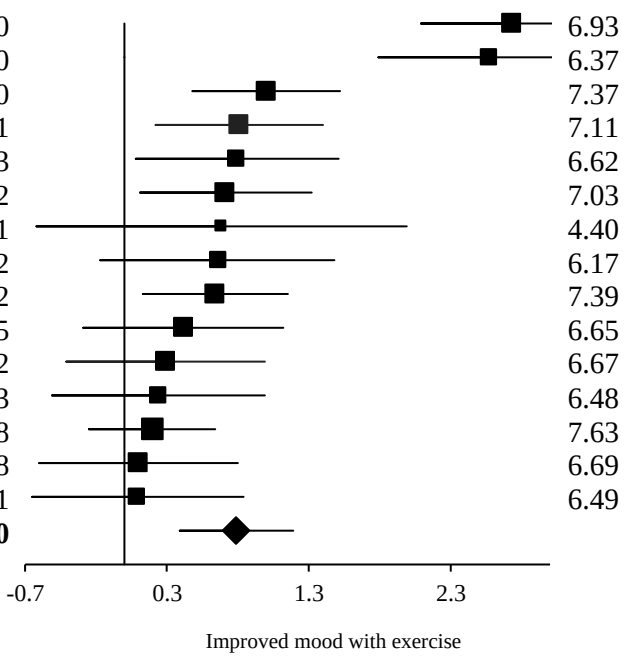


Table 3

Standardised Mean Group Differences (Hedges' g) by Study With Forest Plot

Lead author (date)	*N*	g/g_w	95% CI LL	95% CI UL	*p*	Forest plot of *g* ± 95% CI	Weight (random)
Razazian (2015)	54	2.73	2.09	3.36	.00		6.93
Petajan (1996)	46	2.57	1.79	3.35	.00		6.37
Briken (2014)	42	1.00	.48	1.52	.00		7.37
Bahmani (2019)	47	.81	.22	1.40	.01		7.11
Dalgas (2010)	31	.79	.08	1.51	.03		6.62
Ahmadi (2013)	31	.71	.11	1.32	.02		7.03
Fleming (2019)	9	.68	-.62	1.99	.31		4.40
Sutherland (2001)	22	.66	-.17	1.48	.12		6.17
Negaresh (2019)	61	.64	.13	1.15	.02		7.39
Learmonth (2011)	32	.42	-.29	1.12	.25		6.65
Çakit (2010)	30	.29	-.41	.99	.42		6.67
Tollár (2020)	26	.24	-.51	.99	.53		6.48
Oken (2004)	57	.20	-.25	.64	.38		7.63
Miller (2011)	30	.10	-.60	.80	.78		6.69
Hebert (2011)	26	.09	-.65	.84	.81		6.49
Total	**544**	**.79**	**.39**	**1.19**	**.00**		



Note. *N* = total sample size; *g* = Hedges' *g*; Hedges' g_w = weighted Hedges' *g* (note: weighting only applies when ≥ 2 studies are combined); CI = 95% confidence interval (with lower and upper limits); *p* = significance level.

Further examination of individual study results, grouped by outcome, identified large and positive effects associated with both aerobic and nonaerobic programs (Table 4). These findings were based on mood subscales (POMS-D) but also targeted measures of depression (BDI, BDI-FS, IDS-SR_{30}). The large N_{fs} values indicate that these findings are robust. Four studies in particular reported large ($g > .80$) to very large ($g \geq 1.00$) highly significant overall effects. This included large individual effects associated with nonaerobic yoga (Razazian et al., 2015), as well as endurance aerobic exercise such as aquatic or rowing, cycling, and arm ergometry training (Bahmani et al., 2019; Briken et al., 2014; Petajan et al., 1996; Razazian et al., 2015). Interestingly, Fleming and colleagues (2019) also identified a very large and significant individual effect using the POMS-D scale; although, negligible (QIDS-SR_{16}) and small (HADS-D) non-significant effects were also found in this study. Moreover, medium effects were reported by studies that examined intermittent styles of training, whereby active and comparative rest periods were incorporated. For example, resistance training with rests between weightlifting efforts (Dalgas et al., 2010) and cycling intervals (Negaresh et al., 2019). A further five studies reported small effects ($g_{range} = .04–.29$) with wide 95% CIs. The moderate and positive overall effects reported by Ahmadi et al. (2013), Fleming et al. (2019), and Sutherland et al. (2001) were non-significant, likely due to their small Ns ($N < 32$). The mixed findings are further highlighted by the heterogeneity statistics.

Longer-Term Effects

Four studies examined the longer-term impact of exercise, from 4 to 52 weeks post-intervention (see Table 5). The weighted g was negligible, negative, and not statistically significant. Given the small number of studies contributing to these data, no firm conclusions about the longer-term effects of exercise can be drawn. Interestingly, Learmonth and colleagues (2011) identified a reverse effect at 26-weeks following their aerobic-based

Table 4

Short-Term (Baseline to Post-Intervention) Effects Associated With Exercise Training

Lead author (year)	Exercise type	Measure	$N_{studies}$	$N_{participants}$	g_w	g	95% CI		p	N_{fs}	Heterogeneity	
							Lower	Upper			I^2	τ
Petajan (1996)	Aerobic	POMS-D	1	46		2.57	1.80	3.35	.00			
Fleming (2019)	Nonaerobic		1	9		1.59	.16	2.03	.03			
Oken (2004)	Aerobic		1	57		.31	-.35	.96	.36			
	Nonaerobic					.04	-.56	.63	.90			
		Total POMS-D	3	112	1.41		-.35	3.17	.12	18	93.18	1.48
Briken (2014)	Aerobic	$IDS\text{-}SR_{30}$	1	42		1.00	0.48	1.52	.00			
Razazian (2015)	Nonaerobic	BDI	1	54		2.83	1.91	3.74	.00			
	Aerobic					2.63	1.75	3.51	.00			
Ahmadi (2013)	Nonaerobic		1	31		.78	-.07	1.64	.07			
	Aerobic					.65	-.22	1.47	.14			
Negaresh (2019)	Aerobic		1	61		.64	.13	1.15	.02			
Çakit (2010)	Aerobic		1	30		.29	-.41	.99	.42			
Tollár (2020)	Aerobic		1	26		.24	-.51	.99	.53			
		Total BDI	5	202	.93		.045	1.81	.04	18	89.74	.95
Bahmani (2019)	Aerobic	BDI-FS	1	47		.81	.22	1.40	.01			
Dalgas (2010)	Nonaerobic	MDI	1	31		.79	.08	1.51	.03			
Sutherland (2001)	Aerobic	POMS-SF-D	1	22		.66	-.17	1.48	.12			
Learmonth (2012)	Aerobic	HADS-D	1	32		.42	-.29	1.12	.25			
Fleming (2019)	Nonaerobic		1	9		.41	-.83	1.66	.52			
Miller (2011)	Nonaerobic		1	30		.10	-.60	.80	.78			
		Total HADS-D	3	71	.28		-.18	.74	.24	1	.00	.00
Oken (2004)	Nonaerobic	CES-D-10	1	57		.37	-.23	.97	.22			
	Aerobic					.07	-.59	.72	.84			
Hebert (2011)	Aerobic	BDI-II	1	26		.09	-.65	.84	.81			
Fleming (2019)	Nonaerobic	$QIDS\text{-}SR_{16}$	1	9		.05	-1.18	1.28	.94			

Note. $N_{studies}$ = number of studies providing this data; $N_{participants}$ = number of participants providing this data; Hedges' g_w = weighted standardised mean difference (note: weighting only applies when ≥ 2 studies are combined); p = significance level; CI = 95% confidence interval for g/g_w; N_{fs} = fail-safe N statistic; τ (Tau) = estimate of variance; I^2 = proportional estimate of true effect variance over sampling error observed; Measures: POMS-D = Profile of Mood States – Depression Subscale; $IDS\text{-}SR_{30}$ = Self-Report Quick Inventory of Depression Symptomology; BDI = Beck Depression Inventory; BDI-FS = Beck Depression Inventory – Fast Screen for Medical Patients; MDI = Major Depression Inventory; POMS-SF-D = Profile of Mood States Short Form – Depression Subscale; HADS-D = Hospital Anxiety and Depression Scale – Depression Subscale; CES-D-10 = Centre for Epidemiologic Studies Depression Scale – Boston (Short form); BDI-II = Beck Depression Inventory – Second Edition; $QIDS\text{- }SR_{16}$ = Quick Inventory of Depressive Symptomology.

Table 5

Longer-Term (Post-Intervention to Follow-Up) Effects Associated With Exercise Training

Lead author (year)	Exercise type	Measure	Time (weeks)	$N_{studies}$	$N_{participants}$	g_w	g	95% CI		p	N_{fs}	Heterogeneity	
								Lower	Upper			I^2	τ
Hebert (2011)	Aerobic	BDI-II	4	1	26		.28	-.47	1.03	.46			
Miller (2011)	Nonaerobic	HADS-D	8	1	28		-.03	-.75	.69	.93			
Dalgas (2010)	Nonaerobic	MDI	12	1	31		-.36	-1.05	.33	.31			
Learmonth (2011)	Aerobic	HADS-D	26	1	32		-1.06	-1.80	-.32	.01			
			52	1	32		.25	-.45	.95	.48			
		Total		4	117	-.18		-.50	.14	.28	0	.00	.00

Note. $N_{studies}$ = number of studies providing this data; $N_{participants}$ = number of participants providing this data; Hedges' g_w = weighted standardised mean difference (note: weighting only applies when ≥ 2 studies are combined); p = significance level; CI = 95% confidence interval for g/g_w; N_{fs} = fail-safe N statistic; τ (Tau) = standard deviation of g_w; I^2 = proportional estimate of true effect variance over sampling error observed.

intervention: depression severity scores worsened for exercise participants, whereas controls reported an improvement in mood. However, at 52-weeks, these group differences were no longer significant (Learmonth et al., 2011).

Sensitivity and Subgroup Analyses

Sensitivity analyses revealed there were no outliers: the overall effect remained large and statistically significant even after removing individual studies. Moreover, the contribution of exercise type to study heterogeneity was examined by comparing aerobic ($N_{studies}$ = 8) to nonaerobic interventions ($N_{studies}$ = 3). To ensure data independence, the four studies delivering multiple exercise interventions with the same control group were excluded from this analysis (Ahmadi et al., 2013; Briken et al., 2014; Razazian et al., 2015; Oken et al., 2004). Aerobic interventions were associated with large and significant group differences (g_w = .71, 95% CI [.25, 1.17], $p < .01$); although, these findings were characterised by significant between-study heterogeneity (I^2 = 75.44, τ = 61). In comparison, nonaerobic interventions consistently produced non-significant effects (g_w = .50, 95% CI [-.31, 1.31], p = .23, I^2 = 0, τ = 0). However, differences between the two subgroups were not statistically significant ($Q_{between}$ (1) = .19, p = .67).

Meta-Regressions

The contribution of exercise volume and baseline depression scores to heterogeneity were explored with two univariate meta-regressions (see Table 6 and Appendix H for regression scatterplots). Both variables emerged as significant moderators of depression severity, with Q_{model} statistics indicating that each explained 31% and 37% of the inter-study variance in the pooled effect of exercise, respectively. That is, higher volumes of exercise were associated with greater reductions in depression severity scores. Participants who self-reported more severe symptoms of depression at baseline also demonstrated larger resulting

improvements after exercise programs. However, results of the Q_{resid} statistic for each regression indicated that some unexplained heterogeneity in the pooled effect size remained.

Table 6

Univariate Meta-Regressions Examining the Impact of Covariates on Pre to Post-Treatment Change in Depression Severity

Variable	*B*	R^2	Test of model			Goodness of fit		
			Q_{model}	df	*p*	Q_{resid}	df	*p*
Exercise volume	.07	.31	7.32	1.0	.01	53.40	13.0	.00
Baseline depression	.06	.37	7.89	1.0	.01	49.32	13.0	.00

Note. *B* = standardised beta coefficient of regression model; R^2 = proportion of between-study variance explained; Q_{model} = variability associated with regression model; Q_{resid} = variability unaccounted for by the regression model; df = degrees of freedom; *p* = statistical significance.

Chapter 4

Discussion

Key Findings

The present systematic review evaluated the effects of exercise training interventions on depression severity scores in adults with MS. Fifteen RCTs, comprising of 544 participants, were meta-analysed to generate pooled effect size data. The combined findings are promising: exercise training contributed to immediate and large statistically significant improvements in depression severity. Further research is, however, needed to determine whether these treatment gains are maintained over time. Although no significant differences between aerobic and nonaerobic types of exercise were noted, increased exercise volume and higher baseline depression severity scores were both significantly associated with depression alleviation. The findings of the present study are critically evaluated below, including a discussion of clinical and research implications alongside associated methodological limitations.

Effectiveness of Physical Exercise

The finding that exercise training was associated with large and immediate reductions in depression severity compared with non-exercise controls, is consistent with evidence for the psychological and hormonal mechanisms underlying the alleviation of depression via exercise (Antunes et al., 2016; Haller et al., 2018; Legrand, 2014). The noted treatment effects were also larger than the small to moderate pooled effects previously reported by meta-analyses examining exercise, which included uncontrolled studies using active controls, unsupervised interventions, or multidisciplinary programs, and their effectiveness in MS-related depression (Dalgas et al., 2015; Ensari et al., 2014; Herring et al., 2017). The present finding may, in part, be attributable to the inclusion of newly published studies. Namely, two recently published RCTs (Bahmani et al., 2019; Negaresh et al., 2019) contributed moderate

to large, positive effects, which may reflect advances in training methods or the provision of increased guidance/technique advice from supervising therapists, to improve treatment effectiveness. This large finding is, however, tempered by substantial heterogeneity and dispersion in effect sizes among the included studies, which suggests some imprecision in this estimate. More specifically, seven studies found moderate to large significant effects, while the remaining eight studies found non-significant results. The mixed evidence may be explained by use of varied sample sizes. For example, the eight studies producing non-significant results were typically pilot or feasibility trials and used smaller sample sizes in comparison to those finding significant effects. The small sample sizes used in these eight studies meant they were potentially underpowered to detect the true treatment effects of exercise.

Given that prior research indicates depression is a modifiable symptom of MS (Fiest et al., 2016; Herring et al., 2017), the results of this meta-analysis have some important clinical implications. The results provide some direction for tailored starting and advanced level treatment recommendations. In particular, the majority of interventions included in the present study were targeted at sedentary exercise levels. For a chronic condition such as MS, improvements in depression might require that a person starts at a minimum level of training (i.e., twice or thrice weekly, 30-45 minute sessions; Bahmani et al., 2019; Briken et al., 2014; Dalgas et al., Negaresh et al., 2019) and then consider increasing the duration and volume of sessions over time, as their physical and mental endurance improves to show larger improvements (Razazian et al., 2015). This might include a combination of either aerobic exercise, such as interval (i.e., efforts complemented with rest periods) or endurance training, or less cardiovascular intense nonaerobic exercise (e.g., yoga or resistance training), for at least 8-weeks (Bahmani et al., 2019; Briken et al., 2014; Dalgas et al., 2010; Negaresh et al., 2019; Petajan et al., 1996; Razazian et al., 2015).

No firm conclusions can be drawn regarding longer-term effectiveness of exercise training, with mixed findings noted by the four studies that provided this information. Interestingly, the large and significant increase in depression severity shown at a 26-week follow-up (Learmonth et al., 2011) is consistent with the results of a prior review in healthy adults, whereby depression exacerbated after ceasing an exercise program (Morgan et al., 2018). This individual finding may indicate that exercise does not promote lasting improvement and requires sustained participation to maintain depression benefits, especially considering the control condition reported an improvement in mood comparatively. It follows that the evidence found for the exercise-mediated psychological and hormone mechanisms for alleviating depression, may only result in short-term benefits (Amorosi, 2014; Antunes et al., 2016; Haller et al., 2018; Legrand, 2014; Pickett et al., 2012). Notably, the large and significant increase in depression seen at 26-weeks, reversed to a small improvement in depression at 52-weeks (Learmonth et al., 2011). Although this latter follow-up result was non-significant, it might indicate that depression and its symptoms can take time to alleviate due to the wide range of psychological, physical, and social contributors to low mood (Gay et al., 2010). Therefore, routine assessment of longer-term (i.e., beyond 12 months) depression outcomes is required in future research to draw firm conclusions regarding the longer-term effectiveness of exercise training and enable more specific treatment guidelines (e.g., how long the ‘exercise effect’ lasts before further treatment is required to sustain results).

Interventions providing long-term improvements will offer the greatest benefit to persons with MS as they navigate their lifelong condition, particularly if exercise completed in earlier stages of the disease can protect against depression in later stages when physical activity may not be feasible due to exacerbated disability. Treatment methods that are effective over time are also likely to appeal most to health providers when considering

implementation, not only for their low-maintenance nature, but also to minimise consultation frequency with patients (Kidd et al., 2017).

Moderators of Treatment Effectiveness

Between-group differences for aerobic and nonaerobic exercise were non-significant: the type of exercise did not affect the magnitude of depression reductions. Although, aerobic interventions were associated with significant and moderate depression reductions. This finding is consistent with prior research in the general population and may potentially be attributable to the 'runner's high' phenomenon (production of endorphin hormones) associated with aerobic exercise to alleviate depression (Béland et al., 2020; Morres et al., 2019).

In all, this subgroup analysis was likely underpowered, which may have impeded production of reliable results. Specifically, only three and eight studies were included in the nonaerobic and aerobic subgroups each respectively, which does not concord with recommendations of a minimum of 10 studies per subgroup from The Cochrane Collaboration (Higgins et al., 2019). Notably, one study that was excluded from this analysis, due to sample overlap (i.e., a shared control group), administered a nonaerobic yoga intervention and produced a very large, significant depression severity reduction (Razazian et al., 2015). Notably, this yoga intervention produced the largest individual effect of the study. There is, then, a need for further comparative research examining the relative benefits of each exercise type to better inform program design. In particular, it is clinically important to determine if other forms of nonaerobic exercise, such as yoga, are as effective as strenuous cardiovascular styles of aerobic endurance training. More specifically, lower impact training methods may appeal more to a typically sedentary population and enhance program adherence to ultimately improve mood (Martinsen et al., 1989). Future research might also

consider the use of multiple control groups, particularly if more than one exercise intervention is being examined, to avoid issues of sample dependency.

Exercise volume, operationalised as the total number of sessions over the course of a treatment, emerged as a key feature of program design that explained a significant portion of variance. That is, larger volumes of exercise training were associated with larger depression reductions and indicative of a 'dose-response' relationship (Wasfy & Baggish, 2016). This finding confirms preliminary evidence in the MS population (Herring et al., 2017). The two studies associated with the largest reductions in depression administered exercise thrice weekly for 24 to 45 sessions over the course of treatment (Petajan et al., 1996; Razazian et al., 2015). It follows that exercise prescription for persons with MS should involve larger volumes of exercise, over time, to optimise depression outcomes. Further controlled research is required to identify the optimal volume for aerobic and nonaerobic exercise to further tailor recommendations.

Baseline depression severity was also a crucial patient characteristic: higher scores at baseline were associated with larger immediate reductions in depression. Taken at face value, those experiencing more severe depression exhibit greater capacity for change. Comparatively, those with lower severity scores may have reduced capacity for improvement due to an already 'healthy' status. This finding is consistent with meta-analytic evidence for exercise in clinical samples (i.e., high levels of depression at baseline; Silveira et al., 2013). Although prior MS reviews suggest that baseline depression does not impact on resulting treatment change (Dalgas et al., 2015; Ensari et al., 2014; Herring et al., 2017), the present study included a higher relative proportion of candidates with increased capacity for change (i.e., those more severely depressed). For example, 49% of candidates in the present study were classed as 'depressed' at baseline according to depression measure cut-off scores,

compared to 41%, 15%, and 41% in prior MS reviews, each respectively (Dalgas et al., 2015; Ensari et al., 2014; Herring et al., 2017).

Although engagement with the interventions was generally high, as evidenced by the low overall rates of attrition, individual study results highlight a need to consider strategies to promote accessibility and potentially engagement. Namely, two studies reporting higher rates of withdrawal (i.e., > 20%) revealed reasons related to issues of access, such as lack of time to commit to the program, competing family commitments, or the classes being located too far away (Learmonth et al., 2011; Oken et al., 2004). Therefore, exercise could potentially be administered as a technology-based intervention, such as a smartphone or online app with supervised therapist sessions for live feedback, to increase the accessibility of guided and structured programs (i.e., avoid travel issues, increase choice of class times to work around commitments). The effectiveness of internet-based exercise programs, including mobile health apps is demonstrated in the general population (Jahangiry et al., 2017). More recently, use of the 'Zoom' app has become popular during the Covid-19 pandemic to deliver supervised and online group training interventions (Ng, 2020; Tison et al., 2020). This method of delivery therefore appears to be a worthwhile endeavor to consider long-term in the MS population.

Limitations

The methodological limitations of the current findings must be considered in the context of data sourcing and extraction. Firstly, the strict eligibility criteria may have inhibited the inclusion of all relevant studies, including that of grey literature. Furthermore, potential studies must have made explicit reference to 'multiple sclerosis' or 'exercise'—or their relevant synonyms—to be identified during the initial screening process. This approach enabled objective, replicable, and systematic screening. However, given the various terms used to describe specific exercise training methods, it is likely that relevant studies

administering exercise without making explicit reference to key terms were missed as a consequence (e.g., “aerobics” rather than “exercise”). In an attempt to mitigate this limitation, broad search terms and their synonyms were adopted; however, more specific exercise terms were unable to be covered in searches due to the vast number of different training styles.

The potential confounding role of sample characteristics also needs to be considered. For example, no study investigated MS patients with clinically confirmed diagnoses of MDD; rather, most patients falling in ‘severe’ symptom categories were excluded. This likely resulted in an underpowered sample of depressed candidates with a greater capacity for change (Ensari et al., 2014), which may have underestimated the already large effect magnitude identified. In addition, the proportion of variance explained by baseline depression may also have been underestimated. Future studies should include more participants with MDD, or those falling in severe symptom ranges to increase generalisability of the results to clinical cases of MS-related depression.

A further limitation was the failure of some studies to control for participant exercise levels prior to the intervention. Although the majority of the MS population is sedentary (Blikman et al., 2015), most studies included participants with mild to moderate disease severities of whom prior exercise activity was possible. For example, if participants were engaged in recent, prior exercise programs that were effective at reducing depression, additional improvements may not have been produced from the study interventions. Future RCTs should routinely consider prior activity levels when defining the eligible sample to enable more reliable treatment effect estimates.

The operationalisation of ‘exercise volume’ in this review might also have been problematic. Ideally, exercise dose is determined by three variables: duration, frequency, and intensity (Wasfy & Baggish,2016). Despite the ability of the present study to quantify

duration and frequency, the majority of studies did not formally report exercise intensity - which research has shown may influence depression reduction (Balchin et al., 2016). Therefore, future intervention research should make explicit all exercise dosage variables in order to provide more complete exercise recommendations.

It follows that there were limitations in the context of intervention delivery. In all, the majority of studies administered exercise as a group intervention. Despite the clear benefits for maximising resources (i.e., use of a therapist for one as opposed to many individual sessions), it is possible the socialisation afforded by group training may have confounded resulting depression scores, considering that social support is a variable also linked to improved depression (Cruwys et al., 2013). Social support may have augmented the large results found and promoted exercise adherence as a result (Çakit et al., 2010). Future RCTs should consider using attention controls or measures of social support to control for and enable investigation of the potential impact of socialisation (Bahmani et al., 2019; Razazian et al., 2015; Sutherland et al., 2001),

Lastly, there was also some variation in the post-test data relating to socialisation effects. Although most studies generated immediate effect estimates from mean differences between baseline and post-intervention depression scores, Sutherland and colleagues (2001) took measurements at baseline and week-8 of the 10-week program. This was an attempt to control for the negative effect on psychological health shown in prior research from anticipation of the end of the program and removal of social support from exercising in a group (Petajan et al., 1996). However, this may have led to an underestimated effect estimate—as reinforced by the trim-and-fill results indicating the pooled effect was underestimated—if the already discovered moderate reductions continued to improve in the remaining 2-weeks. Instead, future studies may consider referring participants to support groups upon nearing study completion in attempt to mitigate the social aspect of this concern.

Conclusion

The results of the present study provide evidence for the effectiveness of exercise training to improve depression symptom severity, at least in the short-term, among persons with MS. In particular, the effect of exercise is maximised by targeting those with higher baseline depression: a subgroup most prone to respond due to their increased capacity for change. In addition, improvements in depression severity are augmented by bolstering the session volume and duration; specifically, sessions as frequent as thrice weekly for 8-weeks reported the largest improvements. Future research can extend on the present findings by focusing on clinically depressed persons with MS, further comparing aerobic and nonaerobic exercise types, and including routine long-term assessments and measures of exercise intensity.

References

Note. References marked with an asterisk indicate studies included in the meta-analysis.

Agoritsas, T., Merglen, A., Courvoisier, D. S., Combescure, C., Garin, N., Perrier, A., & Perneger, T. V. (2012). Sensitivity and predictive value of 15 PubMed search strategies to answer clinical questions rated against full systematic reviews. *Journal of Medical Internet Research, 14*(3), e85.

*Ahmadi, A., Arastoo, A. A., Nikbakht, M., Zahednejad, S., & Rajabpur, M. (2013). Comparison of the effect of 8 weeks aerobic and yoga training on ambulatory function, fatigue and mood status in MS patients. *Iranian Red Crescent Medical Journal, 15*(6), 449-454.

Amato, M. P., Derfuss, T., Hemmer, B., Liblau, R., Montalban, X., Soelberg Sørensen, P., Miller, D. H., & 2016 ECTRIMS Focused Workshop Group. (2017). Environmental modifiable risk factors for multiple sclerosis: Report from the 2016 ECTRIMS focused workshop. *Multiple Sclerosis, 24*(5), 590-603.

American College of Sports Medicine. (2014). *ACSM's guidelines for exercise testing and prescription* (L. S. Pescatello, R. Arena, D. Riebe, & P. D. Thompson, Eds. 9 ed.). Wolters Kluwer; Lippincott Williams & Wilkins.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5 ed.).

Amorosi, M. (2014). Correlation between sport and depression. *Psychiatria Danubina, 26*, 208-210.

Amtmann, D., Kim, J., Chung, H., Bamer, A. M., Askew, R. L., Wu, S., Cook, K. F., & Johnson, K. L. (2014). Comparing CESD-10, PHQ-9, and PROMIS depression instruments in individuals with multiple sclerosis. *Rehabilitation Psychology, 59*(2), 220-229.

Antunes, H. K., Leite, G. S., Lee, K. S., Barreto, A. T., Santos, R. V., Souza, H., Tufik, S., & de Mello, M. T. (2016). Exercise deprivation increases negative mood in exercise-addicted subjects and modifies their biochemical markers. *Physiology & Behavior, 156*, 182-190.

Australian Bureau of Statistics. (2012). *Profiles of Disability, Australia, 2009* (Cat. no. 4429.0). ABS. https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/4429.0Main+Features100182009

*Bahmani, D. S., Razazian, N., Farnia, V., Alikhani, M., Tatari, F., & Brand, S. (2019). Compared to an active control condition, in persons with multiple sclerosis two different types of exercise training improved sleep and depression, but not fatigue, paresthesia, and intolerance of uncertainty. *Multiple Sclerosis and Related Disorders, 36*, 1-9.

Balchin, R., Linde, J., Blackhurst, D., Rauch, H. L., & Schönbächler, G. (2016). Sweating away depression? The impact of intensive exercise on depression. *Journal of Affective Disorders, 200*, 218-221.

Bahmani, D. S., Razazian, N., Motl, R. W., Farnia, V., Alikhani, M., Pühse, U., Gerber, M., & Brand, S. (2020). Physical activity interventions can improve emotion regulation and dimensions of empathy in persons with multiple sclerosis: An exploratory study. *Multiple Sclerosis and Related Disorders, 37*, 1-8.

Bech, P., Rasmussen, N. A., Olsen, L. R., Noerholm, V., & Abildgaardm, W. (2001). The sensitivity and specificity of the major depression inventory, using the present state examination as the index of diagnostic validity. *Journal of Affective Disorders, 66*(2-3), 159-164.

Bech, P., Timmerby, N., Martiny, K., Lunde, M., & Soendergaard, S. (2015). Psychometric evaluation of the major depression inventory (MDI) as depression severity scale using

the LEAD (longitudinal expert assessment of all data) as index of validity. *BMC Psychiatry, 15*, 1-7.

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). *Manual for the Beck depression inventory-II.* Harper and Row.

Beck, A. T., Steer, R. A., & Brown, G. K. (2000). *BDI-FastScreen for medical patients, manual.* The Psychological Corporation, A Harcourt Assessment Company.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Archives of General Psychiatry, 4*, 561-571.

Béland, M., Lavoie, K. L., Briand, S., White, U. J., Gemme, C., & Bacon, S. L. (2020). Aerobic exercise alleviates depressive symptoms in patients with a major non-communicable chronic disease: A systematic review and meta-analysis. *British Journal of Sports Medicine, 54*(5), 272-278.

Benedict, R. H., Fishman, I., McClellan, M. M., Bakshi, R., & Weinstock-Guttman, B. (2003). Validity of the Beck depression inventory-fast screen in multiple sclerosis. *Multiple Sclerosis, 9*(4), 393-396.

Björgvinsson, T., Kertz, S. J., Bigda-Peyton, J. S., McCoy, K. L., & Aderka, I. M. (2013). Psychometric properties of the CES-D-10 in a psychiatric sample. *Assessment, 20*(4), 429-436.

Blikman, L. J., van Meeteren, J., Horemans, H. L., Kortenhorst, I. C., Beckerman, H., Stam, H. J., & Bussmann, J. B. (2015). Is physical behavior affected in fatigued persons with multiple sclerosis? *Archives of Physical Medicine and Rehabilitation, 96*(1), 24-29.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2013). *Comprehensive Meta-Analysis.* In (Version 3.3.070) [Computer Software]. Biostat. https://www.meta-analysis.com/

Borg, G. (1998). *Borg's perceived exertion and pain scales*. Human Kinetics.

*Briken, S., Gold, S. M., Patra, S., Vettorazzi, E., Harbs, D., Tallner, A., Ketels, G., Schulz, K. H., & Heesen, C. (2014). Effects of exercise on fitness and cognition in progressive MS: a randomized, controlled pilot trial. *Multiple Sclerosis, 20*(3), 382-390.

Brown, J. C., Huedo-Medina, T. B., Pescatello, L. S., Ryan, S. M., Pescatello, S. M., Moker, E., LaCroix, J. M., Ferrer, R. A., & Johnson, B. T. (2012). The efficacy of exercise in reducing depressive symptoms among cancer survivors: A meta-analysis. *PloS One, 7*(1), e30955.

Browne, P., Chandraratna, D., Angood, C., Tremlett, H., Baker, C., Taylor, B. V., & Thompson, A. J. (2014). Atlas of multiple sclerosis 2013: A growing global problem with widespread inequity. *Neurology, 83*(11), 1022-1024.

Burns, P. B., Rohrich, R. J., & Chung, K. C. (2011). The levels of evidence and their role in evidence-based medicine. *Plastic and Reconstructive Surgery, 128*(1), 305-310.

*Çakit, B. D., Nacir, B., Genc, H., Saraçoglu, M., Karagöz, A., Erdem, H. R., & Ergün, U. (2010). Cycling progressive resistance training for people with multiple sclerosis. *American Journal of Physical Medicine & Rehabilitation, 89*(6), 446-457.

Campbell, J. A., Simpson, S., Jr, Ahmad, H., Taylor, B. V., van der Mei, I., & Palmer, A. J. (2019). Change in multiple sclerosis prevalence over time in Australia 2010-2017 utilising disease-modifying therapy prescription data. *Multiple Sclerosis,* 1-14.

Cetin, K., Johnson, K. L., Ehde, D. M., Kuehn, C. M., Amtmann, D., & Kraft, G. H. (2007). Antidepressant use in multiple sclerosis: Epidemiologic study of a large community sample. *Multiple Sclerosis, 13*(8), 1046-1053.

Chari, D. M. (2007). Remyelination in multiple sclerosis *International Review of Neurobiology, 79*, 589-620.

Choobforoushzadeh, A., Neshat-Doost, H. T., Molavi, H., & Abedi, M. R. (2015). Effect of neurofeedback training on depression and fatigue in patients with multiple sclerosis. *Applied Psychophysiology and Biofeedback, 40*(1), 1-8.

Chwastiak, L., Ehde, D. M., Gibbons, L. E., Sullivan, M., Bowen, J. D., & Kraft, G. H. (2002). Depressive symptoms and severity of illness in multiple sclerosis: Epidemiologic study of a large community sample. *The American Journal of Psychiatry, 159*(11), 1862-1868.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2 ed.). Erlbaum.

Compston, A., & Coles, A. (2008). Multiple sclerosis. *Lancet, 372*(9648), 1502-1517.

Confavreux, C., Aimard, G., & Devic, M. (1980). Course and prognosis of multiple sclerosis assessed by the computerized data processing of 349 patients. *Brain, 103*(2), 281-300.

Confavreux, C., Vukusic, S., Moreau, T., & Adeleine, P. (2000). Relapses and progression of disability in multiple sclerosis. *The New England Journal of Medicine, 343*(20), 1430-1438.

Craft, L. L., & Perna, F. M. (2004). The benefits of exercise for the clinically depressed. *Primary Care Companion to the Journal of Clinical Psychiatry, 6*(3), 104-111.

Cruwys, T., Dingle, G. A., Haslam, C., Haslam, S. A., Jetten, J., & Morton, T. A. (2013). Social group memberships protect against future depression, alleviate depression symptoms and prevent depression relapse. *Social Science & Medicine, 98*, 179-186.

Cumming, G. (2012). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge.

Curren, S. L., Andrykowski, M. A., & Studts, J. L. (1995). Short-form of the profile of moods states (POMS-SF): Psychometric information. *Psychological Assessment, 7*(1), 80-83.

da Costa, B. R., Hilfiker, R., & Egger, M. (2013). PEDro's bias: Summary quality scores should not be used in meta-analysis. *Journal of Clinical Epidemiology, 66*(1), 75-77.

*Dalgas, U., Stenager, E., Jakobsen, J., Petersen, T., Hansen, H. J., Knudsen, C., Overgaard, K., & Ingemann-Hansen, T. (2010). Fatigue, mood and quality of life improve in MS patients after progressive resistance training. *Multiple Sclerosis, 16*(4), 480-490.

Dalgas, U., Stenager, E., Sloth, M., & Stenager, E. (2015). The effect of exercise on depressive symptoms in multiple sclerosis based on a meta-analysis and critical review of the literature. *European Journal of Neurology*, *22*(3), 443-e434.

Dauwan, M., Begemann, M., Slot, M., Lee, E., Scheltens, P., & Sommer, I. (2019). Physical exercise improves quality of life, depressive symptoms, and cognition across chronic brain disorders: A transdiagnostic systematic review and meta-analysis of randomized controlled trials. *Journal of Neurology*, 1-25.

Dettmers, C., Sulzmann, M., Ruchay-Plössl, A., Gütler, R., & Vieten, M. (2009). Endurance exercise improves walking distance in MS patients with fatigue. *Acta neurologica Scandinavica*, *120*(4), 251-257.

Doyne, E. J., Ossip-Klein, D. J., Bowman, E. D., Osborn, K. M., McDougall-Wilson, I. B., & Neimeyer, R. A. (1987). Running versus weight lifting in the treatment of depression. *Journal of Consulting and Clinical Psychology, 55*(5), 748-754.

Duval, S. J., & Tweedie, R. L. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463.

Ellis, P. D. (2010). *The essential guide to effect sizes: Statistical power, meta-analyses and the interpretation of research results*. Cambridge University Press.

Ensari, I., Motl, R. W., & Pilutti, L. A. (2014). Exercise training improves depressive symptoms in people with multiple sclerosis: Results of a meta-analysis. *Journal of Psychosomatic Research, 76*(6), 465-471.

Estrada, E., Ferrer, E., & Pardo, A. (2019). Statistics for evaluating pre-post change: Relation between change in the distribution center and change in the individual scores. *Frontiers in Psychology, 9*, 1-12, Article 2696.

Feinstein, A., Magalhaes, S., Richard, J. F., Audet, B., & Moore, C. (2014). The link between multiple sclerosis and depression. *Nature Reviews Neurology, 10*(9), 507-517.

Fiest, K. M., Walker, J. R., Bernstein, C. N., Graff, L. A., Zarychanski, R., Abou-Setta, A. M., Patten, S. B., Sareen, J., Bolton, J. M., Marriott, J. J., Fisk, J. D., Singer, A., Marrie, R. A., & CIHR Team Defining the Burden and Managing the Effects of Psychiatric Comorbidity in Chronic Immunoinflammatory Disease. (2016). Systematic review and meta-analysis of interventions for depression and anxiety in persons with multiple sclerosis. *Multiple Sclerosis and Related Disorders*, *5*, 12-26.

Fischer, A., Fischer, M., Nicholls, R. A., Lau, S., Poettgen, J., Patas, K., Heesen, C., & Gold, S. M. (2015a). Diagnostic accuracy for major depression in multiple sclerosis using self-report questionnaires. *Brain and Behaviour*, *5*(9), 1-8.

Fischer, A., Schröder, J., Vettorazzi, E., Wolf, O. T., Pöttgen, J., Lau, S., Heesen, C., Moritz, S., & Gold, S. M. (2015b). An online programme to reduce depression in patients with multiple sclerosis: A randomised controlled trial. *The Lancet Psychiatry, 2*(3), 217-223.

Fisniku, L. K., Brex, P. A., Altmann, D. R., Miszkiel, K. A., Benton, C. E., Lanyon, R., Thompson, A. J., & Miller, D. H. (2008). Disability and T2 MRI lesions: A 20-year

follow-up of patients with relapse onset of multiple sclerosis. *Brain: A Journal of Neurology, 131*, 808-817.

*Fleming, K. M., Coote, S. B., & Herring, M. P. (2019). The feasibility of Pilates to improve symptoms of anxiety, depression, and fatigue among people with multiple sclerosis: An eight-week randomized controlled pilot trial. *Psychology of Sport and Exercise, 45*, 1-9.

Gay, M. C., Vrignaud, P., Garitte, C., & Meunier, C. (2010). Predictors of depression in multiple sclerosis patients. *Acta Neurologica Scandinavica, 121*(3), 161-170.

Gharakhanlou, R., Wesselmann, L., Rademacher, A., Lampit, A., Negaresh, R., Kaviani, M., Oberste, M., Motl, R. W., Sandroff, B. M., Bansi, J., Baker, J. S., Heesen, C., Zimmer, P., & Javelle, F. (2020). Exercise training and cognitive performance in persons with multiple sclerosis: A systematic review and multilevel meta-analysis of clinical trials. *Multiple Sclerosis*, 1-17.

Gibson, S. J. (1997). The measurement of mood states in older adults. *Journal of Gerontology: Psychological Sciences, 52B*, 167-174.

Goldman Consensus Group. (2005). The goldman consensus statement on depression in multiple sclerosis. *Multiple Sclerosis, 11*(3), 328–337.

Grazioli, E., Tranchita, E., Borriello, G., Cerulli, C., Minganti, C., & Parisi, A. (2019). The effects of concurrent resistance and aerobic exercise training on functional status in patients with multiple sclerosis. *Current Sports Medicine Reports, 18*(12), 452-457.

Habibzadeh, F., Habibzadeh, P., & Yadollahie, M. (2016). On determining the most appropriate test cut-off value: the case of tests with continuous results. *Biochemia Medica, 26*(3), 297-307.

Haller, N., Lorenz, S., Pfirrmann, D., Koch, C., Lieb, K., Dettweiler, U., Simon, P., & Jung, P. (2018). Individualized web-based exercise for the treatment of depression: Randomized controlled trial. *JMIR Mental Health, 5*(4), e10698.

Hamer, M., Stamatakis, E., & Steptoe, A. (2009). Dose-response relationship between physical activity and mental health: The Scottish Health Survey. *British Journal of Sports Medicine, 43*(14), 1111-1114.

Hariton, E., & Locascio, J. J. (2018). Randomised controlled trials - the gold standard for effectiveness research: Study design: randomised controlled trials. *BJOG: An International Journal of Obstetrics and Gynaecology, 125*(13), 1716.

Haussleiter, I. S., Brüne, M., & Juckel, G. (2009). Psychopathology in multiple sclerosis: Diagnosis, prevalence and treatment. *Therapeutic Advances in Neurological Disorders, 2*(1), 13-29.

*Hebert, J. R., Corboy, J. R., Manago, M. M., & Schenkman, M. (2011). Effects of vestibular rehabilitation on multiple sclerosis-related fatigue and upright postural control: A randomised controlled trial. *Physical Therapy, 91*(8), 1166-1183.

Hemmer, B., Kerschensteiner, M., & Korn, T. (2015). Role of the innate and adaptive immune responses in the course of multiple sclerosis. *The Lancet Neurology, 14*(4), 406-419.

Herring, M. P., Fleming, K. M., Hayes, S. P., Motl, R. W., & Coote, S. B. (2017). Moderators of exercise effects on depressive symptoms in multiple sclerosis: A meta-regression. *American Journal of Preventative Medicine, 53*(4), 503-518.

Hicks, S. D., Jacob, P., Perez, O., Baffuto, M., Gagnon, Z., & Middleton, F. A. (2019). The transcriptional signature of a runner's high. *Medicine and Science in Sports and Exercise, 51*(5), 970-978.

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2019). *Cochrane Handbook for Systematic Reviews of Interventions* (2 ed.). John Wiley & Sons.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557-560.

Hind, D., Kaklamanou, D., Beever, D., Webster, R., Lee, E., Barkham, M., & Cooper, C. (2016). The assessment of depression in people with multiple sclerosis: A systematic review of psychometric validation studies. *BMC Psychiatry, 16*, 1-18.

Hind, D., O'Cathain, A., Cooper, C. L., Parry, G. D., Isaac, C. L., Rose, A., Martin, L., & Sharrack, B. (2010). The acceptability of computerised cognitive behavioural therapy for the treatment of depression in people with chronic physical disease: A qualitative study of people with multiple sclerosis. *Psychology & Health, 25*(6), 699-712.

Hoffmann, T., Glasziou, P., Boutron, I., Milne, R., Perera, R., Moher, D., Altman, D., Barbour, V., Macdonald, H., Johnston, M., Lamb, S., Dixon-Woods, M., McCulloch, P., Wyatt, J., Chan, A., & Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *BMJ (Clinical research ed.), 348*(g1687), 1-12.

Høglund, R., A., & Maghazachi, A., A. . (2014). Multiple sclerosis and the role of immune cells. *World Journal of Experimental Medicine, 4*(3), 27-37.

Howard, J., Trevick, S., & Younger, D. S. (2016). Epidemiology of multiple sclerosis. *Neurologic Clinics, 34*(4), 919-939.

Hsu, C. S., Chang, S. T., Nfor, O. N., Lee, K. J., Lee, S. S., & Liaw, Y. P. (2019). Effects of regular aerobic exercise and resistance training on high-density lipoprotein cholesterol levels in Taiwanese adults. *International Journal of Environmental Research 16*(11), 1-8.

Jahangiry, L., Farhangi, M. A., Shab-Bidar, S., Rezaei, F., & Pashaei, T. (2017). Web-based physical activity interventions: A systematic review and meta-analysis of randomized controlled trials. *Public Health, 152*, 36-46.

Julian, L. J., & Mohr, D. C. (2006). Cognitive predictors of response to treatment for depression in multiple sclerosis. *The Journal of Neuropsychiatry and Clinical Neurosciences, 18*(3), 356-363.

Kalb, R., Feinstein, A., Rohrig, A., Sankary, L., & Willis, A. (2019). Depression and suicidality in Multiple Sclerosis: Red flags, management strategies, and ethical considerations. *Current Neurology and Neuroscience Reports, 19*(10), 1-8.

Kargiotis, O., Paschali, A., Messinis, L., & Papathanasopoulos, P. (2010). Quality of life in multiple sclerosis: Effects of current treatment options. *International Review of Psychiatry, 22*(1), 67-82.

Kidd, T., Carey, N., Mold, F., Westwood, S., Miklaucich, M., Konstantara, E., Sterr, A., & Cooke, D. (2017). A systematic review of the effectiveness of self-management interventions in people with multiple sclerosis at improving depression, anxiety and quality of life. *PloS One*, *12*(10), e0185931.

Kliem, S., Möble, T., Zenger, M., & Brähler, E. (2014). Reliability and validity of the Beck depression inventory-fast screen for medical patients in the general German population. *Journal of Affective Disorders, 156*, 236-239.

Klineova, S., & Lublin, F. D. (2018). Clinical course of multiple sclerosis. *Cold Spring Harbor Perspectives in Medicine, 8*(9), 1-12.

Koch, M. W., Glazenborg, A., Uyttenboogaart, M., Mostert, J., & De Keyser, J. (2011). Pharmacologic treatment of depression in multiple sclerosis. *The Cochrane Database of Systematic Reviews* (2), Article CD007295.

Kohout, F. J., Berkman, L. F., Evans, D. A., & Cornoni-Huntley, J. (1993). Two shorter forms of the CES-D depression symptoms index. *Journal of Aging & Health, 5*(2), 179-193.

Kurtzke, J. F. (1983). Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology, 33*(11), 1444–1452.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology, 4*(863), 1-12.

Lamoureux, B. E., Linardatos, E., Fresco, D. M., Bartko, D., Logue, E., & Milo, L. (2010). Using the QIDS-SR16 to identify major depressive disorder in primary care medical patients. *Behavior Therapy, 41*(3), 423-431.

Latimer-Cheung, A. E., Martin Ginis, K. A., Hicks, A. L., Motl, R. W., Pilutti, L. A., Duggan, M., Wheeler, G., Persad, R., & Smith, K. M. (2013). Development of evidence-informed physical activity guidelines for adults with multiple sclerosis. *Archives of Physical Medicine and Rehabilitation, 94*(9), 1829-1836.

Learmonth, Y. C. (2012). *Theapeutic exercise for those moderately affected with multiple sclerosis* [Doctoral Thesis, University of Glasgow]. http://theses.gla.ac.uk/3595/1/2012Learmonthphd.pdf

*Learmonth, Y. C., Paul, L., Miller, L., Mattison, P., & McFayden, A. K. (2011). The effects of a 12-week leisure centre-based, group exercise intervention for people moderately affected with multiple sclerosis: a randomized controlled pilot study. *Clinical Rehabilitation, 26*(7), 579-593.

Legrand, F. D. (2014). Effects of exercise on physical self-concept, global self-esteem, and depression in women of low socioeconomic status with elevated depressive symptoms. *Journal of Sport & Exercise Psychology, 36*(4), 357-365.

Lim, G., L. , Tam, W., W. , Lu, Y., Ho, C., S. , Zhang, M. W., & Ho, R. C. (2018). Prevalence of depression in the community from 30 countries between 1994 and 2014. *Scientific Reports 8*(1), 1-10.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications.

Lublin, F. D., Baier, M., & Cutter, G. (2003). Effect of relapses on development of residual deficit in multiple sclerosis. *Neurology, 61*(11), 1582-1532.

Lublin, F. D., & Reingold, S. C. (1996). Defining the clinical course of multiple sclerosis: results of an international survey. *Neurology, 46*(4), 907-911.

Lundby, C., Montero, D., & Joyner, M. (2017). Biology of VO2 max: Looking under the physiology lamp. *Acta Physiologica, 220*(2), 218-228.

Maher, C. G., Sherrington, C., Herbert, R. D., Moseley, A. M., & Elkins, M. (2003). Reliability of the PEDro scale for rating quality of randomized controlled trials. *Physical Therapy, 83*(8), 713-721.

Marrie, R. A., Elliott, L., Marriott, J., Cossoy, M., Blanchard, J., Leung, S., & Yu, N. (2015). Effect of comorbidity on mortality in multiple sclerosis. *Neurology, 85*(3), 240-247.

Marrie, R. A., Horwitz, R., Cutter, G., Tyry, T., Campagnolo, D., & Vollmer, T. (2009). The burden of mental comorbidity in multiple sclerosis: Frequent, underdiagnosed, and undertreated. *Multiple Sclerosis, 15*(3), 385-392.

Marrie, R. A., Zhang, L., Lix, L. M., Graff, L. A., Walker, J. R., Fisk, J. D., Patten, S. B., Hitchon, C. A., Bolton, J. M., Sareen, J., El-Gabalawy, R., Marriott, J. J. & Bernstein, C. N. (2018). The validity and reliability of screening measures for depression and anxiety disorders in multiple sclerosis. *Multiple Sclerosis and Related Disorders, 20*, 9-15.

Martinsen, E. W., Hoffart, A., & Solberg, O. (1989). Comparing aerobic with nonaerobic forms of exercise in the treatment of clinical depression: A randomized trial. *Comprehensive Psychiatry, 30*(4), 324-331.

McCabe, M. (2012). *A needs analysis of Australians with MS*. Multiple Sclerosis Research Australia. https://msra.org.au/wp-content/uploads/2016/03/National-MS-Needs-Analysis-2012-1.pdf

McCoy, E. C. (2017). Understanding the intention-to-treat principle in randomized controlled trials *Western Journal of Emergency Medicine, 18*(6), 1075-1078.

McDonald, W. I., Compston, A., Edan, G., Goodkin, D., Hartung, H. P., Lublin, F. D., McFarland, H. F., Paty, D. W., Polman, C. H., Reingold, S. C., Sandberg-Wollheim, M., Sibley, W., Thompson, A., van den Noort, S., Weinshenker, B. Y., & Wolinsky, J. S. (2001). Recommended diagnostic criteria for multiple sclerosis: Guidelines from the International Panel on the diagnosis of multiple sclerosis. *Annals of Neurology, 50*(1), 121-127.

McNair, D., Lorr, M., & Doppleman, L. (1971). *POMS manual for the profile of mood states*. Educational and Industrial Testing.

Meyer-Moock, S., Feng, Y. S., Maeurer, M., Dippel, F. W., & Kohlmann, T. (2014). Systematic literature review and validity evaluation of the expanded disability status scale (EDSS) and the multiple sclerosis functional composite (MSFC) in patients with multiple sclerosis. *BMC Neurology, 14*, 1-10.

*Miller, L., Paul, L., Mattison, P., & McFayden, A. K. (2011). Evaluation of a home-based physiotherapy programme for those with moderate to severe multiple sclerosis: A randomized controlled pilot study. *Clinical Rehabilitation, 25*(8), 720-730.

Milo, R., & Miller, A. (2014). Revised diagnostic criteria of multiple sclerosis. *Autoimmunity Reviews, 13*(4), 518-524.

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Open Medicine: A Peer-Reviewed, Independent, Open-Access Journal, 3*, e123.

Mohr, D. C., Hart, S. L., Fonareva, I., & Tasch, E. S. (2006). Treatment of depression for patients with multiple sclerosis in neurology clinics. *Multiple Sclerosis, 12*(2), 204-208.

Morgan, J. A., Olagunju, A. T., Corrigan, F., & Baune, B. T. (2018). Does ceasing exercise induce depressive symptoms? A systematic review of experimental trials including immunological and neurogenic markers. *Journal of Affective Disorders, 234,* 180-192.

Morres, I. D., Hatzigeorgiadis, A., Stathi, A., Comoutos, N., Arpin-Cribbie, C., Krommidas, C., & Theodorakis, Y. (2019). Aerobic exercise for adult patients with major depressive disorder in mental health services: A systematic review and meta-analysis. *Depression and Anxiety, 36*(1), 39-53.

Motl, R. W. (2014). Benefits, safety, and prescription of exercise in persons with multiple sclerosis. *Expert Review of Neurotherapeutics, 14*(12), 1429-1436.

Multiple Sclerosis Australia. (2017). *2017 federal pre-budget submission*. https://treasury.gov.au/sites/default/files/2019-03/C2016-052_Multiple-Sclerosis-Australia.pdf

Multiple Sclerosis Research Australia. (2018). *Health economic impact of multiple sclerosis in Australia in 2017*. https://msra.org.au/wp-content/uploads/2018/08/health-economic-impact-of-ms-in-australia-in-2017_ms-research-australia_web.pdf

Nebiker, L., Lichtenstein, E., Minghetti, A., Zahner, L., Gerber, M., Faude, O., & Donath, L. (2018). Moderating effects of exercise duration and intensity in neuromuscular vs.

endurance exercise interventions for the treatment of depression: A meta-analytical review. *Frontiers in Psychology, 9*, 1-18.

*Negaresh, R., Motl, R. W., Mokhtarzade, M., Ranjbar, R., Majdinasad, N., Khodadoost, M., Zimmer, P., Baker, J. S., & Patel, D. (2019). Effect of short-term interval exercise training on fatigue, depression, and fitness in normal weight vs. overweight person with multiple sclerosis. *Explore, 15*(2), 134-141.

Ng, K. (2020). Adapted physical activity through COVID-19 [Editorial]. *European Journal of Adapted Physical Activity, 13*, 1.

*Oken, B. S., Kishiyama, S., Zajdel, D., Bourdette, D., Carlsen, J., Haas, M., Hugos, C., Kraemer, D. F., Lawrence, J., & Mass, M. (2004). Randomized controlled trial of yoga and exercise in multiple sclerosis. *Neurology, 62*(11), 2058-2064.

Orwin, R. G. (1983). A fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics, 8*, 157-159.

Paul, L., Coulter, E. H., Miller, L., McFadyen, A., Dorfman, J., & Mattison, P. G. (2014). Web-based physiotherapy for people moderately affected with Multiple Sclerosis; Quantitative and qualitative data from a randomized, controlled pilot study. *Clinical Rehabilitation, 28*(9), 924-935.

*Petajan, J. H., Gappmaier, D., White, A. T., Spencer, M. K., Mino, L., & Hicks, R. W. (1996). Impact of aerobic training on fitness and quality of life in multiple sclerosis. *Annals of Neurology, 39*(4), 432-441.

Petajan, J. H., & White, A. T. (1999). Recommendations for physical activity in patients with multiple sclerosis *Sports Medicine 27*(3), 179-191.

Pickett, K., Yardley, L., & Kendrick, T. (2012). Physical activity and depression: A multiple mediation analysis. *Mental Health and Physical Activity, 5*(2), 125-134.

Pilutti, L. A., Platta, M. E., Motl, R. W., & Latimer-Cheung, A. E. (2014). The safety of exercise training in multiple sclerosis: A systematic review. *Journal of the Neurological Science, 343*(1-2), 3-7.

Poser, C. M., Paty, D. W., Scheinberg, L., McDonald, W. I., Davis, F. A., Ebers, G. C., Johnson, K. P., Sibley, W. A., Silberberg, D. H., & Tourtellotte, W. W. (1983). New diagnostic criteria for multiple sclerosis: Guidelines for research protocols. *Annals of Neurology, 13*(3), 227-231.

Raichlen, D. A., Foster, A. D., Gerdeman, G. L., Seillier, A., & Giuffrida, A. (2012). Wired to run: Exercise-induced endocannabinoid signaling in humans and cursorial mammals with implications for the 'runner's high'. *The Journal of Experimental Biology, 215,* 1331-1336.

Raskind, M. A. (2008). Diagnosis and treatment of depression comorbid with neurologic disorders. *American Journal of Medicine, 121*(11), S28-S38.

*Razazian, N., Yavari, Z., Farnia, V., Azizi, A., Kordavani, L., Bahmani, D. S., Holsboer-Trachsler, E., & Brand, S. (2015). Exercising impacts on fatigue, depression, and paresthesia in female patients with multiple sclerosis. *Medicine and Science in Sports and Exercise, 48*(5), 796-803.

Reilly, T. J., MacGillivray, S. A., Reid, I. C., & Cameron, I. M. (2015). Psychometric properties of the 16-item quick inventory of depressive symptomology: A systematic review and meta-analysis. *Journal of Psychiatric Research, 60*, 132-140.

Ribbons, K., Lea, R., Tiedeman, C., Mackenzie, L., & Lechner-Scott, J. (2017). Ongoing increase in incidence and prevalence of multiple sclerosis in Newcastle, Australia: A 50-year study. *Multiple Sclerosis, 23*(8), 1063-1071.

Romberg, A., Virtanen, A., & Ruutiainen, J. (2005). Long-term exercise improves function impairment but not quality of life in multiple sclerosis. *Journal of Neurology, 252*(7), 839-845.

Rovaris, M., Judica, E., Gallo, A., Benedetti, B., Sormani, M. P., Caputo, D., Ghezzi, A., Montanari, E., Bertolotto, A., Mancardi, G., Bergamaschi, R., Martinelli, V., Comi, G., & Filippi, M. (2006). Grey matter damage predicts the evolution of primary progressive multiple sclerosis at 5 years. *Brain: A Journal of Neurology, 129*, 2628-2634.

Rush, A. J., Carmody, T., & Reimitz, P. E. (2000). The inventory of depressive symptomatology (IDS): Clinician (IDS-C) and self-report (IDS-SR) ratings of depressive symptoms. *International Journal of Methods in Psychiatric Research, 9*, 45-59.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., Markowitz, J. C., Ninan, P. T., Kornstein, S., Manber, R., Thase, M. E., Kocsis, J. H., & Keller, M. B. (2003). The 16-item quick inventory of depressive symptomology (QIDS) clinician rating (QIDS-C) and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry, 54*, 573-583.

Sacco, R., Santangelo, G., Stamenova, S., Bisecco, A., Bonavita, S., Lavorgna, L., Trojano, L., D'Ambrosio, A., Tedeschi, G., & Gallo, A. (2016). Psychometric properties and validity of Beck depression inventory II in multiple sclerosis. *European Journal of Neurology, 23*(4), 744-750.

Sadovnick, A. D., Remick, R. A., Allen, J., Swartz, E., Yee, I. M., Eisen, K., Farquhar, R., Hashimoto, S. A., Hooge, J., Kastrukoff, L. F., Morrison, W., Nelson, J., Oger, J., & Paty, D. W. (1996). Depression and multiple sclerosis. *Neurology, 46*(3), 628-632.

Schuch, F. B., Dunn, A. L., Kanitz, A. C., Delevatti, R. S., & Fleck, M. P. (2016). Moderators of response in exercise treatment for depression: A systematic review. *Journal of Affective Disorders, 195*, 40-49.

Shacham, S. (1983). A shortened version of the profile of mood states. *Journal of Personality Assessment, 47*(3), 305-306.

Siegert, R. J., & Abernethy, D. A. (2005). Depression in multiple sclerosis: A review. *Journal of Neurology, Neurosurgery & Psychiatry 76*(4), 469-475.

Siengsukon, C. F., Aldughmi, M., Kahya, M., Bruce, J., Lynch, S., Ness Norouzinia, A., Glusman, M., & Billinger, S. (2016). Randomized controlled trial of exercise interventions to improve sleep quality and daytime sleepiness in individuals with multiple sclerosis: A pilot study. *Multiple Sclerosis Journal - Experimental, Translational and Clinical, 2*, 1-9.

Silveira, H., Moraes, H., Oliveira, N., Coutinho, E. S., Laks, J., & Deslandes, A. (2013). Physical exercise and clinically depressed patients: A systematic review and meta-analysis. *Neuropsychobiology, 67*(2), 61-68.

Simpson, S., Wang, W., Otahal, P., Blizzard, L., van der Mei, I. A. F., & Taylor, B. V. (2019). Latitude continues to be significantly associated with the prevalence of multiple sclerosis: An updated meta-analysis. *Journal of Neurology, Neurosurgery & Psychiatry, 90*(11), 1193-2000.

Skokou, M., Soubasi, E., & Gourzis, P. (2012). Depression in multiple sclerosis: A review of assessment and treatment approaches in adult and pediatric populations. *ISRN Neurology*, 1-6, Article 427102.

Solaro, C., Gamberini, G., & Masuccio, F. G. (2018). Depression in multiple sclerosis: Epidemiology, aetiology, diagnosis and treatment. *CNS Drugs, 32*(2), 117-133.

Stanton, R., & Reaburn, P. (2014). Exercise and the treatment of depression: A review of the exercise program variables. *Journal of Science and Medicine in Sport, 17*(2), 177-182.

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S. J., Altman, D. G., Moher, D., & Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Clinical research ed.), 343*, d4002.

Strober, L. B., & Arnett, P. A. (2015). Depression in multiple sclerosis: The utility of common self-report instruments and development of a disease-specific measure. *Journal of Clinical and Experimental Neuropsychology, 37*(7), 722-732.

*Sutherland, G., Andersen, M. B., & Stoové, M. A. (2001). Can aerobic exercise training affect health-related quality of life for people with multiple sclerosis? *Journal of Sport & Exercise Psychology, 23*, 122-135.

Temple, R., & Ellenberg, S. S. (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Annals of Internal Medicine, 133*(6), 455-463.

Thompson, A. J., Banwell, B. L., Barkhof, F., Carroll, W. M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M. S., Fujihara, K., Galetta, S. L., Hartung, H. P., Kappos, L., Lublin, F. D., Marrie, R. A., Miller, A. E., Miller, D. H., Montalban, X., Mowry, E. M., Soelberg Sorensen, P., Tintoré, M., Traboulsee, A. L., Trojano, M., Utidehaag, B. M. J., Vukusic, S., Waubant, E., Weinshenker, B. G., Reingold, S. C., & Cohen, J. A. (2018). Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology, 17*(2), 162-173.

Tison, G. H., Avram, R., Kuhar, P., Abreau, S., Marcus, G. M., Pletcher, M. J., & Olgin, J. E. (2020). Worldwide effect of COVID-19 on physical activity: A descriptive study. *Annals of Internal Medicine*, M20-2665.

*Tollár, J., Nagy, F., Tóth, B. E., Török, K., Szita, K., Csutorás, B., Moizs, M., & Hortobágyi, T. (2020). Exercise Effects on Multiple Sclerosis Quality of Life and Clinical-Motor Symptoms. *Medicine and Science in Sports and Exercise, 52*(5), 1007-1014.

Velikonja, O., Curić, K., Ozura, A., & Jazbec, S. S. (2010). Influence of sports climbing and yoga on spasticity, cognitive function, mood and fatigue in patients with multiple sclerosis. *Clinical Neurology and Neurosurgery, 112*(7), 597-601.

Verhagen, A. P., de Vet, H. C., de Bie, R. A., Kessels, A. G., Boers, M., Bouter, L. M., & Knipschild, P. G. (1998). The Delphi list: A criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *Journal of Clinical Epidemiology, 51*(12), 1235-1241.

Veritas Health Innovation. *Covidence systematic review software* In [Computer software]. https://www.covidence.org

Viera, A. J., & Garrett, J. M. (2005). Understanding interobserver agreement: The kappa statistic *Family Medicine 37*(5), 360-363.

Warren, S., & Warren, K. G. (2001). *Multiple sclerosis*. World Health Organisation.

Wasfy, M. M., & Baggish, A. L. (2016). Exercise dose in clinical practice. *Circulation, 133*(23), 2297-2313.

Watson, T. M., Ford, E., Worthington, E., & Lincoln, N. B. (2014). Validation of mood measures for people with multiple sclerosis. *International Journal of MS Care, 16*(2), 105-109.

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. *Acta Psychiatrica Scandinavica, 67*, 361-370.

Appendices

Appendix A: Logic Grids

Table A.1

Logic Grid for Cochrane Library

Exercise	*AND*	*Multiple Sclerosis*
exercise		multiple sclerosis
OR		OR
human physical conditioning		demyelinating autoimmune disease
OR		OR
physical activity		demyelinating disease
OR		OR
endurance training		demyelinating disorder
OR		OR
resistance training		disseminated sclerosis
OR		
high-intensity interval training		
OR		
athletic performance		
OR		
movement therapy		
OR		
occupational therapy		
OR		
physical exertion		
OR		
physiotherapy		
OR		
physical therapy		

Note. The ‘title, abstract, keyword’, ‘in trials’, and ‘word variation’ options were selected. No date or language filters were applied. Search yielded 1210 trials.

Table A.2

Logic Grid for Embase

Exercise	*AND*	*Multiple Sclerosis*
"exercise"/exp		"multiple sclerosis"/de
OR		OR
exercis*:ti,ab		"multiple sclerosis":ti,ab
OR		OR
"athletic performance":ti,ab		"disseminated sclerosis":ti,ab
OR		OR
"endurance training":ti,ab		demyelinat*:ti,ab
OR		OR
"high intensity interval training":ti,ab		"demyelinating disease"/de
OR		OR
"athletic training":ti,ab		"demyelination"/de
OR		
"movement therapy"/de		
OR		
"movement therapy":ti,ab		
OR		
"occupational therapy"/de		
OR		
"occupational therapy":ti,ab		
OR		
"physical activity"/exp		
OR		
"physical activity":ti,ab		
OR		
"physical exertion":ti,ab		
OR		
physiotherapy/exp		
OR		
physiotherapy:ti,ab		
OR		
plyometrics:ti,ab		
OR		
"resistance training":ti,ab		

Note. Under 'mapping' options, only 'map to preferred term in Emtree' was selected. No date or language filters were applied. Search yielded 8155 results.

Table A.3

Logic Grid for PEDro

Exercise	*AND*	*Multiple Sclerosis*
exercise		multiple sclerosis

Note. The advanced search option was used, and "exercise AND multiple sclerosis" was inserted into the 'Abstract and title' search option. The 'clinical trial' option was selected for the 'method' category. No date or language filters were applied. Search yielded 189 results.

Table A.4

Logic Grid for PsycINFO

Exercise	*AND*	*Multiple Sclerosis*
exp exercise		exp multiple sclerosis
OR		OR
exercis* mp		multiple sclerosis.mp
OR		OR
movement therap*.mp		disseminated sclerosis.mp
OR		OR
movement therapy.sh		demyelinating disorder*.mp
OR		OR
physical activit*.mp		demyelinating disease* mp
OR		OR
physical activity.sh		exp demyelination
OR		OR
athletic performance.sh		demyelinating autoimmune disease*.mp
OR		
athletic performance mp		
OR		
athletic training.mp		
OR		
athletic training.sh		
OR		
exp occupational therapy		
OR		
occupational therap*.mp		
OR		
physical therap*.mp		
OR		
physical therapy.sh		
OR		
exp physical endurance		
OR		
physical endurance mp		
OR		
physical exertion.mp		
OR		
exp physical therapy		
OR		
physical therap*.mp		
OR		
physiotherap*.mp		

Note. Two basic searches were conducted for each category and combined using the "AND" feature. The 'map search term to search heading' filter was selected, while 'produced the same results' was deselected. No date or language filters were applied. Search yielded 784 results.

Table A.5

Logic Grid for PubMed

Exercise	*AND*	*Multiple Sclerosis*
"exercise"[mh]		"multiple sclerosis"[mh]
OR		OR
exercis*[tiab]		multiple sclerosis[tiab]
OR		OR
physical conditioning[tiab]		"demyelinating diseases"[mh:noexp]
OR		OR
resistance training[tiab] OR		"demyelinating autoimmune diseases, CNS"[mh]
"athletic performance"[mh]		OR
OR		demyelinat*[tiab]
athletic performance[tiab]		OR
OR		demyelinating disorder[tiab]
physical endurance[tiab]		OR
OR		demyelinating disease[tiab]
physical activit*[tiab]		
OR		
"physical exertion"[mh]		
OR		
physical exertion[tiab]		
OR		
"recreation therapy"[mh]		
OR		
recreation therap*[tiab]		
OR		
"physical therapy modalities"[mh:noexp]		
OR		
"exercise therapy"[mh]		
OR		
physical therap*[tiab]		
OR		
endurance training[tiab]		
OR		
resistance training[tiab]		

Note. The advanced search option was used, applying the 'clinical trials' filter. No date or language filters were applied. Search yielded 470 results.

Appendix B: Depression Measures

Table B

Included Validated Depression Measures, Severity Classifications and Clinical Cut-Off Scores

Measure	Abbreviation	Range	Classification	Clinical cut-off
Beck Depression Inventory	BDI	0–63	0–9: Minimal 10–19: Mild 20–30: Moderate-severe ≥ 31: Severe	12
Beck Depression Inventory – Second Edition	BDI-II	0–63	0–13: Minimal 14–19: Mild 20–28: Moderate-severe 29–63: Severe	19
Beck Depression Inventory – Fast Screen for Medical Patients	BDI-FS	0-21	0–3: Minimal 4–6: Mild 7–9: Moderate-severe 10–21: Severe	4
Center for Epidemiologic Studies Depression Scale – Boston (short form)	CES-D-10	0–30	-	15
Hospital Anxiety and Depression Scale – Depression subscale	HADS-D	0–21	0–7: Normal 8–10: Borderline 11–21: Abnormal	11
Self-Rated Inventory of Depressive Symptoms	$IDS\text{-}SR_{30}$	0–84	0–13: None 14–25: Mild 26–38: Moderate 39–48: Severe ≥ 49: Very severe	28

Measure	Abbreviation	Range	Classification	Clinical cut-off
Major Depression Inventory	MDI	0–50	20–24: Mild 25–29: Moderate ≥ 30: Severe	21
Profile of Mood States – Depression subscale	POMS-D	0–60	-	-
Profile of Mood States Short Form – Depression Subscale	POMS-SF D	0–32	-	-
Self-Report Quick Inventory of Depressive Symptomology	QIDS-SR_{16}	0–27	≥ 5: None 6–10: Mild 11–15: Moderate 16–20: Severe ≥ 21: Very severe	13

Appendix C: PRISMA Checklist

Table C

Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Checklist (Moher et al., 2009)

Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	Title pg.
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	v
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	6-9
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	9-10
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	-
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	11-13
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	11, 18
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	64-67
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	11-13
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	13, 73

Section/topic	#	Checklist item	Reported on page #
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	14
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	14-15
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	15
Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	15-16
Risk of bias across studies	15	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).	16-17
Additional analyses	16	Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.	17
RESULTS			
Study selection	17	Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.	18-19
Study characteristics	18	For each study, present characteristics for which data were extracted (e.g., study size, PICOS, follow-up period) and provide the citations.	18, 20-24. 42-62
Risk of bias within studies	19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	24-25, 77
Results of individual studies	20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	26-31
Synthesis of results	21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.	26-31
Risk of bias across studies	22	Present results of any assessment of risk of bias across studies (see Item 15).	24-25, 77
Additional analysis	23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).	31-32
DISCUSSION			

Section/topic	#	Checklist item	Reported on page #
Summary of evidence	24	Summarize the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	33-38
Limitations	25	Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	38-40
Conclusions	26	Provide a general interpretation of the results in the context of other evidence, and implications for future research.	33-38, 40
FUNDING			
Funding	27	Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	N/A

Appendix D: Data Extraction Sheet

Title:

Lead author:

Year:

Study characteristics

Total N

Intention to treat:_____

Completer analysis:_____

Country: _____

Attrition: _____

Diagnosis method: _____

_____

_____

_____

_____

Intervention demographics

Gender (% or *n*)

Male: _____

Female: _____

Age

M: _____ *SD:* _____

Intervention MS characteristics

Time since diagnosis (years):

M: _____ *SD:* _____

Type: CIS, RR, P (SP, PP)

CIS (N):

RR (N):

P (N):

EDSS

M: _____ *SD:* _____

Exercise intervention:

Sessions (per week): _____

Session duration (mins): _____

Treatment length (weeks): _____

Mode: _____

_____

Activity level prior to study:

_____

_____

_____

_____

Intervention effect size data

Outcome: _____

Sample size: _____

Pre intervention outcome score:

M: _____ *SD:* _____

Post intervention outcome score:

M: _____ *SD:* _____

Other data (e.g. *g*, 95% CI's if mean difference scores unavailable): _____

_____

_____

_____

Control demographics

Control type: _____

Gender (% or *n*)

Male: _____

Female: _____

Age

M: _____ *SD:* _____

MS characteristics

Time since diagnosis (years):

M: _____ *SD:* _____

Type: CIS, RR, P (SP, PP)

CIS (N):

RR (N):

P (N):

EDSS

M: _____ *SD:* _____

Control effect size data

Outcome: _____

Sample size: _____

Pre intervention outcome score:

M: _____ *SD:* _____

Post intervention outcome score:

M: _____ *SD:* _____

Other data (e.g. *g*, 95% CI's if mean difference scores unavailable): _____

_____

_____

_____

Appendix E: PEDro Scale Criteria

PEDro scale

1.	eligibility criteria were specified	no ❑ yes ❑ where:
2.	subjects were randomly allocated to groups (in a crossover study, subjects were randomly allocated an order in which treatments were received)	no ❑ yes ❑ where:
3.	allocation was concealed	no ❑ yes ❑ where:
4.	the groups were similar at baseline regarding the most important prognostic indicators	no ❑ yes ❑ where:
5.	there was blinding of all subjects	no ❑ yes ❑ where:
6.	there was blinding of all therapists who administered the therapy	no ❑ yes ❑ where:
7.	there was blinding of all assessors who measured at least one key outcome	no ❑ yes ❑ where:
8.	measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups	no ❑ yes ❑ where:
9.	all subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome was analysed by "intention to treat"	no ❑ yes ❑ where:
10.	the results of between-group statistical comparisons are reported for at least one key outcome	no ❑ yes ❑ where:
11.	the study provides both point measures and measures of variability for at least one key outcome	no ❑ yes ❑ where:

The PEDro scale is based on the Delphi list developed by Verhagen and colleagues at the Department of Epidemiology, University of Maastricht (*Verhagen AP et al (1998). The Delphi list: a criteria list for quality assessment of randomised clinical trials for conducting systematic reviews developed by Delphi consensus. Journal of Clinical Epidemiology, 51(12):1235-41*). The list is based on "expert consensus" not, for the most part, on empirical data. Two additional items not on the Delphi list (PEDro scale items 8 and 10) have been included in the PEDro scale. As more empirical data comes to hand it may become possible to "weight" scale items so that the PEDro score reflects the importance of individual scale items.

The purpose of the PEDro scale is to help the users of the PEDro database rapidly identify which of the known or suspected randomised clinical trials (ie RCTs or CCTs) archived on the PEDro database are likely to be internally valid (criteria 2-9), and could have sufficient statistical information to make their results interpretable (criteria 10-11). An additional criterion (criterion 1) that relates to the external validity (or "generalisability" or "applicability" of the trial) has been retained so that the Delphi list is complete, but this criterion will not be used to calculate the PEDro score reported on the PEDro web site.

The PEDro scale should not be used as a measure of the "validity" of a study's conclusions. In particular, we caution users of the PEDro scale that studies which show significant treatment effects and which score highly on the PEDro scale do not necessarily provide evidence that the treatment is clinically useful. Additional considerations include whether the treatment effect was big enough to be clinically worthwhile, whether the positive effects of the treatment outweigh its negative effects, and the cost-effectiveness of the treatment. The scale should not be used to compare the "quality" of trials performed in different areas of therapy, primarily because it is not possible to satisfy all scale items in some areas of physiotherapy practice.

Notes on administration of the PEDro scale:

All criteria	**Points are only awarded when a criterion is clearly satisfied**. If on a literal reading of the trial report it is possible that a criterion was not satisfied, a point should not be awarded for that criterion.
Criterion 1	This criterion is satisfied if the report describes the source of subjects and a list of criteria used to determine who was eligible to participate in the study.
Criterion 2	A study is considered to have used random allocation if the report states that allocation was random. The precise method of randomisation need not be specified. Procedures such as coin-tossing and dice-rolling should be considered random. Quasi-randomisation allocation procedures such as allocation by hospital record number or birth date, or alternation, do not satisfy this criterion.
Criterion 3	*Concealed allocation* means that the person who determined if a subject was eligible for inclusion in the trial was unaware, when this decision was made, of which group the subject would be allocated to. A point is awarded for this criteria, even if it is not stated that allocation was concealed, when the report states that allocation was by sealed opaque envelopes or that allocation involved contacting the holder of the allocation schedule who was "off-site".
Criterion 4	At a minimum, in studies of therapeutic interventions, the report must describe at least one measure of the severity of the condition being treated and at least one (different) key outcome measure at baseline. The rater must be satisfied that the groups' outcomes would not be expected to differ, on the basis of baseline differences in prognostic variables alone, by a clinically significant amount. This criterion is satisfied even if only baseline data of study completers are presented.
Criteria 4, 7-11	*Key outcomes* are those outcomes which provide the primary measure of the effectiveness (or lack of effectiveness) of the therapy. In most studies, more than one variable is used as an outcome measure.
Criterion 5-7	*Blinding* means the person in question (subject, therapist or assessor) did not know which group the subject had been allocated to. In addition, subjects and therapists are only considered to be "blind" if it could be expected that they would have been unable to distinguish between the treatments applied to different groups. In trials in which key outcomes are self-reported (eg, visual analogue scale, pain diary), the assessor is considered to be blind if the subject was blind.
Criterion 8	This criterion is only satisfied if the report explicitly states *both* the number of subjects initially allocated to groups *and* the number of subjects from whom key outcome measures were obtained. In trials in which outcomes are measured at several points in time, a key outcome must have been measured in more than 85% of subjects at one of those points in time.
Criterion 9	An *intention to treat* analysis means that, where subjects did not receive treatment (or the control condition) as allocated, and where measures of outcomes were available, the analysis was performed as if subjects received the treatment (or control condition) they were allocated to. This criterion is satisfied, even if there is no mention of analysis by intention to treat, if the report explicitly states that all subjects received treatment or control conditions as allocated.
Criterion 10	A *between-group* statistical comparison involves statistical comparison of one group with another. Depending on the design of the study, this may involve comparison of two or more treatments, or comparison of treatment with a control condition. The analysis may be a simple comparison of outcomes measured after the treatment was administered, or a comparison of the change in one group with the change in another (when a factorial analysis of variance has been used to analyse the data, the latter is often reported as a group × time interaction). The comparison may be in the form hypothesis testing (which provides a "p" value, describing the probability that the groups differed only by chance) or in the form of an estimate (for example, the mean or median difference, or a difference in proportions, or number needed to treat, or a relative risk or hazard ratio) and its confidence interval.
Criterion 11	A *point measure* is a measure of the size of the treatment effect. The treatment effect may be described as a difference in group outcomes, or as the outcome in (each of) all groups. *Measures of variability* include standard deviations, standard errors, confidence intervals, interquartile ranges (or other quantile ranges), and ranges. Point measures and/or measures of variability may be provided graphically (for example, SDs may be given as error bars in a Figure) as long as it is clear what is being graphed (for example, as long as it is clear whether error bars represent SDs or SEs). Where outcomes are categorical, this criterion is considered to have been met if the number of subjects in each category is given for each group.

Extracted from Maher et al. (2003).

Appendix F: Included Study Details

Table F

Study Characteristics and Demographics

Lead author (year)	Country	Measure	*N*	Age (*years*) (*M* ± *SD*)	Gender *(% F)*	MS Subtype (% RRMS)	Disease Duration (years) *(M* ± *SD)*	EDSS *(M* ± *SD)*	Depressed (%)
Ahmadi (2013)	Iran	BDI	31	35.26 (9.01)	100.0	-	5.12 (4.09)	2.22 (1.16)	35.48
Bahmani (2019)	Iran	BDI-FS	47	37.93 (9.15)	100.0	-	7.07 (6.63)	2.24 (1.65)	100.0
Briken (2014)	Germany	IDS-SR_{30}	42	49.75 (7.83)	-	-	15.85 (7.34)	4.95 (.84)	0.0
Çakit (2010)	Turkey	BDI	30	35.95 (10.53)	-	-	7.9 (4.07)	-	100.0
Dalgas (2010)	Denmark	MDI	31	48.4 (9.29)	-	100.0	7.35 (5.9)	3.8 (.89)	0.0
Fleming (2019)	Ireland	1) HADS-D 2) QIDS 3) POMS-D	9	52.55 (6.8)	100.0	-	-	-	1) 33.33 2) 0.0 3) -
Hebert (2011)	US	BDI-II	26	46.4 (10.37)	84.62	88.46	7.1 (5.89)	-	0.0
Learmonth (2011)	Scotland	HADS-D	32	51.6 (7.91)	71.88	-	13.0 (6.97)	5.98 (.44)	0.0
Miller (2011)	Scotland	HADS-D	30	54.6 (7.83)	63.33	0.0	15.85 (8.95)	7.05 (.66)	0.0
Negaresh (2019)	Iran	BDI	61	31.12 (3.06)	65.57	100.0	7.33 (3.53)	1.6 (.93)	100.0
Oken (2004)	US	1) CES-D-10 2) POMS-D	57	49.0 (8.74)	92.98	-	-	3.07 (1.82)	1) 0.0 2) -
Petajan (1996)	US	POMS-D	46	40.05 (2.11)	67.39	-	7.75 (2.05)	3.35 (.54)	-
Razazian (2015)	Iran	BDI	54	33.94 (6.92)	100.0	66.67	6.93 (.82)	3.53 (1.09)	100.0
Sutherland (2001)	Australia	POMS-SF D	22	46.32 (4.87)	54.55	-	6.59 (4.61)	-	-
Tollár (2020)	Hungary	BDI	26	46.25 (6.34)	92.31	65.38	13.6 (4.21)	-	100.0
		Total	**544**	**43.94 (7.55)**	**82.77**	**73.68**	**9.34 (3.77)**	**3.78 (1.74)**	**48.74**

Note. *M* = mean; *SD* = standard deviation; F = female gender; RRMS = Relapsing-remitting MS; EDSS = Expanded Disability Status Scale; % Depressed = percentage of participants exceeding clinical cut-off score at baseline; (-) = data not provided. Depression measures: POMS-D = Profile of Mood States – Depression Subscale; IDS-SR_{30} = Self-Report Quick Inventory of Depression Symptomology; BDI = Beck Depression Inventory; BDI-FS = Beck Depression Inventory – Fast Screen for Medical Patients; MDI = Major Depression Inventory; POMS-SF-D = Profile of Mood States Short Form – Depression Subscale; HADS-D = Hospital Anxiety and Depression Scale – Depression Subscale; CES-D-10 = Centre for Epidemiologic Studies Depression Scale – Boston (Short form); BDI-II = Beck Depression Inventory – Second Edition; QIDS- SR_{16} = Quick Inventory of Depressive Symptomology.

Appendix G: Individual Risk of Bias Assessment

Table G

Reporting Quality of Included Studies Based on the PEDro Scale ($N_{studies}$ = 15)

Lead author (date)	1: Eligibility criteria	2: Random allocation	3: Concealed allocation	4: Baseline comparability	5: Blind subjects	6: Blind therapists	7: Blind assessors	8: Adequate follow-up	9: Intention-to-treat analysis	10: Between-group comparisons	11: Point estimates and variability
Ahmadi (2013)	1	1	0	1	0	0	0	1	1	1	1
Bahmani (2019)	1	1	1	1	0	0	1	0	0	1	1
Briken (2014)	1	1	1	1	0	0	0	1	0	1	1
Çakit (2010)	1	1	0	1	0	0	1	0	0	1	1
Dalgas (2010)	1	1	1	1	0	0	0	1	0	1	1
Fleming (2019)	1	1	1	1	0	0	0	0	0	0	1
Hebert (2011)	1	1	1	1	0	0	0	1	1	1	1
Learmonth (2011)	1	1	0	1	0	0	1	0	1	1	1
Miller (2011)	1	1	1	1	0	0	1	1	0	1	1
Negaresh (2019)	1	1	1	1	0	0	1	1	0	1	0
Oken (2004)	1	1	0	1	0	0	1	0	0	1	1
Petajan (1996)	1	1	0	1	0	0	0	1	0	1	1
Razazian (2015)	1	1	1	1	0	0	1	0	1	1	1
Sutherland (2001)	1	1	0	1	0	0	0	1	1	1	1
Tollár (2020)	1	1	1	1	0	0	1	1	0	1	1

Scoring note: 1 = criterion met, 0 = criterion not met.

Appendix H: Univariate Meta-Regression Scatterplots

Figure H.1

Univariate Meta-Regression to Examine Exercise Volume

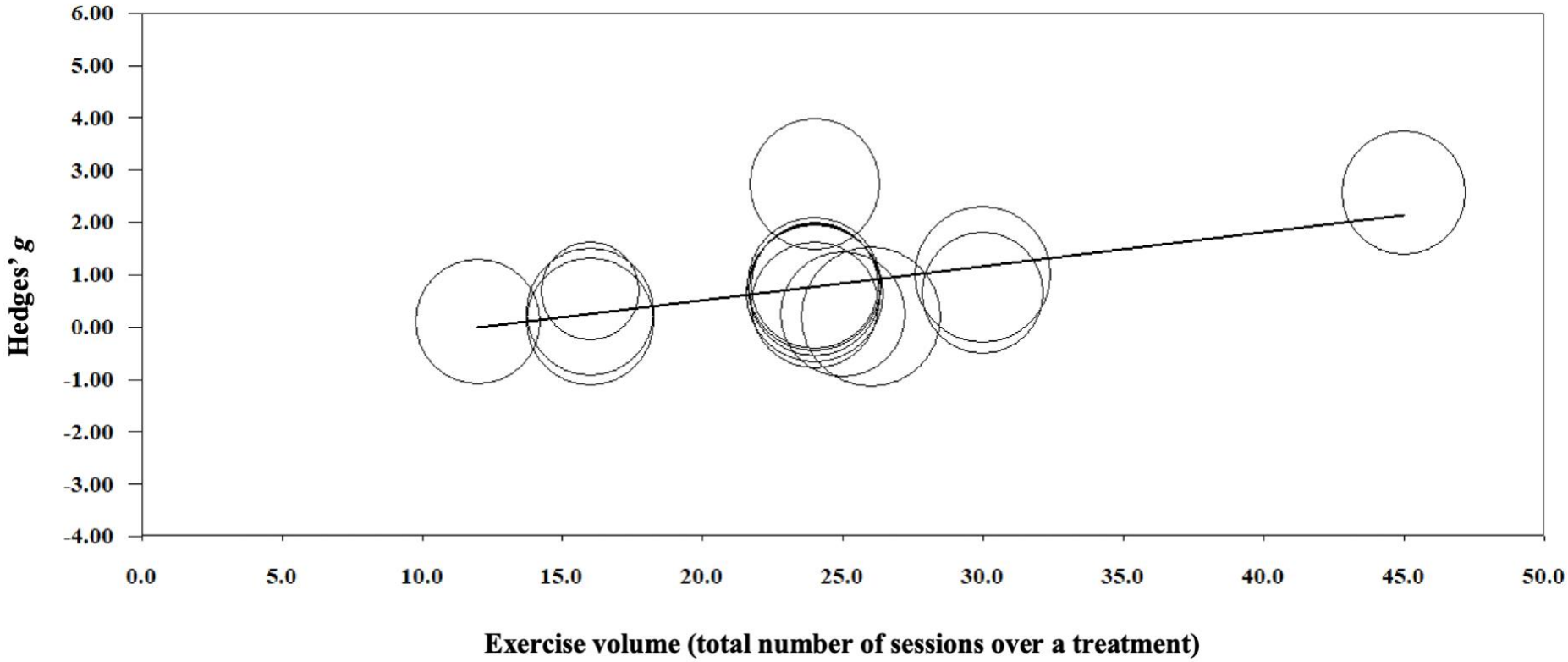


Figure H.2

Univariate Meta-Regression to Examine Baseline Depression Scores