The Effect of Exercise in Clinically Depressed Adults: Systematic Review and Meta-Analysis of Randomized Controlled Trials

Jesper Krogh, MD; Merete Nordentoft, MD; Jonathan A. C. Sterne, PhD; and Debbie A. Lawlor, PhD

Objective: To assess the effectiveness of exercise in adults with clinical depression.

Data Sources: The databases CINAHL, Embase, Cochrane Database of Systematic reviews, Cochrane Controlled Trials Register, MEDLINE, and PsycINFO were searched (1806–2008) using medical subject headings (MeSH) and text word terms *depression*, *depressive disorder* and *exercise*, *aerobic*, *non-aerobic*, *physical activity*, *physical fitness*, *walk**, *jog**, *run**, *bicycling*, *swim**, *strength*, and *resistance*.

Study Selection: Randomized trials including adults with clinical depression according to any diagnostic system were included.

Data Extraction: Two investigators evaluated trials using a prepiloted structured form.

Data Synthesis: Thirteen trials were identified that fulfilled the inclusion criteria. Eight had adequate allocation concealment, 6 had a blinded outcome, and 5 used intention-to-treat analyses. The pooled standardized mean difference (SMD) calculated using a random-effects model was −0.40 (95% CI, −0.66 to −0.14), with evidence of heterogeneity between trials ($I^2 = 57.2\%$, $P = .005$). There was an inverse association between duration of intervention and the magnitude of the association of exercise with depression ($P = .002$). No other characteristics were related to between-study heterogeneity. Pooled analysis of 5 trials with long-term follow-up (ie, that examined outcomes beyond the end of the intervention) suggested no long-term benefit (SMD, −0.01; 95% CI, −0.28 to 0.26), with no strong evidence of heterogeneity in this pooled analysis ($I^2 = 23.4\%$, $P = .27$). There was no strong statistical evidence for small study bias ($P > .27$). Only 3 studies were assessed as high quality (adequately concealed random allocation, blinded outcome assessment, and intention-to-treat analysis). When we pooled results from these, the estimated beneficial effect of exercise was more modest (SMD, −0.19; 95% CI, −0.70 to 0.31) than the pooled result for all 13 studies, with no strong evidence of benefit.

Conclusions: Our results suggest a short-term effect of exercise on depression: on average, depression scores 0.4 of a standard deviation lower in clinically depressed patients randomly assigned to an exercise intervention at the end of that intervention compared to those randomly assigned to a none exercise group. There is little evidence of a long-term beneficial effect of exercise in patients with clinical depression.

J Clin Psychiatry 2011;72(4):529–538

Submitted: *November 27, 2008; accepted October 5, 2009.*
Online ahead of print: *October 19, 2010 (doi:10.4088/JCP.08r04913blu).*
Corresponding author: *Jesper Krogh, MD, Psychiatric Center Bispebjerg, Bispebjerg University Hospital, DK-2400, Denmark (Jesper.Krogh@regionh.dk).*

The yearly incidence of depression in adults is estimated to be between 3% and 5%,[1–3] with a lifetime prevalence of 17% in Western societies.[4] The Global Burden of Disease report by the World Health Organization found that unipolar depressive disorders were the fourth leading cause of disease burden in terms of lost years of healthy life, and that major depression accounted for 12% of all total years lived with disability in 2000 worldwide.[5] Compliance with antidepressant treatment is poor, and, in clinical trials, the dropout rate is reported to be between 12%–40% within the initial 6 to 8 weeks of treatment.[6,7] This has resulted in an interest in the use and evaluation of alternative or complementary therapies, with exercise in particular being the subject of a number of randomized controlled trials to test its effectiveness as a treatment for patients with depression.

There are a number of biologically plausible reasons why exercise might be an effective antidepressant. Experiments on animal models suggest an increase in neurogenesis[8] and an increased serotonergic drive[9] in response to exercise and that these effects result in an antidepressant action. Nonbiologic pathways have also been proposed: depressed patients taking regular exercise might get positive feedback from other people (particularly in societies where being physically active is seen as a virtue) and thus increased self-esteem,[10] or exercise might act as a diversion from negative thoughts.[11]

Since 2001, five systematic reviews and meta-analyses of the effect of exercise as an antidepressant have been published.[12–16] Two of these, 1 of trials in general adult study populations[12] and the other of trials in depressed patients aged over 60 years,[13] concluded that it was impossible to determine whether exercise was an effective antidepressant because of methodological weaknesses of the available trials. The recent Cochrane systematic review[15] found a moderate nonsignificant effect of exercise when analyses were restricted to the methodologically robust trials. The inclusion criteria in all 5 previous reviews may have limited their clinical usefulness. Lawlor and Hopker,[12] Rethorst et al,[16] and the recent Cochrane review[15] had broad inclusion criteria, meaning that a number of included trials were on volunteers (not recruited through clinical settings) who were defined as being depressed on the basis of cutoff scores in self-administered psychometric testing (eg, Beck Depression Inventory, Center for Epidemiologic Studies Depression Scale) as opposed to individuals with a diagnosis of depression obtained after presenting to clinical services. Since these studies do not necessarily reflect the clinical situation in which clinicians might consider "prescribing" exercise, it is important to perform a meta-analysis of exercise and depression, limited to studies in which depression was diagnosed by a health professional

in a clinical setting. This is the only way to address the question of whether "prescribing exercise" in clinical practice is effective. The review by Sjosten and Kivela[13] was of older adults (> 60 years) only, and a recent review by Stathopoulou et al[14] included in one meta-analysis participants with mixed psychiatric diagnosis,[17] a trial with exercise as part of a multiintervention program,[18] a nonrandomized trial,[19] and a trial comparing different exercise intensities.[10] Thus, to date it remains unclear whether exercise is an effective antidepressant in the general population of adults who are diagnosed with depression by trained health professionals.

In order to determine whether health services should provide exercise as a treatment for patients who are diagnosed with depression, we have undertaken a systematic review and meta-analysis of the effectiveness of exercise in adults diagnosed with depression in a clinical setting.

METHOD

Data Sources

We searched MEDLINE (1966–2008), Embase (1980–2008), PsycINFO (1806–2008), CINAHL (1982–2008), Cochrane Controlled Trials Register, and the Cochrane Database of Systematic reviews using medical subject headings (MeSH) and text word terms *depression*, *depressive disorder* and *exercise*, *aerobic*, *non-aerobic*, *physical activity*, *physical fitness*, *walk**, *jog**, *run**, *bicycling*, *swim**, *strength*, and *resistance*. We looked through reviews and contacted authors in the field for knowledge of additional trials. We also searched trial registers at the following Web sites to identify unpublished trials: www.controlled-trials.com and www.clinicaltrials.gov. Furthermore, we hand searched content lists of volumes published between June and September 2008 in the following journals: *Journal of the American Medical Association*, *BMJ*, *Lancet*, *American Journal of Preventive Medicine*, *New England Journal of Medicine*, *Archives of Internal Medicine*, *British Journal of Psychiatry*, *Comprehensive Psychiatry*, *Journal of the Royal Society of Medicine*, *Acta Psychiatrica Scandinavica*, and *British Journal of Sports Medicine*. The main search for trials was completed between April and September 2006 and was updated for the last time on September 12, 2008.

Study Selection

Only randomized controlled trials containing participants with clinical depression as the primary disease were included. A trial was defined as a randomized controlled trial if the allocation of participants to intervention groups was described as randomized (including terms such as *randomly*, *random*, and *randomization*) and compared exercise with "no treatment" or a control group (controls on a waiting list; placebo intervention; or when exercise was an adjunct, with both treatment and control groups receiving an identical established treatment).

Because of our aim to identify trials that had recruited participants in clinical practice, we initially planned to include only trials of participants recruited after a diagnosis of depression by a trained health worker in a clinical setting. However, our initial search revealed that only 4 trials fulfilled these criteria. We therefore relaxed our inclusion criteria to include trials if the participants were diagnosed as having depression according to a diagnostic system (eg, *International Classification of Diseases, Tenth Revision*; Research Diagnostic Criteria), even if this system had been applied to volunteers who were not seen in a health care setting or were referred from doctors who had made a diagnosis of depression. Studies had to include participants who were aged 18 years or above and had depression/depressive symptoms (assessed by any means) as an outcome measure, and we included trials published in any language. We excluded studies without a nonexercise control group and those that measured outcomes immediately before and after a single exercise session. Based on titles and abstracts, one reviewer (J.K.) retrieved potentially relevant studies. Two reviewers (J.K. and M.N., see Figure 1) then determined whether a particular study fulfilled the inclusion criteria. None of the reviewers were blinded to the names of authors, institutions, or journals.

Data Extraction

Two authors independently extracted data (quality criteria, participant details, intervention details, outcome measures, baseline and postintervention results, and main conclusion) using a prepiloted structured form. Any discrepancies in the data extraction were resolved by referring to the original articles and discussion with all authors of this article. All of the trials fulfilling our inclusion criteria measured depressive symptoms on a continuous scale at the end of the intervention period, with most referring to this as the primary outcome. Because the scales used for these continuously measured symptom outcomes varied from one trial to another and were in line with the 3 previous meta-analyses, we used the mean standardized difference in depression symptoms measured on a continuous scale as our primary outcome, and we were able to include all eligible trials in the primary meta-analysis. Some studies assessed this using more than 1 standard tool. For studies that included results from several tools for assessing depressive symptoms, we used the one that the authors described as their primary outcome. If the authors did not clearly state which was the primary outcome, the outcome reported first in the abstract was taken to be the primary outcome.

Because of our original aim of wanting to examine the effects of exercise prescription to participants diagnosed in a clinical setting, we conducted 2 separate primary meta-analyses: one including the small number of trials that fitted these strict criteria (ie, patients recruited from a clinical setting) and the second including all trials identified with our more relaxed criteria (ie, trials that recruited volunteers who were diagnosed with depression using diagnostic criteria and trials from clinical settings). In addition to these primary analyses we also pooled results from those studies that provided data on 2 secondary outcomes: (1) remission, ie, a binary outcome of the proportion of participants in each arm of the trial who were defined as being free of depressive

© Copyright 2010 Physicians Postgraduate Press, Inc.

symptoms and (2) long-term results, defined as outcomes measured at some time after completion of the exercise intervention.

Study Quality

We assessed the quality of trials by noting whether the following aspects of the trial had been addressed in the report: sample size calculation, allocation concealment, intention-to-treat analysis, blinding, and interrater reliability for outcomes that were not self-report. The sample size calculation was considered adequate if the authors clearly stated the minimum effect that they considered to be of clinical importance and the required sample size to achieve this effect at a given power and statistical test (α) level. Concealment of allocation was considered to be adequate if it had been achieved by any of the following methods: central randomization at a site remote from the study; computerized allocation in which records are in a locked, unreadable file that could be assessed only after entering patient details; and the drawing of sealed and opaque sequentially numbered envelopes. Studies that did not use these methods for random allocation or that did not describe how randomization was achieved (including lack of response from contacted authors to clarify the method of randomization) were defined as not adequately concealed. We defined trials as having used intention-to-treat analysis if all the patients were analyzed in the groups to which they were randomly allocated. If only those who started treatment or only those who completed treatment were included in the analysis, we defined the study as not using intention-to-treat analysis. The outcome assessment was defined as blinded when it was undertaken by an assessor who was unaware of the treatment allocation; if the assessment was done by the patients or an assessor who was aware of which intervention the patient was receiving, the assessment was defined as unblind. Interrater reliability was considered adequate if there was only 1 outcome assessor or, in the situation of more than 1 outcome assessor, if an interrater-reliability calculation was reported and was considered inadequate if several investigators assessed outcomes on different participants but no interrater-reliability calculation was provided in the article or obtained from the author.

Contact With Authors

We contacted authors by e-mail or post (sending 3 reminders to nonresponders) to establish missing details in the methods and results sections of the written reports, and to determine the authors' knowledge of or involvement in any current work in the area.

Statistical Analysis

A number of different psychometric instruments were used to measure depression at the outcome assessment in different studies. In order to be able to include all of the studies in our meta-analysis, we estimated a standardized mean difference (SMD) for each individual study. This is the mean difference in depression score between the exercise and control groups divided by the pooled standard deviation (of the distribution of the score used in the study). The result is an effect size on the standard deviation scale. By convention, effect sizes of 0.2–< 0.5, 0.5–< 0.8, and ≥ 0.8 are considered small, medium, and large, respectively. We used Hedges' g[20] to calculate the SMD in each study because this method includes a correction for small study size. The Hedges' g can be interpreted as a conservative estimate of the Cohen d method for estimating SMD. For the dichotomous outcome (remission), we pooled the odds ratios from each study.

We anticipated that systematic differences between studies (heterogeneity) would be likely because trials differed in the type and intensity of exercise used in the intervention, and we anticipated that trials would differ in methodological quality. We therefore used a random-effects model to calculate the overall pooled effect size.[21] The extent of heterogeneity between studies was determined by calculating I^2, which is a measure of the percentage of total variation across studies that is due to heterogeneity between studies rather than sampling variation.[22] We used meta-regression analysis to explore the possible effect of exercise type (aerobic or nonaerobic), exercise context (group or alone), duration of exercise intervention (in weeks), adherence to exercise intervention (% adhering), control group (no treatment, placebo, standard treatment, or some other intervention), and each of the quality indicators as characteristics that might explain any heterogeneity (difference in results between studies). A previous meta-analysis[12] reported that type of publication (abstract only versus journal article or thesis) was related to heterogeneity. In our systematic review, none of the included studies were published only as abstracts and we therefore did not explore publication type as a source of heterogeneity. In order to address multiple testing of study characteristics that may be related to heterogeneity, we used the Monte Carlo permutation test (using 1,000 permutations) to estimate P values in the meta-regression analysis.[23]

For trials that had more than 1 intervention group, we decided a priori to compare the group with the "strongest dose" of exercise to the control group so that we would not try to minimize any effect a priori. Two trials had several intervention groups defined by increasing intensity or increased demand of total energy expenditure.[24,25] In the meta-analysis, we included the comparison of the group allocated to the highest intensity or energy expenditure with the control group. One trial contained both a supervised and a home-based exercise intervention, and we included the supervised since we believed it to result in the highest intensity.[26] Three trials had 2 interventions of different exercise types (eg, aerobic and nonaerobic) as well as a control group. Since different types of exercise do not equate to different doses, for these trials, we randomly selected which of the 2 types of exercise to include in the main meta-analysis. This resulted in selection of the nonaerobic exercise intervention from Mutrie et al[27] and the aerobic exercise intervention in Doyne et al[28] and Krogh et al.[29] In sensitivity analysis, we repeated the meta-analysis using different intervention types from these studies. The results from these sensitivity analyses do not differ from those presented here. We used

STATA (version 9.2) (StataCorp LP, College Station, Texas) statistical software for all analyses.

RESULTS

Study Inclusion

Figure 1 summarizes the process of inclusion of the studies for review and analysis. Sixteen references[24–39] reporting 13 trials fulfilled the broad inclusion criteria, and these are described in Table 1. Of these 13 trials, 4 were considered to fulfill our strict criteria of diagnosis via health care settings.[27,29–31] Of these only 2 primarily recruited from primary care,[27,29] and the other 2 recruited participants from inpatient or outpatient psychiatric services.[30,31] Five of the 13 trials fulfilling our broad criteria had long-term follow-up, which we defined as follow-up that extended beyond the end of the period of the exercise intervention.[29,35–38] The 13 trials included in our review and meta-analysis provided data from 687 patients who were randomly assigned to either an exercise intervention (as monotherapy or as an augmentation) or a nonexercise control group and were included in the final analyses published in the study.

Through contact with authors, we were able to get additional information on quality issues and data needed for pooling from 5 trials included in this review.[26,30–32,38]

Interventions

In 9 trials, the exercise intervention was aerobic,[24,26,28,30–32,34,37,40] and in 3 it was nonaerobic.[25,27,39] One intervention was described as a mixed aerobic/nonaerobic[38] intervention. The median number of exercise sessions per week was 3 (range, 2–5), and the median duration of the intervention was 10 weeks (range, 8–16). In 9 trials, the exercise intervention was a group exercise,[25,26,29–32,34,38,39] and in 4 cases the exercise was on individual basis.[24,27,28,37]

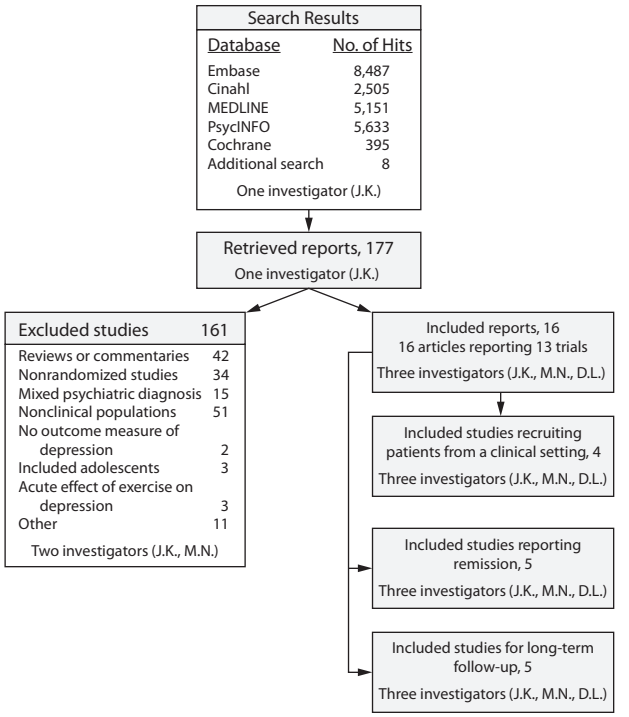
Quality Assessment

Allocation concealment was adequate in 8 of the 13 trials,[24–26,29–31,38,39] the outcome assessment was blinded in 6 trials[24–26,29,32,38] (Singh et al[39] did include a blinded outcome assessment, but their predefined primary outcome was not blinded.), and 5 trials had used intention-to-treat analysis.[24,26,29,32,39] Of the 7 trials for which the primary outcome was not blind, all outcome measures were based on self-report by the patients.[27,28,30,31,34,37,39] Sample size calculations were reported in 6 trials.[24–26,29,38,39] In general, those trials that did well on 1 quality criterion also did well on other criteria, and quality of trials was better in more recently published trials than in earlier published trials (Table 1).

Study Populations

All but 1 study[30] included patients recruited from a community setting, including community volunteers or patients from primary care or psychiatric services in the community. All but 1 study[30] included patients diagnosed with mild to moderate depression. The one exception was a study of patients examined in hospital with moderate to severe depression.[30] One study had the patients referred and diagnosed by general practitioners but did not report a specific diagnostic system.[27] The percentage of females in most trials was greater than that of males.

Figure 1. Study Selection Process[a]



[a]One reviewer examined titles and abstracts from the search result and removed obviously irrelevant reports. Two reviewers then examined full text reports to determine compliance with inclusion criteria. Three reviewers extracted data from included reports.

Exercise Compared With Placebo or as an Adjunct to Treatment

The meta-analysis including only the 4 studies recruited from a clinical setting[27,29–31] resulted in a pooled SMD of −0.47 ([95% CI, −1.13 to 0.18], $I^2 = 79.0\%$, P value for heterogeneity = .003). The pooled estimate for these 4 trials is very similar to that for all 13 trials included in our broader inclusion criteria meta-analysis, and, as far as could be ascertained, causes of heterogeneity were similar between these 4 and all 13 (discussed in full below).

Figure 2 shows the meta-analysis of all 13 studies with our main outcome—ie, depressive symptoms measured on a continuous scale at the end of the duration of the intervention. The pooled SMD, calculated using the random-effects model with Hedges' correction for small trials, was −0.40 (95% CI, −0.66 to −0.14). Thus, on average, depression scores are 0.4 of a standard deviation lower in depressed patients randomly assigned to an exercise intervention at the end of that intervention compared to those randomly assigned to a none exercise group. There was evidence of heterogeneity between the studies ($P = .005$). The I^2 value of 57.2% suggests that just over half of the variation across studies is due to the differences in effect between studies as opposed to chance

© Copyright 2010 Physicians Postgraduate Press, Inc.

Table 1. Characteristics, Results, and Methodological Quality of the Studies Fulfilling Inclusion Criteria for Systematic Review

	Participants				Intervention				Methodological Quality				
Study	n[a]	Type	Age, Mean, y	Baseline Depression	Type	Duration, wk	Adherence, %	Results, Mean Difference (range)	Allocation Concealment	Blind	Intention-to-Treat Analysis	Sample Size Calculation	Interrater Reliability
Klein et al,[38] 1985	22	Community volunteers; depression diagnosed by RDC	30.1 (SD = 6.7)	SCL-D: mean = 2.4; SD = 0.6	Aerobic exercise: Supervised individual running, 2 sessions/wk (n = 14) Control group: Meditation in groups once/wk (n = 8)	12	55.6	SCL-D, 0.2 (−0.5 to 0.9)	No	No	No	No	N/A
Martinsen et al,[31] 1985	43	Psychiatric inpatients; depression diagnosed by *DSM-III*	40 (range, 17–60)	BDI: mean = 28.0; SD = 9.4	Aerobic exercise: Supervised group exercise for 1 hr, 3 sessions/wk at 50%–70% of maximum aerobic capacity (n = 24) Control group: Occupational therapy, 3 sessions/wk (n = 19)	9	85.7	BDI, −10.7 (−16.3 to 5.1)	Yes	No	No	No	N/A
Epstein,[35] 1986	17	Community volunteers; depression diagnosed by *DSM-III*	39.4 (range, 24–60)	BDI: mean = 23.4; SD = 6.90	Aerobic exercise: Supervised group running/walking, 3–5 sessions/wk for 30 min (n = 7) Control group: Waiting list (n = 10)	8	Not provided	BDI, −7.3 (−16.0 to 1.4)	No	No	No	No	N/A
Doyne et al,[29] 1987	19	Community volunteers; depression diagnosed by RDC	28.5 (SD = 4.4)	BDI: mean = 21.1; SD = 6.56	Aerobic exercise: Supervised individual running/walking at 80% of maximal heart rate, 4 sessions/wk (n = 8) Control group: Waiting list (n = 11)	8	60	BDI, −7.1 (−12.3 to −1.9)	No	No	No	No	N/A
Mutrie,[28] 1988	15	Patients diagnosed and referred from general practitioners	42.1 (not provided)	BDI: mean = 22.7; SD = 4.9	Nonaerobic exercise: Supervised individual strength training, 20 min, 3 sessions/wk (n = 8) Control group: Waiting list (n = 7)	4	66	BDI, −11.9 (−16.7 to −7.1)	No	No	No	No	N/A
Veale et al,[32] 1992	75	Outpatients from psychiatric services; depression diagnosed by ClS	35.5 (range, 19–58)	BDI: mean = 24.5; SD = 6.0	Aerobic exercise: Supervised group running, 3 sessions/wk plus routine care (n = 36) Control group: Routine care (n = 29)	12	75	BDI, −3.9 (−9.6 to 1.9)	Yes	No	No	No	N/A
Singh et al,[40] 1997	32	Volunteers from register of people interested in research; depression diagnosed by *DSM-IV*	71 (range, 60–84)	BDI: mean = 19.9; SD = 2.3	Nonaerobic exercise: Supervised progressive resistance training in groups at 80% of repetition maximum; 8 repetitions, 3 sets—3 sessions/wk (n = 17) Control group: Attended seminars on health (n = 15)	10	93	BDI, −4.0 (−10.2 to 2.2)	Yes	No	Yes	Yes	N/A

(continued)

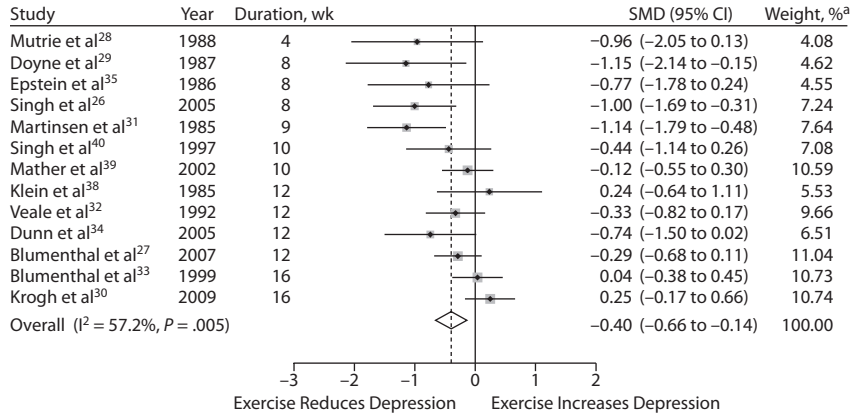
Table 1 (continued). Characteristics, Results, and Methodological Quality of the Studies Fulfilling Inclusion Criteria for Systematic Review

	Participants				Intervention				Methodological Quality				
Study	n[a]	Type	Age, Mean, y	Baseline Depression	Type	Duration, wk	Adherence, %	Results, Mean Difference (range)	Allocation Concealment	Blind	Intention-to-Treat Analysis	Sample Size Calculation	Interrater Reliability
Mather et al,[39] 2002	85	Treatment-resistant community volunteers; depression diagnosed by *ICD-10*	64.9 (range, 53–91)	HDRS-17: mean = 17.1; SD = 6.3	Exercise: Mixed aerobic and nonaerobic supervised group exercise of 45 min, 2 sessions/wk (n = 43) Control group: Attended seminars on health (n = 43)	10	100	HDRS-17, −1.1 (−3.9 to 1.6)	Yes	Yes	No	Yes	No
Dunn et al,[34] 2005	29	Community volunteers; depression diagnosed by *DSM-IV*	35.9 (SD = 6.4)	HDRS-17: mean = 19.7; SD = 2.3	Aerobic exercise: Individual supervised exercise; exercise amount estimated to 17.5 kcal/kg/wk, 5 sessions/wk (n = 16) Control group: Flexibility exercise 3 sessions/wk (n = 13)	12	72	HDRS-17, −4.0 (−7.8 to −0.2)	Yes	Yes	Yes	Yes	No
Singh et al,[26] 2005	37	Community volunteers; depression diagnosed by *DSM-IV*	69 (range, 60–85)	HDRS-17: mean = 18.9; SD = 4.2	Nonaerobic exercise: Supervised progressive resistance training in groups; 80% of repetition maximum, 3 sets, 8 repetitions—3 sessions/ wk (n = 18) Control group: Standard treatment by general practitioner (n = 19)	8	90	HDRS-17, −5.9 (−9.6 to −2.2)	Yes	Yes	No	Yes	Yes
Blumenthal et al,[27] 2007	100	Community volunteers; depression diagnosed by *DSM-IV*	52 (SD = 7.5)	HDRS-17: mean = 16.8; SD = 4.2	Aerobic exercise: Supervised group exercise; training at 70%–80% of maximum heart rate reserve in 30 min, 3 sessions/ wk (n = 51) Control group: Placebo pills provided daily; patients were attended by a psychiatrist 6 times during the trial (n = 49)	12	82	HDRS-17, −1.1 (−3.6 to 1.4)	Yes	Yes	Yes	Yes	No
Krogh et al,[30] 2009	110	Patients from general practice; depression diagnosed by *ICD-10*	37.7	HDRS-17: mean = 17.6; SD = 3.9	Aerobic exercise: Supervised group exercise for 60 min working at 70%–89% of maximal heart rate reserve at 2 sessions/wk Control group: Supervised group exercise for 60 min at 2 sessions/wk; relaxation and light physical exercises	16	42	HDRS-17, 0.4 (−2.0 to 2.9)	Yes	Yes	Yes	Yes	Yes

[a]Number of participants included in the final analysis of the article and used in our meta-analysis.

Abbreviations: BDI = Beck Depression Inventory; CIS = Clinical Interview Schedule; *DSM-IV* = *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition; *DSM-III*=*Diagnostic and Statistical Manual of Mental Disorders*, Third Edition; HDRS-17 = 17-item Hamilton Depression Rating Scale; *ICD-10* = *International Classification of Diseases*, *Tenth Revision*; N/A = not applicable; RDC = Research Diagnostic Criteria; SCL-D = Symptom Check List-depression subscale; SSRI = selective serotonin reuptake inhibitor.

Figure 2. Meta-Analysis of Randomized Controlled Trials Assessing the Effect of Exercise in Patients With Clinically Diagnosed Depression



[a]Weights are from random-effects analysis.
Abbreviation: SMD = standardized mean difference.

Table 2. Difference in Depression Scores at Long-Term Follow-Up Between Exercise Group and Control Group in Randomized Controlled Trials of the Effect of Exercise in Patients With Clinically Diagnosed Depression[a]

Study	Included in Postintervention Analysis, n	Time of Follow-Up From Baseline, Mo	Outcome Mean Difference Between Exercise and Control Groups (95% CI)
Klein et al,[38] 1985	22	9	−0.04 (−0.7 to 0.8)[b]
Singh et al,[36] 2001	32	26	−1.4 (−7.5 to 4.7)[c]
Babyak et al,[37] 2000	89	10	−0.7 (−2.9 to 0.6)[d]
Mather et al,[39] 2002	85	9	−2.2 (−0.6 to 4.9)[d]
Krogh et al,[30] 2009	110	12	0.6 (−1.9 to 3.1)[d]

[a]A negative outcome should be interpreted as a reduction in depression scores. For details of exercise interventions, control groups, and quality of trials, see Table 1; note that Babyak et al[37] report the long-term outcomes of Blumenthal et al.[33]
[b]Symptom Check List—depression subscale.
[c]Beck Depression Inventory.
[d]Hamilton Depression Rating Scale.

due to sampling variation. There was little evidence that differences in effect between studies was explained by any of the indicators of study quality (all *P* values > .18, from permutation tests), by adherence to exercise intervention (*P* = .08), by exercise type (*P* = .25), by whether the participants exercised alone or in a group (*P* = .37), or by the nature of the treatment provided to those in the control group (*P* = .15). However, the duration of the intervention appeared to be related to differences in effect size between studies, such that studies of longer duration tended to have weaker effects (ratio of SMD per 1 week longer duration: 0.12 [95% CI, 0.05 to 0.18], *P* value from permutation test = .002). This can be seen in Figure 2, in which the individual studies are ordered from top to bottom by duration (shortest to longest). The pooled SMD for the 5 studies of less than 10 weeks duration suggested a mean difference of 1 standard deviation (−1.03 [95% CI, −1.40 to −0.66], $I^2 = 0\%$, *P* value for heterogeneity = .98), whereas there was little evidence of exercise having an effect in studies of 10 or more weeks duration (7 studies, pooled SMD: −0.12 [95% CI, −0.30 to 0.05], $I^2 = 21.3\%$, *P* value for heterogeneity = .26).

In a post hoc analysis, requested by the reviewers, we restricted pooling of studies to the 3 trials assessed as having adequate allocation concealment, blinded assessment of the outcome, and intention-to-treat analysis.[24,26,29] The estimated beneficial effect of exercise was more modest (SMD: −0.19 [95% CI, −0.70 to 0.31], $I^2 = 68.0\%$) than for the pooled result for all 13 studies (SMD: −0.40), with no strong evidence of benefit.

Long-Term Effect of Exercise

Only 5 of the 13 studies had long-term follow-up of the participants to examine the effect of the exercise intervention after its completion. These are described in Table 2. The pooled analysis of these studies (Figure 3) suggested that exercise had little effect on depression scores in patients with depression in the longer term beyond cessation of the exercise program: pooled SMD was −0.01 (95% CI, −0.28 to 0.26), $I^2 = 23.4\%$, *P* value for heterogeneity = .27.

The Effect on Remission

Five of the included trials included a dichotomous measure of remission as an outcome representing the random assignation and evaluation of 340 patients.[24,26,28,29,32] The way in which remission was defined and the results from these 5 studies are presented in Table 3. All studies examined remission at the end of the exercise intervention. Figure 4 shows the pooled analyses of these studies. The pooled odds ratio for remission at the very end of the exercise intervention comparing those randomly assigned to exercise with those in the control arm and using a random-effects model was 1.31 (95% CI, 0.63 to 2.71), with I^2 estimated to 53.5% (*P* = .07). As with the associations for the continuous measure of depression symptoms, the effect of exercise on remission varied by duration of the exercise intervention (*P* = .03), such that studies with longer duration found little effect of exercise on remission (Figure 4).

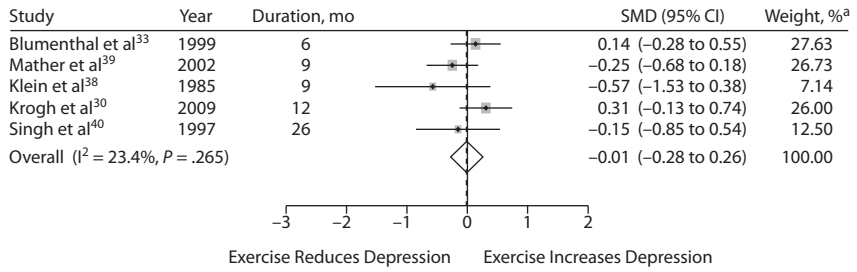
We found little evidence of small study bias, often indicative of publication bias, in either of our meta-analyses (*P* > .27 for both Begg and Egger tests).

DISCUSSION

The results of this systematic review and meta-analysis suggest that exercise at most has a small benefit in relieving symptoms of depression in patients with clinically diagnosed depression in the short term, based on the SMD of −0.4, which is within the range considered to represent a small effect (0.2 to 0.5). Furthermore, we found no evidence that this small effect lasted beyond the duration of the exercise

© Copyright 2010 Physicians Postgraduate Press, Inc.

Figure 3. Meta-Analysis of Randomized Controlled Trials Assessing the Long-Term (beyond the end of the exercise intervention) Effect of Exercise in Patients With Clinically Diagnosed Depression



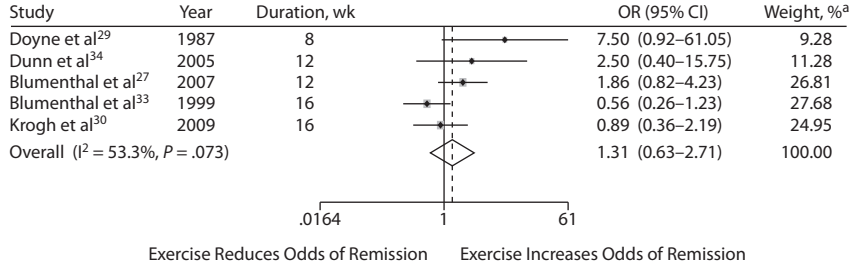
[a]Weights are from random-effects analysis.
Abbreviation: SMD = standardized mean difference.

Table 3. Data From Systematic Review Fulfilling Inclusion Criteria and Reporting Remission Status[a]

Study	Definition of Remission	Remission, n/n: Aerobic Exercise Group	Remission, n/n: Control Group	Odds Ratio (95% CI)	*P* Value
Doyne et al,[29] 1987	BDI score < 9	5/8	2/11	7.5 (0.9–61.0)	.06
Blumenthal et al,[33] 1999	No longer met *DSM-IV* criteria for major depression	36/55	33/48	0.9 (0.4–2.0)	.7
Dunn et al,[34] 2005	HDRS-17 score < 8	5/16	2/13	2.5 (0.4–15.7)	.3
Blumenthal et al,[27] 2007	No longer met *DSM-IV* criteria for major depression and HDRS-17 score < 8	23/51	15/49	1.9 (0.8–4.2)	.2
Krogh et al,[30] 2009	No longer met *ICD-10* criteria for depression and HDRS-17 score < 8	14/48	13/41	0.9 (0.4–2.1)	.8

[a]For details of exercise interventions, control groups, and quality of trials, see Table 1.
Abbreviations: BDI = Beck Depression Inventory; *DSM-IV* = *Diagnostic and Statistical Manual of Mental Disorders*, Fourth Edition; HRSD-17 = 17-item Hamilton Depression Rating Scale; *ICD-10* = *International Classification of Diseases, Tenth Revision*.

Figure 4. Meta-Analysis of Randomized Controlled Trials Assessing the Odds of Remission Comparing Participants Randomized to Exercise or Not in Patients With Clinically Diagnosed Depression



[a]Weights are from random-effects analysis.
Abbreviations: OR = odds ratio, SMD = standardized mean difference.

program. The pooled estimate suggested that those allocated to exercise had 31% greater odds of being in remission at the end of the exercise intervention compared to those in the control arm, but only 5 studies examined this outcome and the estimate was therefore imprecise, with wide confidence intervals that included the null value. Most trials examined only the effect of exercise as an antidepressant up until the end of the exercise intervention and therefore our main analysis focused on outcomes measured at that stage. Within the 13 trials that fulfilled our inclusion criteria there was evidence that those with a shorter intervention (less than 10 weeks) had a stronger beneficial effect than those trials in which the duration of the intervention was longer. In the 5 trials of 10 or more weeks' duration, patients randomly assigned to the exercise intervention had similar scores on depression symptoms as those allocated to control groups.

Our initial aim was to complete a systematic review and meta-analysis that could contribute to answering the important question of whether prescribing exercise when patients present to clinical services and are diagnosed with depression is beneficial. It was disappointing that only 4 trials were identified that recruited participants from health care settings and of these 4, only 2 were in primary care. However, results from the meta-analysis of these 4 trials did not differ from those found in all 13 trials, which included these 4 as well as any trials of community volunteers in whom depression was diagnosed with a clinical diagnostic system rather than a depression symptom score such as the Beck Depression Inventory.

While our review suggests that trial quality in this area of research is improving, with more recent trials generally being of higher quality, it also highlights the need for more high quality trials in this area. When we limited our pooling to trials that had adequately concealed allocation, blinding of outcome assessment, and intention-to-treat analysis, several issues emerge. First, this restriction resulted in just 3 of the 13 trials being included in the meta-analysis. Second, the point estimate from the pooling of these 3 trials suggested no important clinical benefit of exercise in patients with depression, with the point estimate being below the lower threshold of the range of SMD defined as *small effect*. However, the pooled estimate was imprecise, with 95% confidence intervals consistent with a possible small beneficial or detrimental effect. Third, with only 3 trials, it is clearly impossible to examine issues related to the effect of different types of exercise and duration of exercise intervention in high-quality trials.

The reduced effect in trials of longer duration might suggest that any effect of exercise is largely placebo in nature, since placebo effects tend to diminish with time. The weakening of effect with longer duration could point to plausible mechanisms. For example, if the main pathway by which

exercise exerts its effect is related to increased self-esteem due to doing something that in many cultures is seen as virtuous or to increased socializing, then the initial effect of these might be anticipated to wane over time. It is possible that with longer interventions adherence to the exercise program decreases with time and that the effect of exercise is related to adherence. In our meta-regression analyses, the percentage of participants who reported adhering to the exercise program was not related to heterogeneity between studies. However, adherence was defined differently in the various trials, and it is possible that participant report of adherence is measured with error. It is also possible that the association of duration of intervention with size of the effect of exercise on depression in these trials is explained by other features of the study design, for example, trial quality, that are the real drivers of this difference in effect. However, there was no association between trial duration and indicators of quality in the trials, and quality indicators did not explain heterogeneity between studies. Our post hoc subgroup analysis, including only those trials with adequate concealment of allocation, blinding of outcome assessment, and intention-to-treat analysis, included too few studies (only 3 of the 13) to adequately address the influence of quality on the results.

In the 5 trials that assessed the long-term effect of random allocation to exercise beyond the end of the exercise intervention, depression scores were the same in those who had been allocated to exercise and those who had been allocated to control interventions. Taken together these findings suggest that any beneficial effect of exercise as an antidepressant is small, short-lived, and does not extend beyond the end of the exercise intervention. Given the plausible mechanisms that have been suggested for exercise having a beneficial effect on depression (ie, increased levels of endorphins and neutrophins and increased self-esteem, achievement, and socializing), one would not anticipate long-term effects beyond the period of exercise intervention, unless one assumed that, after the intervention, patients were motivated to continue to exercise. Our results suggest that this does not occur, and it is important for clinicians to recognize this since the implication is that any potential benefit of an exercise intervention will only be maintained if the patient continues to adhere to the program after the intervention/prescription.

It could be argued that a more appropriate outcome would be to compare the proportion of individuals who are in remission between intervention groups at the end of the trial. Such studies would require considerably larger sample sizes because of the lower level of statistical power with a given sample size for a binary outcome compared to a continuous outcome, and this may be the reason why only 5 of the 13 trials identified by our search present this binary outcome. The pooled odds ratio for remission suggested some positive effect, but, even after pooling these 5 studies, the estimate was imprecise, with a very wide 95% confidence interval, and consistent with the null hypothesis. This result highlights the need for considerably larger trials in this area that have adequate statistical power to determine the effect of exercise on remission in the short and the longer term. As with our pooled analyses for the continuously measured outcome, studies with longer duration showed very little benefit of exercise on remission.

Quality of Studies

In some of the trials, random allocation was not adequately concealed, intention-to-treat analysis was not conducted, and/or the outcome assessment was not blinded to which treatment the patient had received. These biases in general will tend to result in an exaggeration of the true effect. However, it is important to recognize the special case of blinding in this context. It is increasingly recognized that patient-reported outcomes are important in any test of effectiveness of a treatment. Since it is impossible to blind a patient when the intervention is exercise, any patient-reported outcomes will, by definition, be unblinded. Ideally, future studies should determine both patient and clinician outcomes, with the clinician clearly blinded. One would not want to omit important patient perceptions on the basis of inability to blind them to their intervention. Our systematic review and meta-analysis included 5 new trials,[24–26,29,38] with an additional 361 participants included for analysis, that have been published since the earlier meta-analysis by Lawlor and Hopker.[12] Of note, these more recent trials, in general, were of better quality than earlier published trials.

Strength and Limitations of This Systematic Review and Meta-Analysis

The strength of this systematic review is the inclusion of trials that included only patients diagnosed with clinical depression and also included studies of adults of all ages. The advantage of this approach over previous reviews is that it provides information on whether it is appropriate to "prescribe" exercise to adults who are diagnosed with depression. We employed an extensive search strategy that should have identified all relevant trials. However, only 13 trials comprising results from 687 participants fulfilled our inclusion criteria, and these small numbers limit the power of our meta-regression analysis. We were careful to use a permutation test to determine the *P* values for the meta-regression analysis, which takes multiple testing into account.

CONCLUSION

Our results suggest that exercise interventions may have a small short-term antidepressant effect, with no effect seen for exercise interventions that continue for 10 weeks or more and no long-term effect beyond the end of the exercise intervention. Thus, the available evidence does not support the use of exercise for long-term benefit in patients with clinically diagnosed depression. Both the small number of trials in this area and the small number of participants included in each trial mean that very large, high quality trials, with long-term follow-up, are required to be confident about whether exercise has an important antidepressant effect. Most important, there is a clear need for trials to be conducted in health

care settings where exercise might be plausibly prescribed to patients and for those trials to have adequate statistical power to determine differences in the proportions of individuals who have sustained remission. Additional small trials in this area are unlikely to be able to address these issues.

Author affiliations: Psychiatric Center Bispebjerg, Faculty of Health Sciences, Copenhagen University Hospital, Copenhagen, Denmark (Drs Krogh and Nordentoft); and Department of Social Medicine, University of Bristol, Bristol, United Kingdom (Mr Sterne and Ms Lawlor).
Potential conflicts of interest: The authors report no financial or other relationship relevant to the subject of this article.
Funding/support: Dr Krogh was supported by a grant from AstraZeneca Denmark (£1,500) to cover fees and related travel expenses to attend a course at the University of Bristol on systematic review and meta-analysis and is funded by Assurance and Pension, which is an organization promoting Danish insurance companies. Ms Lawlor was funded by a UK Department of Health Career Scientist Award during the time that this work was undertaken.
Role of sponsors: The sponsors had no role in the design and conduct of the analysis and interpretation contained in this systematic review, which is the independent work of the authors.
Acknowledgments: The authors thank James Blumenthal, Michael Babyak, Egil Martinsen, David Veale, and Marion McMurdo for providing additional information from their studies that enabled them to complete the meta-analysis.

REFERENCES

1. Andrews G, Henderson S, Hall W. Prevalence, comorbidity, disability and service utilization: overview of the Australian National Mental Health Survey. *Br J Psychiatry*. 2001;178(2):145–153.
2. Blazer DG, Kessler RC, McGonagle KA, et al. The prevalence and distribution of major depression in a national community sample: the National Comorbidity Survey. *Am J Psychiatry*. 1994;151(7):979–986.
3. Spaner D, Bland RC, Newman SC. Epidemiology of psychiatric disorders in Edmonton: major depressive disorder. *Acta Psychiatr Scand Suppl*. 1994;89(s376):7–15.
4. Lépine JP, Gastpar M, Mendlewicz J, et al. Depression in the community: the first pan-European study DEPRES (Depression Research in European Society). *Int Clin Psychopharmacol*. 1997;12(1):19–29.
5. Ustün TB, Ayuso-Mateos JL, Chatterji S, et al. Global burden of depressive disorders in the year 2000. *Br J Psychiatry*. 2004;184(5):386–392.
6. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. *Lancet*. 2009;373(9665):746–758.
7. Kirsch I, Deacon BJ, Huedo-Medina TB, et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med*. 2008;5(2):e45.
8. Bjørnebekk A, Mathé AA, Brené S. The antidepressant effect of running is associated with increased hippocampal cell proliferation. *Int J Neuropsychopharmacol*. 2005;8(3):357–368.
9. Gomez-Merino D, Béquet F, Berthelot M, et al. Site-dependent effects of an acute intensive exercise on extracellular 5-HT and 5-HIAA levels in rat brain. *Neurosci Lett*. 2001;301(2):143–146.
10. Bosscher R. Running and mixed physical exercise with depressed psychiatric patients. *Int J Sport Psychol*. 1993;24:170–184.
11. Nolen-Hoeksema S, Morrow J. Effects of rumination and distraction on naturally occurring depressed mood. *Cogn Emot*. 1993;7(6):561–570.
12. Lawlor DA, Hopker SW. The effectiveness of exercise as an intervention in the management of depression: systematic review and meta-regression analysis of randomised controlled trials. *BMJ*. 2001;322(7289):763–767.
13. Sjösten N, Kivelä SL. The effects of physical exercise on depressive symptoms among the aged: a systematic review. *Int J Geriatr Psychiatry*. 2006; 21(5):410–418.
14. Stathopoulou G, Powers M, Berry AC, et al. Exercise interventions for mental health: a quantitative and qualitative review. *Clin Psychol*. 2006; 13:179–193.
15. Mead GE, Morley W, Campbell P, et al. Exercise for depression. *Cochrane Database Syst Rev*. 2008;(4):CD004366.
16. Rethorst CD, Wipfli BM, Landers DM. The antidepressive effects of exercise: a meta-analysis of randomized trials. *Sports Med*. 2009;39(6):491–511.
17. Sexton H, Maere Å, Dahl NH. Exercise intensity and reduction in neurotic symptoms: a controlled follow-up study. *Acta Psychiatr Scand*. 1989; 80(3):231–235.
18. Armstrong K, Edwards H. The effects of exercise and social support on mothers reporting depressive symptoms: a pilot randomized controlled trial. *Int J Ment Health Nurs*. 2003;12(2):130–138.
19. Pinchasov BB, Shurgaja AM, Grischin OV, et al. Mood and energy regulation in seasonal and non-seasonal depression before and after midday treatment with physical exercise or bright light. *Psychiatry Res*. 2000; 94(1):29–42.
20. Egger M, George DS, Altman D, et al.Systematic reviews in health care: investigating and dealing with publication and other biases in meta-analysis. *BMJ* 2001;323:101–105
21. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177–188.
22. Higgins JPT, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–560.
23. Higgins JPT, Thompson SG. Controlling the risk of spurious findings from meta-regression. *Stat Med*. 2004;23(11):1663–1682.
24. Dunn AL, Trivedi MH, Kampert JB, et al. Exercise treatment for depression: efficacy and dose response. *Am J Prev Med*. 2005;28(1):1–8.
25. Singh NA, Stavrinos TM, Scarbek Y, et al. A randomized controlled trial of high versus low intensity weight training versus general practitioner care for clinical depression in older adults. *J Gerontol A Biol Sci Med Sci*. 2005;60(6):768–776.
26. Blumenthal JA, Babyak MA, Doraiswamy PM, et al. Exercise and pharmacotherapy in the treatment of major depressive disorder. *Psychosom Med*. 2007;69(7):587–596.
27. Mutrie N. Exercise as a treatment for moderate depression in the UK National Service**.** Proceedings of sport, health psychology and exercise symposium. London, England: Sports Council and Health Education Authority. 1988;96–105
28. Doyne EJ, Ossip-Klein DJ, Bowman ED, et al. Running versus weight lifting in the treatment of depression. *J Consult Clin Psychol*. 1987;55(5): 748–754.
29. Krogh J, Saltin B, Gluud C, et al. The DEMO trial: a randomized, parallel-group, observer-blinded clinical trial of strength versus aerobic versus relaxation training for patients with mild to moderate depression. *J Clin Psychiatry*. 2009;70(6):790–800.
30. Martinsen EW, Medhus A, Sandvik L. Effects of aerobic exercise on depression: a controlled study. *Br Med J (Clin Res Ed)*. 1985;291(6488):109.
31. Veale D, Le Fevre K, Pantelis C, et al. Aerobic exercise in the adjunctive treatment of depression: a randomized controlled trial. *J R Soc Med*. 1992; 85(9):541–544.
32. Blumenthal JA, Babyak MA, Moore KA, et al. Effects of exercise training on older patients with major depression. *Arch Intern Med*. 1999;159(19): 2349–2356.
33. Dunn AL, Trivedi MH, Kampert JB, et al. The DOSE study: a clinical trial to examine efficacy and dose response of exercise as treatment for depression. *Control Clin Trials*. 2002;23(5):584–603.
34. Epstein D. *Aerobic Activity Versus Group Cognitive Therapy: An Evaluative Study of Contrasting Interventions for the Alleviation of Clinical Depression* [dissertation]. Reno, NV: University of Nevada: 1986
35. Singh NA, Clements KM, Singh MAF. The efficacy of exercise as a long-term antidepressant in elderly subjects: a randomized, controlled trial. *J Gerontol A Biol Sci Med Sci*. 2001;56(8):M497–M504.
36. Babyak MA, Blumenthal JA, Herman S, et al. Exercise treatment for major depression: maintenance of therapeutic benefit at 10 months. *Psychosom Med*. 2000;62(5):633–638.
37. Klein MH, Greist JH, Gurman AS, et al. A comparative outcome study of group psychotherapy vs exercise treatments for depression. *Int J Ment Health*. 1985;13:148–177.
38. Mather AS, Rodriguez C, Guthrie MF, et al. Effects of exercise on depressive symptoms in older adults with poorly responsive depressive disorder: randomised controlled trial. *Br J Psychiatry*. 2002;180(5):411–415.
39. Singh NA, Clements KM, Fiatarone MA. A randomized controlled trial of progressive resistance training in depressed elders. *J Gerontol A Biol Sci Med Sci*. 1997;52(1):M27–M35.
40. Krogh J, Petersen L, Timmermann M, et al. Design paper: the DEMO trial: a randomized, parallel-group, observer-blinded clinical trial of aerobic versus non-aerobic versus relaxation training for patients with light to moderate depression. *Contemp Clin Trials*. 2007;28(1):79–89.

 © Copyright 2010 Physicians Postgraduate Press, Inc.