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

Jesper Krogh, M.D.; Bengt Saltin, M.D.; Christian Gluud, M.D.; and Merete Nordentoft, M.D.

Objective: To assess the benefit and harm of exercise training in adults with clinical depression.

Method: The DEMO trial is a randomized pragmatic trial for patients with unipolar depression conducted from January 2005 through July 2007. Patients were referred from general practitioners or psychiatrists and were eligible if they fulfilled the *International Classification of Diseases, Tenth Revision*, criteria for unipolar depression and were aged between 18 and 55 years. Patients (N = 165) were allocated to supervised strength, aerobic, or relaxation training during a 4-month period. The primary outcome measure was the 17-item Hamilton Rating Scale for Depression ($HAM\text{-}D_{17}$), the secondary outcome measure was the percentage of days absent from work during the last 10 working days, and the tertiary outcome measure was effect on cognitive abilities.

Results: At 4 months, the strength measured by 1 repetition maximum for chest press increased by a mean (95% CI) of 4.0 kg (0.8 to 7.2; p = .014) in the strength training group versus the relaxation group, and maximal oxygen uptake increased by 2.7 mL/kg/min (1.2 to 4.3; p = .001) in the aerobic group versus the relaxation group. At 4 months, the mean change in $HAM\text{-}D_{17}$ score was –1.3 (–3.7 to 1.2; p = .3) and 0.4 (–2.0 to 2.9; p = .3) for the strength and aerobic groups versus the relaxation group. At 12 months, the mean differences in $HAM\text{-}D_{17}$ score were –0.2 (–2.7 to 2.3; p = .8) and 0.6 (–1.9 to 3.1; p = .6) for the strength and aerobic groups versus the relaxation group. At 12 months, the mean differences in absence from work were –12.1% (–21.1% to –3.1%; p = .009) and –2.7% (–11.7% to 6.2%; p = .5) for the strength and aerobic groups versus the relaxation group. No statistically significant effect on cognitive abilities was found.

Conclusion: Our findings do not support a biologically mediated effect of exercise on symptom severity in depressed patients, but they do support a beneficial effect of strength training on work capacity.

Trial Registration: clinicaltrials.gov Identifier: NCT00103415

J Clin Psychiatry 2009;70(6):790–800

Received April 15, 2008; accepted Aug. 14, 2008. From the Psychiatric Center Bispebjerg (Drs. Krogh and Nordentoft); Copenhagen Muscle Research Center, Rigshospitalet (Dr. Saltin); and Copenhagen Trial Unit, Center for Clinical Intervention Research, Rigshospitalet (Dr. Gluud), Faculty of Health Sciences, Copenhagen University Hospital, Copenhagen, Denmark.

The trial was sponsored by "Assurance and Pension," a Danish organization promoting insurance companies. The sponsor wanted absence from work to be included as an outcome measure. Otherwise, the sponsor had no role in trial design, data execution, data collection, data analysis, data interpretation, or the writing of the report. The corresponding author had full access to all data and, together with the coauthors, made the decision to submit the results for publication.

The study results have not previously been presented in public.

Acknowledgments appear at the end of the article.

The authors report no financial or other relationship relevant to the subject of this article.

Corresponding author and reprints: Jesper Krogh, M.D., Psychiatric Center Bispebjerg, Bispebjerg University Hospital, DK-2400, Denmark (e-mail: Jk27@bbh.regionh.dk).

The satisfaction and well-being experienced after a brisk exercise session are familiar to most physically active people. Generally, lay people as well as health care workers accept that exercise renders a plethora of positive effects on physical health. However, the common belief that exercise also improves mental health remains controversial.

The yearly incidence of depression is estimated at 3% to 5%,[1–3] with a lifetime prevalence of 17% in Western societies.[3] Projecting the current development in disease patterns, the World Health Organization (WHO) expects unipolar depression to become the second highest global disease burden in 2030.[4] The economic burden of depression in the United States in 2000 was estimated to be $83 billion (62% were workplace costs, 31% were direct medical costs, and 7% were suicide-related mortality costs).[5]

As medical antidepressant therapy has remission rates of approximately 50%,[6,7] the search for alternative or augmentation therapies is a key issue. In this context, exercise is an interesting intervention with an acceptable price and adverse event profile. However, a meta-analysis published in 2001 concluded that the effectiveness of exercise on depressive symptoms in adults could not be determined because of a lack of good quality research.[8] The

authors found that the majority of trials were without blinded outcome assessment, lacked intention-to-treat analysis, and had short follow-up. Two smaller trials, one comparing strength training versus standard general practitioner treatment[9] and another comparing aerobic training versus no treatment,[10] in an adult population suggested that strength and aerobic training had a positive intensity-related effect on depressed patients. This was supported by a more recently published trial that showed a tendency toward higher remission rates for patients in an exercise program compared with those receiving placebo medication.[11] Only a few trials have compared strength training versus aerobic training.[12,13] Strength and aerobic training could influence depressive symptoms through different biological mechanisms, such as enhanced serotonergic activity, neurotrophic factors, or endorphin levels.[14–16] The cognitive abilities of this patient population are affected by the illness,[17] and it is possible that a positive effect of exercise is partly mediated through increased cognitive skills.[18]

The primary objective for the randomized DEMO trial is to assess the benefits and harms of strength versus aerobic versus relaxation training in patients with depression diagnosed according to the *International Classification of Diseases*, *Tenth Revision* (ICD-10). The primary outcome was the effect on depressive symptoms as measured with the 17-item Hamilton Rating Scale for Depression ($HAM\text{-}D_{17}$)[19] after 4 months of intervention. The secondary outcome was the effect on absence from work, and the tertiary outcome was the effect on cognitive abilities.

METHOD

The protocol was approved by the local ethics committee (KF 01–213) and the Danish Data Protection Agency (J.no. 2004–54-1587) and registered at ClinicalTrials.gov (NCT00103415). A detailed description of the trial design has previously been published.[20]

Participants

We informed the public, general practitioners, and psychiatric institutions about the trial through meetings and advertisements. Patients were considered eligible if they were referred by a medical doctor or a psychologist, fulfilled the ICD-10 criteria for unipolar depression, were 18 to 55 years old, lived in the Greater Copenhagen catchment area, and were able to read and understand the informed consent statement. Patients were considered ineligible if they engaged in any regular sports activity for more than 1 hour per week, had ongoing alcohol or substance abuse, were judged to be at risk of suicide, had poor Danish language skills, had a medical condition that contraindicated physical exercise, or had been on sickness leave for more than 24 consecutive months. All participants gave written informed consent.[20]

ICD-10 (and DSM-IV) diagnosis was established using the Major Depression Inventory[21] as an interviewer-based diagnostic tool. All 3 raters were certified in using the Schedules for Clinical Assessment in Neuropsychiatry (SCAN) interview.[22] However, due to limited resources, we used the Symptom Checklist (SCL-92)[23] to screen for psychotic symptoms, and if psychotic symptoms were suspected, the patients were excluded.

Trial Design

This randomized, parallel-group, observer-blinded, superiority trial was carried out at a single location at Copenhagen University Hospital in Denmark. If the patients were considered eligible for inclusion, they were referred to randomization and exercise testing. Patients were randomly assigned to strength, aerobic, or relaxation training. Randomization was centralized and stratified according to medicine status: (1) not receiving antidepressant medication, (2) having received antidepressant medication for less than 6 weeks, or (3) having received antidepressant medication for more than 6 weeks. DEMO trial staff contacted the Copenhagen Trial Unit (CTU) by phone. Randomization was carried out by the CTU using computerized restricted randomization with a block size of 6. The block size and thus the allocation sequence were unknown to the DEMO trial staff.

Interventions

All patients were scheduled to meet twice per week during a 4-month period for a total of 32 sessions. The participants in each group could freely choose between a session starting at 2 or 5 p.m. All sessions lasted 1.5 hours and were supervised by 1 physiotherapist experienced in instructing psychiatric inpatients. The same 2 physiotherapists were used throughout the trial period. The type and number of exercise interventions were distributed evenly between the two, and thus the physiotherapists were not blinded to allocation. The number of participants per group varied between 3 and 10.

According to guidelines from the American College of Sports Medicine,[24] it could be argued that the interventions should have taken place 3 times per week. However, first, training twice per week should be sufficient to increase both strength[25] and aerobic capacity.[26] Second, we aimed at increasing adherence through fewer expected injuries.[27]

The program in the strength group was designed to increase muscular strength, initially with 12 repetitions of 50% of repetition maximum (RM) 2 or 3 times per exercise. As the patients progressed, the numbers of repetitions were reduced to 10 and 8, with an increase of RM to 75%. The training was a circuit-training program with 6 exercises on machines involving large muscle groups: leg extension, leg press, total abdominal, lower back, chest press, and vertical traction. As a supplement to this, free

weights and sandbags were used for exercising the calf muscles, the arm abductors, the triceps muscles, and the hip abductors.

The aerobic group program was designed to increase fitness as measured by maximal oxygen uptake ($\dot{V}O_2max$). The exercise program involved 10 different aerobic exercises using large muscle groups. Machines were used for cycling, running, stepping, abdominal exercises, and rowing. Additional exercises were sliding movements on small carpets, trampoline, step bench, jump rope, and Ski Fitter (Fitter International; Calgary, Alberta, Canada). During the first 8 sessions, each exercise was done twice for 2 minutes with a 2-minute rest at an intensity level of 70% of maximal heart rate. This gradually increased to a level at which each exercise was done for 3 minutes with a 1-minute rest at an intensity level of 89% during the last 8 sessions.

In the relaxation group, the goal was to avoid muscular contractions or stimulation of the cardiovascular system, and the patients did not engage in activities perceived higher than 12 on the Borg scale.[28] The first 20 to 30 minutes were used for exercises on mattresses or Bobath Balls (Ledregomma; Udine, Italy) or back massage using a Ball Stick Ball (Select; Glostrup, Denmark). This was followed by light balance exercises for 10 to 20 minutes and by relaxation exercises with alternating muscle contraction and relaxation in different muscle groups while lying down for 20 to 30 minutes.

Outcome Measures

Depressive symptoms. Patients were evaluated 4 and 12 months from entry. The primary outcome was the HAM-D_{17} score at 4-month follow-up. The assessor was blinded to intervention group, and the patients were instructed not to reveal their group assignment. After assessment, the assessor was requested to guess which group the patient had been assigned to, making it possible to examine if the blinding was successful. Kappa values for agreement between the right allocation and the guessed allocation at 4 and 12 months were 0.15 and 0.05. This indicated that the blinding of the assessors was successful. Three assessors were used during the trial. Two of these were trained and certified at a WHO collaborating center for mental health in Denmark. Intraclass correlation coefficients between raters 1 and 2 and raters 1 and 3 were 0.93 (0.71 to 0.99) and 0.95 (0.46 to 0.99), respectively.[29] Additional psychometric evaluation was done covering depression (Beck Depression Inventory) and quality of life (World Health Organization-5 Well-Being Index).

Remission was defined as not fulfilling the ICD-10 criteria for depression and having a HAM-D_{17} score < 8.

Absence from work. The secondary outcome was the self-reported percentage of days absent from work during the last 10 working days at 4 and 12 months. Additional work-related information, such as sick leave, unemployment, and job status (i.e., full-time, half-time, or less), was collected.

Cognitive function. Tertiary outcomes related to cognitive function at 4-month and 12-month follow-up included 4 cognitive domains:

Attention. In the Digit Span Test,[30] patients repeat orally given strings of digits of increasing lengths in straight and reversed order. The number of correctly repeated strings is the score. Subtracting Serial Sevens[31] requires the patients to subtract 7 from 100 and continue to subtract 7 until around zero. The score (1–10) is a combination of time and number of errors. The higher the score, the better the performance.

Visuomotor speed. In the A part of the Trail Making Test,[32] patients are asked to connect numbered circles on a sheet in consecutive order. In the B part, they are asked to connect numbers and letters in alternating sequence (A-1-B-2-C-3, and so on). The patients are told to work as fast as possible. The score on each test is the time to complete. The Digit Symbol Test[30] is a symbol/number substitution test in which patients are presented with the numbers 1–9 written at the top of a piece of paper. Each number is represented by a symbol. Below, 100 symbols are listed, but without the corresponding number. The patient is asked to fill in the corresponding numbers in 90 seconds. The number of correct matches is used to calculate the score.

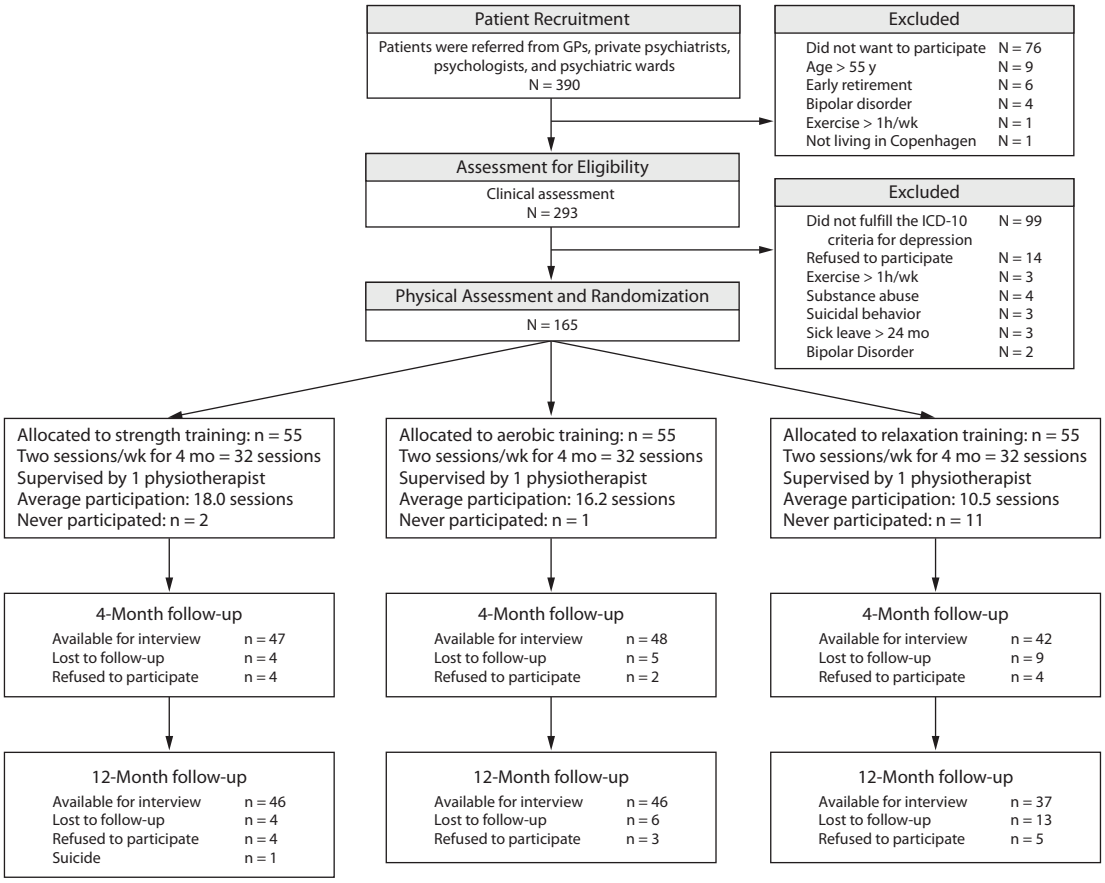
Language. In the S part of Verbal Fluency S and Animals,[33] patients are asked to name as many words beginning with the letter *S* as possible. They are not allowed to use proper nouns. In the Animal part, patients are asked to mention as many animals as they can think of. In both tests, the patients have 60 seconds to name as many as possible. The score in each test corresponds to the number of correct words minus the number of incorrect words.

Memory. In the Buschke Test,[34] a list of 10 unrelated words is read aloud to the patient. The patient is then asked to recall the list. The interviewer repeats the words that the patient misses, and the patient is asked to try again until all 10 words can be said, or until 10 attempts. The score is the total number of omissions or mistakes. In the Rey Complex Figure Test,[35] the patient is shown a geometrically complex figure on a sheet of paper and asked to copy it to another sheet of paper. When this is done, the drawings and the original are put away, and after 3 minutes the patient is asked to draw as much of the figure as he or she can recall. The score is calculated based on the 3-minute recall drawing. The higher the score, the better the performance.

Physical Outcome

The patients' maximal oxygen uptake was measured at entry, at the 4-month follow-up, and at 12-month follow-up using a bicycle cardiopulmonary exercise test.[36] Repetition maximum for knee extension, chest press, and leg

Figure 1. Flowchart of 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



press was measured at entry and postintervention using standardized procedures.

Statistical Analysis

With a Bonferroni-adjusted α of .05/3 = .0166 to allow for comparison of the 3 arms of the trial, as well as a minimal clinically relevant response on the HAM-D_{17} of 4 (δ), a standard deviation of 6, a β = .1, and a dropout of 40%, we calculated that we needed a minimum of 186 patients. However, entry data on the first 50 patients revealed a standard deviation of 3.9, which was confirmed by the international literature.[10,37] We therefore adjusted our sample size calculation to a standard deviation on the HAM-D_{17} of 4. On this basis of this analysis, 28 patients were required in each group, for a total of 84 patients. At 12-month follow-up, the 60% were not available for analysis (based on the first 50 patients), and thus our aim was to include a total of 135 patients.

We analyzed data with SPSS software (version 11.0) (SPSS, Inc.; Chicago, Ill.) on the basis of the intention-to-treat principle, and all patients included were analyzed according to their original group assignment. For continuous outcomes, we used a repeated-measurement, likelihood-based mixed-effect model with an unstructured variance matrix available in SPSS (MIXED procedure). This model approach is recommended for clinical trials with repeated measurements.[38,39] This approach uses data from all included patients (intention-to-treat), handles entry differences, and is able to handle missing data (restricted maximum likelihood procedure) with higher precision and power compared to more traditional methods such as last observation carried forward.[38,40,41] One condition for using this method is that data are "missing at random" or "missing completely at random."[39] Dichotomous outcomes were assessed with odds ratios and χ^2 tests. Attrition, follow-up time, and demographic data were analyzed using analysis of variance (ANOVA) or χ^2 tests.

RESULTS

We received 390 referrals from January 2005 to July 2006. One hundred sixty-five patients were eligible for randomization, and 55 patients were allocated to each training group. Figure 1 shows the trial profile, and Table 1

Table 1. Baseline Data From Participants With Mild to Moderate Depression Randomly Assigned to Strength, Aerobic, or Relaxation Training in a 4-Month Intervention

	Strength (N = 55)	Aerobic (N = 55)	Relaxation (N = 55)	All Participants (N = 165)
Female, N (%)	45 (81.8)	43 (78.2)	34 (61.8)	122 (73.9)
Age, mean (SD), y	41.9 (8.7)	38.1 (9.0)	36.7 (8.7)	38.9 (9.46)
Ethnicity, N (%)				
Caucasian	50 (90.9)	51 (92.7)	50 (90.9)	151 (91.5)
Other	5 (9.9)	4 (7.3)	5 (9.1)	14 (8.5)
Referred from, N (%)				
General practitioner	35 (63.6)	32 (58.2)	31 (56.4)	98 (59.4)
Private practice psychiatrist	15 (27.3)	11 (20.0)	16 (29.1)	42 (25.5)
Outpatient department	5 (9.1)	12 (21.8)	8 (14.5)	25 (15.2)
Depression				
17-Item Hamilton Rating Scale for Depression, mean (SD)	18.2 (3.6)	18.2 (3.8)	16.7 (3.8)	17.8 (3.8)
Montgomery-Asberg Depression Rating Scale, mean (SD)	22.0 (5.6)	22.9 (5.5)	21.6 (4.7)	22.1 (5.3)
DSM-IV criteria for major depressive disorder, N (%)	39 (70.9)	38 (69.1)	35 (63.6)	112 (67.9)
Beck Depression Inventory,[42] mean (SD)	30.6 (8.8)	30.5 (6.9)	31.8 (8.3)	31.0 (8.1)
14-Item Hamilton Rating Scale for Anxiety, mean (SD)	15.1 (5.7)	15.1 (5.6)	14.7 (5.1)	14.6 (4.7)
WHO-5,[43] quality of life, mean (SD)	20 (12.3)	20 (10.1)	23 (11.5)	21.7 (11.3)
Using antidepressant medication, N (%)	39 (70.9)	37 (67.3)	38 (69.1)	114 (69.1)
Having used antidepressant medication > 6 wk, N (%)	35 (63.6)	35 (63.6)	32 (58.2)	102 (61.8)
Receiving psychotherapy, N (%)	24 (43.6)	28 (50.9)	25 (45.5)	77 (46.7)
Previous episodes of depression, mean (SD)	1.3 (2.0)	1.3 (1.9)	1.0 (1.7)	1.2 (1.9)
Time since diagnosis of current depression, mean (SD), mo	13.2 (21.7)	20 (37.4)	20.8 (30.2)	18.2 (30.5)
Work				
Unemployed, N (%)	23 (41.8)	30 (54.5)	20 (36.4)	73 (44.2)
Working full-time ~37 h/wk, N (%)	22 (40.0)	18 (32.7)	23 (41.8)	63 (38.2)
Working part-time ~20 h/wk, N (%)	8 (14.5)	6 (10.9)	10 (18.2)	24 (14.5)
Working < 20 h/wk, N (%)	2 (3.6)	1 (1.8)	2 (3.6)	5 (3.0)
Sick leave, N (%)	29 (52.7)	23 (41.8)	24 (43.6)	76 (46.1)
Percentage of days absent from work in last 10 d, mean (SD)	17.8 (31.5)	30 (34.7)	26.6 (35.3)	25.6 (34.0)
Cognitive skills, mean (SD)				
Verbal intelligence				
Danish Adult Reading Test	33.4 (9.2)	34.2 (8.7)	32.9 (8.1)	33.5 (8.6)
Attention				
Digit Span	11.5 (3.9)	11.2 (2.8)	11.5 (3.3)	11.4 (3.3)
Subtracting Serial Sevens	7.0 (3.8)	7.8 (2.6)	7.6 (2.9)	7.4 (3.0)
Visuomotor speed				
Trail Making A, s	31.0 (9.1)	30.2 (11.7)	28.5 (9.6)	29.84 (10.2)
Trail Making B, s	72.3 (24.3)	70.9 (34.7)	76.0 (33.9)	73.2 (31.4)
Digit Symbol Test	45.7 (9.6)	50.6 (10.6)	48.6 (13.5)	48.4 (11.4)
Design Fluency	27.3 (7.5)	27.9 (7.2)	27.5 (8.4)	27.6 (7.7)
Language				
Verbal Fluency—Animal, no. of words	23.2 (6.8)	22.7 (6.0)	24.0 (6.0)	23.3 (6.2)
Verbal Fluency—S words, no. of words	15.3 (6.0)	14.0 (4.5)	14.3 (5.0)	14.8 (5.1)
Memory				
Buschke Selective Reminding Test, total score	13.9 (8.1)	13.9 (7.3)	14.3 (8.4)	14.4 (7.8)
Rey Complex Figure Test, 3-min recall	22.1 (5.0)	20 (6.4)	21.1 (6.6)	21.0 (6.1)
Physical assessment, mean (SD)				
Body mass index (kg/m^2)	26.1 (4.9)	27.5 (5.7)	26.5 (6.3)	26.5 (5.4)
$\dot{V}O_2$max (mL/kg/min)	27.1 (6.6)	26.0 (6.3)	28.8 (7.7)	27.9 (7.1)
1 RM chest press, kg	33.1 (13.6)	38.8 (15.3)	45.7 (22.2)	38.4 (17.4)
1 RM knee extension, kg	45.3 (16.2)	56.7 (20.3)	57.1 (22.2)	52.3 (20.1)
1 RM leg press, kg	91.0 (35.1)	109.6 (41.1)	110.8 (44.0)	102.7 (40.5)

Abbreviations: RM = repetition maximum, WHO-5 = World Health Organization-5 Well-Being Index.

shows entry demographics and clinical characteristics of each intervention group. At entry, the 3 groups looked balanced, although the HAM-D_{17} score was lower in the relaxation group, which also had a higher proportion of male participants. The 12-month follow-up ended in July 2007.

Compliance, Follow-Up, and Effect on Physical Outcomes

The mean participation was 18.0 (56.2%), 16.2 (50.6%), and 10.5 (32.8%) sessions of the 32 sessions in the strength, aerobic, and relaxation training groups, respectively ($F = 10.05$, $df = 2,162$; $p < .001$). At 4-month follow-up, 137/165 (83.0%) were available for follow-up, with 47/55 (85.4%), 48/55 (87.3%), and 42/55 (76.4%) in the strength, aerobic, and relaxation training groups, respectively ($\chi^2 = 2.667$, $df = 2$, $p = .264$). Eighteen of 165 patients were lost to follow-up, and 10/165 refused to participate. At 12-month follow-up, 129/165 (78.2%) were available for follow-up, with 46/55, 46/55, and 37/55 in the strength, aerobic, and relaxation training groups,

respectively ($\chi^2 = 5.756$, df = 2, p = .056); 23/165 were lost to follow-up, 12/165 refused to participate, and 1/165 committed suicide. The mean (SD) number of days by which the 4-month follow-up was delayed was 32.7 (27.0), 41.3 (25.7), and 54.3 (52.4) for the strength, aerobic, and relaxation training groups, respectively (F = 4.5, df = 2,134; p = .014). The mean number of days by which the 12-month follow-up was delayed was 44.9 (32.6), 56.5 (53.6), and 56.3 (59.3) days for the strength, aerobic, and relaxation training groups, respectively (F = 0.8, df = 2,127; p = .44). Nonparametric testing (Kruskal-Wallis h) confirmed the ANOVA results.

Analysis of age, sex, HAM-D_{17} score, or absence from work during the last 10 working days at entry did not suggest any significant differences between missing participants and participants included in the analysis at either 4 months or 12 months. It is then plausible to consider the missing data as "missing at random," making the mixed-effect model a plausible approach to estimate the effect, based on the total sample with the missing cases included.

The mean (95% CI) increases in 1 RM for chest press, knee extension, and leg press at 4 months were 4.0 kg (0.8 to 7.2; p = .014), 7.5 kg (2.1 to 12.8; p = .007), and 18.3 kg (6.3 to 30.3; p = .003) in the strength group compared with the relaxation group. When the strength group was compared to the aerobic group at 4 months, the increases in 1 RM for chest press, knee extension, and leg press were 4.4 kg (1.5 to 7.2; p = .003), 7.9 kg (3.2 to 12.5; p = .001), and 16.8 kg (6.2 to 27.3; p = .002).

The mean (95% CI) increase in maximal oxygen uptake at 4 months was 2.7 mL/kg/min (1.2 to 4.3; p = .001) when the aerobic group was compared with the relaxation group. At 4 months, the mean difference between the aerobic group and the strength group was 1.74 mL/kg/min (0.3 to 3.2; p = .02).

Primary Outcome

The HAM-D_{17} scores did not differ significantly between the 3 groups at 4-month follow-up (Table 2). The mean (95% CI) HAM-D_{17} scores at 4 months for participants not on medication at entry were –1.4 (–5.8 to 3.0; p = .5) and –0.2 (–4.6 to 4.1; p = .9) for the relaxation group versus the strength and aerobic groups, respectively. The primary outcome was not significantly influenced by including sex as main effect or as an interaction effect with time in the model (time × sex, p = .43). Some participants attended 50% or more of the sessions (strength N = 36, aerobic N = 28, relaxation N = 18), and for them, the estimates were –1.2 (–4.6 to 2.0; p = .4) and 0.2 (–3.2 to 3.7; p = .9) for the relaxation group versus the strength and aerobic groups, respectively. The numbers of patients in remission in the strength, aerobic, and relaxation groups were 19/47 (40.4%), 14/48 (29.2%), and 13/41 (31.7%) ($\chi^2 = 1.462$, df = 2, p = .48) at 4 months and 19/47 (40.4%), 15/46 (32.6%), and 14/37 (37.8%) ($\chi^2 = .628$, df = 2, p = .73) at 12 months.

The mean (95% CI) difference in HAM-D_{17} score between the strength and aerobic groups was –1.7 (–4.1 to 0.6; p = .15) at 4 months and –0.8 (–3.2 to 1.6; p = .50) at 12 months. The analysis did not reveal statistically significant differences in any psychometric outcome at either the 4-month or 12-month follow-up. See Tables 2 and 3 for details.

When both exercise groups were combined and compared with the relaxation group, the differences in HAM-D_{17} score were –0.4 (–2.6 to 1.7; p = .70) and 0.2 (–2.0 to 2.6; p = .86) for 4 and 12 months, respectively.

To investigate whether there was an antidepressant effect of the intervention applied in the control group (relaxation), we divided this group's participants post hoc into 2 groups on the basis of median attendance: those who attended the intervention fewer than 8 times (N = 25) and those who attended 8 times or more (N = 30). When we entered this dichotomized participation term in the model as a main effect and an interaction term with time, we found no statistically significant differences in the primary outcome at either 4 or 12 months between these groups (time × more or fewer than 8 sessions, F = 0.678, df = 1,53; p = .42).

When patients with a score of 18 or higher on the HAM-D_{17} at entry (strength N = 31, aerobic N = 31, and relaxation N = 21) were analyzed, the mean (95% CI) difference on the HAM-D_{17} at 4 months was –2.00 (–6.3 to 2.23; p = .35) and –0.78 (–5.1 to 3.5; p = .72) for strength and aerobic versus relaxation training. For 12-month follow-up, the difference was –1.06 (–5.3 to 3.2; p = .62) and 0.08 (–4.2 to 4.3; p = .97) for strength and aerobic versus relaxation training.

Secondary Outcome

Throughout the trial, absence from work decreased in all of the intervention groups. At 4 months, the group differences were estimated to be –12.5% (–28.9% to 4.0%; p = .13) and –7.9% (–24.1% to 8.3%; p = .30) for strength and aerobic training, respectively, versus relaxation training. At 12 months, there was a significant reduction in absence from work in the strength group compared with the relaxation group, which was confirmed by a nonparametric analysis (Mann-Whitney U: z = –2.5; p = .012).

For the strength group compared to the aerobic group, the difference was –4.6% (–20.2% to 11.0%; p = .55) and –9.4% (–17.7% to –1.0%; p = .03; Mann-Whitney U: z = –2.49; p = .013) at 4 and 12 months.

When the 2 exercise groups were compared with the relaxation group, the difference in absence from work was –10.1% (–24.6% to 4.1%; p = .16) at 4 months and –7.3% (–15.4% to 0.9%; p = .81) at 12 months. No other significant differences in employment status or other

Table 2. Postintervention (4-month) Outcome for Participants With Mild to Moderate Depression Randomly Assigned to Strength, Aerobic, or Relaxation Training in a 4-Month Intervention[a]

				Strength vs Relaxation[b]		Aerobic vs Relaxation[b]	
	Strength (N = 47)	Aerobic (N = 48)	Relaxation (N = 42)	Difference (95% CI)	p	Difference (95% CI)	p
Depression							
17-item Hamilton Rating Scale for Depression, mean (SD)	10.0 (6.4)	12.1 (6.4)	10.6 (5.6)	–1.3 (–3.7 to 1.2)	.3	0.4 (–2.0 to 2.9)	.3
Montgomery-Asberg Depression Rating Scale, mean (SD)	12.1 (8.0)	14.9 (8.7)	13.9 (8.2)	–1.5 (–4.9 to 1.9)	.4	0.2 (–3.2 to 3.6)	.9
Beck Depression Inventory, mean (SD)	13.8 (12.0)	19.2 (11.8)	18.0 (12.7)	–3.2 (–8.1 to 1.6)	.2	0.4 (–4.4 to 5.2)	.9
14-item Hamilton Rating Scale for Anxiety, mean (SD)	8.1 (6.0)	11.5 (7.3)	9.9 (6.9)	–1.8 (–4.5 to 0.8)	.2	1.3 (–1.3 to 4.0)	.3
WHO-5, quality of life, mean (SD)	52.8 (21.8)	41.7 (24.0)	45.2 (20.8)	8.3 (–0.7 to 17.3)	.07	–1.0 (–10.0 to 8.0)	.8
				OR (95% CI)		OR (95% CI)	
Using antidepressant medication, N (%)	31 (66.0)	31 (64.6)	22 (52.4)	1.8 (0.7 to 4.1)	.2	1.7 (0.7 to 3.8)	.2
Receiving psychotherapy, N (%)	23 (48.9)	23 (47.9)	20 (47.6)	1.1 (0.4 to 2.4)	.8	0.9 (0.4 to 2.1)	.8
Work							
Unemployed, N (%)	20 (42.6)	18 (37.5)	17 (40.5)	1.1 (0.5 to 2.6)	.8	0.9 (0.4 to 2.1)	.8
Sick leave, N (%)	16 (34.0)	13 (27.1)	13 (31.0)	1.1 (0.5 to 2.8)	.8	1.0 (0.4 to 2.7)	.9
				Difference (95% CI)		Difference (95% CI)	
Percentage of days absent from work in last 10 d, mean (SD)	4.3 (9.0)	10.0 (17.6)	16.9 (33.3)	–12.5 (–28.9 to 4.0)	.1	–7.9 (–24.1 to 8.3)	.3
Cognitive skills, mean (SD)							
Attention							
Digit Span	12.7 (3.8)	11.6 (3.1)	12.1 (3.6)	0.3 (–0.8 to 1.3)	.6	–0.6 (–1.6 to 0.5)	.3
Subtracting Serial Sevens	7.78 (3.1)	8.31 (2.0)	8.23 (2.8)	–0.3 (–1.4 to 0.9)	.7	0.2 (–1.0 to 1.4)	.8
Visuomotor speed							
Trail Making A, s	27.2 (8.1)	28.6 (11.0)	26.7 (9.4)	0.4 (–3.0 to 3.8)	.8	1.2 (–2.2 to 4.7)	.5
Trail Making B, s	62.9 (18.5)	64.5 (23.7)	67.1 (28.8)	1.9 (–6.0 to 9.8)	.6	1.0 (–6.9 to 9.0)	.8
Digit Symbol Test	51.1 (9.1)	56.0 (10.8)	50.1 (12.8)	1.4 (–2.8 to 5.7)	.5	3.5 (–0.6 to 7.5)	.1
Design Fluency	31.1 (5.8)	30.5 (4.9)	30.5 (7.1)	0.7 (–1.5 to 3.0)	.5	0.4 (–1.9 to 2.7)	.7
Language							
Verbal Fluency—Animal, no. of words	24.9 (9.9)	23.2 (4.73)	31.0 (39.9)	–4.9 (–15.0 to 5.2)	.3	–6.4 (–16.6 to3.7)	.2
Verbal Fluency—S words, no. of words	16.5 (6.1)	15.2 (5.9)	15.6 (6.0)	–0.0 (–2.1 to 2.1)	.9	–0.0 (–2.1 to 2.0)	.9
Memory							
Buschke Selective Reminding Test, total score	10.3 (8.0)	11.1 (7.8)	10.4 (7.2)	0.6 (–3.0 to 4.1)	.7	1.0 (–2.5 to 4.4)	.6
Rey Complex Figure test, 3-min recall	23.7 (6.3)	23.3 (5.6)	23.5 (6.3)	–1.2 (–3.4 to 1.1)	.3	0.4 (–1.8 to 2.6)	.7
Physical assessment							
Vo_2max (mL/kg/min)	29.2 (7.4)	28.9 (6.6)	30.5 (6.8)	1.0 (–0.5 to 2.5)	.2	2.7 (1.2 to 4.3)	.001
1 RM chest press, kg	41.4 (13.8)	40.0 (16.5)	53.3 (24.1)	4.0 (0.8 to 7.2)	.014	–0.3 (–3.7 to 3.1)	.8
1 RM knee extension, kg	57.7 (17.8)	59.5 (19.4)	63.0 (17.7)	7.5 (2.1 to 12.8)	.007	–0.4 (–6.1 to 5.3)	.9
1 RM leg press, kg	117 (41.9)	117.0 (42.3)	129.0 (45.3)	18.3 (6.3 to 30.3)	.003	1.5 (–11.3 to 14.4)	.8

[a]Group Ns represent the numbers of patients available at follow-up.
[b]Group differences are estimated using a likelihood-based mixed-model analysis with repeated measurements for continuous outcomes, and dichotomous outcomes are presented with odds ratios (ORs). The estimate is the mean difference between groups in change from baseline.
Abbreviations: RM = repetition maximum, WHO-5 = World Health Organization-5 Well-Being Index.

work-related outcome were found at either the 4-month or the 12-month follow-up (see Tables 2 and 3).

Tertiary Outcomes

No statistically significant differences in cognitive outcomes were found at either the 4-month or 12-month follow-up (see Tables 2 and 3).

After 4 months, the number of patients admitted to a psychiatric ward was 0 in the exercise groups versus 2 in the relaxation group (Fisher exact test: p = .092). After 12 months, 5 patients in the exercise groups versus 5 patients in the relaxation group had been admitted to hospital (Fisher exact test: p = .15). One patient allocated to strength training committed suicide between the 4- and 12-month follow-ups. Otherwise, we did not see indications of potential harm.

DISCUSSION

Our results suggest that a 4-month exercise intervention intense enough to significantly increase strength and maximal oxygen uptake did not have a biologically

Table 3. Twelve-Month Follow-Up Outcome for Participants With Mild to Moderate Depression Randomly Assigned to Strength, Aerobic, or Relaxation Training in a 4-Month Intervention[a]

				Strength vs Relaxation[b]		Aerobic vs Relaxation[b]	
	Strength (N = 46)	Aerobic (N = 46)	Relaxation (N = 37)	Difference (95% CI)	p	Difference (95% CI)	p
Depression							
17-item Hamilton Rating Scale for Depression, mean (SD)	11.0 (7.1)	11.9 (6.5)	10.0 (5.6)	–0.2 (–2.7 to 2.3)	.8	0.6 (–1.9 to 3.1)	.6
Montgomery-Asberg Depression Rating Scale, mean (SD)	14.2 (9.8)	14.8 (8.3)	13.0 (7.6)	0.9 (–2.7 to 4.4)	.6	1.3 (–2.2 to 4.8)	.5
Beck Depression Inventory, mean (SD)	15.2 (12.6)	16.1 (11.0)	16.0 (11.1)	–0.3 (–5.1 to 4.3)	.9	0.1 (–4.7 to 4.9)	.9
14-item Hamilton Rating Scale for Anxiety, mean (SD)	10.3 (6.9)	11.8 (7.4)	9.2 (5.5)	0.6 (-2.1 to 3.3)	.7	2.2 (–0.5 to 4.9)	.1
WHO-5, quality of life, mean (SD)	46.0 (26.1)	46.9 (22.7)	47.8 (23.4)	1.2 (–8.7 to 11.0)	.8	1.7 (–8.3 to 11.7)	.7
				OR (95% CI)		OR (95% CI)	
Using antidepressant medication, N (%)	28 (60.9)	29 (63.0)	22 (59.5)	1.0 (0.4 to 2.4)	.9	1.2 (0.5 to 2.8)	.7
Receiving psychotherapy, N (%)	15 (32.6)	17 (37.0)	15 (40.5)	0.7 (0.3 to 1.7)	.4	0.8 (0.3 to 2.1)	.7
Work							
Unemployed, N (%)	15 (32.6)	15 (32.6)	7 (18.9)	1.7 (0.6 to 5.0)	.3	1.8 (0.6 to 5.0)	.3
Sick leave, N (%)	9 (19.6)	13 (28.3)	9 (24.3)	0.8 (0.3 to 2.4)	.7	1.2 (0.5 to 3.4)	.7
				Difference (95% CI)		Difference (95% CI)	
Percentage of days absent from work in last 10 d, mean (SD)	1.4 (4.4)	11.2 (19.2)	14.5 (20.7)	–12.1 (–21.1 to –3.1)	.009	–2.7 (–11.7 to 6.2)	.5
Cognitive skills, mean (SD)							
Attention							
Digit Span	11.75 (3.7)	12.2 (3.1)	11.6 (3.4)	0.0 (–1.0 to 1.0)	.9	0.2 (–0.9 to 1.2)	.8
Subtracting Serial Sevens	7.5 (2.9)	9.0 (1.5)	8.0 (3.0)	–0.2 (–1.3 to 1.0)	.8	0.7 (–0.5 to 1.8)	.2
Visuomotor speed							
Trail Making A, s	26.5 (8.4)	26.1 (11.6)	26.8 (11.0)	–0.8 (–4.8 to 3.2)	.7	–0.4 (–4.5 to 3.7)	.9
Trail Making B, s	72.5 (52.8)	61.2 (25.3)	68.4 (30.6)	4.4 (–6.5 to 15.4)	.4	–2.4 (–13.4 to 8.6)	.7
Digit Symbol Test	52.1 (10.1)	54.4 (11.0)	52.1 (13.1)	2.6 (–1.3 to 7.0)	.2	2.3 (–1.9 to 6.5)	.3
Design Fluency	31.1 (5.0)	31.6 (4.3)	30.0 (7.0)	2.1 (–0.2 to 4.3)	.08	2.1 (–0.2 to 4.4)	.08
Language							
Verbal Fluency—Animal, no. of words	25.0 (7.4)	25.4 (5.0)	24.9 (7.1)	1.8 (–0.7 to 4.4)	.2	1.5 (–1.0 to 4.1)	.2
Verbal Fluency—S words, no. of words	15.0 (5.2)	15.3 (6.6)	14.9 (5.7)	0.8 (–1.4 to 3.0)	.5	0.5 (–1.6 to 2.7)	.6
Memory							
Buschke Selective Reminding Test, total score	10.1 (7.0)	6.5 (4.9)	9.4 (7.2)	1.1 (–1.7 to 3.8)	.4	–2.3 (–5.2 to 0.6)	.1
Rey Complex Figure Test, 3-min recall	23.2 (6.0)	22.6 (4.6)	23.4 (7.0)	–0.9 (–3.0 to 1.1)	.4	–0.4 (–2.5 to 1.6)	.7

[a]Group Ns represent the numbers of patients available at follow-up.
[b]Group differences are estimated using a likelihood-based mixed-model analysis with repeated measurements for continuous outcomes, and dichotomous outcomes are presented with odds ratios (ORs). The estimate is the mean difference between groups in change from baseline.
Abbreviations: RM = repetition maximum, WHO-5 = World Health Organization-5 Well-Being Index.

mediated effect on symptom severity of clinical depression postintervention or at long-term follow-up. However, a tendency was shown that strength training decreased absence from work, a difference that was statistically significant at the 12-month follow-up visit.

The strengths of our trial were the centralized randomization, which provided adequate generation of the allocation sequence and adequate allocation concealment; blinding of the assessors of the primary outcome to allocation; the use of intention-to-treat analysis[20,42]; high interrater reliability; and well-conducted interventions that increased the physical parameters they were designed to improve. Inclusion of both medicated and unmedicated patients increases the generalizability in terms of the expected effect in a real community setting. Furthermore, post hoc analysis suggested that the intervention used in our control group (relaxation training) had no impact on depression scores.

Limitations include confounding due to a possible antidepressant effect of the intervention used in our control group (relaxation training). However, the mean effect observed in the relaxation training group (–6.1 on HAM-D_{17} after 4 months) was comparable to the placebo effect observed in a similar trial providing no training at all.[11] The lack of blinding of treatment allocation for patients and physiotherapists could lead to collateral interventions, possibly confounding our results. There was skewed attrition, and the follow-up assessment was significantly later than 4 months in the control group. The latter could be explained by the depression itself, or by the participants' being less willing to return for assessment because they had not received the intervention they had hoped for. A possible type 2 error would occur if patients in the relaxation training group postponed their assessment due to high depression scores at the ideal time for follow-up, or if the general delay in assessment diminished a possible

antidepressant effect of exercise. Other trials[11] have used weekly assessments to exclude early responders from their analysis, but we believe that this violates the principles of intention-to-treat analysis. Furthermore, post hoc analysis did not reveal a greater effect for patients who had a HAM-D_{17} rating above 17 at entry. We recognize that there are some good arguments for using weekly assessments in trials due to increased suicide risk in this population. However, our physiotherapists all had experience with psychiatric inpatients and were instructed to consult trial investigators if they had suspicion of increased suicide risk for any particular participant. It could be argued that we should have used the SCAN interview[22] as a diagnostic tool, but due to limited resources this was not possible. Since 2 of the interviewers had several years of experience from assessing patients with schizophrenia using the SCAN interview,[22] we consider this to have had minimal impact on inclusion.

On the basis of the first 50 patients, we changed our sample size calculation, which is usually not recommended. This change was based only on the observed standard deviation and was not propelled by any intervention effect. However, the confidence intervals of the primary outcome do not suggest that the lack of effect was caused by a power problem at this level of clinical relevance (4 points on the HAM-D_{17}).

We included both medicated and nonmedicated patients and patients who received psychotherapy. The majority of medicated patients had received medication for more than 6 weeks and still fulfilled the ICD-10 criteria for depression and could in some sense be considered treatment resistant. This weakens the trial result in terms of a direct efficacy measurement, i.e., an explanatory aspect of the trial. However, our intention was to conduct a pragmatic trial in order to make the results relevant for clinical practice.

As mentioned, we chose to create a control group with the same level of social interaction and contact with instructors to control for a possible social effect, instead of having a wait-list control group. This control group should be exposed to the same social stimuli as well as regression toward the mean as the experimental groups. We wanted to offer an intervention to the control group that would result in high compliance and involve some sort of physical activity that would interest the participants.

The attendance was only approximately 50% in our experimental groups at 4-month follow-up. In comparison, antidepressant treatment adherence in the primary care setting has been reported to be 56% after 3 months.[43] However, a similar trial of the effect of exercise on depression had a median attendance of 77% in a supervised exercise group.[11]

The increase in strength measured by 1 RM was from 25% to 28% in the strength group, and the increase in maximal oxygen uptake of 11% in the aerobic group is comparable to similar trials.[9,44]

Recent smaller trials comparing an exercise group with a placebo/control group suggested that exercise intensity had an effect on depressive symptoms.[9,10] In our trial, we had 2 training sessions per week, which might have been too few. The majority of our patients had received antidepressive medication for more than 6 weeks at entry, and for this group, the interventions were an addition to the pharmacologic treatment. In relation to this aspect, our result is consistent with those of other large trials that did not find an additional effect of exercise in patients using antidepressive medication.[45,46] Furthermore, a recent trial comparing exercise with placebo did not find a statistically significant effect of exercise on clinical depression.[44] Contrary to our trial, other studies have found a tendency toward higher remission rates in depressed patients engaging in physical activity.[44] The reasons for these differences are unclear.

The 4-month results do not suggest that exercise training had an effect on absence from work. At 12 months, the data suggest a markedly lower work absence in the strength group. This result should be interpreted with caution, since it was not a primary outcome and none of the other work-related outcomes suggest any effect of the intervention—the numbers of participants who were unemployed or on sick leave were the same in all groups. However, there was a tendency toward less absence in both exercise groups compared with the relaxation group, and our trial might not have had enough power to reveal a true significant effect.

There is growing evidence from animal research that exercise increases cognitive function partly by up-regulating neurotrophic factors,[47,48] and meta-analytic findings suggest that exercise has a positive effect in humans who are already cognitively impaired, such as those with dementia.[18] The DEMO trial is to our knowledge the first randomized trial to test whether an exercise intervention has an effect on cognitive abilities in clinically depressed patients. Our results suggest that an exercise intervention twice per week has no significant effect on cognitive abilities in clinically depressed individuals. It is possible that the patients' cognitive abilities measured in this trial were not affected at entry, even though a modest cognitive impairment has previously been detected in depressed outpatients.[49,50] Another explanation could be that part of our patient population had taken medication for some time already that had caused an up-regulation in neurotrophic activity,[51] suggesting that exercise and antidepressant medication mediate a physiologic effect on the brain through the same systems.

Even though it seems increasingly questionable that exercise has a clinically important effect on depressive symptoms, it should be acknowledged that depression is associated with an increased risk of cardiovascular

disease.[52,53] It is important to investigate which exercise programs are most effective in attenuating this risk among patients with depression. The primary financial burden of depression is work-related,[5] and the effect of exercise on this parameter needs further investigation in large-scale trials.

SUMMARY

In conclusion, our trial does not provide evidence for a biologically mediated effect of exercise on clinical depression in a pragmatic outpatient setting. Exercise recommendations suggest that the intervention should have been offered 3 times per week. Furthermore, due to low compliance, our participants received the intervention only once per week. Although this was sufficient to increase both strength and aerobic capacity, this may not have been enough to modify mood. We also observed a tendency toward less absence from work, but no improvement in cognitive skills.

Acknowledgments: We thank the trial participants; Michael Timmermann (physiotherapist) and Linda Kahr (physiotherapist), of Rigshospitalet, Copenhagen, Denmark, for their contribution to the intervention; and Nader Sales, M.Sc. (engineer and data manager), of the Copenhagen Trial Unit, Rigshospitalet, Copenhagen, Denmark, for his contribution to randomization. Furthermore, we thank Lone Pedersen, M.Sc. (psychologist) and Britt Morthorst (nurse), of Psychiatric Center Bispebjerg, Copenhagen, Denmark, for their contribution to clinical assessment of trial participants; Mahdi Mohammadinezhad, M.Sc., of the Institute of Exercise and Sports Sciences, University of Copenhagen, Copenhagen, Denmark, for assistance with exercise testing; and Anders Frøkjær Thomsen, M.D., Ph.D., of Psychiatric Center Bispebjerg, Copenhagen, Denmark, and Kate Whitfield, M.Sc., of the Copenhagen Trial Unit, Rigshospitalet, Copenhagen, Denmark, for commenting on the manuscript. None of the acknowledged persons report conflicts of interest.

REFERENCES

1. Andrews G, Henderson G, Hall W. Prevalence, comorbidity, disability and service utilisation: overview of the Australian National Mental Health Survey. Br J Psychiatry 2001;178:145–153
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. Major depressive disorder. Acta Pscyhiatr Scand 1994;376:7–15
4. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. PLoS Med 2006;3(11):2011–2030
5. Greenberg PE, Kessler RC, Birnbaum HG, et al. The economic burden of depression in the United States: how did it change between 1990 and 2000? J Clin Psychiatry 2003;64(12):1465–1475
6. Machado M, Iskedjian M, Ruiz I, et al. Remission, dropouts, and adverse drug reaction rates in major depressive disorder: a meta-analysis of head-to-head trials. Curr Med Res Opin 2006;22(9):1825–1837
7. Thase ME, Haight BR, Richard N, et al. Remission rates following antidepressant therapy with bupropion or selective serotonin reuptake inhibitors: a meta-analysis of original data from 7 randomized controlled trials. J Clin Psychiatry 2005;66(8):974–981
8. Lawlor D, Hopker S. 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
9. 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
10. Dunn AL, Trivedi MH, Kampert K, et al. Exercise treatment for depression: efficacy and dose response. Am J Prev Med 2005;28(1):1–8
11. Blumenthal J, Babyak M, Doraiswamy M, et al. Exercise and pharmacotherapy in the treatment of major depressive disorder. Psychosom Med 2007;69(7):587–596
12. Martinsen EW, Hoffart A, Solberg Ø. Comparing aerobic with nonaerobic forms of exercise in the treatment of clinical depression: a randomized trial. Compr Psychiatry 1989;30(4):324–331
13. 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
14. Chaouloff F. Effects of acute physical exercise on central serotonergic systems. Med Sci Sports Exerc 1997;29(1):58–62
15. Persson A, Naylor A, Jonsdottir I, et al. Differential regulation of hippocampal progenitor proliferation by opioid receptor antagonists in running and non-running spontaneously hypertensive rats. Eur J Neurosci 2004;19(7):1847–1855
16. Ernst C, Olsen AK, Pinel JP, et al. Antidepressant effects of exercise: evidence for an adult-neurogenesis hypothesis? J Psychiatry Neurosci 2006;31(2):84–92
17. Ravnkilde B, Videbech P, Clemmensen K, et al. Cognitive deficits in major depression. Scand J Psychol 2002;43(3):239–251
18. Heyn P, Abreu B, Ottenbacher K. The effects of exercise training on elderly persons with cognitive impairment and dementia: a meta-analysis. Arch Phys Med Rehabil 2004;85(10):1694–1704
19. Hamilton M. A rating scale for depression. J Neurol Neurosurg Psychiatry 1960;23:56–62
20. 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
21. Bech P, Rasmussen N-A, Olsen LR, et al. The sensitivity and specificity of the Major Depression Inventory, using the Present State Examination as the index of diagnostic validity. J Affect Disord 2001;66(2-3):159–164
22. World Health Organization. Schedules for Clinical Assessment in Neuropsychiatry, Version 2.1. Geneva, Switzerland: Division of Mental Health, World Health Organization; 2008
23. Olsen LR, Mortensen EL, Bech P. The SCL-90 and SCL-90R versions validated by item response models in a Danish community sample. Acta Psychiatr Scand 2004;110(3):225–229
24. Haskell WL, Lee IM, Pate RR, et al. Physical activity and public health: updated recommendation for adults from the American College of Sports Medicine and the American Heart Association. Circulation 2007;116(9): 1081–1093
25. Braith R, Graves J, Pollock M, et al. Comparison of 2 vs 3 days/week of variable resistance training during 10 and 18-week programs. Int J Sports Med 1989;10(6):450–454
26. Atomi Y, Ito K, Iwasaki H, et al. Effects of intensity and frequency of training on aerobic work capacity of young females. J Sports Med Phys Fitness 1978;18(1):3–9
27. Pollock ML, Gettman LR, Milesis CA, et al. Effects of frequency and duration of training on attrition and incidence of injury. Med Sci Sports 1977;9(1):31–36
28. Borg G. Perceived exertion as an indicator of somatic stress. Scand J Rehabil Med 1970;2(2):92–98
29. Bartko J, Carpenter W. On the methods and theory of reliability. J Nerv Ment Dis 1976;163(5):307–317
30. Wechscler D. The Wechsler Adult Intelligence Scale-Revised. San Antonio, Tex: The Psychological Corporation;1981
31. Smith A. The Serial Sevens Subtraction Test. Arch Neurol 1967;17(1):78–80
32. Reitan M. The relation of the Trail Making Test to organic brain damage. J Consult Psychol 1955;193:393–394
33. Borkowski JG, Benton AL, Spreen O. Word fluency and brain damage. Neuropsychologia 1967;5:135–140
34. Buschke H, Fuld PA. Evaluating storage, retension and retrieval in disordered memory and learning. Neurology 1974;24(11):1019–1025
35. Meyers J, Meyers K. Rey Complex Figure and Recognition Trial: Professional Manual. Odessa, Fla: Psychological Assessment Resources; 1995
36. Andersen LB. A maximal cycle exercise protocol to predict maximal oxygen uptake. Scand J Med Sci Sports 1995;5(3):143–146
37. Fava M, Alpert J, Nierenberg AA, et al. A double-blind, randomized trial

of St John's wort, fluoxetine, and placebo in major depressive disorder. J Clin Psychopharmacol 2005;25(5):441–447
38. Mallinckrodt C, Watkin J, Molenberghs G, et al. Choice of the primary analysis in longitudinal clinical trials. Pharm Stat 2004;3:161–169
39. Gueorguieva R, Krystal J. Move over ANOVA: progress in analyzing repeated-measures data and its reflection in papers published in the *Archives of General Psychiatry*. Arch Gen Psychiatry 2004;61(3): 310–317
40. Mallinckrodt C, Sanger T, Dube S, et al. Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. Biol Psychiatry 2003;53(8):754–760
41. Mallinckrodt C, Clark S, David S. Accounting for dropout bias using mixed-effects models. J Biopharm Stat 2001;11(1-2):9–21
42. Beck AT, Ward CH, Mendelson M, et al. An inventory for measuring depression. Arch Gen Psychiatry 1961;4:561–571
43. Bech P, Olsen LR, Kjoler M, et al. Measuring well-being rather than the absence of distress symptoms: a comparison of the SF-36 Mental Health subscale and the WHO-Five Well-Being Scale. Int J Methods Psychiatr Res 2003;12:385–401
44. Gluud LL. Bias in clinical intervention research. Am J Epidemiol 2006;163(6):493–501
45. Lin E, Von Korff M, Katon W, et al. The role of the primary care physician in patients' adherance to antidepressant therapy. Med Care 1995; 33(1):67–74
46. Blumenthal J, Babyak M, Doraiswamy M, et al. Exercise and pharmacotherapy in the treatment of major depressive disorder. Psychosom Med 2007;69(7):587–596
47. 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
48. Mather AS, Rodriguez C, Guthrie M, et al. Effects of exercise on depressive symptoms in older adults with poorly responsive depressive disorder: a randomised controlled trial. Br J Psychiatry 2002;180: 411–415
49. Vaynman S, Ying Z, Gomez-Pinilla F. Hippocampal BDNF mediates the efficacy of exercise on synaptic plasticity and cognition. Eur J Neurosci 2004;20(10):2580–2590
50. Zheng H, Liu Y, Li W, et al. Beneficial effects of exercise and its molecular mechanisms on depression in rats. Behav Brain Res 2006;168(1): 47–55
51. Den Hartog H, Derix M, Bemmel V, et al. Cognitive functioning in young and middle-aged unmedicated out-patients with major depression: testing the effort and cognitive speed hypotheses. Psychol Med 2003; 33(8):1443–1451
52. Grant M, Thase M, Sweeney J. Cognitive disturbance in outpatient depressed younger adults: evidence of modest impairment. Biol Psychiatry 2001;50(1):35–43
53. Peng Q, Masuda N, Jiang M, et al. The antidepressant sertraline improves the phenotype, promotes neurogenesis and increases BDNF levels in the R6/2 Huntington's disease mouse model. Exp Neurol 2008;210(1): 154–163
54. Ford D, Mead L, Chang P, et al. Depression is a risk factor for coronary artery disease in men. Arch Intern Med 1998;158(13):1422–1426
55. Ferketich A, Schwartzbaum J, Frid D, et al. Depression as an antecedent to heart disease among women and men in the NHANES I study. Arch Intern Med 2000;160(9):1261–1268

