Journal of Consulting and Clinical Psychology
1987, Vol. 55, No. 5, 748–754

0022-006X/87/$00.75

Running Versus Weight Lifting in the Treatment of Depression

Elizabeth J. Doyne, Deborah J. Ossip-Klein, Eric D. Bowman, Kent M. Osborn,
Ilona B. McDougall-Wilson, and Robert A. Neimeyer
University of Rochester

We compared the effectiveness of an aerobic and nonaerobic exercise in the treatment of clinical depression in women. A total of 40 women, screened on the Research Diagnostic Criteria for major or minor depressive disorder, were randomly assigned to an 8-week running (aerobic), weight-lifting (nonaerobic), or wait-list control condition. Subjects were reassessed at mid- and posttreatment, and at 1-, 7-, and 12-month follow-ups. Depression was monitored by the Beck Depression Inventory, Lubin's Depression Adjective Check List, and the Hamilton Rating Scale for Depression; fitness level was assessed using submaximal treadmill testing. Results were remarkably consistent across measures, with both exercise conditions significantly reducing depression compared with the wait-list control condition, and generally appearing indistinguishable from each other. No significant between-group fitness changes were noted. These findings indicate that both types of exercise conditions significantly reduce depression and that these results are not dependent on achieving an aerobic effect.

There is increasing interest in the relation between physical fitness and mental health. Hughes (1984) estimated that more than 1,000 studies have been done in this area, with particular emphasis on exercise as a treatment for depression. Anecdotal reports about the psychological benefits of exercise abound, and numerous studies have reported that aerobic exercise may be an effective treatment for depression (e.g., Doyne, Chambless, & Beutler, 1983; Greist et al., 1978; Klein et al., 1985; McCann & Holmes, 1984). However, enthusiastic claims about the therapeutic use of exercise have often gone beyond the available data. The positive findings to date need to be interpreted with caution because of many methodological weaknesses in most investigations. These include nonrandom treatment assignment, lack of appropriate control groups, small sample size, analogue populations, experimenter bias, and lack of statistical rigor (Hughes, 1984; Simons, McGowan, Epstein, & Kupfer, 1985).

Nevertheless, a critical evaluation of the literature at least suggests antidepressant effects of aerobic exercise (i.e. oxygen-consuming exercises that increase the endurance of the pulmonary and cardiovascular systems). In fact, physiological theories often require the acquisition of a fitness effect as a prerequisite for the psychological benefits. Theoretically, decreased cardiovascular response to stress produces a built-in protection against feeling stress (Cooper, 1968) and provides fewer physical cues to which depressed attributions may be made. On a biochemical level, the antidepressant effects may be explained by increased aminergic transmission (Ransford, 1982) and short-term increases in norepinephrine levels (e.g., Dimsdale & Moss, 1980); however, the data on these and other biological theories are inconclusive.

More psychologically based models do not depend as heavily on acquiring an aerobic training effect. One theory is that exercise resembles a graded-task assignment (e.g., Beck, Rush, Shaw, & Emery, 1979) yielding a sense of mastery as concrete achievements are noted. A more behavioral explanation may be that exercise is an activity that increases response-contingent positive reinforcement (e.g., Lewinsohn & Hoberman, 1982) and thereby reduces depression.

In sum, more empirical support is needed to determine if exercise is a viable treatment for depression, and if it is, what the salient factors are that contribute to the therapeutic effect. Both physiological and psychological mechanisms have been proposed, with physiological explanations usually requiring an aerobic training effect. The purpose of the present study was twofold: to provide a randomized controlled trial of exercise treatment for depression with a clinically depressed population and to evaluate whether an aerobic training effect is a necessary condition for change.

Method

Subjects

Subjects were 40 depressed women recruited through mass media to participate in a study on the effects of physical exercise on depression. All of the subjects met the following criteria: (a) 18–35-year-old women; (b) diagnosed as having a major or minor depressive disorder, using the Research Diagnostic Criteria (RDC; Spitzer, Endicott, & Robins,

Appreciation is expressed to Stas Lotyczewski for his assistance with statistical analyses, to Albert B. Craig for providing guidelines and equipment for exercise testing, to Al Altobello for his assistance in data recording, and to the University of Rochester Sports Center for use of their indoor track and universal weight-lifting equipment.

Eric Bowman is now at the Department of Leadership and Law, U.S. Naval Academy, Annapolis, Maryland 21402. Robert Neimeyer is now at the Department of Psychology, Memphis State University, Memphis, Tennessee 38152.

Correspondence concerning this article should be addressed to Deborah J. Ossip-Klein, Department of Psychology, Meliora Hall, University of Rochester, Rochester, New York 14627.

Table 1
Subject Characteristics in Groups Studied for Effects of Exercise on Depression

Characteristic	Track	Universal	Wait list	Total
Age				
M	28.58	27.67	29.46	28.52
SD	3.48	4.81	4.68	4.36
Weight				
M	143.29	156.53	161.65	154.22
SD	27.95	43.50	43.53	39.24
Years of education				
M	14.58	14.80	14.50	14.64
SD	1.98	2.11	1.68	1.90
Marital status (%)				
Single	22.22	63.64	50.00	46.67
Married	55.56	18.18	40.00	36.67
Separated/divorced	22.22	18.18	10.00	16.67
Employed (%)	87.50	91.67	87.50	89.29
RDC diagnosis (%)				
Major	72.73	91.67	69.23	77.78
Minor	27.27	8.33	30.77	22.22
No. of depression episodes				
M	4.30	4.44	4.90	4.55
SD	2.95	6.37	4.09	4.45

Note. RDC = Research Diagnostic Criteria; Major = major depressive disorder; Minor = minor depressive disorder.

1978), (c) agreed to participate in any of the three conditions, (d) agreed not to receive any other form of treatment for depression or engage in exercise other than that prescribed in this study, (f) had not been hospitalized for depression within the preceding year, (g) showed no manic–depressive disorder, imminent suicide threat, or physical contraindications to exercise, and (h) had received written permission from a health-care provider to participate in either of the two exercise programs.

A total of 285 women responded to the advertisements for the study; 146 attended initial screening meetings. In all, 57 women met the criteria for participation, 40 completed treatment (or control period), and 32 completed all of the required follow-up assessments.

Nearly 78% of completers had a diagnosis of major depressive disorder and an average of 4.55 previous depression episodes. Although generalization may be limited to women who participate in an exercise program, this sample of solicited subjects is generalizable to a clinical population (Last, Thase, Hersen, Bellack, & Himmelhoch, 1984). Subject characteristics for this final group are summarized in Table 1. A series of analyses of variance (ANOVAS) and chi-squares revealed no significant between-group differences on any demographic variables ($p \geq .10$).

Screening. Potential subjects were first screened by telephone and then scheduled for a group information meeting, at which time the program, random assignment, and a $60 deposit requirement were described. Deposit refunds in $5 allotments were made contingent on attending exercise and assessment sessions and completing all of the forms. Interested participants signed an informed consent, completed the baseline psychological assessment battery described ahead, and were scheduled for the Schedule for Affective Disorders and Schizophrenia (SADS; Endicott & Spitzer, 1978) interview. Subjects who met the RDC for major or minor depressive disorder scheduled a pretreatment physical fitness test within 1 week. Individuals who were not included in the study were provided referral assistance.

Subjects were matched on baseline Beck Depression Inventory (BDI; Beck & Beamesderfer, 1974) scores in intervals of ≤19, 20–29, and ≥30, and then randomly assigned to conditions. At the pretreatment fitness test, subjects were informed of their group assignment, completed a pretreatment psychological assessment battery, and were interviewed using the Hamilton Rating Scale for Depression (HRSD; Hamilton, 1960).

Treatment overview. All exercise sessions were conducted individually at the University of Rochester Sports Center, under the supervision of 12 trained undergraduate exercise monitors. As a precaution, all monitors completed the Red Cross CPR (cardiopulmonary resuscitation) course. Monitors were rotated between treatment groups and across sessions to minimize nonspecific therapeutic effects due to continued contact with any particular monitor.

Following standard guidelines for exercise prescription (American College of Sports Medicine, 1980), subjects were advised to attend four exercise sessions per week (3 sessions required for reimbursement) over an 8-week period. Each session included an initial 5–10 min warm-up period consisting of specified stretching exercises. The subject then began her exercise program, which was followed by a 5–10 min cool-down period.

Running condition. Subjects walked or ran around a ⅛-mile indoor track. At 7-min intervals, the exercise monitor took the subject's pulse rate and instructed her to adjust her speed in order to achieve and maintain a prescribed target heart rate of 80% of estimated work capacity (220 bpm minus subject's age; American College of Sports Medicine, 1980).

Weight-lifting condition. Subjects used a Universal Exercise Machine and went through a standard 10-station program paced to allow their heart rate to remain at or below 50%–60% of estimated maximum heart rate (i.e., nonaerobic exercise). Subjects' heart rate was monitored as described, to ensure that it remained below the target range.

Wait-list control group. Wait-list subjects were told that their exercise program would be delayed for 8 weeks, after which they could choose their type of program. To control for potential effects of assessment, per se, during this period, subjects were assessed at 4 and 8 weeks, using procedures identical to those for the exercise groups. In addition, to monitor severity of depression, subjects completed mailed self-report psychological batteries at 2-week intervals between assessment sessions; no threatening deteriorations in depression were found.

Assessment. All of the subjects were assessed at baseline (screening) and pre-, mid-, and posttreatment; exercise groups were also assessed at 1-, 7-, and 12-month follow-up intervals.

Measures

SADS. Semistructured interviews using the SADS (Endicott & Spitzer, 1978) were conducted to determine diagnoses of major or minor depressive disorder according to the Research Diagnostic Criteria. Interviews were administered during screening by two clinical psychologists. Interrater reliability across all symptoms was .89 (calculated as agree/[agree + disagree]); reliability for inclusion in the study (major or minor depressive disorder vs. all other disorders, based on 12 cases) was 100%.

BDI. The BDI (Beck & Beamesderfer, 1974) is a self-administered questionnaire. Higher scores indicate more depression.

Depression Adjective Check Lists, Form A (DACL). The DACL (Lubin, 1965) consists of self-report checklists containing both depressed adjectives (*plus* words) and happy adjectives (*minus* words). Higher scores indicate more depression.

HRSD. The HRSD (Hamilton, 1960) is a clinician-rated scale for depression, based on a semistructured interview. Higher scores indicate more depression. Raters were four trained undergraduates (O'Hara & Rehm, 1983) and one master's level assistant. Raters were not informed of subjects' condition assignments; interviews were taped and cross scored for determination of reliabilities. Interrater reliabilities, based on at least nine interviews per rater, ranged from .75 to .94 for all rater pairs.

Physical fitness. Cardiovascular fitness was assessed at pre- and posttreatment and at follow-ups using a modified submaximal graded exercise treadmill test (Astrand & Rodahl, 1977) on a Quinton Instruments "Cardio-Exercise Treadmill" Model 14-44A. Level of cardiovascular fitness was determined by the estimated physical work capacity in METS (work metabolic rate/resting metabolic rate; American College of Sports Medicine, 1980). Prior to the test, subjects' weight was obtained; then, all of the subjects were tested according to the following protocol: 5 min at 0% grade (2.5 METS), 3 min at 5% grade (4.2 METS), 3 min at 10% grade (5.9 METS), 1-min rest period, and 5 min at 20% grade (9.3 METS). Treadmill speed was 2.5 mph. Heart rate was monitored using a Novateck Pulserate Monitor Model 18-0100; individual work capacity was translated into METS.

Results

One-way ANOVAS on baseline and prescores showed no significant differences between groups ($p \geq .11$) on any dependent measure except METS, with universal lower than track and wait list on pre-METS score (11.34 ± 1.78, 14.32 ± 3.35, and 14.46 ± 3.75, respectively), $F(2, 35) = 5.24$, $p < .01$.

Outcome data were generally analyzed using either multivariate analyses of variance (MANOVAS), ANOVAS, or analyses of covariance (ANCOVAS; METS only) across three conditions (track, universal, and wait list) during treatment, and two conditions (track and universal) through the 12-month follow-up. Post hoc analyses were run using the Scheffé method, and F tests for simple effects were calculated for interactions ($p \leq .05$ for all reported). Two sets of analyses were run to deal with missing data, one with missing values replaced with cell means when at least 90% of the scores were available, and one without replacement. The former set is reported, as no discrepancies were found. Mean scores for all depression measures are presented in Table 2.

Beck Depression Inventory and Hamilton Rating Scale for Depression

The BDI and HRSD scores were analyzed using a 3 × 3 (Condition × Time [pre, mid, post]) mixed MANOVA (see Figures 1 and 2). A significant time effect, $F(4, 138) = 14.98$, $p < .01$, and Condition × Time interaction, $F(8, 138) = 4.78$, $p < .01$, were found. Scheffé tests showed significant reductions in depression scores over time for both exercise groups on both measures and no changes for the wait-list condition. The F tests for simple effects showed significant time effects for track and universal groups on the BDI and for all three groups on the HRSD. At posttreatment, track and universal were significantly lower than wait list on both measures.

Changes for the two exercise groups through the 1-month follow-up were analyzed using a 2 × 4 (Condition × Time [pre–1 month]) mixed MANOVA. Results showed a significant time effect, $F(6, 142) = 16.88$, $p < .01$, with pre- significantly greater than mid- and posttest and 1-month scores for both HRSD and BDI, and no significant differences between exercise groups.

Additional assessment periods were conducted for the BDI at baseline, 7 months, and 12 months. Therefore, univariate ANOVAS were also run for BDI alone. A 3 × 4 (Condition × Time [baseline–post]) ANOVA showed a significant time effect, $F(3, 105) = 46.9$, $p < .01$, and Condition × Time interaction, $F(6, 105) = 5.47$, $p < .01$. Again, baseline and prescores were significantly higher than mid- and postscores, and significant improvements occurred over time for track and universal only, with both track and universal having significantly lower postscores than did wait list. A 2 × 7 (Condition × Time [baseline–12 months]) ANOVA yielded a significant time effect, $F(6, 144) = 43.77$, $p < .01$, with baseline and prescores higher than all others, and no significant differences between exercise groups.

Finally, change in depression on the BDI during treatment was measured by chi-square analyses on number of nondepressed subjects in each group (i.e., BDI score < 9). A 3 × 2 (Condition × Depressed/Nondepressed) chi-square showed no initial differences between groups on pre-BDI scores ($p \geq .25$), but significant differences between groups on postscores, $\chi^2(2, N = 39) = 11.63$, $p < .01$. Percentage nondepressed (i.e., BDI < 9) subjects at post were 67%, 80%, and 17% for track, universal, and wait list, respectively. A 2 × 2 (Condition Depressed/Nondepressed) chi-square on post-BDI showed no significant differences between the two exercise groups ($p > .50$). Percentage nondepressed at follow-up for track and universal, respectively, were 82% and 79% (1 month), 75% and 80% (7 months), and 29% and 80% (12 months). Again, chi-square analyses (using Yates correction where appropriate) were not significant. (Note that although the 12-month differences appear significant, they must be viewed cautiously, as cell sizes were quite small at this measurement point.)

Depression Adjective Check Lists, Form A

Changes in total DACL scores during treatment were assessed using a 3 × 4 (Condition Time [baseline–post]) mixed ANOVA. Significant effects were found for condition, $F(2, 28) = 5.8$, $p < .01$, time, $F(3, 84) = 29.76$, $p < .01$, and Condition × Time, $F(6, 84) = 3.85$, $p < .01$. Post hoc analyses showed that both track and universal were significantly lower than wait list, and baseline and prescores were significantly higher than mid, which were significantly higher than post. The F tests for simple effects showed significant time effects for all three groups and significant condition effects at mid and post, with tract and universal lower than wait list at both points.

Changes in total DACL through follow-up were analyzed using a 2 × 7 (Condition × Time [baseline–12 months]) mixed ANOVA. Significant time, $F(6, 144) = 24.00$, $p < .01$, and Condition × Time interaction, $F(6, 144) = 2.28$, $p < .05$, effects were found, with baseline and prescores higher than scores at all other periods, significant improvements over time for both track and universal, and a significant condition effect at 12 months only, with scores lower for universal than for track.

Changes in plus and minus components of the DACL were analyzed using MANOVAS. A 3 × 4 (Condition × Time [baseline–post]) mixed MANOVA showed significant effects for condition, $F(4, 66) = 4.94$, $p < .01$, and time, $F(6, 202) = 11.67$, $p < .01$. Post hoc tests showed that track was lower than wait list for plus scores, and both track and universal were lower than wait

Table 2
Mean Scores and Standard Deviations Across All Depression Measures

Group	Baseline	Pre	Mid	Post	1 month	7 months	12 months
			Beck Depression Inventory[a]				
Track							
M	22.27	19.27	11.00	8.18	5.46	9.14	10.45
SD	6.36	5.61	3.38	5.27	4.32	6.55	5.16
Universal							
M	22.07	19.50	10.96	5.93	6.13	5.37	6.47
SD	7.45	9.18	8.69	8.44	8.31	8.20	6.68
Wait list							
M	20.17	16.06	15.72	15.25			
SD	6.87	5.12	5.06	6.30			
			Hamilton Rating Scale for Depression[a]				
Track							
M		13.36	8.82	6.64	3.46		
SD		4.43	3.68	3.61	2.84		
Universal							
M		13.80	7.78	5.13	5.73		
SD		3.82	4.38	3.44	4.23		
Wait list							
M		12.58	10.23	13.58			
SD		4.58	3.16	5.14			
			Depression Adjective Check List, Form A[a]				
Track							
M	15.09	13.55	10.91	7.55	6.91	9.09	10.64
SD	3.27	4.30	3.96	2.38	2.34	4.92	4.03
Universal							
M	16.07	16.51	9.13	8.01	8.52	7.93	7.70
SD	4.48	4.32	4.88	4.84	3.46	4.25	2.17
Wait list							
M	17.45	15.81	16.57	12.92			
SD	3.64	2.88	4.39	5.73			
			Weight				
Track							
M		143.8		141.9	137.0		
SD		28.99		28.65	22.68		
Universal							
M		153.9		154.5	151.6		
SD		43.84		43.39	42.66		
Wait list							
M		163.5		166.9			
SD		44.94		45.82			
			METS				
Track							
M		14.32		14.60	16.12	17.07	18.15
SD		3.18		3.59	4.16	4.71	4.81
Universal							
M		11.35		13.20	14.03	13.48	12.82
SD		1.72		3.03	3.00	1.91	1.64
Wait list							
M		14.46		12.74			
SD		3.57		2.77			

Note. METS = working metabolic rate/resting metabolic rate.
[a] Lower scores indicate improvement.

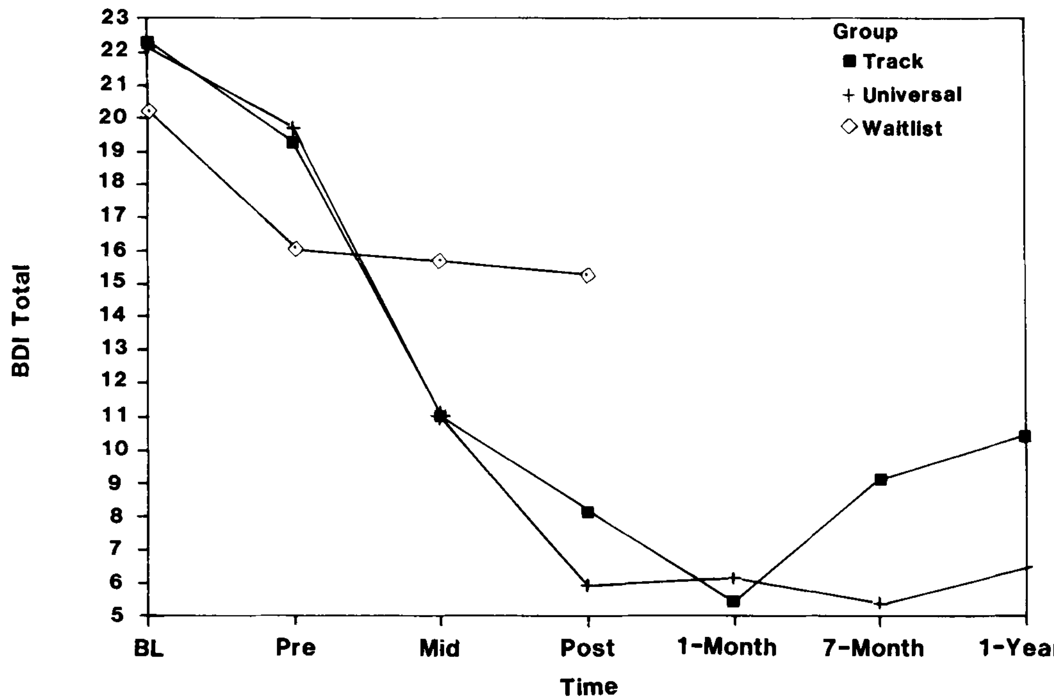

Figure 1. Beck Depression Inventory (BDI) scores through 1-year follow-up.

list for minus scores; baseline and prescores were higher than mid and postscores for plus, and baseline was higher than mid and post, with pre higher than post for minus. A 2 × 7 (Condition × Time [baseline–12 months]) mixed MANOVA showed a significant time effect, $F(12, 286) = 10.56, p < .01$, with baseline and pre scores higher than all others for plus and minus.

Weight

Due to potential confounding effects of weight on treadmill testing, a 3 × 2 (Condition × Pre, Post) ANOVA was run on weights. No significant main effects were found; however, the Condition × Time interaction, $F(2, 35) = 4.03, p = .03$, showed a significant increase in weight for the wait list over time, $F(1, 35) = 6.87, p < .01$.

METS

A 3 × 2 (Condition × Pre, Post) ANCOVA and 2 × 4 (Track, Universal × Post–12 Months) ANCOVA covarying pre-METS showed no significant changes in METS between groups or across time.

Although not statistically significant, weight differences across groups appeared clinically meaningful. Therefore, a second set of ANCOVAs were run on METs, covarying preweights. The 3 × 2 analysis showed a significant interaction, $F(2, 32) = 3.92, p = .03$, with track and universal increasing and wait list decreasing over time. The 2 × 5 (Pre–12 Months) analysis showed a significant time effect, $F(4, 52) = 3.37, p = .02$, with both exercise groups improving somewhat over time.

Attrition

Because differential attrition across groups could confound fitness results, dropout was examined for all of the subjects who completed pretest measures. Dropout rates were 40%, 29%, and 13% for track, universal, and weight list, respectively. A chi-square showed no significant differences across groups. Further, *t* tests and ANOVAs showed no significant differences in depression scores at baseline, pre, or post between dropouts and completers ($p > .22$).

Exercise Frequency

Although subjects were instructed to exercise four times per week, the actual frequency was 2.8 sessions per week (M track = 2.64 ± .79; M universal = 2.95 ± .77, $t = ns$, $p = .31$). Partial correlations between frequency and postscore, controlling for prescore, showed no significant relation on any outcome variable ($p > .25$).

Discussion

The present study showed both statistically and clinically significant decreases in depression in the two exercise groups relative to the wait-list control group. In addition, these improvements were reasonably well maintained through the 1-year follow-up. No significant overall differences in depression were found between the two exercise groups. These findings were remarkably consistent across measures, with a nonsignificant trend on some measures favoring universal. No between-group differences in fitness were found, although there was suggestion

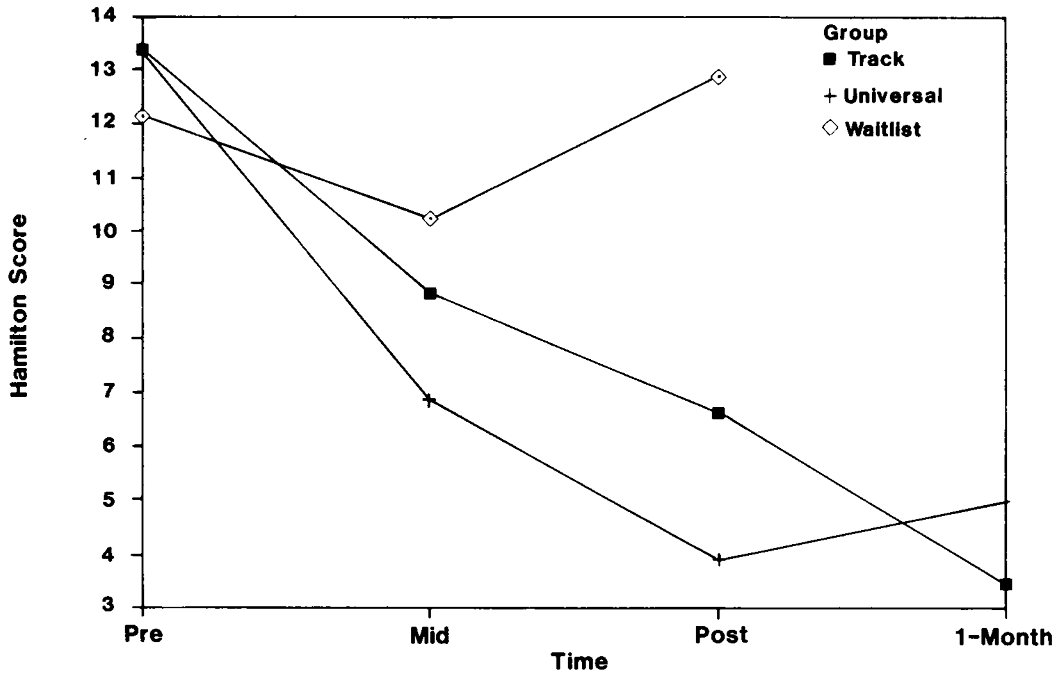

Figure 2. Hamilton Rating Scale for Depression Scores through 1-month follow-up.

of a small improvement in both exercise groups over time. Although it is still possible that an exercise program that achieves an aerobic effect could produce even further improvements in depression, the most conservative conclusion to be drawn from these data is that it is clearly not necessary for exercise to improve cardiovascular fitness in order to demonstrate significant reductions in depression.

The lack of a fitness difference between groups engaged in aerobic and nonaerobic type exercises is somewhat puzzling. It is unlikely that this is due to differential attrition across groups. Although a nonsignificant trend for higher dropout in track was found, this should have biased the group to maintain *more* physically fit subjects and to enhance, not reduce, fitness differences between track and universal. However, several other factors may have contributed to the similarity of the two groups. If universal subjects had engaged in aerobic exercise on their own during treatment, fitness differences between groups could have been attenuated. This did not appear to be the case. At posttreatment, only 2 subjects reported aerobic exercise at least once a week during treatment; ANOVAS eliminating these subjects yielded results identical to those previously reported. It is also possible that the submaximal treadmill test used was simply not sensitive enough to detect real differences. Our concern regarding the safety of using a vigorous exercise test with sedentary, clinically depressed subjects may have led to the choice of a protocol that was "too easy." There are also considerable sources of error in predicting maximum fitness level from submaximal tests (Astrand & Rodahl, 1977), so we may have misjudged subjects' initial levels of fitness, as well as their maximum capacity. A more individually prescribed program (rather than relying on the guideline of 220 bpm–age) may have resulted in more significant fitness changes. Other possibilities are that the intensity, length, or actual frequency of the running program was not sufficient to produce improvements in cardiovascular fitness. However, because frequency was not related to postscore on any outcome measure, the observed frequencies were sufficient to produce changes in depression. Nevertheless, fitness changes might be seen with a longer protocol.

The lack of a requisite fitness effect suggests a renewed focus on behavioral and cognitive mechanisms associated with the act of exercise itself. Engaging in a regular exercise program may produce a sense of accomplishment and enhanced self-efficacy as visible improvements in performance occur. Such changes would be consistent with cognitive/behavioral approaches that underscore the importance of mastery experiences and response-contingent positive reinforcement in the reduction of depression. It is not clear whether there are specific effects of exercise over and above other types of treatment that produce comparable mastery experiences. Further, the role of the regular, personal contact by exercise groups with project staff during the 8-week exercise period cannot be determined. Several attempts were made to minimize personal contact as an alternate explanation for results. First, wait-list subjects attended assessment sessions on the same schedule as did treatment subjects, thus providing some standardized, comparable contact with the control group. Second, treatment subjects exercised individually to avoid a group effect; and finally, exercise monitors were rotated across groups and sessions to minimize the development of personal relationships with subjects. In addition, previous studies suggest improvements in depression for exercise conditions relative to contact-only or placebo control groups (Doyne et al., 1983; McCann & Holmes, 1984). Nevertheless, a nonexercise control group with comparable contact and mastery experiences could help identify potentially unique psychological effects of exercise on depression.

In sum, both running and weight-lifting exercise conditions clearly reduced depression in this population of moderately to severely depressed women. Although it is not possible to isolate the specific mechanisms for change, the present findings do disconfirm the hypothesis that improved cardiovascular fitness is necessary for an exercise treatment effect. Further, the comparable effects of running and weight lifting may allow for more treatment flexibility and individual choice in the selection of an exercise program for depressed women.

References

American College of Sports Medicine. (1980). *Guidelines for graded exercise testing and exercise prescription* (2nd ed.). Philadelphia, PA: Lea & Febiger.

Astrand, P., & Rodahl, K. (1977). *Textbook of work physiology* (2nd ed.). New York: McGraw-Hill.

Beck, A. T., & Beamesderfer, A. (1974). Assessment of depression: The Depression Inventory. In P. Pichot (Ed.), *Psychological measurements in psychopharmacology* (pp. 1–10). Basel, Switzerland: Karger.

Beck, A. T., Rush, A. M., Shaw, B. F., & Emery, G. (1979). *Cognitive therapy of depression.* New York: Guilford Press.

Cooper, K. H. (1968). *Aerobics.* New York: Wiley.

Dimsdale, J. E., & Moss, J. (1980). Plasma catecholamines in stress and exercise. *Journal of American Medical Association, 243,* 340–342.

Doyne, E. J., Chambless, D. L., & Beutler, L. E. (1983). Aerobic exercise as a treatment for depression in women. *Behavior Therapy, 14,* 434–440.

Endicott, J., & Spitzer, R. S. (1978). A diagnostic interview: The Schedule for Affective Disorders and Schizophrenia. *Archives of General Psychiatry, 35,* 837–844.

Greist, J. H., Klein, M. H., Eischens, R. R., Faris, J., Gurman, A. S., & Morgan, W. P. (1978). Running through your mind. *Journal of Psychosomatic Research, 22,* 259–294.

Hamilton, M. (1960). A rating scale for depression. *Journal of Neurological and Neurosurgical Psychiatry, 23,* 56–61.

Hughes, J. R. (1984). Psychological effects of habitual aerobic exercise: A critical review. *Preventive Medicine, 13,* 66–78.

Klein, M. H., Greist, J. H., Gurman, A. S., Neimeyer, R. A., Lesser, D. P., Bushnell, N. J., & Smith, R. E. (1985). A comparative outcome study of group psychotherapy vs. exercise treatments for depression. *International Journal of Mental Health, 13,* 148–175.

Last, C. G., Thase, M. E., Hersen, M., Bellack, A. S., & Himmelhoch, J. H. (1984). Treatment outcome for solicited versus nonsolicited unipolar depressed female outpatients. *Journal of Consulting and Clinical Psychology, 52,* 134.

Lewinsohn, P. M., & Hoberman, H. M. (1982). Depression. In A. S. Bellack, M. Hersen, & A. E. Kazdin (Eds.), *International handbook of behavior modification and therapy* (pp. 397–429). New York: Plenum.

Lubin, B. (1965). Adjective checklist for the measurement of depression. *Archives of General Psychiatry, 12,* 47–62.

McCann, I. L., & Holmes, D. S. (1984). Influence of aerobic exercise on depression. *Journal of Personality and Social Psychology, 46,* 1142–1147.

O'Hara, M. W., & Rehm, L. P. (1983). Hamilton Rating Scale for Depression: Reliability and validity of judgments of novice raters. *Journal of Consulting and Clinical Psychology, 51,* 318–319.

Ransford, C. (1982). A role for amines in the antidepressant effect of exercise: A review. *Medicine and Science in Sports and Exercise, 14,* 1–10.

Simons, A. D., McGowan, M. S., Epstein, L. H., & Kupfer, D. J. (1985). Exercise as a treatment for depression: An update. *Clinical Psychology Review, 5,* 553–568.

Spitzer, R. L., Endicott, J., & Robins, E. (1978). Research Diagnostic Criteria: Rationale and reliability. *Archives of General Psychiatry, 35,* 773–782.

Received March 14, 1986
Revision received January 8, 1987
Accepted January 8, 1987 ■