

**MS #5** 

# Aerobic training to improve fitness and quality of life for clients with multiple sclerosis

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

Level IA3a

Randomized controlled trial, 20 or more participants per condition, low internal validity, high external validity

#### **Clinical bottom line**

For patients with multiple sclerosis, an exercise program **significantly** (*see Glossary*) improved overall fitness, ambulation, mobility, body care, and movement scores. Further, although exercise did **not significantly** (*see Glossary*) improve mood, those who exercised demonstrated a trend toward improved mood.

#### Sample

The study involved 46 participants who were recruited from the community through the Multiple Sclerosis Society and physicians' referrals: 15 were men, 31 women. They averaged 40 years in age. (The original sample consisted of 54 participants; 8 were eventually excluded.) All had a diagnosis of multiple sclerosis; a score of 6.0 or less on Kurtzke's Expanded Disability Status Scale (indicating mild to moderate disability); and no history of a cardiovascular, respiratory, orthopedic, metabolic, or other medical condition that would preclude their participating in the training. None had been involved in a physical activity program within the preceding 6 months.

#### Procedures

The participants were randomly assigned to a treatment or a control group. The treatment group participated in a supervised program of aerobic exercise for 15 weeks, three times a week. This involved 40 minutes of aerobic exercise followed by a 10-minute stretching period. A therapist delivered the intervention in a clinic.

The participants in the control group agreed not to alter their current level of physical activity for the duration of the study.

At the start of the study and at 5, 10, and 15 weeks into it, a graded exercise test and psychological instruments were administered to all the participants. At the start and at 15 weeks, neurological evaluations, blood lipid profiles, and body composition analyses were conducted.

#### Outcomes

Eight outcome areas were of interest to the researchers: *neurological status* (as measured by Kurtzke's Functional System Scales, Kurtzke's Expanded Disability Status Scale, and the Incapacity Status Scale); *workload capabilities*— specifically maximum aerobic capacity (as measured by an arm-and-leg cycle ergometer ["an apparatus for measuring

the work performed by a group of muscles," *Webster's Ninth New Collegiate Dictionary*, s.v.]), oxygen uptake (as measured by an open-circuit indirect calorimeter [an apparatus "for measuring quantities of absorbed or evolved heat or for determining specific heats," *Webster's Ninth New Collegiate Dictionary*, s.v.]), and heart rate (as measured by an electrocardiogram); *strength* (as measured by a computerized force-measurement system); *body composition*— specifically skinfold thickness (as measured with a skinfold caliper) and percentage of body fat (estimated using equations); *blood lipid profiles; affective states*—tension-anxiety, depression-dejection, anger-hostility, and so forth (as measured by the Profile of Mood States); *sickness-related changes in behavior* (as measured by the Sickness Impact Profile); and *the impact of fatigue on daily function* (as measured by the Fatigue Severity Scale).

#### Analyses

The researchers compared scores from the exercise group after the intervention with scores from the control group after the intervention.

#### Results

Compared with the control group, the exercise group showed significant improvement over time for aerobic capacity, physical work capacity, and the physical dimension of the Sickness Impact Profile (assessing ambulation, mobility, and body care and movement). The exercise group also showed a **significant** (*see Glossary*) decrease in skinfold thickness and an improvement in strength.

On the psychological measurements, there was an overall **nonsignficant** (*see Glossary*) trend toward decreased bad moods for the exercise group. Further, on the psychosocial—both psychological and social—dimensions of sickness-related changes in behavior, there was a general nonsignficant trend toward improvement for the exercise group.

Significance and **effect sizes (***r***)** (*see Glossary*) for outcome measures comparing the treatment and control groups for Petajan et al. (1996)

| Outcome              | Significance   | Clinical effect (r) | Size of effect |
|----------------------|----------------|---------------------|----------------|
| VO2                  | Significant    | 0.87                | Large          |
| Work capacity        | Significant    | 0.78                | Large          |
| Max heart rate       | Nonsignificant | 0.45                | Medium         |
| Handgrip             | Nonsignificant | 0.19                | Small          |
| UE strength          | Significant    | 0.81                | Large          |
| LE strength          | Significant    | 0.44                | Medium         |
| POMS                 |                |                     |                |
| Tension              | Nonsignificant | 0.53                | Large          |
| Depression           | Nonsignificant | 0.76                | Large          |
| Anger                | Nonsignificant | 0.81                | Large          |
| Vigor                | Nonsignificant | 0.23                | Small          |
| Fatigue              | Nonsignificant | 0.63                | Large          |
| Confusion            | Nonsignificant | 0.71                | Large          |
| SIP                  |                | 0.57                | Large          |
| Ambulation           | Nonsignificant | 0.86                | Large          |
| Mobility             | Nonsignificant | 0.95                | Large          |
| Body care & movement | Nonsignificant | 0.62                | Large          |
| Physical dimensions  |                |                     |                |
| Social interaction   | Nonsignficant  | 0.73                | Large          |
| Communication        | Nonsignficant  | 0.46                | Medium         |
| Alertness behavior   | Nonsignficant  | 0.42                | Medium         |
| Emotional behavior   | Nonsignficant  | 0.72                | Large          |

| Outcome                 | Significance  | Clinical effect (r) | Size of effect |
|-------------------------|---------------|---------------------|----------------|
| Psychological dimension |               |                     |                |
| Sleep rest              | Nonsignficant | 0.69                | Large          |
| Eating                  | Nonsignficant | 0.15                | Small          |
| Work                    | Nonsignficant | 0.15                | Small          |
| Home management         | Nonsignficant | 0.24                | Small          |
| Recreation & pastimes   | Nonsignficant | 0.81                | Large          |

### Limitations

The exercise group received more attention than the control group, and that may have influenced their outcomes (**attention bias**) (*see Glossary*). Also, the therapists evaluating the participants were not **blinded** (*see Glossary*) to group status. Further, the functional assessment relied on self-recall (**self-recall biases**) (*see Glossary*). Lastly, eight participants dropped out of the exercise program (**lost to follow-up**) (*see Glossary*).

## Glossary

**attention bias**—Also known as the Hawthorne effect, participants who receive some form of attention during treatment will often change their behavior, not because of the treatment per se, but because they are receiving attention. This bias is most frequently seen when the control group is wait listed or receives no treatment.

**blinded/blinding**—Blinding refers to the practice of keeping members of the research study unaware of which group a participant is assigned to (treatment or control) in the study. Single blinding usually refers to keeping study participants unaware of whether they are receiving the experimental or the sham treatment. Double blinding usually refers to keeping the participants and those who are administering the treatment unaware of who is receiving the experimental and who is receiving the sham treatments. In some cases, where it is impossible to blind those administering treatment, the individuals who are administering the outcome measures can be blinded to group status.

Studies in which blinding does not occur can have significant biases. When the participants know that they are receiving the experimental treatment, they often get better because they think they ought to (this is often referred to as the placebo effect). When researchers know that a participant is receiving the experimental treatment, they often subconsciously favor those participants when evaluating them on outcome measures. For instance, when timing a participant in the treatment group, researchers may unknowingly stop the watch a little faster or slower so the treatment participant seems to do better.

**effect sizes** (Cohen's *r*)—An effect size is a measure of clinical significance. It provides information about the magnitude of effect of the treatment. Although related to significance, it is not as influenced by the size of the sample. Therefore, it is possible to have an outcome on which the treatment had a large effect (e.g., the treatment group improved a lot more than the control group) and still have a nonsignificant result. If the results have a large effect but no significance, this means that this effect may be sample specific and not generalizable outside the study. There are many different types of effect sizes. What is reported here is Cohen's *r* can be interpreted in a manner similar to a Pearson's correlation coefficient:

| Effect size r | Size of the effect |
|---------------|--------------------|
| <0.99         | Negligible         |
| 0.10 - 0.29   | Small              |
| 0.30 - 0.49   | Medium             |
| >0.50         | Large              |

Cohen, J. (1977). Statistical power analysis for behavioral sciences. New York: Academic Press.

**lost to follow-up**—Participants who are lost to follow-up may differ from those who remain in the study; for instance, they may be sicker or have family members who are less supportive. If those who were lost to follow-up remained in the study, the results might have been different. In general, a study with more than 20% lost to follow-up is probably biased.

**nonsignificant or no significance**—A statistical term that refers to study findings that are likely to be due to chance differences between the groups rather than to other factors (like the treatment of interest). A nonsignificant result is not generalizable outside the study. Like significance, a nonsignificant result does not indicate the clinical effect. Often studies will show nonsignificant results, yet the treatment group's mean will be better than the control group's. This is usually referred to as a trend in the right direction. Because significance is closely determined by sample size, nonsignificant results would often become significant if the sample size were increased.

**self-recall biases**—*Obsequiousness bias:* Participants may alter their responses to questions so they match the responses that they perceive that the researcher wants. *Unacceptability bias:* Participants may underreport behaviors that they view as unacceptable or embarrassing (e.g., alcohol consumption, inability to complete basic ADLs, etc.).

**significance (or significant)**—A statistical term, this refers to the probability that the results obtained in the study are not due to chance, but to some other factor (such as the treatment of interest). A significant result is likely to be generalizable to populations outside the study.

Significance should not be confused with clinical effect. A study can be statistically significant without having a very large clinical effect on the sample. For example, a study that examines the effect of a treatment on a client's ability to walk may report that the participants in the treatment group were able to walk significantly longer distances than the control group. However, if you read the study you may find that the treatment group was able to walk, on average, 6 feet, whereas the control group was able to walk, on average, 5 feet. Although the outcome may be statistically significant, a clinician may not believe that a 1-foot increase will improve his or her client's function.

Terminology used in this document is based on two systems of classification current at the time the evidence-based literature reviews were completed: Uniform Terminology for Occupational Therapy Practice—Third Edition (AOTA, 1994) and International Classification of Functioning, Disability and Health (ICIDH-2) (World Health Organization [WHO], 1999). More recently, the Uniform Terminology document was replaced by Occupational Therapy Practice Framework: Domain and Process (AOTA, 2002), and modifications to ICIDH-2 were finalized in the International Classification of Functioning, Disability and Health (WHO, 2001).

This work is based on the evidence-based literature review completed by Nancy Baker, ScD, OTR, and Linda Tickle-Degnen, PhD, OTR/L, FAOTA.

For more information about the Evidence-Based Literature Review Project, contact the Practice Department at the American Occupational Therapy Association, 301-652-6611, x 2040.



Copyright 2003 American Occupational Therapy Association, Inc. All rights reserved. This material may be reproduced and distributed without prior written consent.