

Is a Long-Term High-Intensity Exercise Program Effective and Safe in Patients With Rheumatoid Arthritis?

Results of a Randomized Controlled Trial

Zuzana de Jong,¹ Marten Munneke,¹ Aeilko H. Zwinderman,¹ Herman M. Kroon,¹ Annemarie Jansen,² Karel H. Ronday,³ Dirkjan van Schaardenburg,² Ben A. C. Dijkmans,⁴ Cornelia H. M. Van den Ende,⁵ Ferdinand C. Breedveld,¹ Theodora P. M. Vliet Vlieland,¹ and Johanna M. W. Hazes⁶

Objective. There are insufficient data on the effects of long-term intensive exercise in patients with rheumatoid arthritis (RA). We undertook this randomized, controlled, multicenter trial to compare the effectiveness and safety of a 2-year intensive exercise program (Rheumatoid Arthritis Patients In Training [RAPIT]) with those of physical therapy (termed usual care [UC]).

Methods. Three hundred nine RA patients were assigned to either the RAPIT program or UC. The primary end points were functional ability (assessed by the McMaster Toronto Arthritis [MACTAR] Patient Preference Disability Questionnaire and the Health Assessment Questionnaire [HAQ]) and the effects on radiographic progression in large joints. Secondary end points concerned emotional status and disease activity.

The Rheumatoid Arthritis Patients In Training project was supported by grant OG-97-024 from the “College Voor Zorgverzekering” (the Dutch Health Care Insurance Board).

¹Zuzana de Jong, MD, Marten Munneke, MSc, Aeilko H. Zwinderman, PhD, Herman M. Kroon, MD, Ferdinand C. Breedveld, MD, Theodora P. M. Vliet Vlieland, MD, PhD: Leiden University Medical Center, Leiden, The Netherlands; ²Annemarie Jansen, MSc, Dirkjan van Schaardenburg, MD: Jan van Breemen Instituut, Amsterdam, The Netherlands; ³Karel H. Ronday, MD: Leyenburg Hospital, The Hague, The Netherlands; ⁴Ben A. C. Dijkmans, MD: Jan van Breemen Instituut, and Vrije Universiteit Medical Center, Amsterdam, The Netherlands; ⁵Cornelia H. M. Van den Ende, PhD: The Netherlands Institute for Health Services Research, Utrecht, The Netherlands; ⁶Johanna M. W. Hazes, MD, PhD: Erasmus University Medical Center, Rotterdam, The Netherlands, and Leiden University Medical Center, Leiden, The Netherlands.

Address correspondence and reprint requests to Zuzana de Jong, MD, Department of Rheumatology C4-R, Leiden University Medical Center, Post-box 9600, 2300 RC Leiden, The Netherlands. E-mail: z.de_jong@lumc.nl.

Submitted for publication December 2, 2002; accepted in revised form May 1, 2003.

Results. After 2 years, participants in the RAPIT program showed greater improvement in functional ability than participants in UC. The mean difference in change of the MACTAR Questionnaire score was 2.6 (95% confidence interval [95% CI] 0.1, 5.2) over the first year and 3.1 (95% CI 0.7, 5.5) over the second year. After 2 years, the mean difference in change of the HAQ score was -0.09 (95% CI $-0.18, -0.01$). The median radiographic damage of the large joints did not increase in either group. In both groups, participants with considerable baseline damage showed slightly more progression in damage, and this was more obvious in the RAPIT group. The RAPIT program proved to be effective in improving emotional status. No detrimental effects on disease activity were found.

Conclusion. A long-term high-intensity exercise program is more effective than UC in improving functional ability of RA patients. Intensive exercise does not increase radiographic damage of the large joints, except possibly in patients with considerable baseline damage of the large joints.

Regular exercise with a moderate-to-high level of intensity has proven to be effective in improving muscle strength and cardiovascular fitness in both healthy and patient populations (1–12). Cardiovascular risk reduction was found in healthy men and women (5,13) and in patients with hypertension, coronary disease (14), and type 2 diabetes (15) participating in exercise. As a consequence of their disease and its treatment, patients with rheumatoid arthritis (RA) are at greater risk for cardiovascular morbidity and mortality than are their healthy peers (16,17). This is all the more reason long-

Table 1. Eligibility criteria for inclusion in the study*

Age 20–70 years
RA according to ACR 1987 revised criteria (22)
ACR functional classes I–III (23)
Stable DMARD regimen in past 3 months
Able to cycle
Willing to exercise biweekly on fixed schedule
Living within a predefined adherence region of training and/or assessment center
No prosthesis of a weight-bearing joint
No cardiopulmonary disease excluding intensive exercise
No comorbidity causing a short life expectancy
No serious psychiatric disease
Able to complete a questionnaire

* RA = rheumatoid arthritis; ACR = American College of Rheumatology (formerly, the American Rheumatism Association); DMARD = disease-modifying antirheumatic drug.

term regular intensive exercise could be profitable for RA patients. The intuitive approach of giving rest to inflamed joints as well as the fears of damage to the large joints and exacerbation of inflammation often result in advising RA patients against participation in intensive exercise. Evidence is accumulating that intensive weight-bearing exercises improve aerobic fitness and muscle strength of RA patients without any increase in disease activity (1–4,18). An improvement in functional ability with exercise has been demonstrated by the use of functional tests (1,2,4,18–20) and generalized instruments as outcome measures (4,21), but not by means of individualized instruments which mirror the changes in individual function. Also, no reliable data are available on safety with respect to damage of the large joints most loaded by exercise.

The objective of this randomized, controlled trial was to compare the effectiveness and safety of a long-term intensive exercise program with those of physical therapy (termed usual care [UC]).

PATIENTS AND METHODS

Study participants. In 1997, all RA patients registered in 4 outpatient rheumatology clinics and assumed to fulfill the inclusion criteria (Table 1) after their records were screened for demographic and disease-related characteristics were invited by mail to participate in the trial. Patients who were willing to participate, who still met these criteria after a screening by two trained investigators (ZdJ and AJ), and who gave written informed consent were subsequently randomized. The medical ethics committee of each participating center approved the study protocol.

Study protocol. A permutated-blocked randomization (blocks of 4), with stratification for center, age (<50 years and >50 years), and sex, made up by a random digit generator was used to allocate the patients either to a high-intensity exercise program (the Rheumatoid Arthritis Patients In Training

[RAPIT] program) or to UC and to prevent unbalanced distribution. An administrative assistant, not aware of the block size, allocated the interventions.

The patients randomized to the RAPIT group participated in a supervised biweekly group exercise program of 1.25 hours each session. Overall, each session had 3 parts: “bicycle training” (20 minutes), “exercise circuit” (20 minutes), and “sport or game” (20 minutes). Each session was preceded by a “warm-up” and followed by a “cool-down.”

Bicycle load was based on 2 indicators: 1) heart rate during bicycling and 2) rating of perceived exertion (range 0–10). During training, the heart rate was kept at ~70–90% of the predicted maximal heart rate, and the rating of perceived exertion was kept at 4–5.

The “exercise circuit” consisted of 8–10 different exercises intended to improve muscle strength, muscle endurance, joint mobility, and activities of daily living. The proportion of exercise duration/rest duration changed from 90 seconds/60 seconds in the first weeks of the program to 90 seconds/30 seconds after 6 months. Within the exercise circuit, each exercise was repeated 8–15 times.

The “sport or game” section of the program consisted of impact-delivering sporting activities such as badminton, volleyball, indoor soccer, and basketball. The impact loading was also applied during the “warming-up” (stepping aside and jumping) and “exercise circuit” sections of the program.

If necessary, the program was adapted to individual disabilities to reach the same aims. Patients assigned to the UC group were treated by a physical therapist only if this was regarded as necessary by their attending physician.

The attending physicians of patients in each group were informed about the treatment allocation. The physicians had free choice with respect to their medical prescriptions and other treatment strategies, including any form of additional individual physical therapy with the exception of high-intensity weight-bearing exercises. Attendance at any group or individual physical therapy sessions apart from the RAPIT program was recorded in both groups.

At baseline, sociodemographic characteristics were registered along with disease duration, presence of rheumatoid factor, number of disease-modifying antirheumatic drugs (DMARDs) taken since diagnosis except for the current DMARD (“past number of DMARDs”), and radiographic damage of the hands and feet (Larsen/Scott method) (24). The Larsen score of the hands and feet ranges from 0 (no joint space narrowing, no erosions) to 200 (maximal possible damage). The primary end point of effectiveness was functional ability; secondary end points were physical capacity and emotional status. The primary end point of safety was radiographic damage of the large joints, and the secondary end point was disease activity.

Functional ability was assessed with the McMaster Toronto Arthritis (MACTAR) Patient Preference Disability Questionnaire (25) and the Health Assessment Questionnaire (HAQ) (26,27). The MACTAR Questionnaire is a semistructured interview consisting of transitional questions and status questions. Transitional questions concern perceived changes in disease activity and ability to perform previously impaired activities. These activities were elicited from and ranked by the patient at baseline and again at 1 year. The evaluation took place 6 months and 12 months after each elicitation. The

change of weighted score from each baseline could vary from -38 (maximum deterioration) to $+38$ (maximum improvement) (28). The HAQ total score ranged from 0 (no functional limitations) to 3 (serious functional limitations). The change score could thus vary from -3 (maximal improvement) to $+3$ (maximal deterioration).

Physical capacity was determined by aerobic fitness and muscle strength. Aerobic fitness was measured by means of a standardized ergometer test and is given in watts (29). Muscle strength of the knee extensors was measured with an isokinetic dynamometer at an angle velocity of $60^\circ/\text{second}$ and is given in newtons (3).

Emotional status was assessed with the Hospital Anxiety and Depression Scale (HADS). The total HADS score ranges from 0 to 42, and higher scores indicate higher levels of anxiety and/or depression (30).

Radiographic damage of the shoulders, elbows, hips, knees, ankles, and subtalar joints was scored independently by two experienced readers (HMK and ZdJ) using the Larsen method (31) without information about the time sequence, patient's identity, and group allocation. The Larsen score of the large joints (LLJ score) ranges from 0 (no joint space narrowing, no erosions) to 60 (maximal possible damage) and is presented as a mean of the scores by the two readers.

Disease activity was assessed with the original Disease Activity Score with 4 variables (DAS4) (32). The DAS4 is a compiled index based on the number of swollen joints, tender joint score (Ritchie Articular Index [RAI]), erythrocyte sedimentation rate (ESR), and patient's global assessment of general health measured on a visual analog scale. The DAS4 ranges from 0 (no disease activity) to 10 (severe disease activity).

The use of medication in the week preceding the visit ("current use of medication") was registered at baseline and every 3 months along with information on whether participants had a paying job and how many hours per week they spent at this job. Outcome assessments were done at baseline and at 6, 12, 18, and 24 months. Disease activity and physical capacity were assessed every 3 months; radiographic damage of the large joints was assessed only at baseline and at 12 and 24 months.

All clinical outcome assessments were done by 4 research physical therapists who were trained thoroughly before the trial and after 1 year. A manual of procedures and assessment techniques was available in each center. A reproducibility study in 19 patients yielded intraclass correlation coefficients (ICCs) for aerobic fitness, muscle strength, swollen joint count, and RAI of 0.97, 0.98, 0.83, and 0.92, respectively. The ICC based on all readings by the readers of the radiographs of the large joints was 0.95; the mean \pm SD difference in change scores after 2 years between the two readers was 0.030 ± 1.188 .

Clinical outcome assessors were blinded to the treatment allocation and measures were taken throughout the trial to preserve blinding. The patients were instructed repeatedly not to discuss their treatment allocations with the assessor and were given tips on how to avoid unblinding. The rooms in which the assessments took place were located as far as possible from the training location. At the end of the last visit, the assessors were able to guess the treatment allocation correctly in 75% of participants.

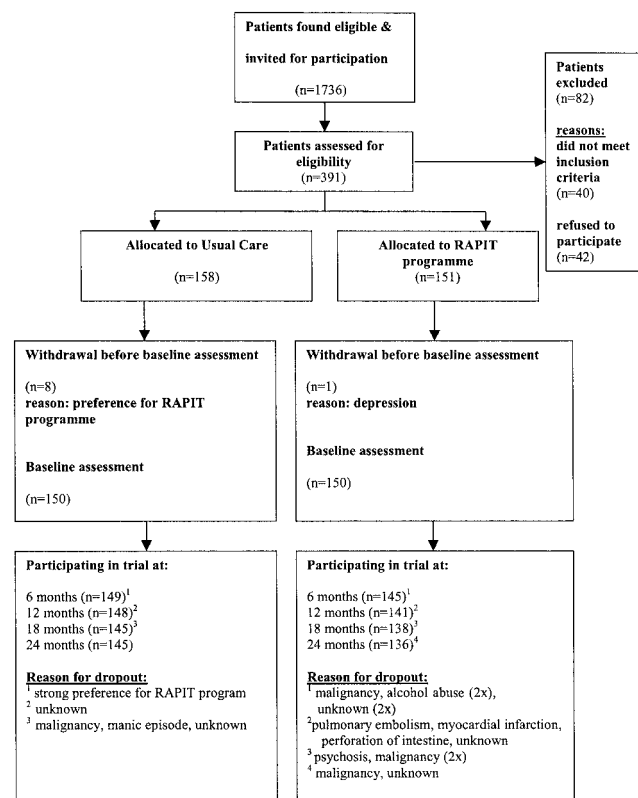


Figure 1. Trial profile. RAPIT = Rheumatoid Arthritis Patients In Training.

Statistical analysis. There are no data on clinically relevant changes in the MACTAR Questionnaire score. The target sample size was based on the ability to detect a difference of 0.20 in the change in the HAQ score, which is assumed to be clinically relevant (33). Based on 0.9 power to detect a significant difference (2-sided $P = 0.05$) and assuming an SD of 0.5, we determined that 119 patients would be required for each study group. To compensate for an expected dropout rate of $\sim 20\%$, we planned to enroll at least 150 patients in each study group. As a threshold for relevant progression in radiographic damage of the large joints and a surrogate for clinically relevant increase in damage, we used the smallest detectable difference (SDD) of the change score calculated according to the method of Lassere et al (34). The analyses are based on intent to treat as initially assigned. All available data were used.

Measures with a Gaussian distribution are expressed as the mean \pm SD, and measures with a non-Gaussian distribution are expressed as the median and interquartile range (IQR; expressed as the net result of 75th percentile $-$ 25th percentile). Differences between the groups at baseline were analyzed by Student's unpaired t -test, Mann-Whitney U test, or chi-square test where appropriate. At each time point, changes from baseline were compared by analysis of variance (ANOVA) and are presented as the mean difference in change between the groups (95% confidence interval [95% CI]). All

Table 2. Baseline demographic and clinical characteristics of the 300 RA patients who were randomized and for whom data were provided*

	UC group (n = 150)	RAPIT group (n = 150)
Age, median (IQR) years	53.5 (18)	54.0 (16)
Female	118 (79)	119 (79)
Duration of RA, median (IQR) years	7.5 (10.8)	5.0 (7)†
RF positive	106 (71)	107 (71)
Radiographic damage of hands and feet, median (IQR)‡	38.5 (54.5)	25.0 (53.8)†
Past number of DMARDs, mean ± SD	2.0 ± 1.2	1.8 ± 1.5
Current treatment		
NSAIDs	110 (73)	102 (68)
DMARDs	134 (89)	117 (78)†
Oral corticosteroids	15 (10)	12 (8)
Intraarticular corticosteroids	7 (5)	17 (11)
Bisphosphonates	4 (3)	3 (2)
Calcium supplement	11 (7)	11 (7)
Vitamin D	3 (2)	2 (1)

* Except where indicated otherwise, values are the number (%) of patients. Interquartile ranges (IQRs) are expressed as the net result of 75th percentile – 25th percentile. UC = usual care; RAPIT = Rheumatoid Arthritis Patients In Training; RF = rheumatoid factor; NSAIDs = nonsteroidal antiinflammatory drugs (see Table 1 for other definitions).

† $P < 0.05$ versus UC group by Mann-Whitney U test or chi-square test, as appropriate.

‡ Larsen score of the small joints (24).

effect analyses were performed after correction for the baseline differences. To compare the effectiveness and safety over the total period of 2 years, repeated measures were analyzed with mixed-effects ANOVA models, with patient number as a random factor and treatment, time, and treatment × time interaction as fixed effects.

RESULTS

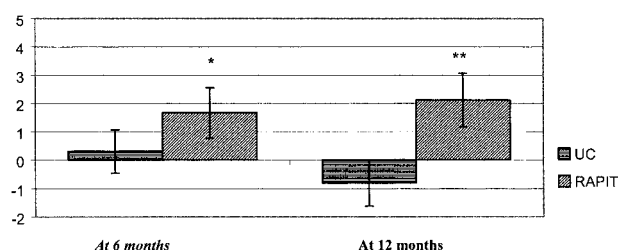
The recruitment took place between September 1997 and April 1998. Of the 1,736 patients who were eligible after screening of the records and invited by mail to participate, 391 patients were assessed for eligibility by the investigators. Finally, 309 patients were randomly assigned for participation (Figure 1). Participants did not differ from eligible nonparticipants in demographic or disease-related characteristics, except that the group of participants was slightly younger (median age 45 years [IQR 16 years] versus 47 years [IQR 16 years]), had a higher proportion of females (79% versus 72%), and had a shorter duration of RA (median 6 years [IQR 9 years] versus 7.5 years [IQR 9 years]).

Nine randomized patients refused participation immediately after randomization. The demographic and disease characteristics of the remaining 300 study participants are shown in Table 2. At baseline, participants in the UC and RAPIT groups were similar in most characteristics except for a slightly longer duration of

RA, a higher frequency of the current use of DMARDs, and more radiographic damage of the hands and feet in the UC group. Also, more participants in the UC group than in the RAPIT group had paying jobs (43% versus 31%), which was a statistically significant difference ($P = 0.05$), but there was no significant difference between the amounts of time spent at these jobs (27.5 hours/week and 26.8 hours/week, respectively).

Over the period of 2 years, 5 patients allocated to the UC group and 14 patients allocated to the RAPIT group withdrew from the trial (Figure 1). Ten of the latter 14 patients had a serious comorbidity not related to RA, and 4 withdrew for other reasons. The 281 completers of the study had not differed at baseline from the 19 who withdrew, in terms of sociodemographic characteristics, disease-related characteristics (disease duration, presence of rheumatoid factor, disease activity, radiographic damage), physical capacity, or functional ability (data not shown).

Change in MACTAR in the first year of treatment



Change in MACTAR in the second year of treatment

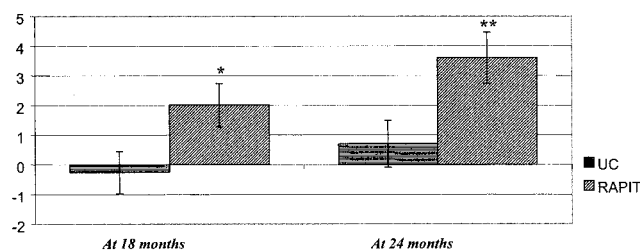


Figure 2. Change in McMaster Toronto Arthritis (MACTAR) Patient Preference Disability Questionnaire score after first and second years of treatment. Values are the mean ± SEM change in MACTAR Questionnaire score. For change at 6 and 12 months: * = $P = 0.320$; ** = $P = 0.042$ versus usual care (UC) group, by mixed-effects analysis of variance (ANOVA). For change at 18 and 24 months: * = $P = 0.020$; ** = $P = 0.015$ versus UC group, by mixed-effects ANOVA. RAPIT = Rheumatoid Arthritis Patients In Training group.

Table 3. Primary end points of effectiveness*

	UC group	RAPIT group	Δ RAPIT group – Δ UC group, mean difference (95% CI) [†]
Functional ability by MACTAR Questionnaire score (n)			
First year			
Baseline (298)	53.0 (5.0)	54.0 (4.8)	
6 months (283)	0.3 \pm 9.3	1.7 \pm 10.5	1.3 (–1.2, 3.7)
12 months (275)	–0.9 \pm 9.8	2.1 \pm 11.2	2.6 (0.1, 5.2)
<i>P</i> for trend [‡]			0.034
Second year			
Baseline 12 months (273)	54.0 (6.0)	54.0 (6.0)	
18 months (273)	–0.3 \pm 8.4	2.0 \pm 8.4	2.4 (0.3, 4.4)
24 months (273)	0.7 \pm 9.4	3.6 \pm 9.8	3.1 (0.7, 5.5)
<i>P</i> for trend [‡]			0.017
Functional ability by HAQ score (n)			
First and second years			
Baseline (299)	0.63 (0.78)	0.69 (0.88)	
6 months (288)	0.00 \pm 0.4	0.03 \pm 0.3	0.01 (–0.08, 0.08)
12 months (284)	0.10 \pm 0.4	0.06 \pm 0.4	–0.04 (–0.13, 0.05)
18 months (271)	0.08 \pm 0.3	0.02 \pm 0.4	–0.07 (–0.16, 0.03)
24 months (276)	0.07 \pm 0.3	0.00 \pm 0.4	–0.09 (–0.18, –0.01)
<i>P</i> for trend [‡]			0.421

* Functional ability was measured by the McMaster Toronto Arthritis (MACTAR) Patient Preference Disability Questionnaire and by the Health Assessment Questionnaire (HAQ). Baseline values are given as the median (interquartile range [expressed as the net result of 75th percentile – 25th percentile]). Followup values are given as the mean \pm SD change from baseline values. See Table 2 for other definitions.

[†] Mean difference (95% confidence interval [95% CI]) between change in the UC group and change in the RAPIT group. Differences are corrected for the baseline differences between the UC and RAPIT groups in duration of rheumatoid arthritis, current use of disease-modifying antirheumatic drugs, and radiographic damage of hands and feet (Larsen score).

[‡] By mixed-effects analysis of variance.

In addition to those patients who withdrew from the trial, 14 other RAPIT participants failed to attend the exercise classes but were regularly evaluated with their group. The patients who failed to attend the exercise classes did not differ in sociodemographic or disease-related characteristics from those who did attend (data not shown). The median percentage of sessions attended was 74% (IQR 27%). Averaged over 2 years, 30% of all participants had a sufficient attendance rate (50–75%), and 49% had a high attendance rate (75–100%). The percentage of participants with a high attendance rate was 65% in the first 6 months, decreased to 49% in the second 6-month period, and remained almost stable thereafter (38% and 43% in the third and fourth 6-month periods, respectively). Over the period of 2 years, no significant changes within or between the groups took place concerning the number of patients with paying jobs and the number of hours per week spent at these jobs (data not shown).

In the UC and RAPIT groups, 55% and 34% of the participants, respectively, were treated individually by a physical therapist at least once, with median cumu-

lative treatment times (for the 2 years of the trial) of 8.2 hours (range 0.5–82.5 hours) and 5.4 hours (range 0.5–37.5 hours), respectively ($P < 0.001$). The physical therapy involved hydrotherapy and different types of physical therapy (passive, active, or applications). None of the patients participated in any high-intensity weight-bearing exercises except in the RAPIT program.

During the study period, no differences in frequency in the use of nonsteroidal antiinflammatory drugs, other painkillers, DMARDs, or corticosteroids (orally or intraarticularly administered) between the groups were noted. Also, no differences in the frequency of changes of DMARD and/or DMARD dosage were found (data not shown).

The results of the intent-to-treat analysis of the outcomes of effectiveness and safety are shown in Tables 3–5.

Primary end points of effectiveness. The mean difference in change in the MACTAR Questionnaire score between the two groups was statistically significant and in favor of the RAPIT group at 12, 18, and 24 months and over the first and second years of the study

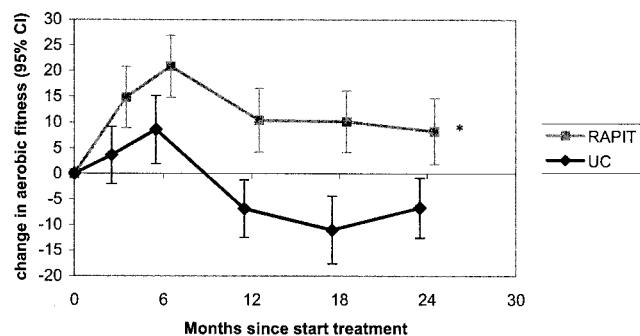


Figure 3. Change in aerobic fitness of the RAPIT and UC groups. Values are the mean (95% confidence interval [95% CI]) change in watts. * = $P < 0.001$ versus UC group, by mixed-effects ANOVA. See Figure 2 for other definitions.

(Figure 2). The mean difference in change in the HAQ score between the UC and RAPIT groups at 24 months was -0.09 (95% CI $-0.18, -0.01$) (Table 3) and thus did not reach the predefined clinically relevant level of 0.2. At all time points except for the 6-month time point, it showed a trend toward better functional ability in the RAPIT group. Over the 2 years of the study, the mean difference in change was not statistically significant ($P = 0.421$).

Secondary end points of effectiveness. Aerobic fitness showed a significantly larger improvement in the RAPIT group than in the UC group at each time point and over the 2 years ($P < 0.001$) (Figure 3). Compared with baseline, after 2 years aerobic capacity decreased in the UC group (mean \pm SD -6.7 ± 35.2 W) and increased in the RAPIT group (mean \pm SD 8.2 ± 37.1 W). After 2 years, muscle strength increased in both groups (mean \pm SD 9.6 ± 52 N in the UC group and 26.1 ± 60.9 N in the RAPIT group). The mean difference in change at all time points (except for the 6-month time point) and over 2 years was statistically significant and in favor of the RAPIT group ($P < 0.001$) (Figure 4). The mean difference in change of emotional status (the HADS score) between the groups was statistically significant at 12 and 24 months as well as over the 2 years ($P = 0.007$) and in favor of the RAPIT group.

Primary end point of safety. In both groups, no change in median radiographic damage of the large joints was found (Table 4 and Figure 5). The mean difference in change of the LLJ score between the groups showed a trend toward more damage in the RAPIT group. However, at all time points and over the 2 years, this difference was not statistically significant ($P = 0.134$). The SDD of the progression of the LLJ score, based on the scores by the two observers,

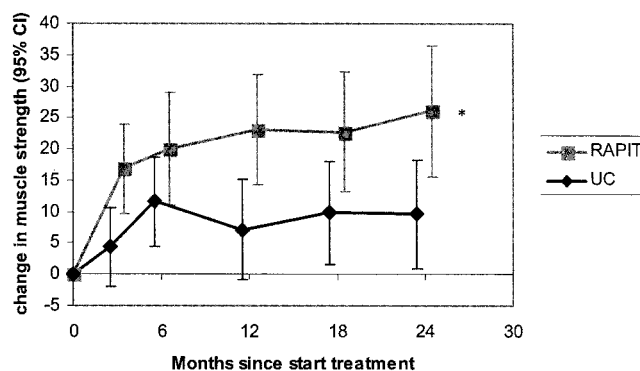


Figure 4. Change in muscle strength of the RAPIT and UC groups. Values are the mean (95% confidence interval [95% CI]) change in newtons. * = $P < 0.001$ versus UC group, by mixed-effects ANOVA. See Figure 2 for other definitions.

amounted to 1.65 points of the LLJ score units. Using the SDD as a threshold for relevant progression in damage of the large joints, we found that it was exceeded in 15 of the UC participants (10.3%) and 20 of the RAPIT participants (14.7%) ($P = 0.284$).

Figure 6 demonstrates that patients with more baseline damage showed slightly more progression in damage, and this was more obvious in the RAPIT group. When we compared the participants with no increase in damage of the large joints with those with any increase in damage, we found that patients with any increase were

Table 4. Primary end point of safety*

	UC group	RAPIT group	Δ RAPIT group – Δ UC group, mean difference (95% CI)†
Radiographic damage of the large joints by LLJ score (n)			
Baseline (293)	2.0 (5.0)	1.5 (4.5)	
12 months (283)	0.0 \pm 0.0	0.0 \pm 0.5	0.2 (0.0, 0.4)
24 months (274)	0.0 \pm 1.0	0.0 \pm 1.0	0.3 (0.0, 0.7)
P for trend‡			0.134

* Radiographic damage of the large joints was measured by the Larsen score of the large joints (LLJ score). Baseline values are given as the median (interquartile range [expressed as the net result of 75th percentile – 25th percentile]). Followup values are given as the mean \pm SD change from baseline values. See Table 2 for other definitions.

† Mean difference (95% confidence interval [95% CI]) between change in the UC group and change in the RAPIT group. Differences are corrected for the baseline differences between the UC and RAPIT groups in duration of rheumatoid arthritis, current use of disease-modifying antirheumatic drugs, and radiographic damage of hands and feet (Larsen score).

‡ By mixed-effects analysis of variance.

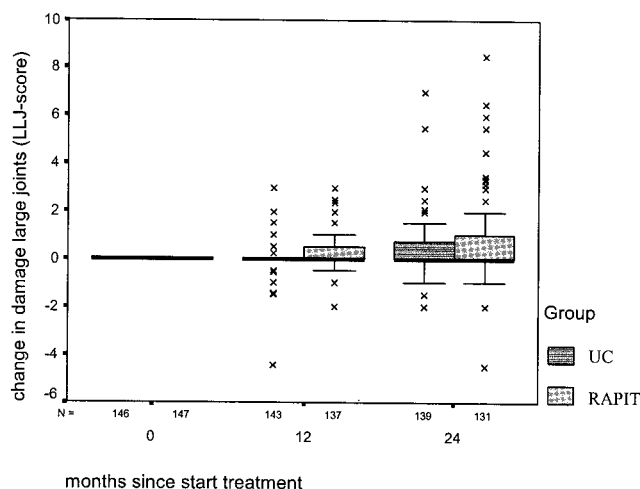


Figure 5. Change in radiographic damage of the large joints. Values are the median (interquartile range) change in Larsen score of the large joints (LLJ score) (x = outlier). N = number of patients with available radiographs at baseline and at 12 and 24 months since start of treatment. See Figure 2 for other definitions.

those with significantly longer duration of RA (median 5 years [IQR 8 years] and 7 years [IQR 9 years], respectively) and with greater baseline damage of the large and small joints (data not shown). No differences were found in baseline physical capacity, disease activity, functional ability, or employment status (percentage of patients

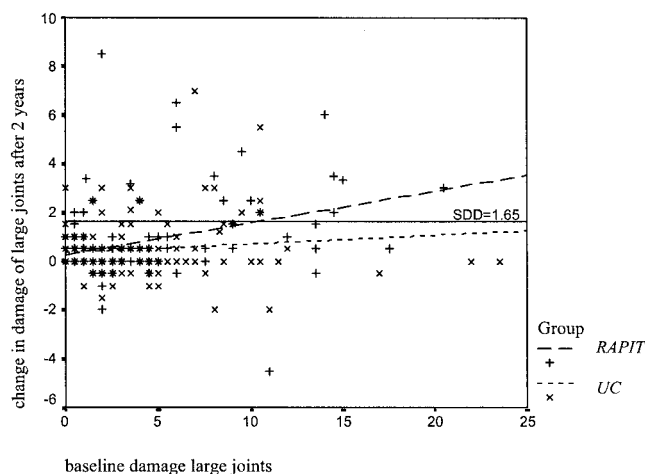


Figure 6. Relationship between radiographic damage of large joints at baseline and change in damage after 2 years. Baseline and change values are given as Larsen scores of the large joints. Shown are best-fitting lines for RAPIT and UC participants. The horizontal line represents the smallest detectable difference (SDD) in change of damage after 2 years. Asterisks indicate superimposition of + and x symbols. See Figure 2 for other definitions.

with a paying job, mean hours per week spent at this job).

Secondary end point of safety. Disease activity (measured by the DAS4) decreased gradually, suggesting a moderate improvement in disease activity in both groups. Baseline values and values at 6, 12, 18, and 24 months are presented in Table 5. The components of the DAS4 (RAI and number of swollen joints) assessed by the research physical therapist showed the same pattern as the DAS4; ESR and general health did not change during the entire study period (data not shown). At all time points and over the 2 years, the mean difference in change in the DAS4 and its components between the two study groups was not statistically significant.

DISCUSSION

Two years of a supervised long-term high-intensity group exercise program resulted in improved functional ability (the MACTAR Questionnaire score) and physical capacity of RA patients compared with RA patients who received UC physical therapy. The participants in the RAPIT group improved significantly in functional ability compared with those in the UC group as measured by the MACTAR Questionnaire, but not by the HAQ. The HAQ's lack of sensitivity to change in exercise trials has already been noted by other investigators (35) who have demonstrated that changes in physical impairments and physical condition are only weakly associated with changes in the HAQ score. Furthermore, when assessing patients who are relatively mildly disabled, as in our study, the HAQ's lack of sensitivity to change can be caused by a floor artifact of the measure. In addition, the HAQ measures the performance in limited areas of health status concerned with the performance of daily tasks. It does not take into account endurance and the ability to perform repetitive and complex tasks, or any individual preferences.

In contrast to the HAQ, the MACTAR Questionnaire assesses the change in activities elicited from a broad number of regions of activities and measures changes in only those that are important to the individual patient. The excellent responsiveness of the MACTAR Questionnaire in an RA clinical trial has already been demonstrated (28), and the change scores we present are comparable with those found by other investigators (36). Surprisingly, in our study, functional ability (measured by the MACTAR Questionnaire) improved in both of the study years, and the magnitude of improvement was approximately the same in each year. This finding contrasts with the steep improvement in

Table 5. Secondary end points of effectiveness and safety*

	UC group	RAPIT group	Δ RAPIT group – Δ UC group, mean difference (95% CI) [†]
Effectiveness (emotional status) by HADS score (n)			
Baseline (296)	11.0 (8.0)	11.0 (8.2)	
6 months (284)	–0.1 \pm 3.9	–0.6 \pm 4.0	–0.5 (–1.5, 0.5)
12 months (283)	0.5 \pm 3.5	–0.8 \pm 4.2	–1.2 (–2.1, –0.3)
18 months (267)	–0.1 \pm 4.1	–1.0 \pm 4.4	–0.8 (–1.9, 0.3)
24 months (275)	0.1 \pm 4.0	–1.2 \pm 4.1	–1.3 (–2.2, –0.3)
<i>P</i> for trend [‡]			0.007
Safety by DAS4 (n)			
Baseline (299)	3.4 (1.9)	3.3 (1.4)	
6 months (283)	–0.4 \pm 0.9	–0.3 \pm 1.0	0.19 (0.0, 0.4)
12 months (286)	–0.4 \pm 1.0	–0.5 \pm 1.1	–0.01 (–0.3, 0.2)
18 months (270)	–0.7 \pm 1.1	–0.6 \pm 1.0	0.07 (–0.2, 0.3)
24 months (277)	–0.7 \pm 1.1	–0.9 \pm 1.2	–0.10 (–0.4, 0.2)
<i>P</i> for trend [‡]			0.851

* Effectiveness (emotional status) was measured by the Hospital Anxiety and Depression Scale (HADS). Safety was measured by the Disease Activity Score with 4 variables (DAS4). Baseline values are given as the median (interquartile range [expressed as the net result of 75th percentile – 25th percentile]). Followup values are given as the mean \pm SD change from baseline values. See Table 2 for other definitions.

[†] Mean difference (95% confidence interval [95% CI]) between change in the UC group and change in the RAPIT group. Differences are corrected for the baseline differences between the UC and RAPIT groups in duration of rheumatoid arthritis, current use of disease-modifying antirheumatic drugs, and radiographic damage of hands and feet (Larsen score).

[‡] By mixed-effects analysis of variance.

physical capacity during the first year of the study, followed by a more leveled course in the second year. This development suggests that participation in the supervised high-intensity exercise classes yields prolonged positive and complex effects on functional ability, and it provides evidence against a direct linear relationship between functional ability and physical capacity.

This study demonstrates that participation in long-term high-intensity exercise classes decreases the level of psychological distress in RA patients as measured by the HADS. The low rate of withdrawal from and high rate of participation in the classes also suggest that the participants in the RAPIT group perceived benefit. Several randomized controlled trials (RCTs) investigated the effectiveness of dynamic training programs on psychological distress in RA patients (3,19,21,37–40). Only two RCTs—each with a duration of <6 months—demonstrated an improvement in the psychological well-being of the participants in the training program (19,38).

In the 2 years of the study, the median radiographic damage of the large joints did not increase in either group. However, the mean difference in change between the groups showed a nonsignificant trend toward a greater increase in damage in the RAPIT group. The value of this observation is difficult to judge, since

the increase in damage was large enough to reach the threshold for relevant progression (SDD) in only a limited number of participants. This is the first study to document extensively the effect of high-intensity joint-loading exercise on radiographic damage of the large joints of RA patients. While in the healthy population moderate recreational physical exercise is associated with a decrease in risk of knee osteoarthritis (41), long-term repetitive loading of the large joints can result in excessive radiographic osteoarthritis (42,43). A well-balanced intensity and frequency of loading is thus essential for the preservation of a healthy joint. The same could apply to an arthritic joint of an RA patient. Therefore, until more research is done, it seems wise to offer individually designed exercises that spare the damaged joints, to RA patients with considerable damage of the large joints who wish to participate in long-term intensive programs.

In the present study, concordant with findings of other investigators, no increase in disease activity was found (1–3). Indeed, we found that the DAS4 decreased during the 2 years of the study, while its assessor-independent components (ESR and patient global assessment) did not change. This phenomenon might be attributed to gradual changes in the performance of the

assessors, but it could also partly reflect the trend toward more aggressive treatment in the past decade.

The following could have biased our results. The randomization procedure created two groups which differed with respect to disease duration, use of DMARDs, and baseline radiographic damage of the hands and feet. This might have occurred because of dropout in the period immediately following the treatment allocation. Therefore, we present the mean differences in change between the groups after correcting for the baseline differences. In addition, the blinding of clinical outcome assessors to the treatment allocation was only partly successful. This could have resulted in a bias of those measurements, which could have been influenced by the opinion of the assessors on the effectiveness and safety of the therapy under study.

Regarding the generalizability of our results, it is noteworthy that we recruited volunteer RA patients who did not have prostheses in any weight-bearing joints. This probably implies that we selected motivated patients with shorter disease durations and thus less damage than those whom we excluded. To carry on exercising is important, since the effects of intensive exercise are lost to a great degree after termination of the training (44). The relatively low withdrawal and high compliance with classes suggest that the RAPIT program was well tolerated in the long term.

Our findings support the view that long-term, high-intensity, weight-bearing exercises improve the functional ability, physical capacity, and emotional status of RA patients. The RA patients were able to perform these exercises without detrimental effects on disease activity, and the exercises were safe for the weight-bearing large joints, probably with the exclusion of patients with considerable baseline damage. Future research should be directed toward the effectiveness and safety of the long-term high-intensity exercises in patients with baseline damage and those with early RA. In early RA, functional ability and physical capacity deteriorate quickly, while the large joints are still relatively spared. The cost:benefit ratio is probably most favorable in these patients.

ACKNOWLEDGMENTS

We gratefully acknowledge the patients for their trust and participation. We thank the assessors, I. Perquin, B. Oud, M. Fluit, and M. van Gulijk, for performing the clinical assessments and Professor A. Cats for reading the radiographs. We also thank Professor D. M. F. M. van der Heijde and Dr. J. D. Macfarlane for critical review of the manuscript.

REFERENCES

- Ekblom B, Lovgren O, Alderin M, Fridstrom M, Satterstrom G. Effect of short-term physical training on patients with rheumatoid arthritis I. *Scand J Rheumatol* 1975;4:80–6.
- Ekdahl C, Andersson SI, Moritz U, Svensson B. Dynamic versus static training in patients with rheumatoid arthritis. *Scand J Rheumatol* 1990;19:17–26.
- Van den Ende CHM, Hazes JMW, le Cessie S, Mulder WJ, Belfor DG, Breedveld FC, et al. Comparison of high and low intensity training in well controlled rheumatoid arthritis: results of a randomised clinical trial. *Ann Rheum Dis* 1996;55:798–805.
- Häkkinen A, Sokka T, Kotaniemi A, Hannonen P. A randomized two-year study of the effects of dynamic strength training on muscle strength, disease activity, functional capacity, and bone mineral density in early rheumatoid arthritis. *Arthritis Rheum* 2001;44:515–22.
- Manson JE, Greenland P, LaCroix AZ, Stefanick ML, Mouton CP, Oberman A, et al. Walking compared with vigorous exercise for the prevention of cardiovascular events in women. *N Engl J Med* 2002;347:716–25.
- Chavannes N, Vollenberg JJ, van Schayck CP, Wouters EF. Effects of physical activity in mild to moderate COPD: a systematic review. *Br J Gen Pract* 2002;52:574–8.
- Jacobsen BK, Bonna KH, Njolstad I. Cardiovascular risk factors, change in risk factors over 7 years, and the risk of clinical diabetes mellitus type 2: the Tromso study. *J Clin Epidemiol* 2002;55:647–53.
- Delagardelle C, Feiereisen P, Krecke R, Essamri B, Beissel J. Objective effects of a 6 months' endurance and strength training program in outpatients with congestive heart failure. *Med Sci Sports Exerc* 1999;31:1102–7.
- Ernstoff B, Wetterqvist H, Kvist H, Grimby G. Endurance training effect on individuals with postpoliomyelitis. *Arch Phys Med Rehabil* 1996;77:843–8.
- Oh-Park M, Fast A, Gopal S, Lynn R, Frei G, Drenth R, et al. Exercise for the dialyzed: aerobic and strength training during hemodialysis. *Am J Phys Med Rehabil* 2002;81:814–21.
- Wiesinger GF, Quittan M, Graninger M, Seeber A, Ebenbichler G, Sturm B, et al. Benefit of 6 months long-term physical training in polymyositis/dermatomyositis patients. *Br J Rheumatol* 1998;37:1338–42.
- Troosters T, Gosselink R, Decramer M. Short- and long-term effects of outpatient rehabilitation in patients with chronic obstructive pulmonary disease: a randomized trial. *Am J Med* 2000;109:207–12.
- Tangney DJ, Brown DB, Brown GJ, Prescott GJ. Cardiovascular risk reduction in men: a nine-year cohort study. *Br J Gen Pract* 2002;52:743–5.
- Franklin B, Bonzheim K, Warren J, Haapaniemi S, Byl N, Gordon N. Effects of a contemporary, exercise-based rehabilitation and cardiovascular risk-reduction program on coronary patients with abnormal baseline risk factors. *Chest* 2002;122:338–43.
- Tudor-Locke CE, Bell RC, Meyers AM. Revisiting the role of physical activity and exercise in the treatment of type 2 diabetes. *Can J Appl Physiol* 2000;25:466–92.
- Bacon PA, Townend JN. Nails in the coffin: increasing evidence for the role of rheumatic disease in the cardiovascular mortality of rheumatoid arthritis. *Arthritis Rheum* 2001;44:2707–10.
- Del Rincón I, Williams K, Stern MP, Freeman GL, Escalante A. High incidence of cardiovascular events in a rheumatoid arthritis cohort not explained by traditional cardiac risk factors. *Arthritis Rheum* 2001;44:2737–45.
- Van den Ende CH, Breedveld FC, le Cessie S, Dijkmans BA, de Mug AW, Hazes JM. Effect of intensive exercise on patients with active rheumatoid arthritis: a randomised clinical trial. *Ann Rheum Dis* 2000;59:615–21.

19. Minor MA, Hewett JE, Webel RR, Anderson SK, Kay DR. Efficacy of physical conditioning exercise in patients with rheumatoid arthritis and osteoarthritis. *Arthritis Rheum* 1989;32:1396–405.
20. Rall LC, Meydani SN, Kehayias JJ, Dawson-Hughes B, Roubenoff R. The effect of progressive resistance training in rheumatoid arthritis: increased strength without changes in energy balance or body composition. *Arthritis Rheum* 1996;39:415–26.
21. Komatireddy GR, Leitch RW, Cella K, Browning G, Minor M. Efficacy of low load resistive muscle training in patients with rheumatoid arthritis functional class II and III. *J Rheumatol* 1997;24:1531–9.
22. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, et al. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;31:315–24.
23. Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. *Arthritis Rheum* 1992;35:498–502.
24. Scott DL, Houssien DA, Laasonen L. Proposed modification to Larsen's scoring methods for hand and wrist radiographs. *Br J Rheumatol* 1995;34:56.
25. Tugwell P, Bombardier C, Buchanan WW, Goldsmith CH, Grace E, Hanna B. The MACTAR Patient Preference Disability Questionnaire—an individualized functional priority approach for assessing improvement in physical disability in clinical trials in rheumatoid arthritis. *J Rheumatol* 1987;14:446–51.
26. Fries JF, Spitz PW, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. *Arthritis Rheum* 1980;23:137–45.
27. Siegert CEH, Vleming LJ, VandenBroucke JP, Cats A. Measurement of disability in Dutch rheumatoid arthritis patients. *Clin Rheumatol* 1984;3:305–9.
28. Verhoeven AC, Boers M, van der Linden S. Validity of the MACTAR questionnaire as a functional index in a rheumatoid arthritis clinical trial. *J Rheumatol* 2000;27:2801–9.
29. Wallin L, Brudin LH. Physical working capacity determined by different types of bicycle exercise tests. *Clin Physiol* 1988;8:529–37.
30. Zigmund AS, Snaith RP. The Hospital Anxiety and Depression Scale. *Acta Psychiatr Scand* 1983;67:361–70.
31. Larsen A, Dale K, Eek M. Radiographic evaluation of rheumatoid arthritis and related conditions by standard reference films. *Acta Radiol Diagn (Stockh)* 1977;18:481–91.
32. Scott DL, van Riel PLCM, van der Heijde DM, Studnicka Benkem A. Assessing disease activity in rheumatoid arthritis. In: van Riel PLCM, van Gestel AM, Scott DL, editors. *EULAR handbook of clinical assessments in rheumatoid arthritis*. CH.2.2 Alphen aan den Rijn (The Netherlands): van Zuiden Communications; 2000. p. 33–7.
33. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993;20:557–60.
34. Lassere M, Boers M, van der Heijde D, Boonen A, Edmonds J, Saudan A, et al. Smallest detectable difference in radiological progression. *J Rheumatol* 1999;26:731–9.
35. Van den Ende CHM, Breedveld FC, Dijkmans BA, Hazes JMW. The limited value of the Health Assessment Questionnaire as an outcome measure in short term exercise trials. *J Rheumatol* 1997;24:1972–7.
36. Tijhuis GJ, Zwiderman AH, Hazes JMW, van den Hout WB, Breedveld FC, Vliet Vlieland TPM. A randomized comparison of care provided by a clinical nurse specialist, an inpatient team, and a day patient team in rheumatoid arthritis. *Arthritis Rheum* 2002;47:525–31.
37. Stenström CH, Arge B, Sundbom A. Home exercise and compliance in inflammatory rheumatic diseases—a prospective clinical trial. *J Rheumatol* 1997;24:470–6.
38. Ekdahl C, Andersson SI, Moritz U, Svensson B. Dynamic versus static training in patients with rheumatoid arthritis. *Scand J Rheumatol* 1990;19:17–26.
39. Harkcom TM, Lampman RM, Banwell BF, Castor CW. Therapeutic value of graded aerobic exercise training in rheumatoid arthritis. *Arthritis Rheum* 1985;28:32–9.
40. Nordemar R. Physical training in rheumatoid arthritis: a controlled long-term study. II. Functional capacity and general attitudes. *Scand J Rheumatol* 1981;10:25–30.
41. Manninen P, Riihimäki H, Heliövaara M, Suomalainen O. Physical exercise and risk of severe knee osteoarthritis requiring arthroplasty. *Rheumatology (Oxford)* 2001;40:432–7.
42. Marti B, Knobloch M, Tschopp A, Jucker A, Howald H. Is excessive running predictive of degenerative hip disease? Controlled study of former elite athletes. *BMJ* 1989;299:91–9.
43. Kujala UM, Kettunen J, Paananen H, Aalto T, Battié MC, Impivaara O, et al. Knee osteoarthritis in former runners, soccer players, weight lifters, and shooters. *Arthritis Rheum* 1995;38:539–46.
44. Häkkinen A, Malkia E, Häkkinen K, Jappinen I, Laitinen L, Hannonen P. Effects of detraining subsequent to strength training on neuromuscular function in patients with inflammatory arthritis. *Br J Rheumatol* 1997;36:1075–81.