Arthritis & Rheumatism (Arthritis Care & Research)
Vol. 53, No. 1, February 15, 2005, pp 39–47
DOI 10.1002/art.20903

ORIGINAL ARTICLE

Cost-Utility and Cost-Effectiveness Analyses of a Long-Term, High-Intensity Exercise Program Compared With Conventional Physical Therapy in Patients With Rheumatoid Arthritis

WILBERT B. VAN DEN HOUT, ZUZANA DE JONG, MARTEN MUNNEKE, JOHANNA M. W. HAZES, FERDINAND C. BREEDVELD, AND THEODORA P. M. VLIET VLIELAND

Objective. To estimate the cost utility and cost effectiveness of long-term, high-intensity exercise classes compared with usual care in rheumatoid arthritis (RA) patients.

Methods. RA patients (n = 300) were randomly assigned to either exercise classes or UC; followup lasted for 2 years. Outcome measures were quality-adjusted life years (QALYs) according to the EuroQol (EQ-5D), Short Form 6D (SF-6D), and a transformed visual analog scale (VAS) rating personal health; functional ability according to the Health Assessment Questionnaire (HAQ) and McMaster Toronto Arthritis Patient Preference Interview (MACTAR); and societal costs.

Results. QALYs in both randomization groups were similar according to the EQ-5D and SF-6D, but were in favor of usual care according to the VAS (annual difference 0.037 QALY; 95% confidence interval [95% CI] 0.002, 0.069). Functional ability was similar according to the HAQ, but in favor of the exercise classes according to the MACTAR (annual difference 2.9 QALY; 95% CI 0.9, 4.9). Annual medical costs of the exercise program were estimated at €780 per participating patient (€1 ≈ $1.05). The increase per patient in total medical costs of physical therapy was estimated at €430 (95% CI €318, 577), and the increase in total societal costs at €602 (95% CI €−490, 1,664). For societal willingness-to-pay equal to €50,000 per QALY, usual care had better cost utility than exercise classes, and significantly so according to the VAS.

Conclusion. From a societal perspective and without taking possible preventive health effects into account, long-term, high-intensity exercise classes provide insufficient improvement in the valuation of health to justify the additional costs.

KEY WORDS. Cost-utility analysis; Cost-effectiveness analysis; Economic evaluation; Physical therapy; Exercise classes; Rheumatoid arthritis.

INTRODUCTION

It has long been accepted that physical activity improves health by preventing mortality and morbidity from cardiovascular diseases, osteoporosis, anxiety, and depression (1–3). Increasingly, exercise programs are developed and implemented for specific patient groups. For rheumatoid arthritis (RA) patients, despite their increased susceptibility for the mentioned comorbidities (4–6), exercise programs have for a long time included only low-intensity training aimed at preserving joint mobility. Patients were discouraged from performing intensive physical activity, for fear of damage to the large joints and exacerbation of joint inflammation (7).

Only relatively recently, it was shown that RA patients benefit from both short-term (8–10) and long-term (11) high-intensity exercise, without worsening of inflammation or progression of joint damage. This was confirmed by the Rheumatoid Arthritis Patients In Training (RAPIT) study (12) that compared a long-term, high-intensity exercise program with usual care consisting of individual physical therapy only if regarded necessary by the attending physician. The study demonstrated that RA patients were able to improve their functional ability, physical capacity, emotional status, and bone mineral density,

Wilbert B. van den Hout, PhD, Zuzana de Jong, MD, Marten Munneke, PT, MSc (current address: University Medical Center St. Radboud, Nijmegen, The Netherlands), Johanna M. W. Hazes, MD, PhD (current address: Rotterdam University Hospital, Rotterdam, The Netherlands), Ferdinand C. Breedveld, MD, PhD, Theodora P. M. Vliet Vlieland, MD, PhD: Leiden University Medical Center, Leiden, The Netherlands.

Address correspondence to Wilbert B. van den Hout, PhD, Department of Medical Decision Making J10-S, Leiden University Medical Center, PO Box 9600, 2300 RC Leiden, The Netherlands. E-mail: Hout@LUMC.NL.

Submitted for publication August 1, 2003; accepted in revised form May 19, 2004.

Table 1. Eligibility criteria for participation in the RAPIT trial*

- Age 20–70 years
- Rheumatoid arthritis according to 1987 ACR criteria (43)
- ACR functional class I–III (44)
- Stable disease-modifying antirheumatic drugs in past 3 months
- Able to cycle
- Willing to exercise biweekly on fixed schedule
- Living within a predefined adherence region of training or assessment center
- No prosthesis of a weight-bearing joint
- No cardiopulmonary disease precluding intensive exercise
- No comorbidity causing a short life expectancy
- No serious psychiatric disease
- Able to complete a questionnaire

* RAPIT = Rheumatoid Arthritis Patients in Training; ACR = American College of Rheumatology (formerly American Rheumatism Association).

without an increase in disease activity or additional damage to the large joints. However, studies have so far not reported on whether these exercise programs provide good value for money from the societal perspective. Here we will present the economic evaluation of the RAPIT study, estimating the cost utility and cost effectiveness of long-term, high-intensity exercise classes compared with usual care, in RA patients.

PATIENTS AND METHODS

The RAPIT study was a multicenter, single-blinded randomized controlled trial (12), comparing safety, effectiveness, and costs of a long-term intensive exercise program (RAPIT) with those of usual care physical therapy. Patients registered in 2 university and 2 nonuniversity outpatient rheumatology clinics in the Netherlands were invited to participate. In May and June 1998, 300 RA patients who fulfilled the eligibility criteria for the trial (Tables 1 and 2) were included and randomized to receive the RAPIT program or usual care.

Intervention. The RAPIT program consisted of long-term, high-intensity weight-bearing exercise classes aimed at maintaining and improving physical ability. Patients participated in the group exercise classes for 2 years, with two 75-minute sessions each week. In general, sessions consisted of warming up, bicycle training, an exercise circuit, sport or game, and cooling down. The classes were supervised by 2 physical therapists and, on average, consisted of 14 patients. Patients were encouraged to communicate possible physical problems with the physical therapists and, if necessary, the program was adapted to individual needs. Patients in the control group received usual care, consisting of individual physical therapy, only if this was regarded necessary by the attending physician.

Assessments of utility. Utility is the valuation of the health of the patient (13), ranging from 0 (as bad as death) to 1 (full health), and was assessed in 3 different ways every 6 months, using questionnaires that were filled out by the patients without supervision. Patients described their general health status using the EuroQol classification system (EQ-5D), consisting of 5 questions on mobility, self care, usual activities, pain/discomfort, and anxiety/depression (14). From the EQ-5D classification system, the EQ-5D utility index was calculated (15). This utility measure reflects how the general public values the health status described by the patient, which is preferred for economic evaluations from a societal perspective.

Quality of life was also assessed using the RAND-36 questionnaire (16). The RAND-36 consists of 36 items on physical and social functioning, role limitations, mental health, vitality, pain, and general health perception. From the RAND-36, the Short Form 6D (SF-6D) utility index was calculated (17). Like the EQ-5D, this SF-6D reflects the general public's valuation of the health described by the patient. The SF-6D is a recent instrument that has not been

Table 2. Trial flow and baseline characteristics of 300 RA patients with baseline assessment*

	RAPIT	UC	P†
Randomized patients, n (%)	151 (100)	158 (100)	—
Baseline assessments, n (%)	150 (99)	150 (95)	0.02
In study after 2 years, n (%)	136 (90)	145 (92)	0.03‡
In RAPIT program after 2 years, n (%)	114 (75)	—	—
Age, median (IQR) years	54.0 (16)	53.5 (18)	0.96
Female, n (%)	119 (79)	118 (79)	1.00
Duration of RA, median (IQR) years	5.0 (7)	7.5 (10.8)	0.01
Rheumatoid factor positive, n (%)	107 (71)	106 (71)	1.00
Radiologic damage hands and feet, median (IQR)	25.0 (53.8)	38.5 (54.5)	0.03
Past number of DMARDs, mean ± SD	1.8 ± 1.5	2.0 ± 1.2	0.77
DMARDs, n (%) of users	117 (78)	134 (89)	0.01
Oral corticosteroids, n (%) of users	12 (8)	15 (10)	0.55
Intraarticular corticosteroids, n (%) of users	17 (11)	7 (5)	0.06

* RA = rheumatoid arthritis; RAPIT = Rheumatoid Arthritis Patients in Training exercise program; UC = usual care; IQR = interquartile range; DMARDs = disease-modifying antirheumatic drugs.
† Mann-Whitney U test or χ^2 test where appropriate.
‡ Tested among patients with baseline assessment.

used much, but its richer classification system could make it a more sensitive utility measure than the EQ-5D.

Patients rated their personal health using a visual analog scale (VAS) ranging from worst imaginable health to best imaginable health. Because the VAS has repeatedly been found to render less favorable valuations than more valid (but also more complicated) utility measures, the obtained VAS values were transformed using the power function $1-(1-VAS)^{1.61}$ (18,19). Because patients experience all the subtleties of their health status, the VAS is potentially more sensitive to change, but it is not preferred for economic evaluations from a societal perspective (20).

Assessments of functional ability. The primary clinical endpoint of the RAPIT study was functional ability, assessed every 6 months by the Health Assessment Questionnaire (HAQ) and by the McMaster Toronto Arthritis Patient Preference Interview (MACTAR). The HAQ consists of 20 questions concerning 8 domains of problems in the activities of daily living (21,22). The total HAQ score ranges from 0 (without any difficulty) to 3 (unable to do). The MACTAR is a semistructured interview, assessed by a trained interviewer, consisting of status and transitional questions (23,24). The status questions elicit the patient's health status through 5 questions on general health; satisfaction with quality of life; and physical, social, and emotional wellbeing. The transitional questions focus on perceived change in disease activity and change in ability to perform a set of 5 impaired activities. These activities were selected and ranked by each individual patient at the beginning of the study. To improve sensitivity to relevant change, a new set of activities could be chosen after a year if the original activities were no longer impaired or no longer of interest to the patient. The MACTAR scores during the second year were calculated by adding the MACTAR score at the end of the first year to the change during the second year. The weighted MACTAR score ranges from 21 to 77 points, with higher scores reflecting better functional ability.

Assessment of costs. Societal costs during the 2-year followup period were assessed in strict accordance to current guidelines for economic evaluations (25). The estimated costs included the costs of the RAPIT program, other medical costs (like individual physical therapy and hospitalizations), and nonmedical costs (like time, travel, and informal care). Costs of the RAPIT program and of disease-modifying antirheumatic drugs were estimated from the study registration. All other costs were estimated from quarterly cost questionnaires filled out by the patients. Costs were not discounted, to facilitate averaging over the short 2-year study period. Costs are reported as annual costs, converted to price level 2002 euros using the price index rate for the Dutch health care sector (obtained from Statistics Netherlands). Euros can be converted to US dollars using the Dutch purchasing power parity index for 2004: €1 = $1.05 (available at: www.oecd.org).

Costs of the RAPIT program. The costing analysis of the RAPIT program was based on experience acquired during the study. The classes were organized by 2 physical therapists. Each class took them 100 minutes, of which 25 minutes were for traveling and preparation and 75 minutes were for the classes themselves. This time was valued at €25.60 per hour, i.e., the income component of the Dutch reimbursement for individual physical therapy (available at: www.ctgzorg.nl). Because physical therapists obtain their income from these reimbursements, they are a reasonable reflection of the true costs for physical therapists. The exercise classes are assumed to take place in gyms, rented at €27 per class. This includes the costs for simple materials and 7 exercise bikes.

An exercise class, on average, consisted of 14 patients, with 94 planned classes per year of 75 minutes each. Patients, on average, attended 72% of the classes. For individual patients, costs for the exercise classes were counted for as long as the patient remained in the program. Time and travel costs were included in analyses from a societal perspective. Time costs were estimated at 95 minutes per patient per class, valued at €5 per hour (26). Travel costs were estimated at €3 per patient per class.

Other medical costs. In the quarterly cost questionnaires, patients reported physical therapy other than the RAPIT program, consultations (general practitioner visits, specialist consultations, paramedical professionals other than physical therapists, and alternative medicine), hospitalizations, home nursing care, and purchased medication. Most cost prices were obtained from Dutch standard prices that were designed to reflect societal costs and to standardize economic evaluations (27,28). Physical therapy other than the RAPIT program was valued at €47 per hour (www.ctgzorg.nl). The general practitioner was valued at €18 per consultation (27), half for telephone, and double for home consultations. Specialist consultations were on average valued at €49 (27), ranging from €25 to €76 for different specialties. Paramedical professionals other than physical therapists were valued at €20 per consultation (27), and alternative medicine, on average, at €29 per consultation. Day care hospitalizations in rheumatology departments were valued at €758 per admission plus €236 per day, and clinical hospitalizations at €909 per admission plus €243 per day (26). Other hospitalizations were valued at €431 per day for nonuniversity hospitals and €599 per day for university hospitals (27,29). Home nursing care was valued at €34 per hour (27). Purchased medication was valued according to the Pharmacotherapeutic Compass (28), plus €6 for each purchase other than over-the-counter medications (27).

Nonmedical costs. Nonmedical costs included time and travel costs required to obtain health care, sports expenses, other out-of-pocket costs (like special appliances and house adaptations), absenteeism, unpaid labor, domestic help, and informal care. Time was valued at €5 per hour (26), with the amount of time varying from 10 minutes per telephone consultation to 8 hours per hospitalization day. Travel was valued at €1.22 plus €0.29 per kilometer, with national averages for the travel distances to different types of medical care. Sports expenses and other out-of-pocket

costs were valued as reported by the patients. Paid labor was valued using the friction cost method in which productivity costs are calculated for at most 4 months, which is the estimated time needed to find a replacement (30). Time spent on unpaid labor was compared with the average over the entire sample (for men and women separately, corrected for the individual amount of domestic and informal care), and the difference was valued at €5 per hour (the value of time). With this method, patients who reported more unpaid labor than average have negative costs (profits). Domestic help was valued at €19 per hour (27), and reported informal care was valued at minimum wages (€9 per hour, with a maximum of 28 hours per week).

Analysis. All patients were evaluated according to intention to treat, with correction for baseline differences (by subtracting from all effectiveness scores the individual baseline score and adding the overall average baseline score). Missing measurements were imputed by carrying forward the last available previous measurement. On average, 4% of the utility measurements, 4% of the HAQ measurements, 6% of the MACTAR measurements, and 8% of the cost questionnaires were missing, with missing measurements about twice as frequent in the RAPIT group.

Because in economic evaluations it is sustained improvement that counts, not the separate measurements but the areas under the measurement curves were used in the analyses. For utility measures, the area under the curves is known as quality-adjusted life years (QALYs). QALYs are an accepted measure for resource allocation decisions involving diverse treatments and patient populations. Cost-effectiveness analyses with QALYs as a measure of effectiveness are known as cost-utility analyses.

In our economic evaluation, both randomization groups were compared with respect to their net benefit (31) by subtracting the costs associated with the interventions from society's willingness to pay (WTP) for the obtained effectiveness:

$$\text{Net benefit} = (\text{effectiveness} \times \text{WTP}) - \text{costs}$$

This net benefit was calculated for each patient individually. The more cost-effective treatment is the one with the higher average net benefit, which also depends on the value of the WTP. If one treatment is more effective than the other at higher costs, then the differences in costs divided by the difference in effectiveness is called the incremental cost-effectiveness ratio (ICER) of the treatment under investigation.

According to the original study protocol, the primary economic evaluation compared 2-year QALYs based on the EQ-5D with 2-year total societal costs. Sensitivity analysis was, however, also performed by considering different utility and effectiveness measures (EQ-5D, SF-6D, VAS, HAQ, and MACTAR) and different cost measures (total societal costs and societal costs of all physical therapy).

For all outcome measures, differences between the randomization groups were tested using double-sided bootstrapping (32), with 1,000,000 replications and 0.05 significance threshold. Reported confidence intervals are the corresponding 95% trimmed asymmetric confidence intervals (95% CIs). Bootstrapping explicitly compares the means in both groups, without making distributional assumptions and thus allowing for skewed distributed costs.

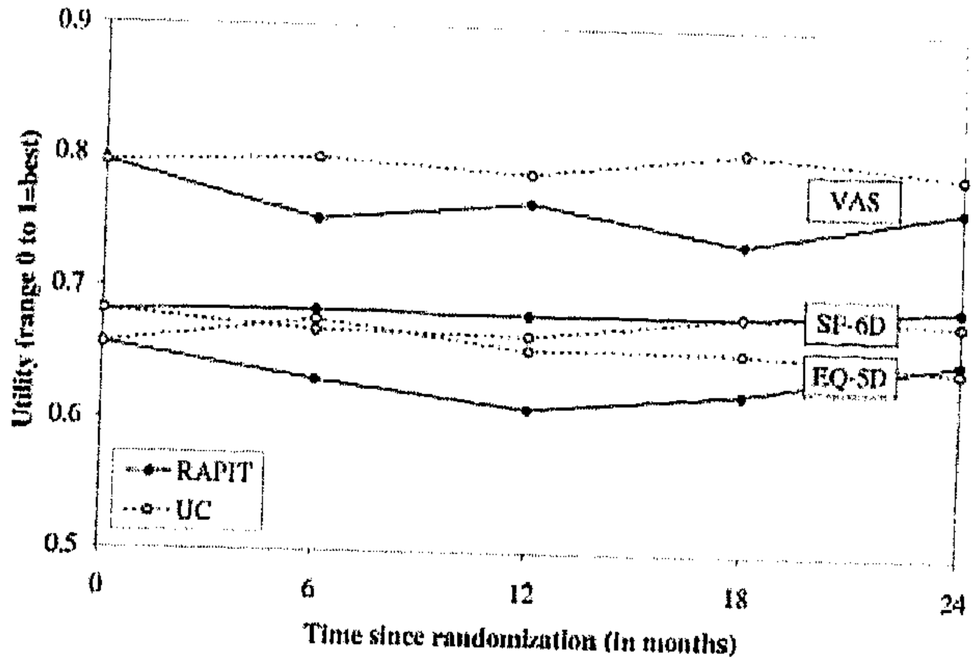

Figure 1. Utility measures. VAS = visual analog scale; SF-6D = Short Form 6D; EQ-5D = EuroQol 5D; RAPIT = Rheumatoid Arthritis Patients In Training; UC = usual care.

RESULTS

Utility and effectiveness analyses. Compared with the usual care group, the average valuation of health (utility) in the RAPIT group was somewhat more favorable according to the SF-6D measure, but was less favorable according to the EQ-5D and the VAS measures (Figure 1). Over the 2-year followup period, only the difference according to the VAS was significant (Table 3), with an estimated annual difference of 0.037 QALY in favor of usual care (95% CI 0.002, 0.069).

Patients in the RAPIT group showed a larger improvement in functional ability according to both the HAQ and the MACTAR (Figure 2). Over the 2-year period, the difference on the MACTAR was significant (Table 3), with an estimated annual difference of 2.9 (95% CI 0.9, 4.9).

Costs of the RAPIT program. The medical costs of the RAPIT program were estimated at €10,800 per group annually, of which 24% was for materials and accommodation and 76% for the 2 physical therapists (160 hours per therapist annually). With, on average, 14 patients per group, these estimated costs amount to €780 per patient. Based on 94 planned sessions per year with 72% attendance of patients, the costs per patient per attended class were €11.40. If supervision would be provided by only a single physical therapist, or by a single therapist assisted by a trained lay person valued at minimum wages, then the medical costs of the RAPIT program could be reduced by 35% and 23%, respectively.

From a societal perspective, costs also include the patients' invested time and travel costs. These costs amounted to €7,100 (for 107 hours per patient annually) and €2,600, respectively. This almost doubled the costs of the exercise classes to €20,500 per group annually.

Table 3. Average QALYs and functional ability per patient*

	RAPIT	UC	Difference	*P*†
QALYs, based on EQ-5D				
Year 1	0.633 (0.130)	0.666 (0.142)	−0.033	0.04
Year 2	0.627 (0.192)	0.651 (0.186)	−0.024	0.27
Average	0.630 (0.150)	0.659 (0.153)	−0.029	0.11
QALYs, based on SF-6D				
Year 1	0.682 (0.058)	0.671 (0.078)	0.011	0.15
Year 2	0.683 (0.083)	0.676 (0.103)	0.007	0.55
Average	0.682 (0.065)	0.673 (0.084)	0.009	0.31
QALYs, based on VAS				
Year 1	0.767 (0.127)	0.795 (0.137)	−0.028	0.07
Year 2	0.752 (0.129)	0.797 (0.192)	−0.045	0.04
Average	0.759 (0.145)	0.796 (0.149)	−0.037	0.03
HAQ				
Year 1	0.743 (0.206)	0.740 (0.252)	0.003	0.90
Year 2	0.740 (0.328)	0.793 (0.292)	−0.053	0.14
Average	0.741 (0.251)	0.766 (0.254)	−0.025	0.39
MACTAR				
Year 1	54.4 (7.0)	53.0 (6.2)	1.4	0.08
Year 2	56.8 (12.4)	52.3 (12.0)	4.5	0.002
Average	55.6 (9.2)	52.7 (8.5)	2.9	0.005

* Values are the mean (SD). QALYs = quality-adjusted life years, ranging from 0 (worst) to 1 (best); RAPIT = Rheumatoid Arthritis Patients in Training exercise program; UC = usual care; EQ-5D = EuroQol classification system; SF-6D = Short Form 6D; VAS = transformed visual analog scale rating personal health; HAQ = Health Assessment Questionnaire, ranging from 0 (without any difficulty) to 3 (unable to do); MACTAR = McMaster Toronto Arthritis Patient Preference Interview, ranging from 21 (worst) to 77 (best).
† Double-sided bootstrapping.

Costs of all physical therapy per patient. Because participation gradually decreased over time, the average medical costs per patient for the RAPIT program decreased from €733 in the first year to €634 in the second year (Figure 3, Table 4). In both randomization groups, patients could seek individual physical therapy. During the first quarter, the medical costs for individual physical therapy were comparable, but later on they were considerably lower for the RAPIT patients. Over the 2-year period, the medical costs for individual physical therapy were reduced by 62%, which partly compensated the costs of the RAPIT program. Nevertheless, the annual medical costs of all physical therapy (both the RAPIT program and individual physical therapy) were still €430 in favor of usual care (95% CI €318, 577). Including time and travel costs, the annual difference in societal costs of all physical therapy was €996 (95% CI €860, 1,170).

Total societal costs per patient. All medical and nonmedical cost categories that were not directly associated with physical therapy did not show a significant difference between the randomization groups (Table 4). Patients in the RAPIT group did report considerably more out-of-pocket costs, which was mainly due to house adaptations (annual cost difference €312, 95% CI €−84, 585). On the other hand, they also reported more unpaid labor than the

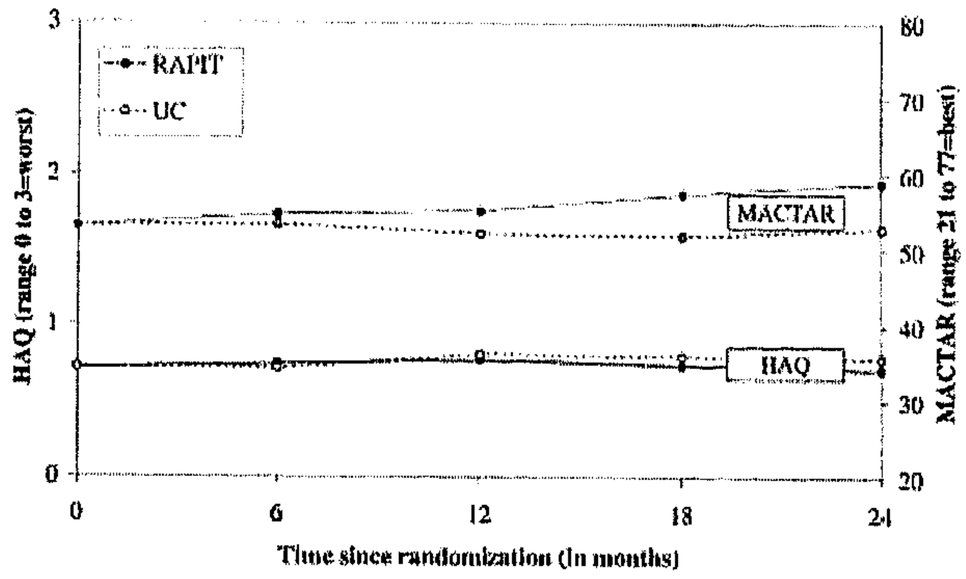

Figure 2. Effectiveness measures. HAQ = Health Assessment Questionnaire; RAPIT = Rheumatoid Arthritis Patients In Training; UC = usual care; MACTAR = McMaster Toronto Arthritis Patient Preference Interview.

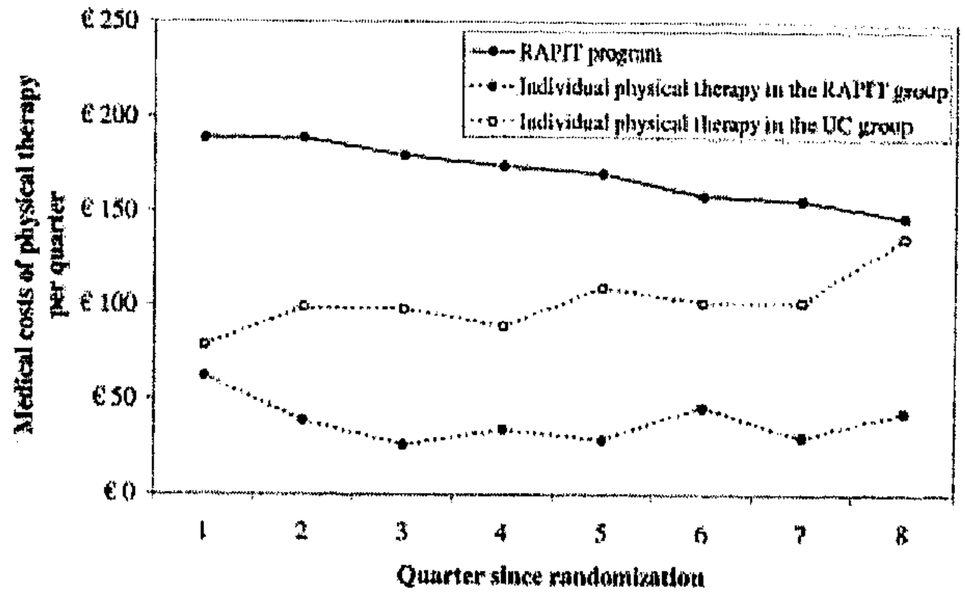

Figure 3. Medical costs of all physical therapy per quarter (purchasing power parity index: €1 = $1.05). RAPIT = Rheumatoid Arthritis Patients In Training; UC = usual care.

Table 4. Average annual health care consumption and costs per patient*

	RAPIT		UC		
	Volume	Cost in €†	Volume	Cost in €†	P‡
Medical costs					
RAPIT program, classes	60	684	—	—	< 0.001
Individual physical therapy, hours, %	3.2, 23	155	8.4, 45	408	0.001
Rheumatologists, visits	4.8	279	4.7	271	0.66
Other specialists, visits	2.4	94	2.5	98	0.80
General practitioner, visits	5.5	91	4.2	72	0.06
Paramedical professionals, visits§	1.0	53	0.9	45	0.60
Alternative medicine, visits	1.0	29	0.9	25	0.73
Hospitalizations, %	20	607	22	602	0.98
Home nursing care, hours	0.3	22	0.1	8	0.35
Medication		102		154	0.07
Total medical costs (SD)		2,115 (2,125)		1,683 (1,968)	0.07
Nonmedical costs					
T&T RAPIT program, hours	95	616	—	—	< 0.001
T&T individual physical therapy, hours	3	31	8	81	< 0.001
T&T other medical care, hours	37	253	36	243	0.76
Sports expenses		8		20	0.06
Other out-of-pocket		730		406	0.06
Paid labor; %, hours/week	33, 24	—	42, 25	—	—
Absenteeism, hours/week	0.7	121	0.7	155	0.44
Unpaid labor, hours/week	26	−270	23	270	0.17
Home help, hours/week	0.7	782	0.9	884	0.54
Informal care, hours/week	0.8	362	0.9	405	0.73
Total nonmedical costs (SD) hours/week	—	2,634 (4,120)	—	2,464 (3,877)	0.71
Total societal costs (SD)	—	4,749 (4,953)		4,147 (4,497)	0.27

* RAPIT = Rheumatoid Arthritis Patients in Training exercise program; UC = usual care; T&T = time and travel.
† Purchasing power parity index: €1 = $1.05.
‡ Double-sided bootstrapping.
§ Paramedical professionals other than physical therapists.

usual care group (annual cost difference €540, 95% CI €−240, 1,310). The estimated annual difference in total societal costs was €602 per patient (95% CI €−490, 1,664).

Cost-utility and cost-effectiveness analyses. The primary economic evaluation of the RAPIT study was the cost-utility analysis comparing the €602 difference in total societal costs to the 0.029-year difference in QALYs based on the EQ-5D. If society values QALYs at (the often quoted) €50,000 per QALY, then the difference in net benefit is 0.029 × €50,000 + €602 = €2,052 per patient annually. Because both costs and QALYs are in favor of usual care, the net benefit is also in favor of usual care, regardless society's WTP. The difference in net benefit did not reach statistical significance for any WTP ($P \geq 0.06$, with $P = 0.06$ for WTP = €50,000 and $P < 0.10$ for €16,000 ≤ WTP ≤ €1,200,000).

In sensitivity analyses, we considered other utility (SD-6D and VAS), effectiveness (HAQ and MACTAR), and costs measures. According to the SF-6D, patients in the RAPIT program had higher QALYs than patients receiving usual care. The additional total societal costs of €602 rendered a 0.009 QALY improvement, with an estimated ICER of €67,000 per QALY. However, the net benefit was again not significantly different for any WTP ($P \geq 0.27$ and $P = 0.83$ for WTP = €50,000).

According to the VAS, patients in the RAPIT program had lower QALYs than patients receiving usual care, so cost utility was in favor of usual care for all levels of society's WTP. Here, the difference was significant for all WTP above €19,000 per QALY ($P \leq 0.05$ and $P = 0.02$ for WTP = €50,000).

Both measures of functional ability were in favor of the RAPIT program. For the HAQ, the ICER was estimated at €24,000 per HAQ point, but the net benefit was not significantly different for any WTP ($P \geq 0.27$). For the MACTAR, the ICER was estimated at €210 per MACTAR point, and the net benefit was significantly different for WTP greater than €1,200 per MACTAR point ($P \leq 0.05$).

We also performed sensitivity analyses including only the costs associated with physical therapy because all other cost categories showed no significant difference. The results basically remained the same, except that the reduced variability in costs considerably increased the statistical significance of the economic differences in favor of usual care, and the estimated ICERs were less favorable for the RAPIT program by 66% (Table 5).

DISCUSSION

In line with previous studies (3), the RAPIT study has shown that for RA patients a long-term, high-intensity exercise program is safe and effective in improving phys-

Table 5. Cost-utility and cost-effectiveness analyses of the RAPIT exercise program compared with usual care for different utility, effectiveness, and cost measures*

	Estimated result	Statistical significance of difference in net benefit†
Total societal costs		
EQ-5D†	UC better cost utility for all WTP	Not significant for any WTP
SF-6D	ICER €67,000 per QALY‡§	Not significant for any WTP
VAS	UC better cost-utility for all WTP	Significant for WTP ≥€19,000 per QALY
HAQ	ICER €24,000 per HAQ point	Not significant for any WTP
MACTAR	ICER €210 per MACTAR point	Significant for WTP ≥€1,200 per MACTAR point
Societal costs of all physical therapy¶		
EQ-5D	UC better cost utility for all WTP	Significant for WTP ≤€160,000 per QALY
SF-6D	ICER €110,000 per QALY	Significant for WTP ≤€37,000 per QALY
VAS	UC better cost utility for all WTP	Significant for all WTP
HAQ	ICER €40,000 per HAQ point	Significant for WTP ≤€12,000 per HAQ point
MACTAR	ICER €340 per MACTAR point	Significant for WTP ≤€190 per MACTAR point and for WTP ≥€1,100 per MACTAR point

* RAPIT = Rheumatoid Arthritis Patients in Training exercise program; EQ-5D = EuroQol classification system; UC = usual care; WTP = willingness to pay; SF-6D = Short Form 6D; VAS = transformed visual analog scale rating personal health; HAQ = Health Assessment Questionnaire; MACTAR = McMaster Toronto Arthritis Patient Preference Interview; ICER = incremental cost-effectiveness ratio.
† Primary economic analysis, all others are sensitivity analyses.
‡ Net benefit = (effectiveness × WTP) − costs. Cost effectiveness is in favor of the RAPIT program for WTP below the ICER, and is in favor of UC for WTP above the ICER.
§ Purchasing power parity index: €1 = $1.05.
¶ RAPIT exercise program and individual physical therapy, including time and travel costs.

ical capacity and emotional status, without an increase in disease activity or radiographic joint damage (12). It also proved to be effective in improving functional ability and in slowing down age-related bone loss. In the economic evaluation reported here, we studied whether from the societal perspective clinical effectiveness is attained at reasonable costs.

The costs of exercise programs can vary considerably, depending on the number of participating patients, the accommodation, and the agreed income of the physical therapists. In the Dutch setting, the annual medical costs were estimated at €10,800 per RAPIT group, which is €780 per participating patient (€1 ≈ $1.05). The additional time and travel costs incurred by the patients almost doubled these costs and can seriously discourage participation by the patients. Costs of the RAPIT program were partly compensated by a decrease in individual physical therapy. As a result, the annual increase in medical costs of all physical therapy for patients participating in the RAPIT program was estimated at €430. The annual increase in total societal costs was estimated at €602 per patient. More research is needed on the development of less expensive exercise interventions. If effectiveness could be maintained with less therapists' supervision and more home-based exercises, this would cut down on costs. However, until now, home-based exercise programs designed for patients with RA did prove effective with respect to the improvement of physical capacity but, perhaps due to their low intensity or lack of impact, did not accomplish significant changes in functional capacity or bone mineral density (11,33,34).

For policy making from a societal perspective, costs need to be weighed against an effectiveness measure that is applicable and comparable for a wide range of diseases and treatments. The 3 QALY measures used in our study satisfy this requirement and, despite their conceptual differences, led to the same conclusion. An often-quoted rule of thumb is that costs up to €50,000 per QALY are acceptable (35). At that threshold, the cost utility of usual care was better than for the RAPIT program, and significantly so according to the VAS. We conclude that, according to current societal standards, the exercise classes did not improve the health valuation of the RA patients sufficiently to provide good value for money.

A number of remarks need to be made that temper the sharpness of this conclusion. Most importantly, our analysis has not taken into account preventive effects on cardiovascular and fracture risks and the associated cost savings (2), because the available data were insufficient for a quantitative extrapolation of future participation, effectiveness, damage, and costs. Second, although corrected for, results may have been influenced by some statistically significant baseline differences that we attribute to early dropout among healthier patients randomized to obtain usual care. Third, the used utility measures may be less appropriate for evaluation of the health of RA patients in clinical trials (36), because trials are usually powered for more sensitive primary outcome measures and patients continually adjust to their worsening health status (37). The few economic evaluations of physical therapy in somewhat similar patient groups either did not include utility measures (38–40) or also found insufficient improvement (41). Finally, the societal perspective ignores important financial considerations. For example, patients could be asked to contribute to the amount of what they would have to pay at a fitness center.

In economic evaluations, effectiveness is preferably estimated using utility measures because they aim to estimate the overall value of a treatment. A remarkable discrepancy revealed in our analysis is that significant and

clinically important improvements on the effectiveness measures targeted by the intervention, like physical capacity and functional ability, did not translate into a significant improvement on the utility measures. There can be several explanations for this finding. First, utility measures may not be sensitive enough to detect true improvements in overall value or, equivalently, true improvements in overall value may be too small to be detected by available utility measures. This explanation does not invalidate their use in economic evaluations, provided that the utility measures include the relevant domains, as was the case in our study. Second, the value of the improvements on the targeted effectiveness measures may be countered by unexpected or unknown changes in other domains, which is suggested here by the significant difference in the VAS.

To compare the efficiency of different treatments in rheumatology, disease-specific outcome measures can be used. We analyzed the primary clinical endpoints of the study, both measuring functional ability. The HAQ is used extensively in rheumatology research, but has been reported to be insensitive to changes due to exercise therapy (7,42). This was confirmed in the RAPIT study. Patients in the RAPIT group on average showed only a 0.01 improvement on a scale from 0 to 3, at the annual costs of €24,000 per HAQ point. The above-mentioned rule of thumb of €50,000 per QALY suggests that the estimated costs per HAQ point are unacceptably high.

The most sensitive measure in our study was the MACTAR: over the 2-year period, patients following the RAPIT program had an average MACTAR score that was 2.9 points better than for patients receiving usual care. Most of the difference in the MACTAR score was attributable to the items that assessed difficulty performing the individually selected impaired activities. Because the weighted MACTAR score, on average, assigns 3 points to these activities, the estimated 2.9 difference can be interpreted as that patients following the RAPIT program had 1 less relevant activity problem, at the annual cost of €630 per problem. Unfortunately, at this moment there are no data on what constitutes a clinically relevant MACTAR change, nor is there previous literature to decide whether the estimated costs per MACTAR point compare favorably with other interventions in rheumatology.

From the individual patient's perspective, costs need not play a role in the decision on whether or not to participate in exercise classes. During the study, most patients reported personal willingness to pay that was insufficient to cover the costs (on average €300 annually). Nevertheless after the study ended, 58% of the patients continued to follow exercise classes at their own costs for at least a year. If funding does not provide a problem, then there is no reason to discourage participation, provided qualified supervision is available to prevent injuries and to adjust the exercises to individual needs. Moreover, the reduction in costs for individual physical therapy could provide some scope for reimbursement.

From an individual patient perspective, there is much to be said for participating in long-term, high-intensity exercise classes. However, from a societal perspective and without taking possible preventive health effects into account, they provide insufficient improvement in the valuation of health to justify the additional costs.

REFERENCES

1. Berryman JW. The tradition of the "six things non-natural": exercise and medicine from Hippocrates through ante-bellum America. Exerc Sport Sci Rev 1989;17:515–59.
2. Macera CA, Hootman JM, Sniezek JE. Major public health benefits of physical activity. Arthritis Rheum 2003;49:122–8.
3. Stenstrom CH, Minor MA. Evidence for the benefit of aerobic and strengthening exercise in rheumatoid arthritis. Arthritis Rheum 2003;49:428–34.
4. Wong JB, Ramey DR, Singh G. Long-term morbidity, mortality, and economics of rheumatoid arthritis. Arthritis Rheum 2001;44:2746–9.
5. 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.
6. Huusko TM, Korpela M, Karppi P, Avikainen V, Kautiainen H, Sulkava R. Threefold increased risk of hip fractures with rheumatoid arthritis in Central Finland. Ann Rheum Dis 2001;60:521–2.
7. Hazes JM, van den Ende CH. How vigorously should we exercise our rheumatoid arthritis patients? Ann Rheum Dis 1996;55:861–2.
8. Van den Ende CHM, Hazes JM, 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.
9. Westby MD. A health professional's guide to exercise prescription for people with arthritis: a review of aerobic fitness activities. Arthritis Rheum 2001;45:501–11.
10. Van den Ende CHM, Vliet Vlieland TPM, Munneke M, Hazes JM. Dynamic exercise therapy for rheumatoid arthritis. Cochrane Database Syst Rev 2000;(2):CD000322.
11. 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.
12. De Jong Z, Munneke M, Zwinderman AH, Kroon HM, Jansen A, Ronday KH, et al. Is a long-term high-intensity exercise program effective and safe in patients with rheumatoid arthritis: results of a randomized controlled trial. Arthritis Rheum 2003;48:2415–24.
13. Torrance GW. Measurement of health state utilities for economic appraisal: a review. J Health Econ 1986;5:1–30.
14. The EuroQol Group. EuroQol: a new facility for the measurement of health-related quality of life. Health Policy 1990;16: 199–208.
15. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997;35:1095–108.
16. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-item health survey 1.0. Health Econ 1993;2:217–27.
17. Brazier J, Roberts J, Deverill M. The estimation of a preference-based measure of health from the SF-36. J Health Econ 2002;21:271–92.
18. Stiggelbout AM, Eijkemans MJ, Kiebert GM, Kievit J, Leer JW, de Haes HJ. The 'utility' of the visual analog scale in medical decision making and technology assessment: is it an alternative to the time trade-off? Int J Technol Assess Health Care 1996;12:291–8.
19. Tijhuis GJ, Jansen SJ, Stiggelbout AM, Zwinderman AH, Hazes JM, Vlieland TP. Value of the time trade off method for measuring utilities in patients with rheumatoid arthritis. Ann Rheum Dis 2000;59:892–7.
20. Wu AW, Jacobson KL, Frick KD, Clark R, Revicki DA, Freedberg KA, et al. Validity and responsiveness of the EuroQol as a measure of health-related quality of life in people enrolled in an AIDS clinical trial. Qual Life Res 2002;11:273–82.
21. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of

patient outcome in arthritis. Arthritis Rheum 1980;23:137–45.
22. Siegert CEH, Vleming LJ, Vandenbroucke JP, Cats A. Measurement of disability in Dutch rheumatoid arthritis patients. Clin Rheumatol 1984;3:305–9.
23. 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.
24. 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.
25. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press; 1996.
26. Van den Hout WB, Tijhuis GJ, Hazes JM, Breedveld FC, Vliet Vlieland TP. Cost effectiveness and cost utility analysis of multidisciplinary care in patients with rheumatoid arthritis: a randomised comparison of clinical nurse specialist care, inpatient team care, and day patient team care. Ann Rheum Dis 2003;62:308–15.
27. Oostenbrink JB, Koopmanschap MA, Rutten FF. Standardisation of costs: the Dutch manual for costing in economic evaluations. Pharmacoeconomics 2002;20:443–54.
28. Dutch Health Insurance Executive Board. Pharmacotherapeutic compass. Amstelveen: Dutch Health Insurance Executive Board; 1999.
29. Rutten FFH, Ineveld BMv, Ommen Rv, Hout BAv, Huijsman R. Kostenberekening bij Gezondheidsonderzoek: richtlijnen voor de praktijk. Utrecht: Jan van Arkel; 1993.
30. Koopmanschap MA, Rutten FF, van Ineveld BM, van Roijen L. The friction cost method for measuring indirect costs of disease. J Health Econ 1995;14:171–89.
31. O'Brien BJ, Briggs AH. Analysis of uncertainty in health care cost-effectiveness studies: an introduction to statistical issues and methods. Stat Methods Med Res 2002;11:455–68.
32. Desgagne A, Castilloux AM, Angers JF, LeLorier J. The use of the bootstrap statistical method for the pharmacoeconomic cost analysis of skewed data. Pharmacoeconomics 1998;13: 487–97.
33. Häkkinen A, Sokka T, Kotaniemi A, Kautiainen H, Jappinen I, Laitinen L, et al. Dynamic strength training in patients with early rheumatoid arthritis increases muscle strength but not bone mineral density. J Rheumatol 1999;26:1257–63.
34. Westby MD, Wade JP, Rangno KK, Berkowitz J. A randomized controlled trial to evaluate the effectiveness of an exercise program in women with rheumatoid arthritis taking low dose prednisone. J Rheumatol 2000;27:1674–80.
35. Ubel PA, Hirth RA, Chernew ME, Fendrick AM. What is the price of life and why doesn't it increase at the rate of inflation? Arch Intern Med 2003;163:1637–1.
36. Wolfe F, Hawley DJ. Measurement of the quality of life in rheumatic disorders using the EuroQol. Br J Rheumatol 1997; 36:786–93.
37. Wilson IB. Clinical understanding and clinical implications of response shift. Soc Sci Med 1999;48:1577–88.
38. Bakker C, Hidding A, Linden S, Doorslaer E. Cost effectiveness of group physical therapy compared to individualized therapy for ankylosing spondylitis: a randomized controlled trial. J Rheumatol 1994;21:264–8.
39. Severens JL, Oerlemans HM, Weegels AJ, 't Hof MA, Oostendorp RA, Goris RJ. Cost-effectiveness analysis of adjuvant physical or occupational therapy for patients with reflex sympathetic dystrophy. Arch Phys Med Rehabil 1999;80:1038–43.
40. Sevick MA, Bradham DD, Muender M, Chen GJ, Enarson C, Dailey M, et al. Cost-effectiveness of aerobic and resistance exercise in seniors with knee osteoarthritis. Med Sci Sports Exerc 2000;32:1534–40.
41. Patrick DL, Ramsey SD, Spencer AC, Kinne S, Belza B, Topolski TD. Economic evaluation of aquatic exercise for persons with osteoarthritis. Med Care 2001;39:413–24.
42. Van den Ende CH, Breedveld FC, Dijkmans BA, Hazes JM. The limited value of the Health Assessment Questionnaire as an outcome measure in short term exercise trials. J Rheumatol 1997;24:1972–7.
43. 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.
44. 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.