The effectiveness of behavioral graded activity in patients with osteoarthritis of hip or knee

Cindy Veenhof

The study presented in this thesis was performed at the Netherlands Institute for Health Services Research (NIVEL), Utrecht, The Netherlands. The study was funded by the Dutch Health Care Insurance Board.
Financial support for the printing of this thesis has been kindly provided by the Netherlands Institute for Health Services Research (NIVEL), Dutch Arthritis Association, and the Foundation 'Annafonds' in Leiden.
Cover: Bianca Veenhof
Printing: Universal Press, Veenendaal

ISBN 90-6905-793-X

http://www.nivel.nl
nivel@nivel.nl
Telefoon 030 2 729 700
Fax 030 2 729 729

VRIJE UNIVERSITEIT

The effectiveness of behavioral graded activity in patients with osteoarthritis of hip or knee

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Geneeskunde
op donderdag 14 september 2006 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Cindy Veenhof

geboren te Rockanje

promotoren:	prof.dr. J. Dekker
	prof.dr. J.W.J. Bijlsma
copromotor:	dr. C.H.M. van den Ende

we
bewegen
en zullen zelfs
aan het
dansen
gaan.

Contents

C h a p t e r

1

General introduction

Osteoarthritis

Osteoarthritis (OA) is a chronic musculoskeletal disorder, which can affect multiple joints. The most common joints to be affected are the knee (gonarthrosis) and hip (coxarthrosis). The prevalence of OA is increasing, mainly due to increasing life spans, and the consequences of OA have a significant impact on society (1). The incidence in general practice in the Netherlands increases with age and is 0.9/1000 per year for OA of the hip and 1.5/1000 per year for OA of the knee. OA of both hip and knee is more common in female than in male patients (2).

In pathological terms, OA is described as a condition of cartilage degeneration, stiffening of the underlying subchondral bone (sclerosis), and active new bone formation (osteophytes). In later stages of the disease, the pathological changes in cartilage and bone are followed by pathological changes in other tissues of the joint and its surroundings, such as synovial membrane, capsula, ligaments and muscles. This may lead to capsular restriction, instability of the joint, and muscle atrophy (3-5). These changes will often lead to a reduced loadability of joints, which will often result in functional disability.

There is no single accepted classification system for OA. For pragmatic reasons, the clinical criteria for the classification of OA of the hip and knee of the American College of Rheumatology, are used most, also in primary care (6;7). These criteria are presented in Table 1.

Table 1 Clinical criteria of the ACR for the classification of OA of hip/knee

Osteoarthritis of the hip	**Osteoarthritis of the knee**
Hip pain *and* Hip internal rotation < 15º *and* ESR ≤ 45 mm/hour (if no ESR available: substitute hip flexion ≤ 115º)	Knee pain *and* At least 3 of the following 6: - Age > 50 years - Morning stiffness < 30 minutes - Crepitus - Bony tenderness - Bony enlargement - No palpable warmth
OR	
Hip pain *and* Hip internal rotation ≥ 15º *and* Pain on hip internal rotation *and* Morning stiffness of the hip ≤ 60 minutes *and* Age > 50 years	

ESR = Erythrocyte Sedimentation Rate.

Physical consequences of osteoarthritis

The consequences of OA can be grouped in the International Classification of Functioning, Disability, and Health, known as the ICF (8). The ICF is a multipurpose classification and is designed by the World Health Organization to record and organise a wide range of information about health and health-related states. The ICF has two parts, each containing two separate components, described as follows: Part 1 covers functioning and disability and includes body functions, body structure, activities, and participation. Part 2 covers contextual factors, including environmental factors and personal factors. The ICF is illustrated in Figure 1. The term "disability" is defined as "impairment" (dysfunction or loss of "body functions or structure"), "limitation" (the difficulty an individual may experience in involvement in life situations (activities)) or "restriction" (problems an individual may experience in involvement in life situations (participation)).

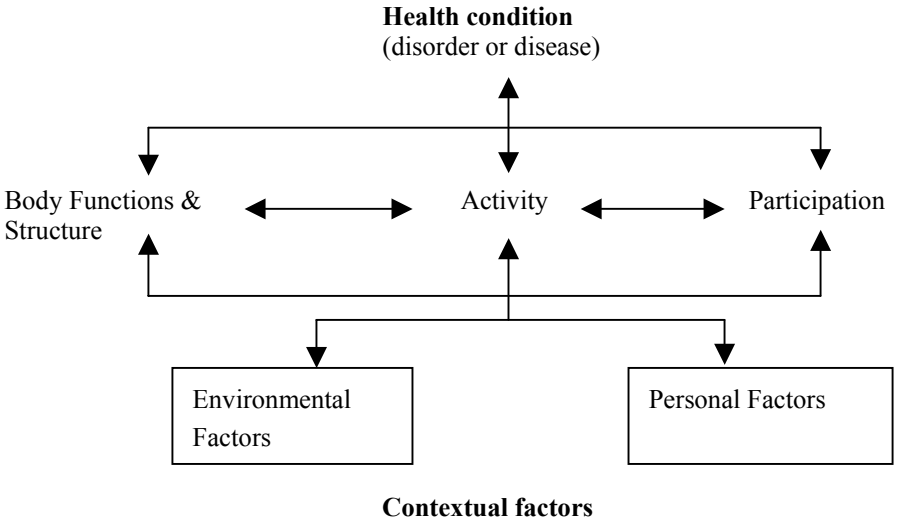

Figure 1 The ICF model (8)

The main impairments of patients with OA of the hip or knee are pain, that typically worsens with weight bearing and activity and improves with rest, and stiffness and gelling of the involved joint after periods of inactivity (9). As the disease progresses, additional impairments are loss of joint mobility, decrease of muscle strength, postural deformities and instability (10;11). Especially in elderly patients, OA has a major impact on their functioning in daily life and frequently leads to limitations of activities (1;9;11). For example, the patient's performance of daily activities such as walking, kneeling, dressing and gardening becomes limited. Also, coping behavior and psychological factors such as depression, fear, and avoidance behavior can influence pain and movement functions, and lead to limitations of activities. OA can lead to absence from work and to a decline in quality of life.

Recently, based on the ICF, a typical spectrum of problems in functioning of patients with OA was defined. Such generally-agreed-on lists of ICF categories can among other things serve as Brief Core Set to be rated in all patients included in a clinical study with OA. The Brief Core Set OA includes 13 items and is presented in Table 2 (12).

Table 2 International Classification of Functioning, Disability and Health (ICF)-categories included in the Brief Core Set for osteoarthritis.

ICF component	**ICF category title**	**ICF code**
Body functions	Sensation of pain	b280
	Mobility of joint functions	b710
	Muscle power functions	b730
Body structures	Structure of lower extremity	s750
	Structure of upper extremity	s730
	Additional musculoskeletal structures related to movement	s770
Activities and participation	Walking	d450
	Dressing	d540
	Hand and arm use	d445
Environmental factors	Immediate family	e310
	Products and technology for personal use in daily living	e115
	Health services, systems and policies	e580
	Design, construction and building products and technology of buildings for public use	e150

Current insights in factors associated with a high risk for limitations of activities in patients with OA, are mainly based on cross-sectional studies. Factors linked to functional decline include pain, obesity, use of passive coping styles (such as worrying, catastrophizing, and avoidance of physical activity), muscle weakness, poor aerobic capacity and, in some studies, radiographic disease severity (1;10;13;14). However, little is known about the course of pain and disability in patients with OA over time. A recent review on the course of functional status and pain in OA of the hip and knee, stated that functional status and pain will deteriorate over time, i.e. after three years of follow-up, in both hip and knee OA. Furthermore, increased laxity, proprioceptive inaccuracy, older age, greater BMI, greater knee pain intensity and increase in knee pain appear to increase the risk of functional deterioration of patients with knee OA, in the first three years of follow-up. On the other hand, protective factors for functional status of knee OA are muscle strength, mental health, self-efficacy, social support and aerobic exercise (15).

Avoidance model

As mentioned before, not only direct causes of OA related to disability, but also psychological factors, such as coping / avoidance of physical activity have been found to be related to a higher level of disability (10;13-15). The avoidance model (Figure 2) is a framework which can be used to illustrate the effect of avoidance of activity on disability (16;17). The theory of this model states that patients no longer perform certain activities because they anticipate that these activities increase pain and suffering. In the short term, pain can be reduced by avoiding activities.
However, in the long term, avoidance of activities can have both physical (loss of mobility, muscle strength and fitness) and psychological (loss of self-esteem, depression) consequences. All these consequences may augment disability. For example, due to muscle weakness, the joints become less stable, which subsequently leads to a higher level of disability. Consequently, the patient avoids activity even more, thus entering a vicious circle towards a higher level of disability.

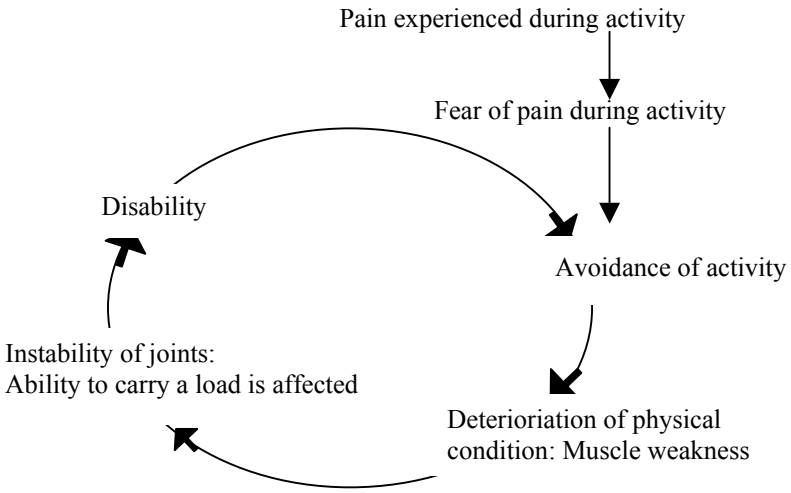

Figure 2 The Avoidance Model (17)

Management of osteoarthritis

The principle objectives of managing OA are to control pain adequately, improve function and reduce disability. Since the status and requirements of patients can change over time, it is necessary to review and adjust treatment regularly. Both pharmacologic therapy and non-pharmacologic interventions can be given to patients with OA of the hip and/or knee (11;18-20). Potential pharmacologic therapies are analgesics, nonsteroidal anti-inflammatory drugs (NSAIDs), coxibs, and intra-articular steroids. Drug therapy for pain management is most effective when combined with non-pharmacologic strategies (18;19;21). The main non-pharmacologic strategies are exercise, bracing and footwear, weight loss, patient education, acupuncture, behavioral interventions and surgical treatment (11;18-20;22).

Exercise therapy in osteoarthritis

Exercise therapy involves the prescription of muscular contraction and bodily movement ultimately to improve the individual's overall function and to help meet the demands of daily living (23;24). The main areas treated by

exercise therapy are: limitations of activities, such as walking and bending down; if necessary, restrictions in household and occupational tasks; the causal impairments, such as loss of muscle strength, pain and loss of mobility; and inadequate coping strategies (25). There is strong evidence that exercise therapy has a short term benefit in patients with OA of the hip and/or knee. Beneficial effects of exercise therapy on pain, physical function and patient global assessment (PGA) have been demonstrated (23;26;27). Therefore, exercise therapy is recommended as intervention to decrease the problems associated with OA and to stimulate the patient's functioning and activities (19;21;28). In line with these recommendations, exercise therapy is of central importance in the Dutch guidelines for clinicians and physiotherapists (25;29). Although exercise therapy is beneficial at the short term, beneficial post-treatment effects of exercise therapy in patients with OA seem to decline over time and finally disappear. This decline is thought to be related to the difficulties people have in maintaining adherence to prescribed exercise. (26;27).

Cognitive-behavioral treatment

An important issue is how patients can be stimulated to adhere to exercise therapy. To enhance long-term benefit, adherence to exercise therapy is of utmost importance. When exercises are integrated into the daily life of a person, it will be easier to sustain these exercises for a longer period.

Traditionally, the management of OA has been based on a disease-oriented biomedical model, which considers pain to be a sign of physiological damages. Biomedical treatments focus mainly on pain relief and the treatment of disability. Moreover, treatment is guided by the amount of pain experienced by the patients, leading to a pain-contingent approach (18-20).

Recently, the focus of attention within physiotherapy has shifted towards a behaviorally-oriented operant conditioning approach, which is based on the biopsychosocial perspective (30-35). According to the biopsychosocial model, pain is not necessarily caused by underlying pathology or impairment; but psychological and social factors may also be important in the development and maintenance of complaints (36;37). Based on this model three cognitive-behavioral modalities exist: operant, cognitive, and respondent (38). Of these three modalities, the focus within the physiotherapy, is mainly on behavioral graded activity. Behavioral graded

activity is based on the operant conditioning principles, introduced to the area of pain by Fordyce (36). According to this theory, overt behavior that accompanies pain (e.g. complaining, avoidance of activities) should be understood as types of behavior. The frequency of these behaviors might be influenced by the consequences of that behavior. Essential features of an operant conditioning approach are positive reinforcement of healthy behavior, and consequent withdrawal of attention towards pain behavior and time-contingent instead of pain-contingent management (31;36;38). Behavioral graded activity always starts with the determination of baseline values of the individually chosen activities. The patient and therapist then gradually increase the level of activities towards a preset goal.

To date, the effectiveness of operant behavioral treatment programmes have been documented in several studies (33). More specifically, graded activity programmes have been shown to improve the level of daily activities, reduce disability, and reduce duration of sick leave in patients with back pain (31;32). However, other studies reported no significant effects on pain and functional status (34;35). So far, no studies were performed on the effectiveness of graded activity in patients with a progressive and specific chronic disease, like hip and/or knee OA. Since patients with hip and/or knee OA tend to avoid pain by decreasing their activities, such intervention seems appropriate to increase their level of activities and to increase patients' adherence to these activities. Another factor which contributes to the maintenance of adherence is to include repeated treatment sessions over a long time period in the graded activity treatment, the so-called booster-sessions (39). These booster-sessions also allow patients more time to realize a behavioral change towards a more active lifestyle.

Outcome measures

Since the effects of physiotherapy/exercise therapy are relatively small, using appropriate instruments which are able to detect differences over time, is of main importance in clinical trials on the effectiveness of such therapies. Nowadays, many different instruments are used, both in clinical practice and research on patients with OA to evaluate outcome of treatment. The

Outcome Measures in Rheumatology Clinical Trials (OMERACT) group recommended the use of the domains pain, physical function and patient global assessment in clinical trials in patients with OA. Stiffness was described as an important optional domain (40). This is in line with the recommendations of several guidelines for outcome measurement in OA trials (European League Against Rheumatism (EULAR) (41), Food and Drug Administration (FDA) (42), and Slow-acting Drugs in Osteoarthritis (SADOA) (43)). However, these guidelines differ in their recommendations of specific instruments or do not include recommendations of instruments at all (42;44). Nowadays, a large number of instruments is available to assess the outcome dimensions of the OMERACT. The question arises which instruments are most appropriate to use.

Aim and scope of this thesis

Exercise therapy has a short term benefit in patients with OA of the hip and/or knee. However, these effects seem to decline over time and finally disappear (26;27). The aim of the current study was to compare the effectiveness of a behavioral graded activity program with usual physiotherapeutic care (UC), according to the Dutch physiotherapy guideline (25), in patients with OA of the hip and/or knee. The hypothesis was that in the long term (> 6 months after treatment) behavioral graded activity results in less pain, less limitations in activities and better patient global assessment compared to treatment according to the guideline. Therefore, we performed a randomized clinical trial, in which we compared both treatments (chapter 2). We also determined the cost-effectiveness of behavioral graded activity (chapter 3). Furthermore, we investigated whether specific subgroups of patients particularly benefit from behavioral graded activity (chapter 4). Additionally, we investigated whether a consistent pattern of factors influence the adherence to behavioral graded activity (chapter 5). Patients were recruited for our study by physiotherapists and by articles in local newspapers. We examined the effect of these two different recruitment methods on the characteristics of the participants and on the efficacy of both treatments (chapter 6). And finally, to facilitate the choice of the most appropriate questionnaires to measure the OMERACT outcome dimensions

in patients with OA of the hip and/or knee, we performed a systematic review on the published self assessment instruments on pain, physical function and patient global assessment for patients with hip and/or knee OA (chapter 7).

Chapter 2 to 7 were originally written as separate articles for publication in scientific journals. Therefore, some overlap between chapters exists, especially with regard to the description of the methodology and interventions.

In summary, the following research questions are being examined in this thesis:

- What is the effectiveness of behavioral graded activity in patients with osteoarthritis of the hip and/or knee? (Chapter 2)
- What is the cost-effectiveness of behavioral graded activity in patients with osteoarthritis of the hip and/or knee? (Chapter 3)
- Which groups of patients with osteoarthritis of the hip and/or knee benefit particularly from behavioral graded activity? (Chapter 4)
- Which factors influence the success of behavioral graded activity in patients with osteoarthritis of the hip and/or knee? (Chapter 5)
- What is the influence of two recruitment strategies on the characteristics of patients and on the outcome of a randomized controlled trial involving patients with osteoarthritis of the hip and/or knee? (Chapter 6)
- What is the evidence available on the psychometric qualities of osteoarthritis questionnaires? (Chapter 7)

References

1. Felson DT, Lawrence RC, Dieppe PA, Hirsch R, Helmick CG, Jordan JM et al. Osteoarthritis: new insights. Part 1: the disease and its risk factors. Ann Intern Med 2000; 133(8):635-646.
2. Van der Waal AJ, Bot SD, Terwee CB, van der Windt DA, Schellevis FG, Bouter LM et al. The incidences of and consultation rate for lower extremity complaints in general practice. Ann Rheum Dis 2006; 65(6):809-815.
3. Dieppe P. The classification and diagnosis of osteoarthritis. In: Kuettner KE, Goldberg VM, editors. Osteoarthritic disorders. Rosemont IL: American Academy of Orthopaedic Surgeons, 1995: 5-12.
4. Duthrie R, Bentley G. Mercer's Orthopaedic surgery. Oxford press, 1996.
5. Threlkeld AJ, Currier DP. Osteoarthritis. Effects on synovial joint tissues. Phys Ther 1988; 68(3):364-370.
6. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Arthritis Rheum 1986; 29(8):1039-1049.
7. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991; 34(5):505-514.
8. World Health Organization. International Classification of Functioning, Disability and Health: ICF. Geneva: WHO, 2001.
9. Recommendations for the medical management of osteoarthritis of the hip and knee: 2000 update. American College of Rheumatology Subcommittee on Osteoarthritis Guidelines. Arthritis Rheum 2000; 43(9):1905-1915.
10. Dekker J, Boot B, van der Woude LH, Bijlsma JW. Pain and disability in osteoarthritis: a review of biobehavioral mechanisms. J Behav Med 1992; 15(2):189-214.
11. Sarzi-Puttini P, Cimmino MA, Scarpa R, Caporali R, Parazzini F, Zaninelli A et al. Osteoarthritis: an overview of the disease and its treatment strategies. Semin Arthritis Rheum 2005; 35(1 Suppl 1):1-10.
12. Dreinhofer K, Stucki G, Ewert T, Huber E, Ebenbichler G, Gutenbrunner C et al. ICF Core Sets for osteoarthritis. J Rehabil Med 2004;(44 Suppl):75-80.
13. Creamer P, Lethbridge-Cejku M, Hochberg MC. Determinants of pain severity in knee osteoarthritis: effect of demographic and psychosocial variables using 3 pain measures. J Rheumatol 1999; 26(8):1785-1792.
14. Keefe FJ, Caldwell DS, Queen K, Gil KM, Martinez S, Crisson JE et al. Osteoarthritic knee pain: a behavioral analysis. Pain 1987; 28(3):309-321.
15. Van Dijk GM, Dekker J, Veenhof C, van den Ende CHM. The course of functional status and pain in osteoarthritis of hip and knee: a systematic review. Accepted for publication in Arthritis Rheum, 2006.

16. Dekker J, Tola P, Aufdemkampe G, Winckers M. Negative affect, pain and disability in osteoarthritis patients: the mediating role of muscle weakness. Behav Res Ther 1993; 31(2):203-206.
17. Steultjens MP, Dekker J, Bijlsma JW. Avoidance of activity and disability in patients with osteoarthritis of the knee: the mediating role of muscle strength. Arthritis Rheum 2002; 46(7):1784-1788.
18. Felson DT, Lawrence RC, Hochberg MC, McAlindon T, Dieppe PA, Minor MA et al. Osteoarthritis: new insights. Part 2: treatment approaches. Ann Intern Med 2000; 133(9):726-737.
19. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR et al. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. Arthritis Rheum 1995; 38(11):1541-1546.
20. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR et al. Guidelines for the medical management of osteoarthritis. Part I. Osteoarthritis of the hip. Arthritis Rheum 1995; 38(11):1535-1540.
21. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P et al. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Ann Rheum Dis 2003; 62(12):1145-1155.
22. Sharma L. Nonpharmacologic management of osteoarthritis. Curr Opin Rheumatol 2002; 14(5):603-607.
23. Smidt N, de Vet HC, Bouter LM, Dekker J, Arendzen JH, de Bie RA et al. Effectiveness of exercise therapy: a best-evidence summary of systematic reviews. Aust J Physiother 2005; 51(2):71-85.
24. Tan JC, Horn SE. Practical Manual of Physical Medicine and Rehabilitation: Diagnostics, Therapeutics, and Basic Problems. St Louis: Mosby, 1998.
25. Vogels EMHM, Hendriks HJM, van Baar ME, Dekker J, Hopman-Rock M, Oostendorp RA et al. Clinical practice guidelines for physical therapy in patients with osteoarthritis of the hip or knee. Amersfoort: KNGF, 2001.
26. Fransen M, McConnel S, Bell M. Therapeutic exercise for people with osteoarthritis of the hip or knee. A systematic review. J Rheumatol 2002; 29:1737-1745.
27. Van Baar ME, Assendelft WJ, Dekker J, Oostendorp RA, Bijlsma JW. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review of randomized clinical trials. Arthritis Rheum 1999; 42(7):1361-1369.
28. Philadelphia Panel evidence-based clinical practice guidelines on selected rehabilitation interventions for knee pain. Phys Ther 2001; 81(10):1675-1700.
29. Bijlsma JW. Behandeling bij osteoartrose. Ned T Reum 2005; 1:27-29.
30. Fordyce WE, Brockway JA, Bergman JA, Spengler D. Acute back pain: a control-group comparison of behavioral vs traditional management methods. J Behav Med 1986; 9(2):127-140.

31. Lindstrom I, Ohlund C, Eek C, Wallin L, Peterson LE, Fordyce WE et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. Phys Ther 1992; 72(4):279-290.
32. Lindstrom I, Ohlund C, Eek C, Wallin L, Peterson LE, Nachemson A. Mobility, strength, and fitness after a graded activity program for patients with subacute low back pain. A randomized prospective clinical study with a behavioral therapy approach. Spine 1992; 17(6):641-652.
33. Morley S, Eccleston C, Williams A. Systematic review and meta-analysis of randomized controlled trials of cognitive behaviour therapy and behaviour therapy for chronic pain in adults, excluding headache. Pain 1999; 80(1-2):1-13.
34. Ostelo RW, de Vet HC, Berfelo MW, Kerckhoffs MR, Vlaeyen JW, Wolters PM et al. Effectiveness of behavioral graded activity after first-time lumbar disc surgery: short term results of a randomized controlled trial. Eur Spine J 2003; 12(6):637-644.
35. Staal JB, Hlobil H, Köke AJA, Twisk JWR, Smid T, van Mechelen W. The effects of graded activity for low back pain in occupational health on sick leave, functional status and pain: a randomised controlled trial. Ann Intern Med 2004; 140:77-84.
36. Fordyce WE. Behavioral methods for chronic pain and illness. St. Louis: Mosby, 1976.
37. Turk DC, Okifuji A. Evaluating the role of physical, operant, cognitive, and affective factors in the pain behaviors of chronic pain patients. Behav Modif 1997; 21(3):259-280.
38. Vlaeyen JW, Haazen IW, Schuerman JA, Kole Snijders AM, Eek van H. Behavioural rehabilitation of chronic low back pain: comparison of an operant treatment, and operant-cognitive treatment and an operant-respondent treatment. Br J Clin Psychol 1995; 34:95-118.
39. Marks R, Allegrante JP. Chronic osteoarthritis and adherence to exercise: a review of the literature. J Aging Phys Act 2005; 13(4):434-460.
40. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997; 24(4):799-802.
41. Guidelines for the clinical investigation of drugs used in rheumatic diseases, European drug guidelines, Series 5. World Health Organization, Regional Office for Europe. Copenhagen: European League Against Rheumatism. 1985.
42. Bellamy N. Osteoarthritis clinical trials: candidate variables and clinimetric properties. J Rheumatol 1997; 24(4):768-78.
43. Lequesne MF, Brandt KF, Bellamy NF, Moskowitz RF, Menkes CJ, Pelletier JP et al. Guidelines for testing slow acting drugs in osteoarthritis. J Rheumatol Suppl 1994; 41:65-71.
44. Bellamy N. Outcome measurement in osteoarthritis clinical trials. J Rheumatol 22[43S], 49-51. 1995.

Chapter

2

Effectiveness of behavioral graded activity in patients with osteoarthritis of hip and/or knee: a randomized clinical trial

Accepted for publication as:
Veenhof C, Köke AJA, Dekker J, Oostendorp RAB, Bijlsma JWJ, Van Tulder MW, Van den Ende CHM. Effectiveness of behavioral graded activity in patients with osteoarthritis of hip and/or knee: a randomized clinical trial. Planned for publication in Arthritis & Rheumatism (Arthritis Care & Research), October 2006.

Abstract

Objective. To determine the effectiveness of a behavioral graded activity program (BGA) compared to usual care (UC; exercise therapy and advice, according to the Dutch guideline for physiotherapy) in patients with osteoarthritis of the hip and/or knee. The BGA intervention aims to increase activity in the long term and consists of an exercise program with boostersessions, using operant treatment principles.

Methods. A cluster randomized trial involving 200 patients with hip and/or knee OA (ACR-criteria). Primary outcome measures were pain (VAS and WOMAC), physical function (WOMAC), and patient global assessment (PGA), assessed at week 0, 13, 39, and 65. Secondary outcome measures comprised tiredness (VAS), patient-oriented physical function (MACTAR), 5-meter walking time, muscle strength, and range of motion. Data were analyzed according to intention-to-treat principle.

Results. Both treatments showed beneficial within-group effects, both at short term and in the long term. The mean differences between the 2 groups for pain and functional status were not statistically significant. Significant differences in favor of BGA were found for the functional scale MACTAR and 5-meter walking time test at week 65.

Conclusion. Since both interventions resulted in beneficial long term effects, the superiority of BGA over UC has not been demonstrated. Therefore, BGA seems an acceptable method to treat patients with hip and/or knee OA, with equivalent results compared with UC.

Introduction

Osteoarthritis (OA) is a common joint disorder, causing pain, joint stiffness, muscle weakness, and joint instability, and threatening mobility and an active lifestyle (1;2). There is strong evidence that exercise therapy has a short term benefit for OA. Exercise therapy involves the prescription of muscular contraction and bodily movement ultimately to improve the individual's overall function and to help meet the demands of daily living (3;4). Beneficial effects of exercise therapy on pain, physical function, and patient global assessment (PGA) have been demonstrated (5;6). Therefore, exercise therapy is recommended as intervention to decrease the problems associated with OA and to stimulate the patient's functioning and activities (7-9). In line with these recommendations, exercise therapy is of central importance in the Dutch guidelines for clinicians and physiotherapists (10;11).

Although exercise therapy is beneficial at the short term, beneficial post-treatment effects of exercise therapy in patients with OA seem to decline over time and finally disappear (5;6). This was found in the study of Van Baar (12;13), comparing treatment from general practitioners (usual care) with treatment from general practitioners plus a 12-week period of exercise therapy. After 12 weeks, exercise therapy was more effective in reducing pain and disability compared to usual care. The size of the effects were medium and small respectively. However, at 24 weeks exercise therapy was only associated with a small to moderate effect on pain, and at 36 weeks no differences were found between intervention groups.

To enhance long term effects of exercise therapy, integration of exercise therapy with daily performed activities, based on cognitive-behavioral principles and additional boostersessions, seems promising (14;15). The graded activity intervention is an exercise program applying operant behavioral principles to increase the time of performance of daily activities. Operant behavioral principles include positive reinforcement of healthy behavior and consequent withdrawal of attention towards pain behavior; pain management is time-contingent instead of pain-contingent (16-18). Several studies investigated the effectiveness of graded activity programs, mainly in workers with subacute non-specific low back pain. Positive effects were found on the duration of sick leave, however, the

effects on pain and functional status were not significant (15;17;19;20). So far, no studies were performed on the effectiveness of graded activity in patients with an irrecoverable chronic disease, like hip and/or knee OA. Since these patients tend to avoid pain by decreasing their activities, such intervention seems appropriate to increase their level of activities.

The objective of the current study was to compare the effectiveness of a behavioral graded activity program (BGA) with usual physiotherapeutic care (UC), according to the Dutch physiotherapy guideline (11), in patients with OA of the hip and/or knee. The hypothesis was that in the long term (> 6 months after treatment) BGA results in less pain, less limitations in activities, and better patient global assessment compared to treatment according to the guideline.

Methods

Study design

A cluster randomized trial was conducted, comparing two interventions, BGA and UC. Assessments were conducted at 0, 13, 39, and 65 weeks. The study was approved by the Medical Ethical Committee of the VU University Medical Center, Amsterdam, The Netherlands.

Study population

Physiotherapists

A random sample of 600 physiotherapists, practicing in the primary care in the same district in the central region of the Netherlands, were invited (by letter) to participate in the study. This sample, of our Institute's-national database of physiotherapists, was representative for all primary care physiotherapists in the Netherlands. A total of 100 physiotherapists responded of whom 87 (divided over 72 practices) were willing and able to participate. To avoid contamination of interventions, cluster randomization was performed at the level of the participating physiotherapeutic practices. The participating practices were randomly assigned to 1 of the 2 treatment regimes by means of a computer generated random sequence table.

Patients

Inclusion criteria were OA of hip or knee according to the clinical criteria of the American College of Rheumatology (21;22). Exclusion criteria were: other pathology explaining the complaints; complaints in <10 out of 30 days; treatment for these complaints with exercise therapy in the preceding six months; <50 or >80 years of age; indication for hip or knee replacement within one year; contraindication for exercise therapy; inability to understand the Dutch language; and a high level of physical function, because patients who perform at a high level of physical function at baseline do not need to increase their level of physical function. A high level of physical function was operationalized on a score of <2 on the walking ability and physical function sections of the Algofunctional index (23).

Patients were recruited in 2 ways. First, patients referred to physiotherapy were recruited by the participating physiotherapists at their first visit to the physiotherapist of their own choice (November 2001-May 2003). Because the recruitment rate was rather slow, a second recruitment strategy was used, i.e. patients responded to articles about the benefit of exercise therapy and the performed study, published in local newspapers (November 2002-May 2003). Patients who responded to the newspaper articles were allocated to a physiotherapist participating in the trial by choosing a physiotherapist from a list of participating physiotherapists. At this time, the patients were not aware of the kind of intervention (BGA or UC) the physiotherapists were assigned to. An extensive description of the recruitment strategies and the influence of these strategies on the study population is published elsewhere (24). The same inclusion procedure was performed for all patients. First, they received oral information by phone, after which a first screening was performed (by phone). If patients were eligible, written information was sent and a final screening visit at home was planned (to control for the ACR criteria and exclusion criteria) and, if patients were willing and eligible, informed consent was signed.

Interventions

Behavioral Graded Activity (BGA)

BGA is a behavioral treatment integrating the concepts of operant conditioning with exercise therapy comprising boostersessions. BGA was based on the time-contingency management as described by Fordyce (16)

and applied by Lindström (17). The intervention is directed at increasing the level of activities in a time-contingent way, with the goal to integrate these activities in the daily living of the patients. Appendix 1 presents a description of the concept and content of the BGA intervention as applied in our study. The BGA treatment was outlined in a complete protocol and included written materials (e.g., education messages, activity diaries, performance charts). The treatment consisted of a 12-week period with a maximum of 18 sessions, followed by 5 pre-set boostermoments with a maximum of 7 sessions (respectively in week 18, 25, 34, 42, and 55).

Usual care (UC)
The physiotherapists in the usual care group, were advised to treat the patients according to the Dutch physiotherapy guideline for patients with hip and/or knee OA. This guideline consists of general recommendations, emphasizing provision of information and advice, exercise therapy, and encouragement of a positive coping with the complaints (Appendix 2) (11). The treatment consisted of a maximum of 18 sessions within a period of 12 weeks. The treatment could be discontinued within the 12-week period if, according to the physiotherapists, all treatment goals had been achieved.

Both BGA and usual care were given individually by physiotherapists in primary care and included home-based exercises. One session in primary care lasted approximately 30 minutes. Physiotherapists in both treatment groups received education on the allocated treatment and the BGA physiotherapists were supported and advised by phone and meetings during the study. BGA physiotherapists received a 2 day- training, which focused on specific skills, necessary to perform a behavioral treatment like BGA. The UC physiotherapists received a 4 hour-training concerning the Dutch guideline. All physiotherapists documented every session on standardized registration forms, including deviations from the protocol.

Outcome assessment
Demographic and clinical data were collected for each patient including age, sex, education, height, weight, location of OA, duration of complaints and the presence of other chronic disorders. X-rays of the hip and/or knee were scored by a rheumatologist following a standardized procedure according to

the Kellgren and Lawrence scale (consisting of five degrees: 0, no OA;1, doubtful OA; 2, minimal OA; 3, moderate OA; and 4, severe OA.) (25;26).
Impairments. Patients rated their pain at the moment of assessment and tiredness in the past week on a visual analog scale (VAS;0-10). Furthermore, pain in the last 48 hours was assessed with the pain subscale of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC, range 0-20) (27). Measurements of assisted active range of joint motion (ROM) of the knee (flexion, extension) and the hip (flexion, extension, external rotation, internal rotation, abduction) were performed with a goniometer, according to a standardized protocol (28). Isometric muscle strength of knee (extension) and hip (extension, abduction) was measured with the MicroFet, a hand-held dynamometer (29). The measurements of both ROM and muscle strength were repeated two times, the average score was used in the analyses (12).
Physical Function. Physical function was assessed with the physical function subscale of the WOMAC (27) and the MACTAR (30). The ability to walk was measured by a 5-meter walking time test (in seconds).
Activity Level. The level of physical activity was determined by the total minutes of performed activities as assessed by the SQUASH (Short Questionnaire to Assess Health Enhancing Physical Activity) (31).
Self perceived change. Patient global assessment (PGA) was assessed by patients on a 8-point scale (1=vastly worsened; 8=completely recovered) (32).
Health-related quality of life. Quality of life was determined with the MOS Short Form 36 (SF-36), all 8 subscales were assessed (33).

All outcome measures had good psychometric qualities. Primary outcome measures were pain (VAS and WOMAC), physical function (WOMAC) and PGA, according to the core set of outcome measures of clinical trials with patients with OA defined by the Outcome Measures in Rheumatology Clinical Trials (OMERACT) group (34). All primary and secondary outcome measures were obtained at baseline, 13 weeks, 39 weeks, and 65 weeks, with the exception of the outcome measures ROM, muscle strength, and walking time which were obtained at baseline, 13 weeks, and 65 weeks.

Assessments were performed on a test location, in the presence of research assistants; the exception was the assessment at 39 weeks, consisting of only questionnaires, which was sent by mail.

Blinding

Three trained research assistants, who were blinded for the assigned treatment, performed all assessments. Patients were instructed not to give information about the allocated treatment to the research assistants. The research assistants were asked to guess the assigned treatment immediately after measurements at week 13 and week 65. Because of the kind of intervention, patients and physiotherapists could not be blinded for the assigned treatment.

Statistical analysis

The target sample size was 200 patients. This number yields to a power of 80% to detect a 25% difference in patient global assessment (PGA) and small to medium-sized effects (effect-size=0.2-0.4) in the outcome measures pain and physical functioning, at two-sided significance level of 0.05 given a maximum loss to follow up of 20% (35).

The ratings of PGA were dichotomized as improved ("completely recovered", "very much improved" and "much improved") versus not improved ("slightly improved", "not changed", "slightly worsened", "much worsened" and "vastly worsened") and odds ratios (OR) were calculated to test differences between groups. Muscle strength data were corrected for body mass by dividing the raw scores by the patient's weight. Next, z-scores were computed and added for knee (containing 2 items, extension both sides) and hip (containing 4 items, extension and abduction, both sides) separately, as described by Steultjens et al. (36). Concerning ROM, mean scores of left and right side were calculated and used for analyses, as described by Steultjens et al. (37).

The statistical analysis was carried out according to the intention-to-treat principle. Patient data were analysed in the intervention group to which they had initially been assigned. This included withdrawals and patients not treated according to the assigned treatment. Baseline comparability was performed by descriptive statistics to examine if randomization was successful. Change scores were calculated by subtracting the baseline scores

from the post-treatment scores (week 13, 39, and 65, respectively) and were compared for the two intervention groups using Student's t-test. In order to adjust for differences in patients' condition, multiple linear regression analyses were performed with the change scores as dependent variable and type of intervention as independent variable. The following characteristics were used as covariates in the adjusted analyses: the baseline score of each outcome measure, duration of complaints, location of OA (hip, knee, or both), age, sex, and recruitment method (physiotherapist or newspaper). P values < 0.05 were considered statistically significant.

In addition, multilevel analyses and per-protocol analyses were performed. With multilevel analysis, it is possible to correct on the one side for dependency of observations within subjects and, on the other side, to take into account the variation between physiotherapists (38;39). To check whether multilevel analyses on the physiotherapist level was required, independency of observations within and among physiotherapists was determined by calculation of intraclass correlation coefficients. Per-protocol analyses were performed, excluding all patients with deviations from the treatment protocol. Deviations were defined as less than 6 sessions physiotherapy within the first 12 weeks (both groups), or less than 2 boostersessions (BGA) after the first 12 weeks, or a total hip / knee replacement during the whole study period (both groups).

Results

Study population

Information regarding the patient flow through the trial is presented in Figure 1. During the study, 110 patients were recruited by participating physiotherapists and 90 patients after publication of articles in local newspapers. A total number of 200 patients were included: 97 patients in the BGA group and 103 patients in the UC group.

The BGA group and the UC group had similar baseline characteristics and baseline values of the outcome measures, as presented in Table 1. Assessment at week 13 was completed by 90 BGA patients and 102 UC patients; 82 BGA patients and 88 UC patients completed the week 39 assessment and 87 BGA patients and 92 UC patients completed the trial up

to 65 weeks. The numbers and reasons of loss to follow up were similar in the intervention groups, with the exception that more UC patients were lost to follow up because of lack of motivation (7 UC patients versus 2 BGA patients). For 10 UC patients treatment deviated from the study protocol, because treatment was terminated within 6 sessions; 12 UC patients received total hip replacement (THR) or total knee replacement ()TKR) during the study period. For 23 BGA patients treatment deviated from the protocol: for 6 BGA patients treatment was terminated within 6 sessions and for 17 BGA patients less than two boostersessions were performed. Seven BGA patients received THR or TKR. Patients who completed treatment according to protocol were similar to patients who did not complete treatment, with the exception of pain at baseline. Patients who did not complete treatment reported significantly more pain at baseline (VAS: 4.8 versus 3.7). One patient of the BGA group reported adverse effects (increase of pain) and withdrew at the end of the therapy (after 3 boostersessions).

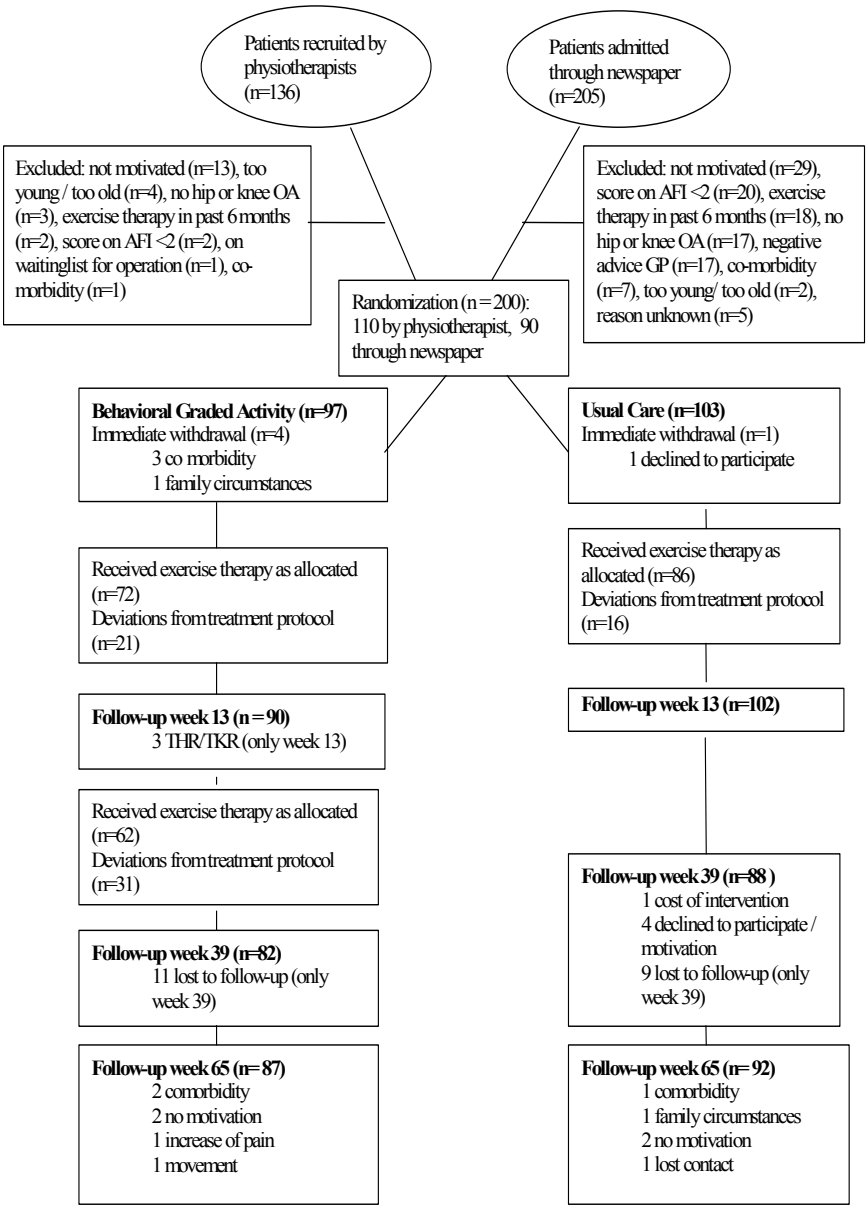

Figure 1 Patient flow diagram through study

Table 1 Baseline characteristics of both intervention groups.

Characteristics	**BGA n=97**		**UC n=103**	
Female, n (%)	73	(75)	81	(79)
Age, mean (sd)	65.1	(7.4)	64.5	(8.3)
Location of OA, n (%)				
knee	67	(69)	63	(61)
hip	22	(23)	28	(27)
both	8	(8)	12	(12)
Duration of complaints, n (%) (n=198)				
< 1 year	23	(24)	24	(23)
1 – 5 years	39	(41)	33	(32)
> 5 years	33	(35)	46	(45)
Radiological evidence OA (K&L ≥ 2) (n=146)				
knee (n=101), n (%)	26	(52)	31	(61)
hip (n=51), n (%)	18	(86)	29	(97)
Comorbidity, n (%)	63	(68)	65	(64)
Body Mass Index, mean (sd)	28.2	(4.2)	28.8	(4.6)

K&L: Kellgren & Lawrence score

Treatment

Fifty-five physiotherapists (26 BGA and 29 UC) treated patients included in the study. The mean number of treated patients per physiotherapist was 3.6 and varied between 1 and 11. The patients who were allocated to the BGA group received on average 14,1 (SD 5,5) treatment sessions versus 11,7 (SD 4,3) in the UC group ($p<0.01$).

BGA treatment was mainly directed towards improvement of activities (e.g., walking, climbing stairs, gardening) / decrease of limitations in activities (in 92% of sessions), reduction of impairment in bodily functions (e.g., muscle strength, range of motion) (53%), and increase of participation (51%). The BGA physiotherapists followed the interventions as described in the BGA protocol for most patients; activities were chosen by 86% of the patients, baseline was set for 70%, and a gradually increasing exercise program was made (84%) and executed (84%).

The main treatment goals of the UC intervention were improvement of activities / decrease of limitations in activities (in 84% of sessions), reduction of impairments in bodily functions (76%), and reduction of pain (64%). Most common UC interventions were exercise therapy of bodily functions (81% of sessions), exercise therapy of activities (74%), and providing information and advice (56%). The UC group applied more

passive interventions compared to the BGA physiotherapists: manipulating joints (41% of sessions compared to 3%), and massage (21% of sessions compared to 2%).

BGA patients reported significantly more frequently to adhere to home activities compared to the UC patients, both at 13 and 65 weeks: at 13 weeks 75% of the BGA patients reported to exercise 'frequently' or 'very frequently' compared to 44% of the UC patients (Chi-square, $p<0.01$); at 65 weeks 56% and 33% respectively reported to adhere to the exercises (Chi-square, $p<0.01$).

Blinding

The research assistants were asked to guess the assigned treatment immediately after week 13 and week 65 assessments. The research assistants guessed the assigned treatment in 57% of the cases after 13 weeks (Cohen's kappa= 0.14) and 46% of the cases after 65 weeks (Cohen's kappa=-0.09).

Outcome

Table 2 presents the results of the intention-to-treat analyses on the effectiveness of treatment of the primary outcome measures. All primary outcome measures showed significant improvements within both groups and improvements increased with time.

For the primary outcome measures pain and physical function, an overview of the changes over the course of the trial is given in Figure 2a and 2b. After 13 weeks, BGA patients improved respectively 25.8% and 20.8% compared to baseline on WOMAC pain and physical function, after 65 weeks the improvement of BGA patients increased to 42.8% and 25.6% compared to baseline. This pattern was similar for the scores of UC patients: 25.3% and 17.9% improvement compared to baseline on WOMAC pain and physical function after 13 weeks and more improvement, 36.8% and 25.1%, respectively compared to baseline, after 65 weeks. The differences between the groups in improvement for pain, physical function, and PGA at all assessments were in favor of the BGA group. However, the differences were small and not statistically significant.

Table 2 Primary outcome measures: improvements and differences between intervention groups.

Outcome measures	**BGA**		**UC**		**Difference BGA-UC [95% CI]**	**Adjusted difference [1] [95% CI]**
	n	**Mean [95% CI]**	**n**	**Mean [95% CI]**		
Pain at assessment, VAS:0-10						
Baseline, mean (sd)	97	4.3 (2.8)	103	3.7 (2.5)		
Δ Week 13 - Baseline	90	-0.61 [-1.2;-.005]	102	-0.47 [-1.0; 0.1]	-0.14 [-0.9; 0.6]	0.26 [-0.4; 0.9]
Δ Week 39 - Baseline	82	-0.15 [-0.8; 0.5]	88	0.62 [0.0; 1.2]	-0.77 [-1.7; 0.1]	-0.26 [-1.0; 0.5]
Δ Week 65 - Baseline	87	-1.01 [-1.7;-0.3]	92	-0.58 [-1.1;-0.03]	-0.44 [-1.3; 0.5]	0.14 [-0.6; 0.8]
Pain, subscale WOMAC:0-20						
Baseline, mean (sd)	97	9.1 (3.3)	103	8.7 (3.1)		
Δ Week 13 - Baseline	90	-2.35 [-3.0;-1.7]	102	-2.20 [-2.9;-1.5]	-0.15 [-1.1; 0.8]	0.17 [-0.6; 1.0]
Δ Week 39 - Baseline	82	-2.30 [-3.3;-1.3]	88	-1.00 [-1.8;-0.2]	-1.30 [-2.5;-0.1]	-0.97 [-2.1; 0.2]
Δ Week 65 - Baseline	87	-3.90 [-4.7;-3.1]	92	-3.20 [-3.9;-2.5]	-0.70 [-1.8; 0.4]	-0.32 [-1.3; 0.6]
Physical function, subscale WOMAC: 0-68						
Baseline, mean (sd)	97	28.7 (12.5)	103	29.1 (9.9)		
Δ Week 13 - Baseline	90	-5.98 [-8.0;-4.0]	102	-5.21 [-6.9;-3.5]	-0.76 [-3.4; 1.8]	-0.56 [-3.0; 1.9]
Δ Week 39 - Baseline	82	-6.94 [-9.6;-4.3]	88	-5.22 [-7.4;-3.0]	-1.72 [-5.1; 1.6]	-1.30 [-4.4; 1.8]
Δ Week 65 - Baseline	87	-7.35 [-10.4;-4.3]	92	-7.29 [-9.3;-5.2]	-0.07 [-3.6; 3.5]	0.27 [-2.9; 3.4]
Patient global assessment					Odds ratio:	Odds ratio:
Week 13, n (%) improved	90	37 (41)	102	37 (36)	0.81 [0.5; 1.5]	0.88 [0.5; 1.6]
Week 39, n (%) improved	79	33 (42)	86	24 (28)	0.54 [0.3; 1.0]	0.55 [0.3; 1.1]
Week 65, n (%) improved	86	48 (56)	88	43 (49)	0.76 [0.4; 1.4]	0.87 [0.5; 1.6]

Negative signs indicate improvement within groups or improvement in favor of BGA (in case of differences in change between groups). Concerning Patient global assessment OR < 1 indicates improvement in favor of BGA.

[1] Analyses are adjusted for the baseline score of each outcome measure, duration of complaints, location of OA (hip, knee, or both), age, sex, and recruitment method (physiotherapist or newspaper).

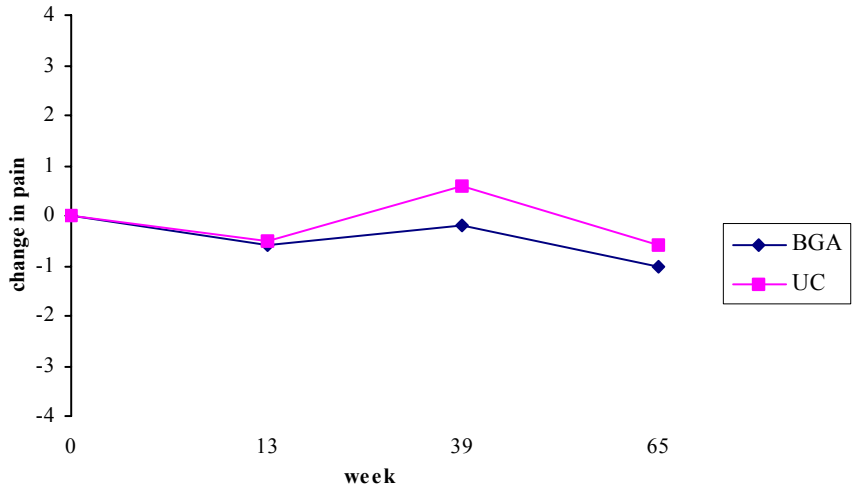

Figure 2a Change in pain (VAS) over time[1]
[1] Negative signs indicate improvement

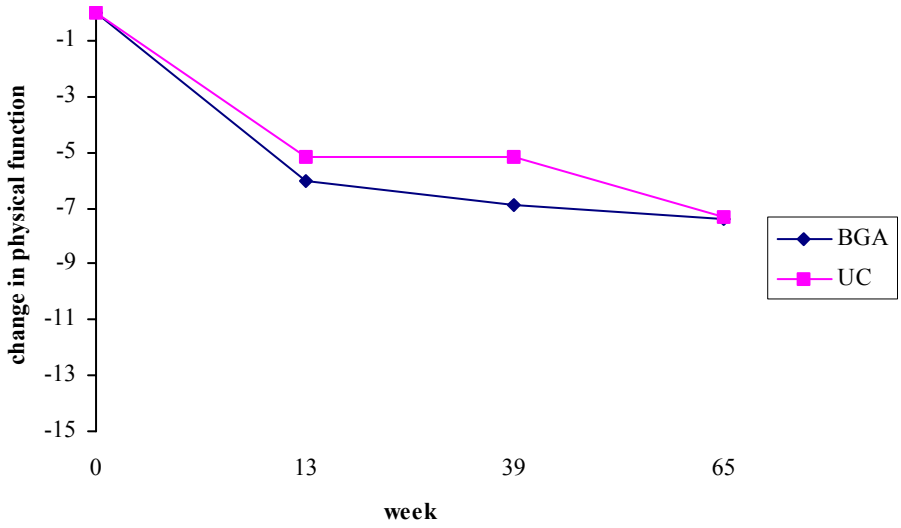

Figure 2b Change in physical function (WOMAC) over time[1]
[1] Negative signs indicate improvement

Table 3 presents the results of the secondary outcome measures. In general, the pattern was the same as for the primary outcome measures: statistically significant improvement within groups; differences between groups were in favor of BGA but mostly not significant. Exceptions are the 5 meter walking time test, and the physical function as assessed on the MACTAR, showing significant beneficial results in favor of BGA, especially in the long term. No differences, either, were found on the 8 subscales of the SF-36, except for a beneficial effect of UC on the 'role physical functioning' subscale.

Alternative analyses (multilevel analysis and per-protocol analysis) yielded similar results (data not shown). Intraclass correlation coefficients among and within physiotherapists were estimated as <0.01; therefore, all multilevel analyses were performed on the patients' level. The per-protocol analysis was restricted to 80, 68, and 72 UC patients, respectively, and 64, 58, and 63 BGA patients in week 13, 39, and 65, respectively.

Table 3 Secondary outcome measures: improvements and differences between intervention groups.

Outcome measures	BGA			UC			Difference BGA-UC [95% CI]		Adjusted difference[1] [95% CI]	
	n	Mean	[95% CI]	n	Mean	[95% CI]				
Tiredness past week, VAS:0-10										
Baseline, mean (sd)	97	5.3	(2.4)	103	4.9	(2.3)				
Δ Week 13 - Baseline	90	-1.28	[-1.8;-0.8]	102	-0.84	[-1.5;-0.6]	-0.44	[-1.2; 0.3]	-0.10	[-0.7; 0.5]
Δ Week 39 - Baseline	82	-0.59	[-1.9;-0.7]	88	0.31	[-1.3; 0.1]	-0.89	[-1.7;-0.1]	-0.47	[-1.2; 0.3]
Δ Week 65 – Baseline	87	-2.13	[-3.1;-2.0]	92	-1.12	[-2.1;-1.0]	-1.01	[-1.9;-0.1]	-0.49	[-1.1; 0.1]
Muscle strength hip, z-score										
Baseline, mean (sd)	97	0.23	(3.8)	103	-0.17	(3.1)				
Δ Week 13 - Baseline	87	1.07	[0.5; 1.6]	98	0.13	[-0.4; 0.6]	0.94	[0.2; 1.7]	0.75	[0.04; 1.5]
Δ Week 65 – Baseline	77	0.29	[-0.3; 0.9]	79	-0.22	[-0.9; 0.5]	0.51	[-0.4; 1.4]	0.12	[-0.6; 0.9]
Muscle strength knee, z-score										
Baseline, mean (sd)	97	0.05	(1.6)	103	-0.02	(1.9)				
Δ Week 13 - Baseline	89	0.40	[0.1; 0.7]	100	0.25	[-0.01; 0.5]	0.15	[-0.3; 0.5]	0.15	[-0.2; 0.5]
Δ Week 65 – Baseline	75	-0.32	[-0.7; 0.03]	86	-0.38	[-0.8; 0.1]	0.06	[-0.5; 0.6]	-0.03	[-0.4; 0.4]
ROM: hip flexion to extension in degrees										
Baseline, mean (sd)	97	129.6	(14.6)	103	126.2	(15.8)				
Δ Week 13 - Baseline	89	1.77	[-0.3; 3.8]	100	2.78	[0.7; 4.8]	-1.01	[-3.9; 1.9]	-0.42	[-3.1; 2.3]
Δ Week 65 – Baseline	82	-0.80	[-3.4; 1.8]	89	-0.29	[-2.9; 2.3]	-0.51	[-4.1; 3.1]	0.10	[-3.2; 3.4]
ROM: knee flexion to extension in degrees										
Baseline, mean (sd)	97	122.2	(15.9)	103	124.3	(12.1)				
Δ Week 13 - Baseline	90	3.17	[1.4; 4.5]	101	2.52	[0.8; 4.3]	0.65	[-1.8; 3.1]	0.47	[-1.9; 2.8]
Δ Week 65 - Baseline	84	4.79	[0.9; 8.6]	91	3.55	[1.8; 5.3]	1.24	[-2.8; 5.3]	1.02	[-2.9; 5.0]

Table 3 continued

Outcome measures	BGA			UC			Difference BGA-UC [95% CI]		Adjusted difference[1] [95% CI]	
Physical function, MACTAR:-15 to 15										
Δ Week 13 - Baseline	90	6.41	[5.2; 7.6]	102	5.30	[4.0; 6.6]	1.11	[-0.6; 2.9]	1.00	[-0.8; 2.8]
Δ Week 39 - Baseline	93	4.28	[2.5; 6.0]	97	2.16	[0.5; 3.8]	2.12	[-0.3; 4.5]	1.96	[-0.4; 4.3]
Δ Week 65 - Baseline	87	6.02	[4.4; 7.6]	92	3.27	[1.6; 4.9]	2.75	[0.5; 5.0]	2.56	[0.3; 4.9]
Physical function, 5 meter walking in sec.										
Baseline, mean (sd)	97	4.8	(1.2)	103	4.8	(1.5)				
Δ Week 13 - Baseline	90	-0.41	[-0.6;-0.2]	102	-0.19	[-0.4; 0.0]	-0.22	[-0.5; 0.04]	-0.20	[-0.4;-0.01]
Δ Week 65 – Baseline	87	-0.44	[-0.7;-0.2]	92	-0.13	[-0.3;-0.04]	-0.31	[-0.6;-0.03]	-0.33	[-0.6; -0.1]
Physical activity (SQUASH, in minutes)										
Baseline, mean (sd)	97	1761	(1221)	103	1664	(984)				
Δ Week 13 - Baseline	80	170	[-44; 383]	89	266	[19; 513]	-96	[-427; 236]	-67.7	[-390; 254]
Δ Week 39 - Baseline	72	-101	[-356; 154]	66	-248	[-615; 9]	148	[-225; 520]	148.1	[-183; 479]
Δ Week 65 - Baseline	77	12	[-247; 271]	78	87	[-150; 324]	-74	[-428; 280]	-47.6	[-368; 273]
SF-36, subscale role physical function:0-100[2]										
Baseline, mean (sd)	97	40.0	(40.7)	103	45.2	(41.7)				
Δ Week 13 - Baseline	90	14.1	[4.9; 23.3]	102	15.2	[5.1; 25.2]	-1.1	[-14.8; 12.5]	-4.88	[-16.5; 6.7]
Δ Week 39 - Baseline	82	12.1	[1.6; 22.5]	88	9.2	[-1.4; 19.9]	2.8	[-2.0; 17.7]	0.88	[-11.6; 13.4]
Δ Week 65 - Baseline	87	8.00	[-3.7; 19.7]	92	17.8	[6.0; 29.5]	-9.8	[-26.3; 6.8]	-14.7	[-27.5;- 1.9]

Positive signs indicate improvement (within groups) or improvement in favor of BGA-patients (in case of differences in change between groups), with the exception of tiredness and 5 meter walking.

[1] Analyses are adjusted for the baseline score of each outcome measure, duration of complaints, location of OA (hip, knee, or both), age, sex, and recruitment method (physiotherapist or newspaper).

[2] No significant differences were found on the other 7 subscales of the SF-36.

Discussion

In a single blind cluster randomized controlled trial, the effectiveness of an operant behavioral graded activity program comprising boostersessions (BGA), compared to usual care according to the Dutch physiotherapy guideline hip and/or knee OA (usual care, UC), was assessed. The interventions were given by physiotherapists in primary care. Both groups reported beneficial effects in the long term. The differences between the groups in improvement for pain, functional status and patient global assessment (PGA) were small and not significant. The differences were, however, in general in favor of the BGA group. This pattern was similar for the scores on the secondary outcome measures, with the exception of the scores on the MACTAR and 5 meter walking test. For these outcome measures significant differences were found in favor of BGA at week 65. Furthermore, BGA patients reported to adhere significantly more to the home exercises and activities.

On basis of literature (5;6), we expected only short term (post-treatment) effects and no long-term effects of UC. Surprisingly, the beneficial effects of UC remained stable and even slightly increased in the long term; as a result, the BGA and UC interventions were about equally effective. An explanation for the unexpected beneficial effects in the long term of the UC-group, is a general change of approach in physiotherapy. First of all, the shift from a more passive approach (e.g., massage therapy and physiotherapy modalities) to an active approach (exercise therapy, education), as advocated in the guidelines, is gaining more and more support in the physiotherapy profession (9;11;40;41). Within this active approach, the exercises become more functional and task oriented recently (42-44). This shift from the level 'impairment of body functions (e.g., muscle strength, range of motion)' to the functional 'activities' level (e.g., walking, climbing stairs) could be confirmed in the registration forms of the physiotherapists. The UC applied in our trial was an adaptation of the exercise therapy protocol of Van Baar et al (12). In Van Baar's study, exercise therapy was mainly directed towards improvement of muscle strength (93% of treatments), improvement of range of motion (85%), and reduction of pain (80%). We expected that the goals of the UC physiotherapists would be comparable to van Baar's study, namely at the

level of ‘impairments of body functions’. However, in the present study improvement of activities / decrease of limitations in activities (84% of treatments) was the most frequently mentioned intervention in the UC group.

Apart from the enhanced functional orientation in the UC group, the contrast between the intervention groups was lower than expected in other respects also. First, we expected that BGA patients would be treated at least 5 sessions more compared to the UC group, since BGA patients received additional boostersessions after the first 12 weeks of treatment. However, the average number of sessions was 14.1 of the BGA group versus 11.7 of the UC group (which was still a significant difference, $p<0.01$). Second, the contrast between intervention was smaller since not all BGA patients were treated according to the protocol. Specific elements of the BGA were described in the registration forms, and baseline was set for only 70% of the BGA patients, and an exercise protocol was only made for 84% of the BGA patients. This might indicate that also other specific elements of BGA, such as reinforcing feedback and extinction of pain behavior (not registered) may not be adequately executed. Possibly, a 2 day-course is too short to completely master the skills which are necessary to treat patients according to a behavioral graded activity protocol, as was suggested by King et al. (45). The results of other studies on the efficacy of behavioral graded activity are comparable: no differences were found in pain and physical functioning reported on self-administered questionnaires. Positive effects were found in patient-specific measures (17) and return to work or other behavior (less sick leave) (15;20). In our study, beneficial effects were also found for the patient specific measure MACTAR and the 5 meter walking time test. Finally, considering the lower contrast, the present study might have been underpowered.

Especially for the outcome measure MACTAR, the significant difference between both groups after 65 weeks, seems to be clinically relevant. Although no research has been done on the minimal clinical important difference / improvement of the MACTAR, a difference of 46% (6.02 for BGA compared to 3.27 for UC) can be interpreted as a clinically relevant difference. The clinical important difference of the walking time test is less convincing (an improvement of 9% compared to baseline in BGA patients, compared to 3% in UC patients). Although both interventions were about equally effective, we have the opinion that the overall improvement

within groups also was clinically relevant. Angst et al. (46) concluded that changes over 12% compared to baseline, in pain and physical function, can be detected as minimal clinically important differences. In our study we found changes in pain after 65 weeks of 43% (BGA) and 37% (UC), and changes in physical function of 25% (both groups) . Tubach et al. (47) stated that the minimal clinically important improvement for relative changes (changes compared to baseline) were between 32% and 40% for pain (VAS) and between 21% and 26% for physical function (WOMAC). Besides this, 56% of the BGA-patients and 49% of the UC-patients reported, after 65 weeks, to have improved on the patient global assessment, a measure which is often used as indicator for clinically relevant differences.

As for the design of the study, some comments can be made. First, patients were recruited in 2 ways, referrals to physiotherapists and respondents to articles in local newspapers. However, the influence of these 2 recruitment strategies on the study population and results of the study were investigated. It appeared that the recruitment method affected clinical characteristics and physical functioning of the patients, but the recruitment method was not an effect modifier since it did not affect treatment outcome (24). Second, a longer follow up might give a more definite answer on the effectiveness of BGA, since the time between the last session (week 55) and last assessment (week 65) was only 10 weeks. Third, since BGA consists of more sessions and requires a training of physiotherapists, it is important that the clinical evaluation of such new treatment program is accompanied by an economic evaluation. This economic evaluation on the cost-effectiveness of BGA will be performed in the near future. Finally, since the aim of the present study was to investigate whether BGA was more effective than UC, there was no reason for including a no-treatment control group. Because we lacked such no-treatment group we are not able to attribute the within-group improvement to either treatment (BGA or UC) or other factors. Considering the slowly deteriorating condition of OA and the limited evidence in literature on long term effects of exercise therapy, improvement in time is very improbable. Therefore, we believe both interventions had beneficial effects over time.

In conclusion, both treatment groups showed beneficial long term effects. The differences between BGA and UC on most outcome measures were in favor of BGA. However, these differences were small and not

significant, with the exception of the patient-oriented physical function and 5 meter walking time test. Since, unexpectedly, UC also showed beneficial effects in the long term, the BGA and UC treatment were about equally effective. It needs to be investigated whether both treatments differ in costs and whether specific groups of OA patients particularly benefit from BGA.

References

1. Dekker J, Boot B, van der Woude LH, Bijlsma JW. Pain and disability in osteoarthritis: a review of biobehavioral mechanisms. J Behav Med 1992; 15(2):189-214.
2. Van Baar ME, Dekker J, Lemmens JA, Oostendorp RA, Bijlsma JW. Pain and disability in patients with osteoarthritis of hip or knee: the relationship with articular, kinesiological, and psychological characteristics. J Rheumatol 1998; 25(1):125-133.
3. Smidt N, de Vet HC, Bouter LM, Dekker J, Arendzen JH, de Bie RA et al. Effectiveness of exercise therapy: a best-evidence summary of systematic reviews. Aust J Physiother 2005; 51(2):71-85.
4. Tan JC, Horn SE. Practical Manual of Physical Medicine and Rehabilitation: Diagnostics, Therapeutics, and Basic Problems. St Louis: Mosby, 1998.
5. Fransen M, McConnel S, Bell M. Therapeutic exercise for people with osteoarthritis of the hip or knee. A systematic review. J Rheumatol 2002; 29:1737-1745.
6. Van Baar ME, Assendelft WJ, Dekker J, Oostendorp RA, Bijlsma JW. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review of randomized clinical trials. Arthritis Rheum 1999; 42(7):1361-1369.
7. Philadelphia Panel evidence-based clinical practice guidelines on selected rehabilitation interventions for knee pain. Phys Ther 2001; 81(10):1675-1700.
8. Hochberg MC, Altman RD, Brandt KD, Clark BM, Dieppe PA, Griffin MR et al. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. Arthritis Rheum 1995; 38(11):1541-1546.
9. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P et al. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis. Ann Rheum Dis 2003; 62(12):1145-1155.
10. Bijlsma JW. Behandeling bij osteoartrose. Ned T Reum 2005; 1:27-29.
11. Vogels EMHM, Hendriks HJM, van Baar ME, Dekker J, Hopman-Rock M, Oostendorp RA et al. Clinical practice guidelines for physical therapy in patients with osteoarthritis of the hip or knee. Amersfoort: KNGF, 2001.
12. Van Baar ME, Dekker J, Oostendorp RA, Bijl D, Voorn TB, Lemmens JA et al. The effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a randomized clinical trial. J Rheumatol 1998; 25(12):2432-2439.
13. Van Baar ME, Dekker J, Oostendorp RA, Bijl D, Voorn TB, Bijlsma JW. Effectiveness of exercise in patients with osteoarthritis of hip or knee: nine months' follow up. Ann Rheum Dis 2001; 60(12):1123-1130.
14. Linton SJ, Hellsing AL, Andersson D. A controlled study of the effects of an early intervention on acute musculoskeletal pain problems. Pain 1993; 54:353-359.
15. Ostelo RW, de Vet HC, Berfelo MW, Kerckhoffs MR, Vlaeyen JW, Wolters PM et al. Effectiveness of behavioral graded activity after first-time lumbar disc surgery: short term results of a randomized controlled trial. Eur Spine J 2003; 12(6):637-644.

16. Fordyce WE, Fowler RS, Lehmann JF, Delateur BJ, Sand PL, Trieschmann RB. Operant conditioning in the treatment of chronic pain. Arch Phys Med Rehabil 1973; 54:399-408.
17. Lindstrom I, Ohlund C, Eek C, Wallin L, Peterson LE, Fordyce WE et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. Phys Ther 1992; 72(4):279-290.
18. Vlaeyen JW, Haazen IW, Schuerman JA, Kole Snijders AM, Eek van H. Behavioural rehabilitation of chronic low back pain: comparison of an operant treatment, and operant-cognitive treatment and an operant-respondent treatment. Br J Clin Psychol 1995; 34:95-118.
19. Lindstrom I, Ohlund C, Eek C, Wallin L, Peterson LE, Nachemson A. Mobility, strength, and fitness after a graded activity program for patients with subacute low back pain. A randomized prospective clinical study with a behavioral therapy approach. Spine 1992; 17(6):641-652.
20. Staal JB, Hlobil H, Köke AJA, Twisk JWR, Smid T, van Mechelen W. The effects of graded activity for low back pain in occupational health on sick leave, functional status and pain: a randomised controlled trial. Ann Intern Med 2004; 140:77-84.
21. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Arthritis Rheum 1986; 29(8):1039-1049.
22. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991; 34(5):505-514.
23. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee. Validation--value in comparison with other assessment tests. Scand J Rheumatol Suppl 1987; 65:85-89.
24. Veenhof C, Dekker J, Bijlsma JW, van den Ende CHM. The influence of various recruitment strategies on the study population and outcome of a randomized controlled trial involving patients with osteoarthritis of hip or knee. Arthritis Rheum 2005; 53(3):375-382.
25. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957; 16(4):494-502.
26. Ravaud P, Dougados M. Radiographic assessment in osteoarthritis. J Rheumatol 1997; 24(4):786-791.
27. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988; 15(12):1833-1840.
28. Norkin CC, White DJ. Measurement of joint motion: a guide to goniometry. Philadelphia: FA Davis Company, 1986.

29. Bohannon RW. Muscle strength testing with hand-held dynamometers. In: Admundsen LR, editor. Muscle strength testing: Instrumented and non-instrumented systems. New York: Churchill Livingstone, 1990: 69-88.
30. 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(3):446-451.
31. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. J Clin Epidemiol 2003; 56(12):1163-1169.
32. Heijden GJMG van der. Shoulder disorder treatment: Efficacy of ultrasoundtherapy and electrotherapy. Maastricht The Netherlands: University Press Maastricht, 1996.
33. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30(6):473-483.
34. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997; 24(4):799-802.
35. Cohen J. Statistical power analysis for the behavioral sciences: revised edition. London: Academic Press Inc, 1976.
36. Steultjens MP, Dekker J, van Baar ME, Oostendorp RA, Bijlsma JW. Muscle strength, pain and disability in patients with osteoarthritis. Clin Rehabil 2001; 15(3):331-341.
37. Steultjens MP, Dekker J, van Baar ME, Oostendorp RA, Bijlsma JW. Range of joint motion and disability in patients with osteoarthritis of the knee or hip. Rheumatology (Oxford) 2000; 39(9):955-961.
38. Bryk AS, Raudenbusch SW. Hierarchical lineair models: Applications and data management methods. Newbury Park: Sage Publications, 1992.
39. Goldstein H. Multilevel statistical models. New York: Halsted Press, 1995.
40. Bekkering GE, Hendriks HJ, van Tulder MW, Knol DL, Hoeijenbos M, Oostendorp RA et al. Effect on the process of care of an active strategy to implement clinical guidelines on physiotherapy for low back pain: a cluster randomised controlled trial. Qual Saf Health Care 2005; 14(2):107-112.
41. Jette AM, Delitto A. Physical therapy treatment choices for musculoskeletal impairments. Phys Ther 1997; 77(2):145-154.
42. Alexander NB, Galecki AT, Grenier ML, Nyquist LV, Hofmeyer MR, Grunawalt JC et al. Task-specific resistance training to improve the ability of activities of daily living-impaired older adults to rise from a bed and from a chair. J Am Geriatr Soc 2001; 49(11):1418-1427.
43. De Vreede PL, Samson MM, van Meeteren NL, Duursma SA, Verhaar HJ. Functional-task exercise versus resistance strength exercise to improve daily function in older women: a randomized, controlled trial. J Am Geriatr Soc 2005; 53(1):2-10.

44. Ketelaar M, Vermeer A, Hart H, Petegem-van Beek E, Helders PJ. Effects of a functional therapy program on motor abilities of children with cerebral palsy. Phys Ther 2001; 81(9):1534-1545.
45. King M, Davidson O, Taylor F, Haines A, Sharp D, Turner R. Effectiveness of teaching general practitioners skills in brief cognitive behaviour therapy to treat patients with depression: randomised controlled trial. BMJ 2002; 324(7343):947-950.
46. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum 2001; 45(4):384-391.
47. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis 2005; 64(1):29-33.

Appendix 1 Description of the Behavioral Graded Activity (BGA) intervention.

Content	BGA consists of three phases: 1. Starting phase: Provision of educational messages, selection of problematic activities and treatment goals, and determination of baseline value. 2. Treatment phase: Increase of the selected activities, gradually and in a time-contingent way, by means of an exercise program, which is reproduced in performance charts. 3. Integration phase: Support and reinforcement of the behavioral change and integration of the increased level of activities in the daily living of the patient (maximum of 7 sessions in five determined boostersessions in week 18, 25, 34, 42, and 55).
Educational messages	- Not pain relief, but improvement of functioning is the primary goal of the treatment. - Exercise and physical activity are recommended. The performance of physical activity should not depend on the amount of pain.
Activities	Problematic activities (maximum of 3) are selected by patients on activity list. Individually tailored exercises, to improve impairments limiting the performance of these activities, are selected.
Goals	For each activity and each exercise, short term and long term goals are set and recorded in a treatment agreement form.
Baseline values	To determine baseline values, patients perform the selected activities until (pain) tolerance during 1 week and record these activities in a diary.
Gradually increasing exercise program	An individually-based scheme is made on a time-contingent basis for each activity and exercise, starting slightly under baseline values and gradually increasing towards the pre-set short term goal. Patients should not under-perform nor over-perform this gradually increasing scheme.
Visual reproduction	Performance charts are used to record and visualise the performance of activities and exercises.
Reinforcement	Positive reinforcement is given towards healthy and active behavior; pain behavior is extinguished.
Stopping rule	The gradual increase of activities has to be interrupted when an active inflammatory process is suspected or diagnosed (e.g. redness of the knee, increase in knee effusion, or comparable complaints). Hereafter, the increase of activities starts at a lower level. In case of recurrent inflammatory processes, the treatment goal needs to be changed and the rate of increasing activities needs to be decelerated.
Duration	Maximum of 18 sessions within first 12 weeks. Additional boostersessions in week 18, 25, 34, 42, and 55

Appendix 2 Description of the Usual Care (UC) intervention (11).

Content	The guideline describes 3 distinct patient profiles, which are based on 6 main problem areas.	
Patient profiles	Patient profile A:	Active inflammatory process in the joint is predominant; most important complaints are pain and impairments related to movement of hip/knee.
	Patient profile B:	Patient has episodes with pain complaints, impairments related to movement, which gradually results in limitations of activities as well as episodes of pain. Patient looks for solutions himself and needs extra guidance during episodes of intense pain.
	Patient profile C:	Patient has long-lasting or chronically recurring complaints ; limitations of activities and possible participation problems are of central concern. Patient experiences little or no control over the situation.
Main problem areas	The guideline describes 6 problem areas for patients with OA: impairments related to active inflammatory processes, pain, impairments related to movement, limitations of activities, participation problems and inadequate coping strategies. On basis of these problem areas physical therapists can classify patients as having one of the three defined patient profiles.	
Central goal	To counter the effect of osteoarthritis by decreasing the patient's pain, limitations of activities and participation problems. In other words, to optimize the patient's level of activity and participation in life.	
Therapeutic approach	For each identified problem area, treatment goals and interventions are advised. Possible interventions are: - providing information and advice - exercise therapy - passive interventions, like traction to the joint and TENS (only in active inflammatory processes) - stimulating and increasing the level of activities - stimulating compliance with therapy	
Duration	Maximum of 18 sessions within 12 weeks.	

Chapter

3

The cost-effectiveness of behavioral graded activity in patients with osteoarthritis of hip and/or knee.

Submitted for publication as:
Coupé VMH, Veenhof C, Van Tulder MW, Dekker J, Bijlsma JWJ, Van den Ende CHM. The cost-effectiveness of behavioral graded activity in patients with osteoarthritis of hip and/or knee. Submitted for publication in Annals of Rheumatic Diseases.

Abstract

*Objective***:** To evaluate whether exercise therapy based on behavioral graded activity comprising boostersessions, is cost-effective compared to usual care, that is, exercise therapy according to the Dutch physiotherapy guideline.
Methods: An economic evaluation from a societal perspective was carried out alongside a randomized trial involving 200 patients with osteoarthritis of the hip and/or knee. Outcome measures were pain, physical functioning, self perceived change, and quality of life, assessed at baseline, 13, 39 and 65 weeks. Costs were measured using cost diaries for the entire follow-up period of 65 weeks. Cost and effect differences were estimated using multilevel analysis. Uncertainty around the cost-effectiveness ratios was estimated by bootstrapping and graphically represented on cost-effectiveness planes.
Results: Ninety-seven patients received behavioral graded activity and 103 usual care. At 65 weeks, no differences between graded activity and usual care in improvement with respect to baseline were found on any of the outcome measures. The mean (95% CI) difference in total costs between the two treatment groups was -€773 (-€2360 ; €772), that is, behavioral graded activity resulted in less costs but this difference was non-significant. Since effect differences were small, a large incremental cost-effectiveness ratio of € 51385 per QALY was found for graded activity versus usual care.
Conclusions: This study provides no evidence that behavioral graded activity is either more effective or less costly than usual care. Yielding similar results to usual care, behavioral graded activity seems an acceptable method to treat patients with osteoarthritis of the hip and/or knee.

Introduction

Osteoarthritis (OA) of the hip and/or knee is a common joint disorder. The incidence in general practices in the Netherlands is 2.1/1000 per year for OA of the hip and 3.6/1000 for the knee (1). Treatment of OA is directed at pain relief and prevention of disability. A systematic review of several randomized trials showed that exercise therapy for irrecoverable chronic diseases such as OA has a short term positive effect on pain and daily functioning (2). However, this positive effect seems to decline over time and finally disappear, resulting in a recurring need for treatment, increased disability, work absenteeism and utilization of health care (2;3). Maintaining the short-term benefits of exercise therapy is therefore important.

The current study is a cluster randomized controlled trial that investigates whether behavioral graded activity (BGA), that is a behavioral treatment integrating the concepts of operant conditioning with exercise therapy comprising booster sessions, consolidates the positive short term effects of usual exercise therapy. The clinical paper shows a positive long term effect of BGA (4). Contrary to previous findings, however, this positive effect is also found for usual exercise therapy, leading to insignificant effect differences between the treatment groups.

As OA may lead to considerable costs, it is important that the clinical evaluation of a new treatment program is accompanied by an economic evaluation. However, hardly any economic evaluation on the cost-effectiveness of exercise therapy for OA of the hip and/or knee has been conducted. In this paper, we present a cost-effectiveness analysis of the BGA program in comparison to usual exercise therapy.

Methods

Study design

An economic evaluation was conducted alongside a cluster randomized controlled trial comparing BGA and usual care according to the Dutch Osteoarthritis guideline of the Royal Dutch College for Physiotherapy (KNGF). Effectiveness and cost-effectiveness over 65 weeks were investigated. The study was approved by the Medical Ethical Committee of the VU University Medical Center, Amsterdam, The Netherlands.

Study population

Eighty-seven physiotherapists, willing and able to participate in the study, were recruited. Participating physiotherapeutic practices were randomly assigned to 1 of the 2 treatment programs. Because recruitment of patients through participating physiotherapists was slow, a second recruitment strategy was used, i.e. patients responded to articles about the benefit of exercise therapy and the performed study, published in local newspapers. Thus, recruited patients were referred to a participating physiotherapist. Patients were included if they fulfilled the clinical criteria for OA of the hip or knee of the American College of Rheumatology (5;6). All patients willing and eligible to participate gave their informed consent. In total, 200 patients were included. For more details on the trial, we refer to the clinical paper (4).

Interventions

The behavioral graded activity group received a treatment integrating the concepts of operant conditioning with exercise therapy comprising boostersessions. BGA was directed at increasing the level of activities in a time-contingent way, with the goal to integrate these activities in the daily living of patients (4;7;8). Treatment consisted of a 12-week period with a maximum of 18 sessions, followed by five pre-set booster moments with a maximum of seven sessions (in week 18, 25, 34, 42, and 55, respectively).

The usual care group received treatment according to the Dutch physiotherapy guideline for patients with OA of hip and/or knee (9). This guideline consists of general recommendations, emphasizing provision of information and advice, exercise therapy, and encouragement when positively coping with the complaints. Treatment consisted of a 12-week period with a maximum of 18 sessions and could be discontinued within this 12-week period if, according to the physiotherapist, all treatment goals had been achieved.

Clinical outcome measures

Patients completed health questionnaires at baseline, 13, 39 and 65 weeks. Primary outcome measures were pain (VAS and WOMAC), physical function (WOMAC) and self perceived change (Patient Global Assessment) according to the core set of outcome measures of clinical trials with patients with OA defined by OMERACT III(10-12). For the cost-effectiveness

analysis also health related quality of life (EuroQol-5D) was measured (13).

Assessment of resource utilization

Over the same period of 65 weeks, patients provided data on the direct costs of OA within and outside the health care sector and on the indirect costs of productivity loss. To this end, patients recorded resource utilization in cost diaries, covering the periods 1-12, 13-24, 25-36, 37-38, 39-50, 51-62, 63-65 weeks. The most important resources used within the health care sector were physiotherapeutic treatment and OA related hospitalization. Other health care resources were: allied health care other than physiotherapy, medical-specialist care, medication and home care. Resource utilization outside the health care sector included alternative therapies and informal care by friends or family members. Indirect costs of productivity loss were estimated by measuring absenteeism from paid and unpaid work.

Valuation of health care consumption; unit costs

The economic evaluation was conducted from a societal perspective. As the study was carried out between 2002 and 2004, 2003 prices were used. Because the follow-up period was limited to 65 weeks, no discounting was applied. Standard prices were used to value resource utilization (14;15). Table 1 shows unit cost prices for most of the resources considered. Prices of medication were obtained from the Royal Dutch Society for Pharmacy (16). Absenteeism from paid work was valued with the friction cost method (17). This approach considers loss of production through work absenteeism only during a friction period that is needed to replace the person. Production loss is valued using mean age- and sex- specific incomes of the Dutch population (14). Absenteeism from unpaid work, such as voluntary work and informal care, was valued with the shadow price method (14;18). In this method, the value of unpaid work is the cost of the required professional if the unpaid workers were unavailable. The shadow price of unpaid work is assumed to be equal to the tariff for cleaning work.

Table 1 Costs per unit health care resource used in the economic evaluation of behavioral graded activity (year 2003).

Healthcare resource (Unit)	Cost per unit (€)
General practice consultation	
General practitioner(visit)	20.20
Physiotherapist (Session)	22.75
Manual therapist (Session)	31.46
Outpatient attendance	
Policlinic care (visit)	56.00
Specialist care (visit)	98.00
Diagnostic procedures	
X-ray	39.00
MRI	255.29
CT-scan	98.00
Hospital inpatient stay	
Hospital admission (day)	337.00
Replacement knee	1962.42
Replacement hip	2061.68
Other	
Absenteeism paid labor (hour)	34.98*
Absenteeism unpaid labor (hour)	8.30†
Professional home care (hour)	21.70
Informal care (hour)	8.30†

* Average cost per hour according to the friction cost method. In the analysis gender and age dependent costs are used.

† Shadow price, being equal to the hour price for cleaning work.

Statistical analysis

Statistical analyses were carried out on an intention-to-treat principle. To avoid loss of information, we imputed missing data for patients with an incomplete set of cost diaries using the Expectation Maximization (EM) Algorithm in SPSS 12.0.1 (19). This algorithm is an iterative optimization method to estimate missing data given the available data.

Patients treated by the same physiotherapist formed clusters within the trial. The appropriate way to analyze such clustered data is by multilevel analysis (20;21). We performed multilevel analysis using MLwiN. The resulting cost and effect differences between treatment groups are corrected for dependence between patients treated by the same physiotherapist.

As cost data is typically skewed, confidence intervals for cost differences cannot be estimated with conventional methods that assume normality. We therefore applied the non-parametric bootstrap with a 1000 replications (22-

24). This basically amounts to drawing 1000 samples of the same size as the original dataset by sampling with replacement from the observed data. These resamples can be used to estimate standard errors and confidence intervals.

For the cost-effectiveness analysis, the difference in total cost between the two treatment groups was compared to the difference at 65 weeks in improvement in VAS and WOMAC scores and in percentage improved according to Patient Global Assessment. For the latter, the ratings of Patient Global Assessment were dichotomized as improved versus not improved. To compare the difference in total cost with the difference in quality of life gained, the scores on the EuroQol-5D were translated into a utility using preferences of the UK general population (25). The utilities of patients at baseline, 13, 39 and 65 weeks were then used to compute Quality Adjusted Life Years (QALYs). Basically, the time spent in each of the three periods in the trial (0-13, 13-39, 39-65 weeks) is weighed by the utility experienced in that period (26;27). Subsequently, the cost difference was compared with the difference in quality-adjusted life years gained over 65 weeks.

Uncertainty around the cost-effectiveness ratios was estimated using the bias corrected and accelerated bootstrapping method (5000 replications) and presented in a cost-effectiveness plane (28;29).

Sensitivity analyses were carried out to study the effect of different imputation strategies. First, we imputed zero costs if data were missing for the last half of the follow-up period. Second, we imputed mean costs for missing data. Also, a complete case analysis was carried out, considering the patients who completed all cost diaries. To investigate the effect of outliers, we performed an analysis in which 5% of patients with total costs more than €11000 were excluded. The threshold of €11000 was chosen after inspection of the data of patients with extremely high costs of absenteeism from either paid or unpaid work.

Both for the complete cases as for the dataset completed by EM imputation, per-protocol analysis were performed. In the per-protocol analysis, all patients with deviations from the treatment protocol were excluded. Deviations were defined as less than 6 sessions physiotherapy within the first 12 weeks (both groups), or less than 2 booster sessions (graded activity group) after the first 12 weeks, or a total hip / knee replacement during the whole study period (both groups).

Results

Clinical outcomes

At baseline, no differences were found between the BGA-group and the usual care group. Both treatment groups showed beneficial effects in the long term. However, no differences in improvement between the two treatment groups were found on any of the outcome measures. Full details on the clinical outcomes are presented in the clinical paper (4).

Resource use

Ten patients never returned any cost diary and were excluded from the evaluation. The remaining 190 patients returned 84% of the cost diaries. Only 64% of the patients completed all diaries. For these 64% of patients, Table 2 lists the utilization of health care resources and absenteeism from paid and unpaid work. Note that we present the results for the EM imputed data in the remainder of the paper.

Behavioral graded activity was associated with less medical-specialist care, hospitalization, hip replacements, and absenteeism from paid work compared to usual care, but with more informal care and help in housekeeping. However, these differences were small and not statistically significant. One interesting detail in Table 2 is the small number of patients with work absenteeism in the BGA group.

Table 2 Reported mean (SD) healthcare utilization for patients with complete cost data over 65 weeks by treatment group. The last column (N) indicates the number of patients with resource utilization.

	Behavioral graded activity(n=56)			**Usual care** (n=66)		
	mean	*(SD)*	*N*	*mean*	*(SD)*	*N*
HEALTH CARE SECTOR						
General practice consultation						
General practitioner (visits)	1.4	(2.2)	*27*	1.5	(2.7)	*35*
Physiotherapist (sessions)	18.0	(16.7)	*53*	20.2	(14.6)	*66*
Other allied health care (sessions)	0.5	(1.8)	*12*	0.2	(1.1)	*10*
Outpatient attendance						
Policlinic care (visits)	0.2	(1.6)	*2*	0.2	(0.9)	*4*
Specialist care (visits)	0.6	(1.6)	*12*	1.4	(2.6)	*28*
Diagnostic procedures						
X-ray (number)	0.6	(1.3)	*22*	0.9	(1.4)	*33*
MRI (number)	0.1	(0.3)	*4*	0.1	(0.4)	*6*
Other (number)	0.1	(0.4)	*5*	0.2	(0.6)	*11*
Hospital inpatient stay						
Hospital admissions (days)	0.1	(0.9)	*2*	1.1	(3.0)	*12*
Replacement knee (number)	0.02	(0.1)	*1*	----	----	*0*
Replacement hip (number)	----	----	*0*	0.2	(0.4)	*9*
OUTSIDE HEALTH CARE SECTOR						
Complementary or alternative therapist (sessions)	0.5	(3.0)	*4*	0.3	(1.6)	*4*
Absenteeism paid labor (hours)	2.6	(17.4)	*3*	5.4	(19.9)	*11*
Absenteeism unpaid labor (hours)	59.1	(156.1)	*15*	49.5	(112.0)	*32*
Informal care (hours)	32.7	(79.0)	*13*	18.9	(45.7)	*20*
Housekeeper (hours)	19.4	(57.3)	*10*	15.2	(60.3)	*10*

Costs

Table 3 shows the mean (standard deviation) costs for the two groups. Compared to the usual care group, we observed lower direct health care costs and higher costs outside the health care sector in the BGA group. Total direct costs were similar. From the direct health care costs, a substantial part was attributable to hospitalization. In the usual care group these costs doubled those in the graded activity group, but this difference was not significant.

Indirect costs in the graded activity group were approximately half those in the usual care group. This difference, caused by differences in work absenteeism, was not significant. The difference in total costs was -€773 (95% CI: -€2360 to €772), €2530 (SD €4888) for BGA and €3341 (SD €5055) for usual care. This difference was not significant.

Table 3 Mean (SD) costs (in euros) over 65 weeks by treatment group for all patients (missing data imputed).

	Behavioral graded activity * *(n=90)*		**Usual care*** *(n=100)*		**Difference** (95% CI) †	
	mean	*(SD)*	*mean*	*(SD)*		
HEALTH CARE SECTOR						
Primary care costs	462	(354)	519	(327)	-57	(-142; 35)
General practitioner	28	(41)	33	(50)	-5	(-18; 9)
Allied health care	433	(341)	486	(318)	-53	(-144; 36)
Secondary care costs	463	(1095)	813	(1761)	- 350	(-742; 83)
Policlinic care	13	(144)	8	(41)	5	(-13; 24)
Specialist care	71	(82)	127	(244)	-55	(-114; 3)
Diagnostic procedures	73	(138)	76	(141)	-3	(-36; 31)
Hospitalization	306	(999)	603	(1522)	-297	(-666; 74)
Medication costs	41	(92)	57	(128)	-16	(-50; 18)
Professional home care costs	67	(314)	178	(650)	-110	(-254; 34)
Direct health care costs	*1033*	*(1334)*	*1567*	*(2160)*	*-534*	*(-226 ; 672)*
OUTSIDE HEALTH CARE SECTOR						
Complementary or alternative therapy	48	(174)	30	(136)	18	(-27; 63)
Informal care	212	(542)	135	(318)	77	(-51; 200)
Housekeeper	524	(1416)	290	(1133)	230	(-130; 598)
Direct costs outside health care sector	783	(1641)	455	(1282)	330	(-80; 723)
Total direct costs	*1816*	*(2628)*	*2022*	*(2671)*	*-205*	*(-958; 516)*
Absenteeism paid labor	462	(2991)	1041	(3705)	-578	(-1596; 334)
Absenteeism unpaid labor	251	(873)	278	(734)	-15	(-442; 447)
Indirect costs	714	(3208)	1319	(4888)	-600	(-1763; 493)
Total costs	***2530***	***(4888)***	***3341***	***(5055)***	***-773***	***(-2360; 772)***

* Raw estimates

† Difference corrected for clustering within the factor physiotherapist with 95 % confidence intervals obtained from a non-parametric bootstrap with 1000 replications.

Cost-effectiveness

Table 4a shows the total costs and effects at 65 weeks for the different outcome measures in the BGA and the usual care group. Table 4b shows the differences in total costs and effects, and the incremental cost-effectiveness ratios.

Table 4a Mean costs (in euros) and effects by treatment group for all patients (missing data imputed).

	Behavioral graded activity		**Usual care**	
Effect measure	*Costs*	*Effects*	*Costs*	*Effects*
QALYs (EuroQol) (n=87/92)	2375	0.71	3440	0.73
WOMAC pain (scale 0-20)(n=83/91)	2418	3.67*	3400	3.14*
WOMAC physical (scale 0-68) (n=77/89)	2284	7.03*	3169	7.29*
VAS pain now (scale 0-10) (n=84/91)	2410	0.85*	3400	0.57*
VAS pain past week (scale 0-10) (n=84/89)	2410	1.92*	3458	1.79*
Patient Global Assessment (% improved) (n=83/87)	2435	54	3483	48

* a positive sign indicates improvement compared to baseline.

Table 4b Mean cost (in euros) and effect differences between treatment groups (BGA-UC) and cost- effectiveness ratios for all patients (missing data imputed).

Effect measure	**Cost difference**[†]	**Effect difference**[†]	**Incremental Cost-ffectiveness ratio**
QALYs (EuroQol) (n=87/92)	-1028	-0.02	51385
WOMAC pain (scale 0-20)(n=83/91)	-942	0.60	-1575
WOMAC physical (scale 0-68) (n=77/89)	-813	-0.17	4701
VAS pain now (scale 0-10) (n=84/91)	-952	0.27	-3476
VAS pain past week (scale 0-10) (n=84/89)	-1016	0.13	-7699
Patient Global Assessment (% improved) (n=83/87)	-1005	6.5	-155

[†] Cost and effect difference corrected for clustering within the factor physiotherapist

Considering the scale of the outcome measures, the effect differences were close to zero. Therefore, large cost-effectiveness ratios were found. The incremental cost-effectiveness ratio for quality of life years gained was €51385 per QALY. As the difference in QALY over 65 weeks was just

about negative, this incremental cost-effectiveness ratio means that implementing BGA yields €51385 per QALY that is lost by not giving usual care. Figure 1a shows the cost-effectiveness plane for QALYs gained. Ninety-two percent of the cost-effect pairs lie below the x-axis, the area where BGA is associated with lower costs.

The cost-effectiveness ratio for improvement in the WOMAC pain score was -€1575 per point, meaning that implementing BGA yields €1575 per point of improvement on the WOMAC pain score. Figure 1b shows the corresponding cost-effectiveness plane. Ninety-one percent of the cost-effect pairs lie in the area where costs of BGA are lower.

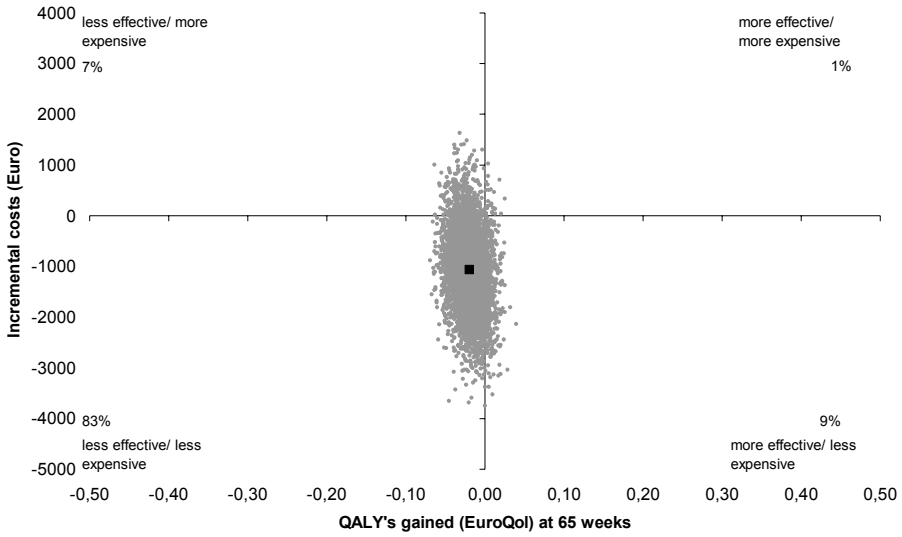

Figure 1a Cost-effectiveness plane for QALYs gained at 65 weeks for behavioral graded activity versus usual care.

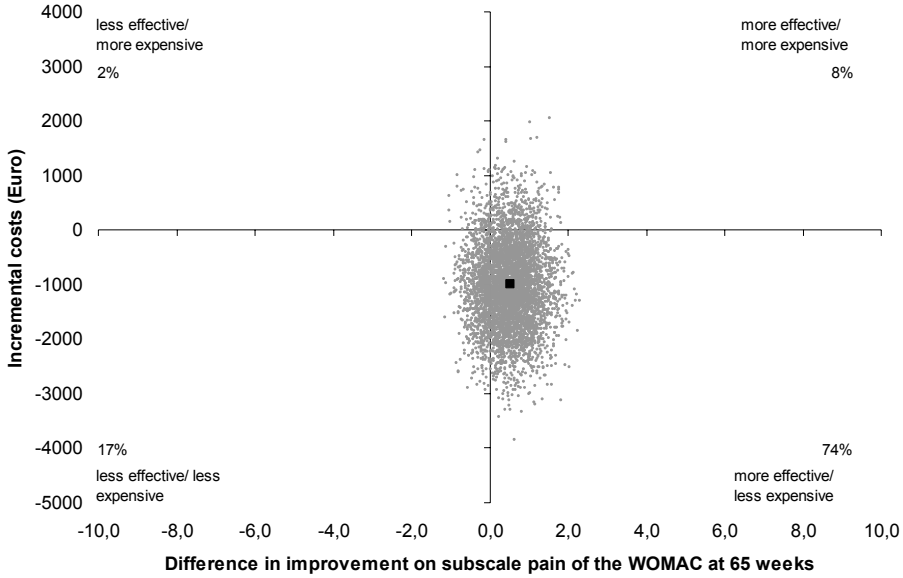

Figure 1b Cost-effectiveness plane for the difference in improvement on the subscale pain of the WOMAC at 65 weeks, for behavioral graded activity versus usual care.

Sensitivity analysis

We presented cost differences corrected for clustering of patient data. Uncorrected differences were similar. For the difference in total costs between the groups, the uncorrected difference was –€811 (95% CI: -€2106 to €946). When excluding from the analysis 11 patients with total costs exceeding €11000, we found a difference in total cost of -€740 (95% CI: -€1447 to €72). Considering only patients with complete follow-up on cost data resulted in a difference in total costs of -€1096 (95% CI: -€3105 to € 819). When zero costs were imputed for missing cost diaries in the last half of the follow-up period, the difference in total costs was -€889 (95% CI: -€2601 to €857). Imputation of mean costs for missing cost diaries resulted in a difference of -€627 (95% CI: -€1846 to €824).

Per-protocol analysis
Twenty patients from the graded activity group and ten from the usual care group were excluded from the per-protocol analysis. Mean difference in total costs between the groups was -€987 (95% CI: -€2777 to €786). For patient with complete follow-up on cost data, this difference was -€1057 (95% CI: -€3308 to €834) (n=45/60).

Discussion

Differences in direct and indirect costs between the BGA group and usual care group were not statistically significant. Due to the skewed distribution of cost data, however, cost-effectiveness studies are generally underpowered to detect differences between groups. With the exception of direct costs outside the health care sector, costs in the BGA group were consistently lower than in the usual care group. This was particularly true for the indirect costs of absenteeism. Possibly, the behavioral component of the graded activity program leads to less avoidance behavior, so that patients are less inclined to refrain from working. Interestingly, both groups are associated with similar costs for allied health care. Apparently, the graded activity protocol, prescribing more treatment sessions than the usual care program, did in practice not result in more treatment sessions. Finally, total costs were lower in the BGA group. However, the difference in total costs between the two groups of -€773 was surrounded by large confidence bounds (95% CI: -€2360 to €772). Therefore, we cannot exclude that this difference is a coincidence.

The incremental cost-effectiveness ratio for BGA compared to usual care was €51385 per QALY. With a negative difference in quality of life, this incremental cost-effectiveness ratio means that implementing BGA yields €51385 per QALY that is lost by not giving usual care. However, the difference between the 2 treatment groups in quality of life is extremely small and the sign of this difference seems of no importance. As such, it is hard to interpret this cost-effectiveness ratio and it seems more reasonable to base conclusions on cost differences rather than on cost-effectiveness.

A limiting factor in this study is the quality of the cost data. Resource utilization was monitored by patients themselves and recorded in 7

cost diaries. It is known that such information may be inaccurate, depending on one's memory and motivation to participate (29). Furthermore, it may be that missing data is caused by illness or hospitalization, thereby possibly leading to an underestimation of costs. However, sensitivity analyses showed that our results were robust to differing assumptions for missing data.

Earlier publications comparing the costs and cost-effectiveness of different types of exercise therapy for OA of the hip and/or knee are not available. In a study by Van Baar et al., exercise therapy in combination with advice and medication was compared to advice and medication only (3). When adjusted to the year 2004, the reported costs of exercise therapy are roughly half those found in our study. This discrepancy may seem remarkable, but is probably largely explained by the fact that they used prices of reimbursements by insurance companies instead of Dutch guideline prices, which where not available at that point in time.

In conclusion, this study provides no evidence that behavioral graded activity for patients with OA of the hip and/or knee is either more effective or less costly than usual care. Yielding similar results, behavioral graded activity seems an acceptable method to treat patients with OA of the hip and/or knee.

References

1. Miedema H. Reuma-onderzoek meerdere echelons (Rome): basisrapport. (Rheumatism-research on different levels of the health care system: basis report). Leiden, The Netherlands: NIPG-TNO, 1994.
2. Fransen M, McConnel S, Bell M. Therapeutic exercise for people with osteoarthritis of the hip or knee. A systematic review. J Rheumatol 2002; 29:1737-1745.
3. Van Baar ME, Assendelft WJ, Dekker J, Oostendorp RA, Bijlsma JW. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review of randomized clinical trials. Arthritis Rheum 1999; 42(7):1361-1369.
4. Veenhof C, Köke AJA, Dekker J, Oostendorp RA, Bijlsma JWJ, van Tulder MW et al. Effectiveness of behavioral graded activity in patients with osteoarthritis of hip and/or knee: a randomized clinical trial. Accepted for publication in Arthritis Rheum 2006.
5. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Arthritis Rheum 1986; 29(8):1039-1049.
6. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991; 34(5):505-514.
7. Fordyce WE, Fowler RS, Jr., Lehmann JF, Delateur BJ, Sand PL, Trieschmann RB. Operant conditioning in the treatment of chronic pain. Arch Phys Med Rehabil 1973; 54(9):399-408.
8. Lindstrom I, Ohlund C, Eek C, Wallin L, Peterson LE, Fordyce WE et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. Phys Ther 1992; 72(4):279-290.
9. Vogels EMHM, Hendriks HJM, Van Baar ME, Dekker J, Hopman-Rock M, Oostendorp RA. Clinical practice guidelines for physical therapy in patients with osteoarthritis of the hip or knee. Amersfoort: KNGF, 2001.
10. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988; 15(12):1833-1840.
11. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997; 24(4):799-802.
12. Heijden GJMG vd. Shoulder disorder treatment: Efficacy of ultrasoundtherapy and electrotherapy. Maastricht The Netherlands: University Press Maastricht, 1996.
13. The EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. The EuroQol Group. Health Policy 1990; 16(3):199-208.

14. Oostenbrink JB, Koopmanschap MA, Van Rutten FFH. Handleiding voor kostenonderzoek. Methoden en richtlijnprijzen voor economische evaluaties in de gezondheidszorg. (Handbook for cost studies, methods and guidelines for economic evaluations in health care). The Hague, The Netherlands: Health Care Insurance Council, 2000.
15. Oostenbrink JB, Koopmanschap MA, Rutten FF. Standardisation of costs: the Dutch Manual for Costing in economic evaluations. Pharmacoeconomics 2002; 20(7):443-454.
16. Z-index. G-standaard. 2002. The Hague, The Netherlands, Z-index.
17. Koopmanschap MA, Rutten FF. A practical guide for calculating indirect costs of disease. Pharmacoeconomics 1996; 10(5):460-466.
18. Busschbach JJ, Brouwer WB, van der DA, Passchier J, Rutten FF. An outline for a cost-effectiveness analysis of a drug for patients with Alzheimer's disease. Pharmacoeconomics 1998; 13(1 Pt 1):21-34.
19. Oostenbrink JB, Al MJ. The analysis of incomplete cost data due to dropout. Health Econ 2005.
20. Manca A, Rice N, Sculpher MJ, Briggs AH. Assessing generalisability by location in trial-based cost-effectiveness analysis: the use of multilevel models. Health Econ 2005; 14(5):471-485.
21. UK Beam Trial Team. United Kingdom back pain exercise and manipulation (UK BEAM) randomised trial: effectiveness of physical treatments for back pain in primary care. BMJ 2004; 329(7479):1377.
22. Barber JA, Thompson SG. Analysis of cost data in randomized trials: an application of the non-parametric bootstrap. Stat Med 2000; 19(23):3219-3236.
23. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. Health Econ 1997; 6(4):327-340.
24. Thompson SG, Barber JA. How should cost data in pragmatic randomised trials be analysed? BMJ 2000; 320(7243):1197-1200.
25. Dolan P. Modeling valuations for EuroQol health states. Med Care 1997; 35(11):1095-1108.
26. Matthews JN, Altman DG, Campbell MJ, Royston P. Analysis of serial measurements in medical research. BMJ 1990; 300(6719):230-235.
27. Richardson G, Manca A. Calculation of quality adjusted life years in the published literature: a review of methodology and transparency. Health Econ 2004; 13(12):1203-1210.
28. Briggs A, Fenn P. Confidence intervals or surfaces? Uncertainty on the cost-effectiveness plane. Health Econ 1998; 7(8):723-740.
29. Briggs AH, Wonderling DE, Mooney CZ. Pulling cost-effectiveness analysis up by its bootstraps: a non-parametric approach to confidence interval estimation. Health Econ 1997; 6(4):327-340.

Chapter

4

Which patients with osteoarthritis of hip and/or knee benefit most from behavioral graded activity?

Accepted for publication as:
Veenhof C, Van den Ende CHM, Dekker J, Köke AJA, Oostendorp RAB, Bijlsma JWJ. Which patients with osteoarthritis of hip and/or knee benefit most from behavioral graded activity? Accepted for publication in International Journal of Behavioral Medicine, 2006.

Abstract

Objective. To investigate whether behavioral graded activity (BGA) has particular benefit in specific subgroups of osteoarthritis (OA)-patients.
Methods. 200 patients with OA of the hip and/or knee (clinical ACR-criteria), participated in a randomized clinical trial on the efficacy of BGA compared to treatment according to the Dutch physiotherapy guideline (usual care, UC). Changes in pain (VAS), physical functioning (WOMAC and MACTAR), and patient global assessment (PGA) were compared for specific subgroups. Subgroups were assigned by the median-split-method and analyzed using analysis of covariance.
Results. Beneficial effects of BGA were found for patients with a relatively low level of physical functioning ($p \leq 0.03$). Furthermore, beneficial effects of BGA in patients with a low level of internal locus of control were marginally significant ($p=0.05$).
Conclusion. Patients with a relatively low level of physical functioning benefit more from BGA compared to UC. Compared to UC, BGA is the preferred treatment option in patients with a low level of physical functioning.

Introduction

Osteoarthritis (OA) is a common joint disorder, which has a major impact on functioning in daily life (1;2). Available evidence indicates beneficial short term effects of exercise therapy on pain, physical function, and patient global assessment (PGA) in patients with OA. However, these short term effects decline over time and disappear in the long term (3;4).

To enhance long term effects of exercise therapy, integration of exercise therapy with daily performed activities based on cognitive-behavioral principles and additional boostersessions seems promising. This treatment is based on the assumption that psychosocial factors interfere with the physical function of patients (5). Indeed, a behavioral graded activity program was found to result in beneficial long term outcomes, however, the outcome was not superior to usual care (6). It remains to be investigated whether specific characteristics of the patients are effect modifiers, and thus, whether specific subgroups of patients would particularly benefit from a treatment based on behavioral principles.

With regard to subgroups, three specific expectations can be formulated. First, the main objective of behavioral graded activity is to realize a more active lifestyle (6). A low level of functional activities can be caused by avoidance behavior of patients. Long-lasting avoidance of activities leads to disuse and increased disability (7). Because of the systematic attempt towards a more active lifestyle, patients with a relative low level of physical functioning are expected to benefit more from behavioral graded activity. Second, it has been demonstrated that especially patients with passive coping strategies, such as retreating, worrying, and resting, have high levels of physical and psychological disability and tend to avoid activity (8;9). Therefore, it can be expected that patients with passive coping strategies have particular benefit from behavioral graded activity. Finally, since patients with high internal locus of control report less pain (8), beneficial effects of behavioral graded activity are expected in patients with low levels of perceived control. A low level of perceived control is operationalized in 2 ways, as a low level of internal locus of control and as a high level of powerful others locus of control.

The aim of the study is to determine whether behavioral graded activity, compared to usual care, has particular benefit in specific subgroups

of patients. Beneficial long term effects are expected in: (1) patients with a relatively low level of physical functioning at the start of the treatment; (2) patients with passive coping styles to pain; (3) patients with a relative low level of internal locus of control or a relative high level of powerful others locus of control.

Methods

Subjects

A cluster randomized controlled trial was conducted, comparing 2 interventions in 200 patients with hip and/or knee OA. Inclusion criteria of eligible patients were OA of the hip or knee according to the clinical criteria of the American College of Rheumatology (10;11). An extensive description of the methods of the trial is published elsewhere (6). All patients completed written informed consent. An assessor, blinded for the allocated treatment, performed assessments at baseline, 13 weeks (post-treatment), and after 39 weeks and 65 weeks follow up. The study was approved by the medical ethics committee of the VU University Medical Center, Amsterdam.

Interventions

Behavioral graded activity

BGA is a behavioral treatment integrating the concepts of operant conditioning with exercise therapy comprising boostersessions. The intervention is directed at increasing the level of activities in a time-contingent way, with the goal to integrate these activities in the daily living of patients. The patient has many responsibilities during this treatment; the physiotherapist has a more coaching role. The treatment consisted of a 12-week period with a maximum of 18 sessions, followed by 5 pre-set boostermoments with a maximum of 7 sessions (respectively in week 18, 25, 34, 42, and 55).

Usual care

The physiotherapists of patients in the usual care (UC) group were advised to treat their patients according to the Dutch physiotherapy guideline for patients with hip and/or knee OA (12). The treatment consisted of a maximum of 18 sessions within a period of 12 weeks.

More specific information on the interventions has been published elsewhere (6). Both BGA and UC were given individually by physiotherapists in primary care.

Outcome measures

Primary outcome measures were pain, physical function, and patient global assessment (PGA), according to the core set of outcome measures of clinical trials with patients with OA defined by OMERACT III (13). Patients rated their pain at assessment and in the past week on a visual analog scale (VAS; 0-10). Physical function was assessed with the physical function subscale of the condition-specific Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC; range: 0-68), and the patient-oriented MACTAR (range: -15 to 15) (14;15). PGA was assessed by patients on a 8-point scale (1=vastly worsened; 8=completely recovered) (16).

Subgroups

For each possible effect-modifying factor, patients were classified into 2 subgroups. Subgroups were determined by the median-split-method. In this method, 2 groups are composed (low, high) for each variable with the median as cut off point. The following factors were studied.

Physical function. Physical function was assessed with the physical function subscale of the WOMAC (0-68) (14). The median for the WOMAC was 29.0. A higher score reflects more limitations in physical function.

Locus of control. Locus of control was assessed by the Multidimensional Health Locus of Control (MHLC), which consists of separate subscales for internal locus of control and powerful others locus of control (range 6-36) (17). A high score reflects a high use of the specific locus of control. The median of internal locus of control was 21.0, the median of powerful others locus of control was 18.5.

Pain coping. The use of passive pain coping strategies is reflected by high levels of retreating, worrying and resting, which are subscales of the Pain Coping Inventory (range 0-4) (18). A high score on a subscale means that the specific strategy is used when in pain. The medians on the subscales worrying, resting, and retreating were 1.67, 2.20 and 1.71 respectively.

Characteristics of patients. In an exploratory analysis, the influence of the following features was studied:

- Demographic data: age (median split: median= 65.0) and sex (male / female).
- Clinical features: location of OA (knee/hip; patients with both knee and hip OA were left out of this subanalysis), duration of complaints (5 years was chosen as a cut-off point for duration of complaints), obesity (Body Mass Index <30 / ≥ 30), pain assessed on a VAS (0-10) (median-split: median=5.0), and radiological score as assessed with Kellgren & Lawrence scale by a radiologist. We chose minimal OA (grade 2 or more) as a cut off point for radiological degeneration.
- Lifestyle: level of physical activity according to the SQUASH (Short Questionnaire to assess health enhancing physical activity) (median split: median= 1530 minutes/week) (19).

Statistical analysis

Statistical analyses were performed according to the intention-to-treat principle. To analyze the effects, change scores for pain and physical function (WOMAC and MACTAR) were calculated (follow up minus baseline scores). Only data on the 65 week follow up were used for these secondary analyses. To study whether differences existed in the effects of BGA between subgroups, the effect modification of treatment was tested using analysis of covariance (ANCOVA). Corresponding with the primary analyses on the efficacy of BGA, the following covariates were included in the subgroup analyses in order to control for differences in patients' condition: duration of complaints, location of OA (hip, knee, or both), age, sex, and recruitment method (physiotherapist or newspaper) (6). In order to avoid overcorrection, this correction was omitted in analyses of subgroups based on these characteristics (e.g., in analyses of effect modification by age, age was not included as a covariate). The significance level for effect modification was set at 0.05.

Results

As presented in Table 1, at baseline there were no relevant differences between the studied groups on the prognostic variables sex, age, location of OA, duration of complaints, radiological degeneration, BMI, pain, physical function, used pain coping style, and locus of control.

Physical function, locus of control and pain coping

In general, both BGA and UC resulted in beneficial long term effects. However, no differences were found between both interventions on the primary outcome measures pain, physical function and patient global assessment (6). In Table 2, the effects of treatment are presented for subgroups of patients, on the basis of their physical function, pain coping, and locus of control. The expected larger beneficial long term effects of BGA in patients with a relatively low level of physical functioning were confirmed, both for the outcome measures pain and physical functioning, with the exception of the outcome measure 'pain at assessment'.

Also, the hypothesis of greater benefit of BGA in patients with a relatively low level of internal locus of control was confirmed for the outcome measure 'pain in past week', which was marginally significant (p=0.05). However, for the outcome measures 'pain at assessment', physical function (WOMAC / MACTAR), and PGA (not in table) this effect modification was not significant.

The remaining hypotheses of greater benefit of BGA in patients with a high level of powerful others locus of control, or in patients using passive coping strategies could not be confirmed.

Table 1 Baseline characteristics of the intervention groups behavioral graded activity (BGA) and usual care (UC).

Characteristics	**BGA n=97**		**UC n=103**	
Gender: Female, n (%)	73	(75)	81	(79)
Age, mean (sd)	65.1	(7.4)	64.5	(8.3)
Location of OA, n (%)				
knee	67	(69)	63	(61)
hip	22	(23)	28	(27)
both	8	(8)	12	(12)
Duration of complaints, n (%)				
< 1 year	23	(24)	24	(23)
1 – 5 years	39	(41)	33	(32)
> 5 years	33	(35)	46	(45)
Radiological evidence OA (K&L ≥ 2) (n=146)[1]				
knee (n=101), n (%)	26	(52)	31	(61)
hip (n=51), n (%)	18	(86)	29	(97)
Body Mass Index, mean (sd)	28.2	(4.2)	28.8	(4.6)
Severity of pain, mean (sd)				
At assessment (VAS)	4.3	(2.8)	3.7	(2.5)
Past week (VAS)	5.7	(2.2)	5.5	(2.2)
Physical function, mean (sd)				
subscale physical function (WOMAC)	28.7	(12.5)	29.1	(9.9)
Physical activity, mean (sd)				
in min/week (SQUASH)	1761	(1221)	1664	(984)
Locus of Control (MHLC), mean (sd)				
Internal health locus of control	21.5	(5.5)	20.1	(5.5)
Powerful others health locus of control	18.5	(5.6)	18.3	(5.3)
Pain Coping Inventory (PCI), mean (sd)				
Retreating	1.8	(0.5)	1.7	(0.4)
Worrying	1.7	(0.5)	1.7	(0.4)
Resting	2.2	(0.5)	2.3	(0.5)

[1] K&L: Kellgren & Lawrence score

Table 2a Mean effects of BGA treatment (Δ Week 65 – Baseline) in subgroups according to level of physical functioning (WOMAC).

Physical functioning					
	High level of physical functioning		Low level of physical functioning		p-value effect modification
	BGA n=48	UC n=46	BGA n=34	UC n=46	
Pain at assessment (VAS)	-0.28	-0.33	-2.21	-0.83	0.12
Pain last week (VAS)	-1.55	-2.07	-2.82	-1.49	0.03
Physical functioning (WOMAC)	-2.79	-4.72	-14.06	-10.04	0.03
Physical functioning (MACTAR)	4.51	4.15	8.03	2.39	0.01

Table 2b Effects of BGA treatment in subgroups according to passive coping strategies.

Passive coping strategies					
Worrying	Low level of worrying		High level of worrying		p-value effect modification
	BGA n=49	UC n=45	BGA n=31	UC n=44	
Pain at assessment (VAS)	-0.98	-0.24	-1.10	-0.75	0.57
Pain last week (VAS)	-2.10	-1.39	-1.77	-2.07	0.28
Physical functioning (WOMAC)	-7.55	-5.50	-6.91	-9.24	0.20
Physical functioning (MACTAR)	6.82	3.60	4.45	3.48	0.40
Resting	Low level of resting		High level of resting		p-value effect modification
	BGA n=49	UC n=50	BGA n=33	UC n=39	
Pain at assessment (VAS)	-0.56	-0.32	-1.45	-0.72	0.52
Pain last week (VAS)	-2.02	-1.78	-1.88	-1.65	0.82
Physical functioning (WOMAC)	-7.39	-6.25	-7.27	-8.66	0.73
Physical functioning (MACTAR)	5.25	3.70	6.67	3.33	0.35

Table 2b continued

Withdrawing	Low level of withdrawing		High level of withdrawing		p-value effect modification
	BGA n=29	UC n=47	BGA n=51	UC N=41	
Pain at assessment (VAS)	-1.55	-1.11	-0.73	-0.20	0.40
Pain last week (VAS)	-2.28	-1.50	-1.80	-1.95	0.33
Physical functioning (WOMAC)	-9.31	-5.59	-6.09	-8.80	0.29
Physical functioning (MACTAR)	7.07	2.81	5.24	4.17	0.44

Table 2C Effects of BGA treatment in subgroups according to locus of control.

	Locus of control				
Internal locus of control	Low level of internal locus of control		High level of internal locus of control		p-value effect modification
	BGA n=37	UC n=57	BGA n=43	UC n=33	
Pain at assessment (VAS)	-0.97	-0.05	-0.74	-1.33	0.09
Pain last week (VAS)	-2.16	-1.43	-1.67	-2.31	0.05
Physical functioning (WOMAC)	-9.88	-7.55	-5.04	-6.88	0.30
Physical functioning (MACTAR)	7.59	2.93	4.14	4.36	0.06
Powerful others locus of control	Low level of powerful others locus of control		High level of powerful others locus of control		p-value effect modification
	BGA n=43	UC n=46	BGA n=38	UC n=44	
Pain at assessment (VAS)	-0.83	-0.33	-0.87	-0.73	0.57
Pain last week (VAS)	-2.31	-1.13	-1.45	-2.40	0.004
Physical functioning (WOMAC)	-7.83	-5.99	-6.78	-8.65	0.35
Physical functioning (MACTAR)	6.98	3.13	4.37	3.80	0.20

BGA: Behavioral graded activity; UC: Usual care

Additional analyses

Exploratory analyses were performed to study effect modification of 8 features for the effectiveness of BGA (sex, age, location of OA, duration of complaints, radiological degeneration, BMI, pain, and level of physical activity), leading to a total number of 40 interactions which were tested (8 features and 5 outcome measures). Three significant effect modifiers were found. In patients without radiological evidence of OA, relatively great beneficial effects of UC were found on 'pain last week' (p=0.05), compared with patients with radiological evidence of OA. In addition, in patients with relatively high score of pain at the start of treatment a beneficial effect of BGA on patient oriented physical function (MACTAR) was found compared with UC (p=0.03). Finally, in patients with obesity (BMI > 30) a beneficial effect of BGA on physical function (WOMAC) was found compared with UC (p=0.03).

Discussion

In the present study 3 hypotheses were tested concerning subgroup analyses of a randomized clinical trial on the effectiveness of behavioral graded activity. As expected, patients with a relatively low level of physical functioning showed larger beneficial effects of BGA compared to UC, both on pain and physical functioning. According to the avoidance model, patients with more limitations in physical functioning, tend to avoid activities. Because of their inactivity, their physical condition deteriorates, resulting in more limitations (7). One of the primary goals of BGA is to gradually increase activity levels despite pain and to educate patients that their disease can be self-managed, which might explain the success of BGA in patients with a low level of physical functioning.

The other hypotheses could not be confirmed, with the exception of the beneficial effects we found for patients with a relative low level of 'internal locus of control' on the outcome measure 'pain in past week' which was marginally significant (p=0.05). Therefore, there is an indication that patients with a relatively low level of 'internal locus of control', assessed at baseline, benefit more from BGA, while patients with a relatively high level of 'internal locus of control' benefit more from UC. A possible explanation

is that BGA patients learn that their pain and limitations in activities are common conditions that can be self-managed, rather than serious conditions that need careful protection (20). Probably, patients with a low level of internal locus of control have the opportunity to develop these skills and therefore make more improvements when treated with BGA.

Exploratory analyses were performed concerning effect modification by characteristics of the patients. Evidence was found of beneficial effects of UC in patients without radiological evidence of OA compared to patients with radiological evidence of OA, which is in line with the findings of Van Baar (21). Also, evidence was found of beneficial effects of BGA on physical function (WOMAC) in patients with obesity. A possible explanation is that BGA interrupts the vicious circle of obesity leading to inactivity (22). Considering the high prevalence of obesity among OA-patients, the particular benefit of BGA in obese patients is of great potential value for OA-patients. However, taking into account the number of exploratory tests (n=40), these results should be interpreted with caution.

On basis of our findings, it can be concluded that both treatments result in beneficial effects which endure in the long term. For patients with a relatively low level of physical functioning at the start of the treatment and, to a lesser degree, patients with a low level of internal locus of control, treatment with BGA (compared to UC) is the preferred treatment option, as particular benefit was found for these specific subgroups of patients.

References

1. Dekker J, Boot B, van der Woude LH, Bijlsma JW. Pain and disability in osteoarthritis: a review of biobehavioral mechanisms. Journal of Behavioral Medicine 1992; 15(2):189-214.
2. van Baar ME, Dekker J, Lemmens JA, Oostendorp RA, Bijlsma JW. Pain and disability in patients with osteoarthritis of hip or knee: the relationship with articular, kinesiological, and psychological characteristics. Journal of Rheumatology 1998; 25(1):125-133.
3. Fransen M, McConnel S, Bell M. Therapeutic exercise for people with osteoarthritis of the hip or knee. A systematic review. Journal of Rheumatology 2002; 29:1737-1745.
4. van Baar ME, Assendelft WJ, Dekker J, Oostendorp RA, Bijlsma JW. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review of randomized clinical trials. Arthritis and Rheumatism 1999; 42(7):1361-1369.
5. Linton SJ, Hellsing AL, Andersson D. A controlled study of the effects of an early intervention on acute musculoskeletal pain problems. Pain 1993; 54:353-359.
6. Veenhof C, Köke A, Dekker J, Oostendorp R, Bijlsma JWJ, van Tulder MW et al. Effectiveness of behavioral graded activity in patients with osteoarthritis of hip and/or knee: a randomized controlled trial. 2006.
7. Vlaeyen JW, Linton SJ. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain 2000; 85(3):317-332.
8. Keefe FJ, Smith SJ, Buffington AL, Gibson J, Studts JL, Caldwell DS. Recent advances and future directions in the biopsychosocial assessment and treatment of arthritis. Journal of Consulting and Clinical Psychology 2002; 70(3):640-655.
9. Steultjens MP, Dekker J, Bijlsma JW. Avoidance of activity and disability in patients with osteoarthritis of the knee: the mediating role of muscle strength. Arthritis and Rheumatism 2002; 46(7):1784-1788.
10. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis and Rheumatism 1986; 29(8):1039-1049.
11. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis and Rheumatism 1991; 34(5):505-514.
12. Vogels E, Hendriks H, van Baar ME, Dekker J, Hopman-Rock M, Oostendorp RA et al. Clinical practice guidelines for physical therapy in patients with osteoarthritis of the hip or knee. Amersfoort: KNGF, 2001.

13. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. Journal of Rheumatology 1997; 24(4):799-802.
14. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. Journal of Rheumatology 1988; 15(12):1833-1840.
15. 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. Journal of Rheumatology 1987; 14(3):446-451.
16. van der Heijden G. Shoulder disorder treatment: Efficacy of ultrasoundtherapy and electrotherapy. Maastricht The Netherlands: University Press Maastricht, 1996.
17. Wallston KA, Wallston BS, DeVellis R. Development of the Multidimensional Health Locus of Control (MHLC) Scales. Health Education Monographs 1978; 6(2):160-170.
18. Kraaimaat FW, Bakker A, Evers A. Pijncoping-strategieën bij chronische pijnpatiënten: De ontwikkeling van de Pijn-Coping-Inventarisatielijst (PCI) (in Dutch). Gedragstherapie 1997; 30:185-201.
19. Wendel-Vos GC, Schuit AJ, Saris WH, Kromhout D. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. Journal of Clinical Epidemiology 2003; 56(12):1163-1169.
20. Vlaeyen JW, Crombez G. Fear of movement/(re)injury, avoidance and pain disability in chronic low back pain patients. Manual Therapy 1999; 4(4):187-195.
21. van Baar ME, Dekker J, Bijl D, Oostendorp RA, Voorn TB, Bijlsma JW. Exercise therapy in patients with osteoarthritis of hip or knee: outcome in specific subgroups of patients. In: van Baar ME, editor. Effectiveness of exercise therapy in osteoarthritis of hip or knee. Utrecht: NIVEL, 1998: 55-73.
22. Petersen L, Schnohr P, Sorensen TI. Longitudinal study of the long-term relation between physical activity and obesity in adults. International Journal of Obesity and Related Metabolic Disorders 2004; 28(1):105-112.

Chapter

5

Factors influencing the success of behavioral graded activity in patients with osteoarthritis of the hip and/or knee: a qualitative study

Submitted for publication as:
Veenhof C, Van Hasselt TJ, Köke AJA, Dekker J, Bijlsma JWJ, Van den Ende CHM. Factors influencing the success of behavioral graded activity in patients with osteoarthritis of the hip and/or knee: a qualitative study. Submitted for publication in Australian Journal of Physiotherapy, 2006.

Abstract

Objective. Recently, a randomized controlled trial was conducted on the effectiveness of behavioral graded activity (BGA) in patients with osteoarthritis of the hip and/or knee. BGA is an exercise program with boostersessions, using operant behavioral principles, which aims at increasing the level of activities and integrating this level of activities in daily life in the long term. The purpose of the present study was to investigate why certain patients treated with BGA, successfully integrate activities in their daily lives and others do not.

Methods. A qualitative study was performed, based on interviews with 12 patients. Patients were selected according to the model of deliberate sampling for heterogeneity. Selection was based on their success of the treatment as assessed on the patient global assessment. The data were coded and analysed using the methods developed in grounded theory.

Results. The findings from this study suggest that 2 factors influence the long term performance of activities. First, patients' initial motivation for long term goals seems to relate to a higher adherence of patients to perform activities in the long term. Second, an active involvement of patients during the treatment process seems to relate to a higher adherence to perform activities in the long term.

Conclusion. Although the involvement of patients in the treatment process is already part of the BGA protocol, it would be beneficial to emphasize the importance of an active involvement of patients right from the start of the treatment. Furthermore, to increase the success of treatment, physiotherapists should gain a clear understanding of the patients' initial motives for treatment.

Introduction

Osteoarthritis (OA), especially of the knee and hip, is the most common joint disorder, causing pain, joint stiffness, muscle weakness, and joint instability, threatening mobility, and an active lifestyle (1;2). There is strong evidence that exercise therapy in patients with OA of the hip and/or knee, has beneficial effects on pain and physical function in short term. However, beneficial post-treatment effects of exercise therapy seem to decline over time and finally disappear (3;4).

To enhance the long term effectiveness, integrating exercise therapy with daily performed activities based on cognitive behavioral principles and additional boostersessions seems promising. This treatment is based on the assumption that psychosocial factors interfere with the physical function of patients (5-7). Recently, such an intervention, namely behavioral graded activity (BGA), was compared with usual physiotherapeutic care (UC, physiotherapy according to the Dutch OA guideline, which mainly consists of providing information and exercise therapy (8)). Both treatments resulted in beneficial long term effects, but, in general, no differences were found between both interventions on the primary outcome measures pain (visual analog scale (VAS); Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)), physical function (WOMAC), and patient global assessment (PGA) (9).

Since patients with OA are very heterogeneous in their complaints and restrictions of activities, it is likely that certain patients will benefit more from BGA than other patients. This was confirmed in subgroup analyses (i.e. comparing effectiveness of BGA and UC in specific subgroups of patients), which demonstrated that patients with a relatively low level of physical functioning benefited more from BGA, compared to UC. Patients with a relatively high level of physical functioning benefited equally from both treatments. Furthermore, limited evidence was found for beneficial effects of BGA (compared to UC) in patients with a low level of internal locus of control (10).

However, these subgroup analyses yield only a limited understanding of the differential effects of BGA. To further investigate why some patients remained to have an active lifestyle, i.e. adhered to perform activities in the long term, after completion of BGA treatment and others did

not, we conducted a qualitative study. To our knowledge, no studies on the success and exercise adherence of treatments from the point of view of patients with OA are available.

The objective of the study was to investigate which factors explain the differences, after BGA treatment, between patients who successfully integrate activities in their daily lives and patients who do not succeed in integrating activities in their daily lives.

Methods

Study design

In this qualitative study, open-ended in-depth interviews were conducted with patients with OA who were treated with BGA. The interviews were conducted 1 to 6 months after the last assessment (which was planned 65 weeks after the start of treatment). This variation in time span was caused by the relatively long inclusion period (1,5 years) of the original trial (9).

Study population

We invited a sample of participants treated with BGA to participate in this qualitative study. Patients were included in the original clinical trial (9) if they were diagnosed to have OA of the hip or knee according to the clinical criteria of the American College of Rheumatology (11;12). Twohundred patients participated in the trial, 97 BGA patients and 103 UC patients. More specific information on the study design, population and intervention is presented in the clinical paper (9).

For the present study, the aim was to select 2 sub-samples of patients from the BGA-group (n=97). Patients were selected according to the 'model of deliberate sampling for heterogeneity', meaning that a wide range of subjects are represented in the sample, which increases the external validity (i.e. results can be generalized to a broader population) (13). Patients were selected on basis of their success of the treatment as assessed on the patients global assessment (PGA). PGA was assessed on an 8-point scale (1=completely recovered; 8=vastly worsened) (14). The score on PGA was taken as a basis to select 2 samples of patients. One group of patients was selected with a low score on PGA (score 1-3, ranging from completely

recovered to improvement), and 1 group of patients with a high score on PGA (score 6-8, ranging from worsened to vastly worsened). The researcher contacted appropriate candidates by phone and send letters to interested patients, informing them about the aim and procedure. Informed consent was obtained prior to each interview.

Behavioral graded activity

BGA is a behavioral treatment integrating the concepts of operant conditioning with exercise therapy comprising boostersessions. BGA is based on the time-contingency-management as described by Fordyce (15) and applied by Lindström (5). The intervention is directed at increasing the level of activities in a time-contingent way, with the goal to integrate these activities in the daily living of patients. Patients have many responsibilities and an active role during this treatment, the physiotherapists have a coaching role. BGA was given individually by physiotherapists in primary care. The BGA treatment is completely protocolized and included written materials (e.g., education messages, activity diaries, performance charts). The treatment consisted of a 12-week period with a maximum of 18 sessions, followed by 5 pre-set boostermoments with a maximum of 7 sessions (respectively in week 18, 25, 34, 42, and 55).

Data collection

The study was based on open-ended in-depth interviews centering on the patient's experiences of the treatment BGA. The interview was conducted as a conversation departing from 3 main themes: aspects related to the content of the BGA treatment, aspects related to experiences with the physiotherapist, and aspects related to the patient. A conversation developed from these themes and subsequent questions were more specific. New topics brought up by the patients were discussed in the following interviews (13). Letters were sent out to 19 patients of whom 13 patients agreed to participate. The quality of 1 interview was poor, because the patient did not consequently respond to the questions; this interview was left out for final analysis. Data collection stopped after 12 interviews; no new relevant data seemed to emerge during the last interviews, which indicates that a saturation point had been reached (13). Each interview lasted about 90 minutes and was performed in the patient's home. The interviews were

recorded on audiotape.

Data analysis

The method of data analysis was based on a grounded theory approach (13). This inductive data analysis was performed using the software package WinMax-Pro98. Initially, interview transcripts were read to identify conceptual themes in the text which were then coded. The codes in each interview were then compared and codes expressing related concepts were grouped together to create broader categories that linked codes across interviews. Constant comparison of emerging issues and searching for deviant or negative cases helped to confirm and further develop a tentative theory to which each patient's experiences could contribute (13). All interviews were analysed by the same researcher (TH); a random sample of interviews was analysed by a second investigator (CV).

To increase the validity of the coding framework, additional strategies were adopted. First, 2 interviews were conducted by 2 researchers to monitor consistency of the process and completeness of data collection. Second, triangulation of researchers, meaning that the researcher and co-researcher first analysed the interviews independently and then compared and discussed the codes and interpretation. Finally, peer debriefing, meaning that interim analyses were discussed in a group of researchers.

Results

Study population

Twelve patients participated in the interviews, resulting in 2 sub-samples of 6 patients. Six patients were identified as being improved and 6 patients as worsened on the PGA. Characteristics of these patients are presented in Table 1.

Exercise adherence

During the interviews, patients were asked whether they integrated the activities in their daily life (or in other words: whether they still adhered to the performance of the activities). Six patients reported that they still performed the same level of activities/exercises as during the treatment

period. Four patients reported that they did not perform the activities/exercises as was agreed during the treatment period and the answer of 2 patients on their performance of activities was doubtful.

When comparing the results on PGA with the exercise adherence of the patients, as reported during the interviews, it appeared that these 2 factors did not agree with each other (Table 1). Some patients who scored high on PGA, e.g., because they perceived less pain, did not continue with their activities, while some patients who scored low on PGA, e.g., because their pain remained the same, reported that their level of activities had increased considerably.

'Because my complaints disappeared, I was no longer motivated to continue with the activities/exercises' (a non-adherent patient with a high score on PGA).

'Although I experience the same level of pain, I have learned to continue with my activities and I realize that I achieve more because of that' (an adherent patient with a low score on PGA).

This disagreement between exercise adherence and PGA can be explained by the level on which the 2 outcome measures were assessed. Exercise adherence is an outcome measure at the level of the process of the treatment, while PGA is an outcome measure at the level of the patients' health. Because the main goal of BGA was to increase the patient's level of activities, which is closer to the process of treatment, we decided to focus on the exercise adherence of the patients.

Table 1 Characteristics of interviewed patients.

Characteristics	**Patients**											
Gender	F	F	F	F	F	F	M	M	F	F	M	M
Age (in years)	69	80	70	71	76	51	71	71	73	75	55	75
Location of OA	Hip	Knee	Knee	Knee	Knee	Hip	Knee	Knee	Knee	Knee + Hip	Knee	Knee
Pain (baseline VAS, 0-10)	5	7	4	4	5	7	4	8	2	4	8	5
Physical function (baseline, Womac 0-68) [1]	13	39	36	44	13	36	37	10	10	30	?	11
Patient global assessment[2]	2	7	7	2	7	1	1	2	6	6	1	7
Active lifestyle after treatment[3]	yes	no	yes	yes	no	?	?	yes	yes	yes	no	no

F: Female; M: Male;

[1] a higher score indicates a worse status of physical function;

[2] measured after 65 weeks on a 8-point scale, 1=completely recovered, 8=vastly worsened;

[3] adherence to perform activities, as reported during the interviews;

? doubtful if patient maintained active lifestyle after treatment.

Factors relating to the exercise adherence of patients

To understand why some patients adhere to their activities / exercises after the BGA treatment, and others do not, we analysed whether factors or combination of factors could be identified which relate to the exercise adherence of patients. Factors were compared on the level of the treatment, physiotherapist and patients. As presented in Table 2, for most factors (e.g., satisfaction of patients, former physiotherapeutic experiences of patients, patients' time, attitude and social support), no relationship was found with the exercise adherence of patients because of a lack of consistency in the results. However, we did find a clear relationship for 2 factors.

Table 2 Identification of factors, influencing the success of treatment, illustrated for adherent, non-adherent and unknown patients[1]

Factors	**Adherent patients (n=6)**	**Non-adherent patients (n=4)**	**Unknown (n=2)**
High satisfaction of treatment	4	3	2
Positive experience with physiotherapist	4	4	2
Former physiotherapeutic experiences	3	3	1
Positive attitude towards physical activity	6	3	1
High self-efficacy	5	2	0
Social support	5	1	2
Time	5	2	1
Motivation for treatment:			
- short term	0	2[2]	2
- long term	6	0	0
Active involvement of patient	6	1	1

[1]For each factor the number of patients in each group is presented.
[2]two patients did not report a motivation for treatment

First, the initial motivation of patients played an important role during the treatment process; some patients were motivated to reach short term goals, e.g., to decrease the pain, while others were motivated to reach long term goals, e.g., to postpone an operation or to live independently as long as possible. It appeared that all patients in the adherent group were initially motivated to reach long term goals with their visits to the physiotherapist, while all patients of the non adherent group reported a short term goal or had no specific goal to visit the physiotherapist. These patients tended to quit

performing their activities as soon as the short term motivation was obtained. Therefore, there seems to be a relationship between the initial motivation of the patients to visit the physiotherapist and the exercise adherence of the patients:

'I wanted to get rid of the pain. If the pain disappears, why would I bother to continue my exercises? I understand it is better to do my exercises to avoid that the pain returns but, in case the pain will return, I will start with my exercises again' (patient with motivation for short term goal).

'I continue with my exercises, they are integrated in my daily living. I really know these exercises have beneficial effects and that motivates me to continue with my exercises. The main motivation to do all this is to prevent an operation to get a new hip' (patient with motivation for long term goal).

Second, patient involvement in treatment differed. Some patients reported that they were actively involved in choosing the performed activities, in gradually increasing these activities and in using the performance charts. On the other hand, other patients reported that main decisions were taken by the physiotherapists and they performed the activities as instructed. It appeared that all patients of the adherent group reported that they were actively involved in the whole process of treatment; the physiotherapists had a coaching function in their treatment process. However, most patients of the non-adherent group reported that the physiotherapists made all decisions (which was sometimes a deliberate choice of the patients). Therefore, it seems that the active involvement of the patients during the treatment process is a facilitator for the success of the treatment:

'The physiotherapist determined the gradual increase of the exercises; he told me, for example, to increase the exercises with 5 minutes. I liked the fact that he told me what to do, nevertheless, he was my physiotherapist' (patient, not actively involved in treatment process).

'The approach of the physiotherapist was very democratic, which I appreciated. Together, we discussed the activities and the increase of the activities. I could indicate to what extent I wanted to increase the activities, in order to main performing the exercises' (actively involved patient in treatment process).

Discussion

It has become increasingly important to identify subgroups of patients, in order to match the applied intervention with the subgroup's clinical presentation (16;17). The qualitative data analysed in this study might provide valuable insight into the specific factors that relate to the success of treatment based on behavioral aspects. The findings from this study suggest that 2 factors influence the success, operationalised as patients' exercise adherence, of the BGA treatment.

First, an initial motivation for long term goals of patients seems to be positively related to the exercise adherence of patients. Apparently, the initial motivation and values of patients play an important role in integrating the activities in daily life and in accepting the pain. According to the model of motivation for pain self-management of Jensen et al. (18), 2 variables influence the readiness to change of patients: the patients' belief about the importance of change (patients' values) and the patients' beliefs about one's ability to engage in this behavioral change. This concurs with the recently developed acceptance approach of McCracken et al. (19), assuming that the acceptance of chronic pain is an active willingness to engage in meaningful activities in life regardless of pain related sensations. Therefore, the focus of treatment is not on reducing pain but on reducing the distress and the disabling influence of pain in chronic pain patients. An important condition is that activities which are valuable for the individual patient are the basis of treatment.

Second, the involvement of patients during the treatment process seems to relate to the success of the treatment. This agrees with literature on exercise adherence; it was concluded that a self-regulation approach, characterized by a mutual-participation relationship between patient and physiotherapist, is of great use in achieving long term exercise adherence

(20). Also, rehabilitation appears to be more effective where goals are explicitly set and where patients participate in setting them (21). Furthermore, the involvement of patients has a positive influence on the self-efficacy of patients, since it enlarges the persons' confidence in their ability to successfully execute and accomplish a given task (22). There is growing evidence that self-efficacy functions as mediator of pain and psychosocial health status (23).

The success of the treatment as assessed by PGA did not correspond with the adherence of patients to perform activities in the long term. Apparently, PGA measures the general health of patients while adherence measures the process of treatment. Since PGA is one of the primary outcome measures of our study, and PGA reflects the opinion of the patients, we believe it was the best measure to select patients. However, the main objective of the BGA treatment is to integrate activities in daily life, and therefore adherence to perform activities in the long term, seems the most appropriate measure when analysing relations between factors and success of treatment. Moreover, PGA reports about the recovery/improvement experienced by patients, imply that patients improve over time, while BGA-treatment concerns coping with complaints: patients are motivated to integrate higher levels of activities in their daily lives in spite of their complaints, e.g., pain.

In the future, it would be interesting to investigate whether differences can be found between patients who integrated activities in daily life and patients who do not integrate these activities in daily life. Although the specific consequences of exercise nonadherence among patients with OA have not been well studied (24), the expectancy is that the lack of performing activities is likely to result in a deconditioned state.

Although this qualitative study can be used as a valuable source of information, there are some limitations that deserve attention. First, recall bias needs to be taken into account, since patients were interviewed quite a long time after the treatment (1 to 6 months after the last assessment). Furthermore, the sample size was small, 12 respondents participated in the interviews. However, during the last two interviews no new relevant information emerged which indicates that a saturation point had been reached (13).

The expanded knowledge and information provided from this study can be useful in the implementation and application of the BGA treatment. Although the involvement of patients in the treatment process is already part of the BGA protocol, based on the results of this study, it would be beneficial to emphasize to physiotherapists (e.g., during the education of the BGA protocol and at the start of the treatment) the importance of the active involvement of patients right from the start of treatment. Furthermore, to increase the success of treatment, physiotherapists need to consider the patients' preferences/values at the start of treatment.

References

1. Dekker J, Boot B, van der Woude LH, Bijlsma JW. Pain and disability in osteoarthritis: a review of biobehavioral mechanisms. Journal of Behavioral Medicine 1992; 15(2):189-214.
2. Van Baar ME, Dekker J, Lemmens JA, Oostendorp RA, Bijlsma JW. Pain and disability in patients with osteoarthritis of hip or knee: the relationship with articular, kinesiological, and psychological characteristics. Journal of Rheumatology 1998; 25(1):125-133.
3. Fransen M, McConnel S, Bell M. Therapeutic exercise for people with osteoarthritis of the hip or knee. A systematic review. Journal of Rheumatology 2002; 29:1737-1745.
4. Van Baar ME, Assendelft WJ, Dekker J, Oostendorp RA, Bijlsma JW. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review of randomized clinical trials. Arthritis & Rheumatism 1999; 42(7):1361-1369.
5. Lindstrom I, Ohlund C, Eek C, Wallin L, Peterson LE, Fordyce WE et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. Physical Therapy 1992; 72(4):279-290.
6. Linton SJ, Hellsing AL, Andersson D. A controlled study of the effects of an early intervention on acute musculoskeletal pain problems. Pain 1993; 54:353-359.
7. Van Baar ME, Dekker J, Oostendorp RA, Bijl D, Voorn TB, Bijlsma JW. Effectiveness of exercise in patients with osteoarthritis of hip or knee: nine months' follow up. Annals of Rheumatic Diseases 2001; 60(12):1123-1130.
8. Vogels EMHM, Hendriks HJM, van Baar ME, Dekker J, Hopman-Rock M, Oostendorp RAB et al. Clinical practice guidelines for physical therapy in patients with osteoarthritis of the hip or knee. Amersfoort: KNGF, 2001.
9. Veenhof C, Köke AJA, Dekker J, Oostendorp RAB, Bijlsma JWJ, van Tulder MW et al. Effectiveness of behavioral graded activity in patients with osteoarthritis of hip and/or knee: a randomized controlled trial. 2005.
10. Veenhof C, Van den Ende CHM, Dekker J, Köke AJA, Oostendorp RAB, Bijlsma JWJ. Which patients with osteoarthritis of hip and/or knee benefit most from behavioral graded activity? Accepted for publication in Arthritis and Rheumatism, 2006.
11. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Arthritis & Rheumatism 1986; 29(8):1039-1049.
12. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis & Rheumatism 1991; 34(5):505-514.
13. Strauss AL, Corbin J. Basics of qualitative research. Grounded theory procedures and techniques. London: Sage, 1990.

14. Van der Heijden GJMG. Shoulder disorder treatment: Efficacy of ultrasoundtherapy and electrotherapy. Maastricht The Netherlands: University Press Maastricht, 1996.
15. Fordyce WE, Fowler RS, Lehmann JF, Delateur BJ, Sand PL, Trieschmann RB. Operant conditioning in the treatment of chronic pain. Archives of Physical Medicine and Rehabilitation 1973; 54:399-408.
16. Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. Annals of Internal Medicine 2004; 141(12):920-928.
17. George SZ, Delitto A. Clinical examination variables discriminate among treatment-based classification groups: a study of construct validity in patients with acute low back pain. Physical Therapy 2005; 85(4):306-314.
18. Jensen MP, Nielson WR, Kerns RD. Toward the development of a motivational model of pain self-management. Journal of Pain 2003; 4(9):477-492.
19. McCracken LM, Carson JW, Eccleston C, Keefe FJ. Acceptance and change in the context of chronic pain. Pain 2004; 109(1-2):4-7.
20. Sluijs EM, Knibbe JJ. Patient compliance with exercises: different theoretical approaches to short-term and long-term compliance. Patient Education and Counseling 1991; 17(3):191-204.
21. Baker SM, Marshak HH, Rice GT, Zimmerman GJ. Patient participation in physical therapy goal setting. Physical Therapy 2001; 81(5):1118-1126.
22. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychological Review 1977; 84(2):191-215.
23. Keefe FJ, Smith SJ, Buffington AL, Gibson J, Studts JL, Caldwell DS. Recent advances and future directions in the biopsychosocial assessment and treatment of arthritis. Journal of Consulting and Clinical Psychology 2002; 70(3):640-655.
24. Marks R, Allegrante JP. Chronic osteoarthritis and adherence to exercise: a review of the literature. Journal of Aging and Physical Activity 2005; 13(4):434-460.

Chapter

6

The influence of various recruitment strategies on the study population and outcome of a RCT involving patients with osteoarthritis of the hip or knee

Published as:
Veenhof C, Dekker J, Bijlsma JWJ, Van den Ende CHM. The influence of various recruitment strategies on the study population and outcome of a RCT involving patients with osteoarthritis of the hip or knee. Arthritis & Rheumatism (Arthritis Care & Research) 2005; 53(3):375-382.

Abstract

*Objective.*To examine the effect of 2 different recruitment methods on the characteristics of participants with osteoarthritis (OA) of the hip or knee and on the efficacy of an exercise program.
Methods. In a clinical trial on the effectiveness of exercise therapy in OA of the hip or knee, 2 groups of patients were recruited: 1 group through referrals by physiotherapists (PT group, n=110) and one group invited by newspaper articles (NP group, n=90). At baseline, demographic, clinical and psychosocial data were collected and compared between the 2 groups using chi-square and Student's t-tests. After 13 weeks of exercise therapy and follow-up assessments at week 39 and 65, the main outcome measures (pain, physical function, and global perceived effect) were assessed and compared by multiple regression analysis.
Results. The NP group reported less pain and tiredness at baseline, although more joints were affected with OA. The PT-group scored higher on the 'powerful-others' scale of locus of control. After adjusting for baseline differences, the effect of treatment after 13, 39, and 65 weeks was comparable for both groups for all outcome measures.
Conclusion. Recruitment method affects clinical characteristics and physical functioning of patients recruited for the study. A mix of recruitment strategies does not seem to affect treatment outcome, on the condition that adjustments are made for baseline differences.

Introduction

To conduct a clinical trial in primary care, recruitment of a sufficient number of study participants is among the most challenging aspects. Various recruitment strategies are used for clinical trials, including referrals by physicians and physiotherapists, advertisements, and mailings. Different recruitment sources may target different types of patients and the recruitment method can influence the results of studies on the effectiveness of a treatment. For example, one may speculate that advertisement preferentially attracts patients whose symptoms are not severe enough for them to seek medical care. The effect of various recruitment strategies (telephone survey versus a media campaign) on the types of subjects entered into a primary prevention clinical trial was investigated by King et al. (1). Few differences between recruitment sources were found for demographic variables (weight, marital status, number of persons in household, interest in making health changes), but counter to expectations, no differences were found in subsequent exercise adherence rates.

Recruitment strategy may affect clinical and psychosocial characteristics of the patient sample, which in turn may affect treatment outcome. In osteoarthritis (OA) of the hip or knee, loss of muscle strength (2-5), obesity (6,7), decreasing educational status (8), use of passive coping styles (such as worrying, catastrophizing, and avoidance of physical activity) (9-11), helplessness (7,8), and radiographic evidence (3,4) are associated with both pain and limitations of activities, which are the primary outcome measures of clinical trials with patients with OA (12). However, the influence of these factors on the efficacy of interventions has been investigated in only a few studies. Sex, age, race, degree of obesity, muscle strength, range of mobility (ROM), level of disabilities, and passive pain coping strategies were shown not to influence treatment outcome (13,14). Limited evidence was found of beneficial effects in patients without radiological OA, in patients with reports of recent onset, and in patients who complied with exercise therapy (13).

In the present study, we examined the effect of 2 recruitment strategies on the characteristics of patients with OA of the hip or knee and on the efficacy of an exercise program in these patients. The first population was recruited through participating physiotherapists referred by general

practitioners (GP's). Difficulties in recruiting sufficient subjects led to the use of a second recruitment strategy. The second population responded to local newspaper articles about the study. Patients responding to the newspaper were hypothesized to have less severe symptoms and to use a more active coping style compared to patients referred by a physiotherapist. Furthermore, it was hypothesized that differences in patient characteristics do not influence the efficacy of exercise therapy on primary outcome measures.

Methods

Study population

To answer the research questions, data from a randomized clinical trial were used. In the randomized clinical trial, 2 interventions for patients with OA of the hip or knee were compared. The intervention group was treated with exercise therapy integrated with the concept of graded activity (15,16) given in intermittent sessions (called behavioral graded activity (BGA)). The treatment of the control group consisted of exercise therapy according to the Dutch guideline for physiotherapists concerning hip or knee OA (Usual Care, UC) (17). Data were collected at baseline and at 13 weeks, 39 weeks, and 65 weeks.

All treatments were given by physiotherapists in primary care. A total of 87 physiotherapists were trained to participate in the study. To avoid exchange of information between participating physiotherapists about the 2 different treatments, randomization was performed at the level of the participating physiotherapeutic practices.

Two hundred patients with hip or knee OA participated in the trial. Inclusion criteria of eligible patients were OA of the hip or knee according to the clinical criteria of the American College of Rheumatology (18,19). Exclusion criteria were as follows: other pathology explaining the symptoms; symptoms in <10 of 30 days; treatment for these symptoms with exercise therapy in the preceding 6 months; <50 or >80 years of age; indication for hip or knee replacement within 1 year; contraindication for exercise therapy; inability to understand the Dutch language; and a high level of physical function, operationalized on a score of <2 on the walking

ability and physical function sections of the Algofunctional Index (AFI) (20). Data of all 200 patients were used in this study.

Recruitment methods

Patients were recruited for the trial in 2 ways, through physiotherapists and through articles in local newspapers. The selection procedure of both recruitment strategies is described.

Patients recruited by physiotherapists (PT group) were selected in the period November 2001 through May 2003. Patients, who had been referred to participating physiotherapists, received oral and written information about the study from the physiotherapist. If the patient was interested, the physiotherapist contacted the research team to enroll the patient. Then, the research team contacted the patient by phone explaining the goal and implications of participating in the study and performing a first screening. If patients were eligible, a final screening visit at home was planned.

Recruitment through newspapers (NP group) took place in the period November 2002 through May 2003. Articles about the benefit of exercise therapy in patients with OA of the hip or knee and information about the BGA study were published in local newspapers of several villages and small towns. The total circulation was about 300.000. Interested patients were asked to contact the research team by telephone. The selection procedure consisted of several steps:

1. Interested patients contacted the research team by telephone. The research team gave oral information. Patients were screened by telephone and were instructed to consult their GP about the treatment by a physiotherapist, the study, and a referral to a physiotherapist. Furthermore, written information about the treatment and implications of participating in the study was sent to them.
2. Two weeks later, eligible patients were contacted to check the opinion of their GP. If patients were still motivated to participate and a referral to a physiotherapist was obtained, an appointment for a screening at home was made.
3. A research assistant visited the patients and performed the final screening. In case of inclusion, the patients chose a physiotherapist participating in the trial. The patients were not aware of the kind of intervention (BGA or UC)

the physiotherapists would give.
If patients were eligible and willing to participate, informed consent was signed and baseline measurements were performed.

Outcome assessment
Demographic and clinical data were collected for each patient including age, sex, education, height, weight, location of osteoarthritis, duration of symptoms, and the presence of other chronic disorders. The body mass index was calculated for each patient (weight/height^2). X-rays of the hip and knee were scored by a rheumatologist following a standardized procedure according to the Kellgren and Lawrence scale (21,22). The Kellgren and Lawrence scale consists of 5 degrees: 0, no OA; 1, doubtful OA; 2, minimal OA; 3, moderate OA; and 4, severe OA.

Impairments in body functions. Patients rated their pain and tiredness at the moment of assessment and in the past week on a visual analog scale (VAS). Pain was assessed with the pain subscale of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) (23) and the pain section of the AFI (20). Stiffness was assessed with the stiffness subscale of the WOMAC. Measurements of assisted active ROM of the hip and knee bilaterally were performed with a goniometer according to a standardized protocol (24). Isometric muscle strength of the hip and knee bilaterally was measured with the MicroFet, a hand-held dynamometer (25). The measurements of both ROM and muscle strength were repeated 2 times, the average score was used in the analyses.

Limitations of activities. Limitations were assessed with the physical function subscale of the WOMAC and the walking ability and physical function sections of the AFI. Global perceived effect (GPE) was assessed by patients on a 8-point scale (1=vastly worsened; 8=completely recovered) (26). The ability to walk was measured by a 5-meter walking time test. The level of physical activity was determined by the Short Questionnaire to Assess Health Enhancing Physical Activity (SQUASH) (27).

Psychosocial variables. Coping was assessed by the Pain Coping Inventory (PCI) (28). Locus of control was measured by the Multidimensional Health Locus of Control (MHLC), reflecting 3 dimensions of health locus of control beliefs: internality, powerful others and chance

(29,30). Patients rated their social support on the Social Support Scale (SOS) (31).

Health-related quality of life. Quality of life was assessed with the MOS Short form 36 checklist (SF-36) (32).

All used instruments are reported to be reliable and valid. Primary outcome measures were pain (VAS and WOMAC), physical function (WOMAC), and global perceived effect, according to the core set of outcome measures of clinical trials with patients with OA defined by Outcome Measures in Rheumatology Clinical Trials III (OMERACT) (12). All measurements were obtained at baseline, 13 weeks, 39 weeks, and 65 weeks by 3 trained research assistants who were blinded for the recruitment condition (physiotherapist or newspaper) and the assigned treatment.

Statistical analysis

The analysis was performed on an intention-to-treat basis. At 13 weeks, 39 weeks, and 65 weeks loss to followup was 8, 30, and 21, respectively. The ratings of GPE were dichotomized as improved ("completely recovered", "very much improved", and "much improved") versus not improved ("slightly improved", "not changed", slightly worsened", "much worsened", and "vastly worsed") and risk ratios (RR) were calculated to test differences between groups. Since GPE could not be determined at baseline, GPE was assessed only at week 13, 39 and 65. Comparisons of the baseline scores between the 2 recruitment groups were conducted using chi-square tests for frequency data and Student's t-tests for continuous data. Change scores were calculated by subtracting the baseline scores from the posttreatment scores (week 13, 39, and 65) and were compared for the 2 recruitment groups using Student's t-test. To adjust for baseline differences, multiple linear regression analyses were performed with the change scores as dependent variable, type of recruitment as independent variable, and the baseline scores of each outcome measure as covariates. In addition, pain at the moment of assessment, location of OA, and duration of symptoms (i.e. the most important variables on which the groups differed at baseline; see Results), were included as covariates. P values < 0.05 were considered statistically significant.

Results

Participant flow

Figure 1 shows the flow of participants through the stage of patient enrollment. During the study, 136 patients with hip or knee OA were recruited by participating physiotherapists. Of these 136 patients, 110 patients participated in the study; 23 patients were not eligible after screening by phone; during the final screening, an additional 3 patients were excluded. A total of 395 patients contacted the research team after publication of articles in local newspapers, of which 205 patients were interested in participating after receiving oral information. Of these 205 patients, 97 were excluded after screening by phone and after reading written information and consulting the GP. Of the remaining 108 patients, an additional 18 patients were excluded during the final screening, leading to a total of 90 patients of the NP group participating in the trial.

In summary, 81% of the recruited patients by physiotherapists appeared to fulfill the inclusion criteria, compared to 23% of the patients who contacted the research group after reading local newspaper articles.

The 13-week followup measurement was completed by 192 patients. Reasons for withdrawal and loss to followup were surgery of patient (n=3), comorbidity (n=2), loss of motivation (n=1), family circumstances (n=1), and death of patient (n=1). At 65 weeks, 179 patients completed the followup. Reasons for loss to followup were loss of motivation (n=9), comorbidity (n=5), family circumstances (n=2), moving house of patient (n=2), costs of treatment (n=1), adverse effects of treatment (n=1), and death of patient (n=1).

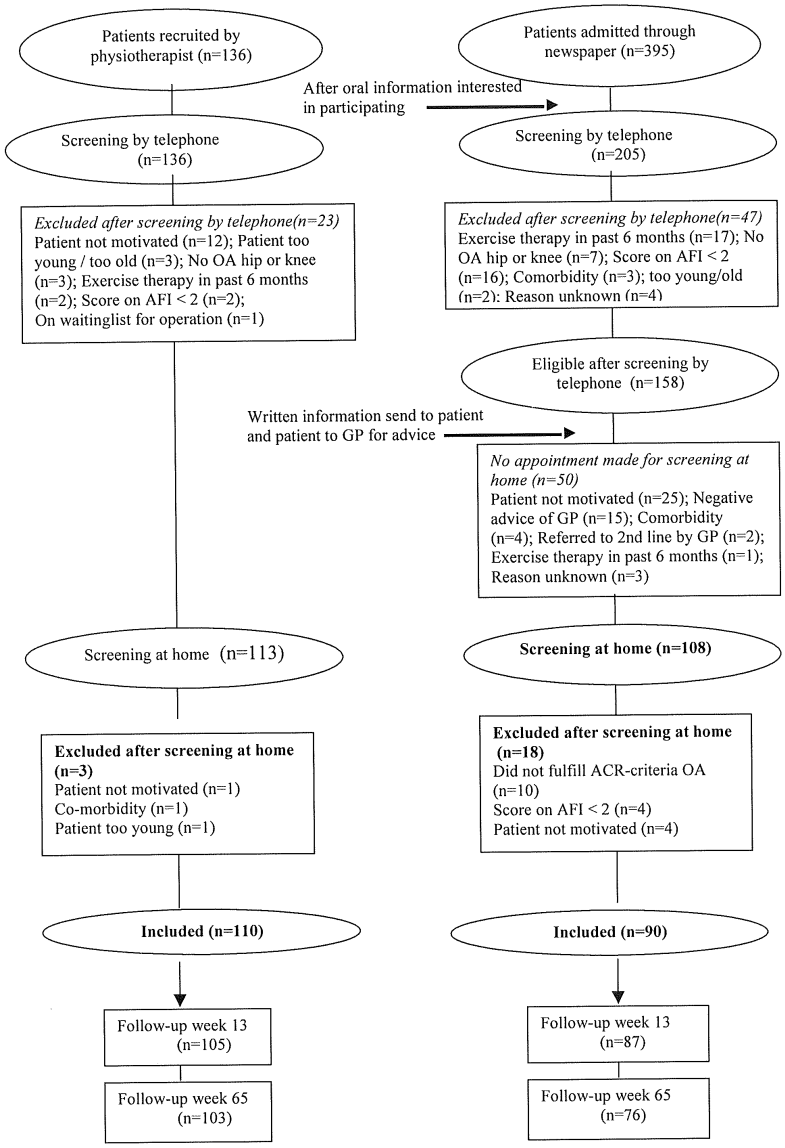

Figure 1 Flow chart of inclusion-procedure for each recruitment strategy.

Demographic and clinical data

The demographic and clinical characteristics of the patients in the 2 recruitment groups differed in 4 variables (Table 1). The NP group consisted of more patients with both knee and hip OA and of more patients with 2 affected sides of knee-OA. Furthermore the NP group reported a higher duration of symptoms of both the knee and hip. No differences, however, were found in radiological evidence of OA. Finally, level of education was higher in the NP group.

Impairments in body functions and limitations of activities

The severity of the most important symptoms, namely pain, stiffness and tiredness, and the restrictions in physical function of the patient are presented in Table 2. The PT group reported more pain and tiredness compared with the NP group. No significant differences were found in ROM, muscle strength of the hip and knee, or pain section of AFI (data not presented). Concerning physical function, no differences were found in either WOMAC score, AFI, or in time spent on physical activity, although the NP group performed the 5-meter walking test faster than the PT group.

Table 1 Demographic and clinical characteristics of the study population at baseline.

Characteristics	**Physiotherapist n=110**		**Newspaper n=90**		**p-value**
Gender: Female, n (%)	81	(74)	73	(81)	0.21
Age, mean (sd)	64.9	(7.9)	64.6	(7.9)	0.76
Location of OA, n (%)					
knee,	74	(67)	56	(62)	< 0.01
hip,	33	(30)	17	(19)	
both,	3	(3)	17	(19)	
Affected side OA, n (%)					
knee (n=150):					
one-side	49	(64)	34	(47)	0.04
two sides	28	(36)	39	(53)	
hip (n=70)					
one side	28	(78)	22	(65)	0.23
two sides	8	(22)	12	(35)	
Duration of complaints of knee (n=148), n (%)					
< 1 year	27	(36)	6	(8)	< 0.01
1 – 5 years	23	(30)	28	(39)	
> 5 years	26	(34)	38	(53)	
Duration of complaints of hip (n=70), n (%)					
< 1 year	13	(36)	2	(6)	< 0.01
1 – 5 years	14	(39)	16	(47)	
> 5 years	9	(25)	16	(47)	
Radiological evidence OA (Kellgren ≥ 2) (n=146)					
knee (n=101), n (%)	26	(52)	31	(61)	0.37
hip (n=51), n (%)	26	(90)	21	(96)	0.45
Comorbidity, n (%)					
no comorbidity	35	(33)	32	(36)	0.9
1 comorbidity	35	(33)	17	(19)	
≥ 2 comorbidities	37	(35)	39	(44)	
Body Mass Index, mean (sd)	28.9	(4.5)	28.0	(4.4)	0.17
Education: ≤ highschool, n (%)	85	(78)	48	(53)	< 0.01

Sd: standard deviation.

Table 2 Impairments in body functions and limitations of activities of the study population at baseline[1].

	Physiotherapist n=110		Newspaper n=90		p-value
Severity of pain					
at assessment (VAS)	4.5	(2.5)	3.4	(2.7)	0.04
past week (VAS)	5.8	(2.1)	5.3	(2.3)	0.14
subscale pain (WOMAC)	9.3	(3.0)	8.4	(3.4)	0.049
Stiffness (WOMAC)	3.9	(1.7)	3.9	(1.5)	0.91
Tiredness (VAS)					
at assessment	4.6	(2.4)	3.4	(2.6)	< 0.01
past week	5.3	(2.2)	4.7	(2.5)	0.08
Physical function					
subscale physical function (WOMAC)	32.5	(13.8)	31.5	(10.7)	0.59
5-m. walking time	5.0	(1.6)	4.5	(0.9)	< 0.01
Physical activity					
in min/week (SQUASH)	1748	(1170)	1665	(1019)	0.60

[1] Values are means with standard deviations between parenthesis.

Psychosocial variables

The psychosocial characteristics of the 2 recruitment groups were similar (Table 3). No differences were found in coping styles, social support, or quality of life (assessed with SF-36). One exception is a higher score of the PT group on the powerful-others scale of locus of control.

Table 3 Psychosocial variables of the study population at baseline[1].

	Physiotherapist n=110		Newspaper n=90		p-value
Pain Coping Inventory (PCI)					
Pain transformation	2.3	(0.5)	2.3	(0.6)	0.87
Distraction	2.4	(0.6)	2.4	(0.5)	0.99
Reduction demands	2.2	(0.5)	2.3	(0.5)	0.16
Retreating	1.7	(0.5)	1.8	(0.5)	0.21
Worrying	1.7	(0.4)	1.8	(0.5)	0.39
Resting	2.2	(0.5)	2.3	(0.5)	0.05
Locus of Control (MHLC)					
Internal health locus of control	21.3	(5.6)	20.1	(5.4)	0.12
Powerful others health locus of control	19.2	(5.6)	17.5	(5.1)	0.03
Chance health locus of control	19.4	(6.2)	19.2	(5.5)	0.79
Social support (SOS)	20.0	(8.5)	20.2	(8.6)	0.86

[1] Values are means with standard deviations between parenthesis.

Effect of treatment

Table 4 presents the results of the effectiveness of treatment after 13 weeks and 65 weeks. After 13 weeks, in both groups almost 40% of the patients rated themselves as improved on dichotomized GPE, with no differences between the groups (risk ratio 0.96, CI: 0.5 to 1.7). The physical function score (WOMAC) improved significantly within both groups (6.6 points in the PT group versus 5.2 points in the NP group). The difference between the PT group and NP group was not significant. The same pattern was observed for the scores on pain: statistically significant improvement within groups, but differences between groups were not significant. One exception was pain scored on the WOMAC scale. The PT group improved significantly more compared with the NP group. After adjusting for the baseline differences in pain (VAS at assessment), localization of OA, duration of symptoms, and the baseline score of the pain scale of WOMAC, all differences between the PT and NP groups were not significant.

The analyses of data at 39 and 65 weeks yielded similar results. The results of the effectiveness of treatment after 65 weeks are presented in Table 4. Compared with baseline, all outcome measures improved significantly within groups, but differences between groups were not found, with the exception of VAS pain at assessment. Again, the PT group improved more compared with the NP group, but after adjusting for baseline differences no significant differences were found.

Table 4 Primary outcome measures: differences between two recruitment groups after 13 and 65 weeks.

Outcome measure	PT-group Improvement	NP-group improvement	Difference between groupsNP - PT	Adjusted # difference between groups NP - PT
Week 13:	n=105	n=87		
Pain				
pain at assessment (VAS)	-0.75 (2.8)	-0.28 (2.7)	0.48 [-0.3; 1.3]	-0.09 [-0.8; 0.6]
pain past week (VAS)	-1.43 (2.6)	-0.90 (2.7)	0.53 [-0.2; 1.3]	0.42 [-0.2; 1.1]
subscale pain WOMAC	-2.70 (3.3)	-1.76 (3.0)	0.94 [0.03; 1.8]*	0.47 [-0.4; 1.3]
Physical function WOMAC	-6.60 (9.5)	-5.17 (9.9)	1.43 [-1.4; 4.4]	0.95 [-1.9; 3.8]
Global perceived effect††	40 (38%)	34 (39%)	0.96 [0.5; 1.7]	0.8 [0.4; 1.5]
Week 65:	n=103	n=76		
Pain				
pain at assessment (VAS)	-1.23 (3.2)	-0.18 (2.6)	1.04 [0.2; 1.9]*	0.16 [-0.6; 0.9]
pain past week (VAS)	-2.04 (2.9)	-1.72 (2.6)	0.32 [-0.5; 1.2]	-0.09 [-0.8; 0.6]
subscale pain WOMAC	-3.94 (3.8)	-3.00 (3.3)	0.93 [-0.1; 2.0]	0.22 [-0.8; 1.3]
Physical function WOMAC	-8.37 (12.9)	-6.83 (12.9)	1.53 [-2.4; 5.5]	0.12 [-3.6; 3.8]
Global perceived effect††	54 (54%)	37 (50%)	1.17 [0.6; 2.1]	0.92 [0.5; 1.8]

Change scores (posttest at 13 resp. 65 weeks – pretest at baseline) with standard deviations. Mean and adjusted difference between groups [95% CI]. Negative signs indicate improvement.

* Difference between groups, $p \leq 0.05$.

\# Adjusted for following baseline variables: duration of complaints, location of OA, VAS pain at assessment and baseline score on the specific outcome measure.

†† Percentage improved; calculated with risk ratio.

Discussion

Several strategies have been used to recruit patients for clinical trials involving patients with OA of the hip and/or knee. The present study is the first to evaluate the influence of recruitment methods (physiotherapist and newspaper) on the study population and the efficacy of interventions in patients with OA. Patients recruited by a physiotherapist reported more pain and tiredness, needed more time to walk 5 meters, and scored higher on the powerful others scale of locus of control. In contrast, the NP group was higher educated, consisted of more patients with multiple affected joints, and reported a higher duration of symptoms. Only small differences were found in the effectiveness of treatment, which disappeared after adjustment for baseline differences.

Different recruitment methods appear to attract different subjects. We hypothesized that patients responding to a newspaper have less severe symptoms and use a more active coping style compared with patients referred by a physiotherapist. As expected, the NP group reported less severe symptoms and limitations of the OA, although more joints were affected by OA. A possible explanation is that the NP group has learned to cope with the pain and limitations due to OA and are less inclined to consult GP's or physiotherapists for these symptoms. This conclusion is confirmed by the higher score of the PT group on the locus of control scale powerful others. No differences, however, were found in coping style.

As hypothesized, the differences between the 2 groups did not influence the effectiveness of the treatment on the primary outcome measures pain, physical function, and GPE. We only found a small difference in effect on the WOMAC pain scale (at 13 weeks) and VAS pain at assessment (at week 65), which disappeared after differences between both groups (pain at assessment (VAS), localization of OA, duration of symptoms and WOMAC pain baseline score) were taken into account while analyzing the effects of treatment. Our results confirm the findings of King et al. (1) on the effect of recruitment method on types of subjects entered into a primary prevention trial. King et al. found that there were few differences between recruitment sources for demographic variables, but these differences did not influence the effect of the intervention. In conclusion, our study included a more heterogeneous group of patients with

OA of the hip and/or knee by using different recruitment sources. The heterogeneity of the study population did not influence the efficacy of the intervention.

The question arises which is the most efficient way to recruit patients for trials in primary care concerning hip or knee OA. Some research has been done about the most successful methods for recruiting adults with OA, but the results seem to be contradictory. According to Spencer et al. (33) the most successful methods for enrolling participants with OA were through recruitment letters and television coverage. Others concluded that recruiting patients with OA through a local newspaper article was more efficient compared with recruiting participants via general practice (34). In contrast, Maurer et al. (35) found physician referrals from affiliated clinics more effective than, for example, advertisements. One has to deliberate the advantages and disadvantages of each method to choose the best strategy. In our trial, publishing articles in local newspapers was a more successful method for recruiting patients with OA, which is in line with the study of Davey et al. (34). In a period of 6 months, 90 patients were enrolled in the study, whereas the physiotherapists enrolled 110 patients in 18 months. However, to recruit 1 patient by newspaper took a lot more effort and time per patient from the investigators. This was due to the more extensive selection procedure for this group and the high number of ineligible persons who contacted the researchers. Only 23% of the persons who admitted through local newspapers participated in the study, as opposed to 81% of the patients who were recruited by physiotherapists. Differences in costs involved with the two strategies were not assessed. This may influence the choice of a specific recruitment method. Unfortunately, many clinical trials have to cope with a severe delay in patient inclusion, as described in Lasagna's law (36). The results of this study justify the use of a mix of recruitment strategies in studies on the effect of OA interventions in primary care. This can improve and quicken the inclusion of patients in clinical trials.

In conclusion, the results of our study suggest that a mix of recruitment strategies can be used to study the effectiveness of exercise therapy in patients with OA of the hip and/or knee. Each recruitment strategy attracts subjects from different segments of the target community, which increases the generalization of the results of a clinical trial.

References

1 King AC, Harris RB, Haskell WL. Effect of recruitment strategy on types of subjects entered into a primary prevention clinical trial. Ann Epidemiol 1994;4(4):312-20.

2. Dekker J, Tola P, Aufdemkampe G, Winckers M. Negative affect, pain and disability in osteoarthritic patients: the mediating role of muscle weakness. Behav Res Ther 1993;31:203-6.

3. McAlindon TE, Cooper C, Kirwan JR, Dieppe PA. Determinants of disability in osteoarthritis of the knee. Ann Rheum Dis 1993;52:258-62.

4. Madsen OR, Bliddal H, Egsmose C. Sylvest J. Isometric and isokinetic quadriceps strength in gonarthrosis; inter-relations between quadriceps trength, walking ability, radiology, subchondral bone density and pain. Clin Rheum 1995;14(3):308-14.

5. Steultjens MP, Dekker J, Van Baar ME, Oostendorp RA, Bijlsma JW. Muscle strength, pain and disability in patients with osteoarthritis. Clin Rehabil 2001;15(3):331-41.

6. Jordan JM, Luta G, Renner JB, Linder GF, Dragomir A, Hochberg MC, Fryer JG. Self-reported functional status in osteoarthritis of the knee in a rural southern community: the role of sociodemographic factors, obesity, and knee pain. Arthritis Care Res 1996;9(4):273-8.

7 Creamer P, Lethbridge-Cejku M, Hochberg MC. Factors associated with functional impairment in symptomatic knee osteoarthritis. Rheumatology 2000;39:490-6.

8. Creamer P, Lethbridge-Cejku M, Hochberg MC. Determinants of pain severity in knee osteoarthritis: Effect of demographic and psychosocial variables using 3 measures. The J Rheumatol 1999;26(8):1785-92.

9. Sullivan MJL, Stanish W, Waite H, Sullivan M, Tripp DA. Catastrophizing, pain and disability in patients with soft-tissue injuries. Pain 1998;77:253-60.

10. Evers AWM, Kraaimaat FW, Geenen R, Bijlsma JW. Psychosocial predictors of functional change in recently diagnosed rheumatoid arthritis patients. Behav Res Ther 1998;36:179-93.

11 Lankveld W van, Näring G, Pad-Bosch P van ‘t, Putte L van de. Behavioral coping and physical functioning: the effect of adjusting the level of activity on observed dexterity. J Rheumatol 1999;26:1058-64.

12. Bellamy N, Kirwan J, Boers M. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Consensus development at OMERACT 3. J Rheumatol 1997;24:799-802.

13. Van Baar ME, Dekker J, Bijl D, Oostendorp RAB, Voorn ThB, Bijlsma JWJ. Exercise therapy in patients with osteoarthritis of hip or knee: outcome in specific subgroups of patients. In: ME van Baar. Effectiveness of exercise therapy in osteoarthritis of hip or knee.Utrecht: NIVEL; 1998. p. 55-74.

14. Ettinger WH, Burns R, Messier SP, Applegate W, Rejeski WJ, Morgan T, Schumaker S, Berry MJ, O'Toole M, Monu J, Craven T. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. JAMA 1997;277:25-31.
15. Lindström I, Öhlund C, Eek C, Wallin L, Peterson LE, Fordyce WE, Nachemson AL. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. Phys Ther 1992;72:279-93.
16. Lindström I, Öhlund C, Eek C, Wallin L, Peterson LE, Nachemson AL. Mobility, strength, and fitness after a graded activity program for patients with subacute low back pain. A randomized prospective clinical study with a behavioral therapy approach. Spine 1992;17:641-52.
17. Vogels EMHM, Hendriks HJM, Baar ME van, et al. KNGF-richtlijn voor het fysiotherapeutisch handelen bij patiënten met artrose van heup en/of knie. Amersfoort: KNGF, 2001.
18. Altman R, Alarcón G, Appelrouth D, Bloch D, Borenstein D, Brandt K, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505-14.
19. Altman R, Asch E, Bloch D, Bole G, Borenstein D, Brandt K, et al. Development of criteria for the classification and reporting of osteoarthritis: classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039-49.
20 Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee. Scand J Rheumatol 1987;65:85-9.
21. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. Ann Rheum Dis 1957;16:494-502.
22. Ravaud P, Dougados M. Radiographic assessment in osteoarthritis. J Rheumatol 1997;24:786-91.
23. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt L. Validation study of WOMAC: a health status instrument for measuring clinically-important patient-relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orth Rheumatol 1988;1:95-108.
24. Norkin CC, White DJ. Measurement of joint motion: a guide to goniometry. Philadelphia: FA Davis Company; 1986.
25. Bohannon RW. Muscle strength testing with hand-held dynamometers. In: Admundsen LR, eds. Muscle strength testing: Instrumented and non-instrumented systems. New York: Churchill Livingstone; 1990: 69-88.
26. Heijden GJMG van der. Shoulder disorder treatment: Efficacy of ultrasoundtherapy and electrotherapy. Maastricht The Netherlands: University Press Maastricht; 1996.
27. Wendel-Vos GCW, Schuit AJ, Saris WHM, Kromhout D. Reproducibility and relative validity of the short questionnaire tot assess health enhancing physical activity. J Clin Epidemiol 2003;56(12):1163-9.

28. Kraaimaat FW, Bakker A, Evers AWM. Pijncoping-strategieën bij chronische pijnpatiënten: De ontwikkeling van de Pijn-Coping-Inventarisatielijst (PCI) (in dutch). Gedragstherapie 1997;30(3):185-201.
29. Wallston KA, Wallston BS, DeVellis R. Development of the Multidimensional Health Locus of Control (MHLC) Scales. Health Educ Manogr 1978;6(2):160-70.
30. Winefield HR. Reliability and Validity of the Health Locus of Control Scale. J Pers Assess 1982;46(6): 614-9.
31 Feij JA, Doorn CD, Kampen D van, Berg PT van den, Resing WCM. Sensation seeking and social support as moderators of the relationship between life events and physical illness / psychological distress. In: JAM Winnubst & S Maes, eds. Lifestyle stress and health. Leiden: DSWO Press; 1992. p.285-302.
32. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36) I. Conceptual framework and item selection. Med Care 1992;30:473-83.
33. Spencer AC, Kinne S, Belza BL, Ramsey S, Patrick DL. Recruiting adults with osteoarthritis into an aquatic exercise class: strategies for a statewide intervention. Arthritis Care Res 1998;11(6):455-62.
34. Davey R, Edwards SM, Cochrane T. Recruitment strategies for a clinical trial of community-based water therapy for osteoarthritis. Br J Gen Pract 2003;53:315-7.
35. Maurer BT, Moreno SI, Pickard AR, Wurst BE, Norden DK, Schumacher HR Jr. A comparison of recruitment methods for an osteoarthritis exercise study. Arthritis Care Res. 1995;8(3):161-6.
36. Cramer JA. Patient recruitment and compliance issues in clinical trials. In: JA French, MA Dichter & IE Leppik, eds. New antileptic drug development: Preclinical and clinical aspects. Amsterdam: Elsevier Science Publishers BV; 1993. p.211-22.
37. Steultjens MP, Dekker J, Bijlsma JW. Coping, pain and disability in osteoarthritis: a longitudinal study. J Rheumatol 2001;28(5):1068-72.

Chapter

7

Psychometric evaluation of osteoarthritis questionnaires: a systematic review of the literature

Published as:
Veenhof C, Bijlsma JWJ, Van den Ende CHM, van Dijk GM, Pisters MF, Dekker J. Psychometric evaluation of osteoarthritis questionnaires: a systematic review of the literature. Arthritis & Rheumatism (Arthritis Care & Research) 2006; 55(3):480-492.

Abstract

Objective. 1. To identify and describe available questionnaires designed to assess pain, physical function and patient global assessment in patients with osteoarthritis (OA) of hip and/or knee. 2. To evaluate evidence for the psychometric qualities of these questionnaires.
Methods. Instruments were identified through systematic literature searches. Information relating to the characteristics and validity, reproducibility, and responsiveness of the questionnaires was systematically extracted from published papers and rated according to a checklist of criteria.
Results. The 32 questionnaires that met the inclusion criteria could be divided in 24 condition-specific instruments, 7 generic instruments and 1 patient-specific scale. Most studies were found for the condition-specific WOMAC and generic SF-36. These instruments also demonstrated, overall, the highest ratings. None of the investigated questionnaires demonstrated satisfactory results for all qualities of the checklist.
Conclusion. Although the final choice of a questionnaire depends on the purpose of assessment, at this moment the WOMAC and SF-36 seem to be the most appropriate instruments to measure pain and physical function in patients with hip and/or knee OA. Therefore, the WOMAC and SF-36 could be recommended in guidelines concerning outcome measurement in OA trials, completed with a measure on patient global assessment.

Introduction

Both in clinical practice and in research on patients with osteoarthritis (OA), outcome is evaluated using many different instruments. The Outcome Measures in Rheumatology Clinical Trials (OMERACT) group defined a core set of outcome dimensions for clinical studies in hip and knee OA, which are pain, physical function (the performance of daily activities), and patient global assessment (1;2). These are in line with the recommendations of several guidelines for outcome measurement in OA trials (European League Against Rheumatism (EULAR) (3), Food and Drug Administration (FDA) (4), and Slow-acting Drugs in Osteoarthritis (SADOA) (5)). However, these guidelines differ in their recommendations of specific instruments or do not include recommendations of instruments at all (2;5).

Nowadays, a large number of instruments are available to assess the outcome dimensions of the OMERACT. The issue arises of which instruments are most appropriate to use. The selection of an instrument should depend on the instrument's psychometric qualities (namely, reproducibility, validity, and responsiveness), and on practical considerations (for example, time to complete, ease of scoring, and mode of administration).

Because the majority of the instruments developed for patients with OA are questionnaires, our focus in this article is on questionnaires. Several reviews of OA questionnaires have been published and recently, a special issue on outcome measurements was published by Arthritis Care & Research (2;6-8). Sun et al. concluded that both the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) and Lequesne Index are recommended as primary measures in treatment studies (8). However, none of these reviews give a complete systematic overview of all available instruments for patients with OA of the hip and/or knee. Before a specific core set of questionnaires can be recommended, a systematic comparison of the descriptive and psychometric qualities of the instruments is required. Recently, systematic reviews of measurement instruments for specific populations have been conducted (9;10).

The objective of this paper was to give an overview of published self-assessed (self-administered and interview-based) instruments on pain, physical function, and patient global assessment for patients with hip and/or

knee OA. To evaluate the selected questionnaires, data on the descriptive and psychometric qualities of the instruments were systematically collected and rated using a checklist (9;11). This overview will facilitate the choice of the most appropriate questionnaires to measure the OMERACT outcome dimensions in patients with OA of the hip and/or knee.

Materials and methods

Study selection

An extensive search was conducted in the Medline (1966 through May 2004), CINAHL (1982 through May 2004) and Embase (1988 through May 2004) databases. The broad computerized search strategy was built on asearch strategy for OA of the hip/knee; search strategy for outcome assessment; search strategy for the outcome dimensions pain, physical function, and patient global assessment; and search strategy for psychometric qualities. Furthermore, references of the retrieved articles were screened for relevant articles.

Inclusion of articles, based on the title and abstract, was decided by 2 independent reviewers (CV, CHME). In case of uncertainty, the full article was read by 2 independent reviewers (CV, MFP). If necessary, disagreements were resolved by a third reviewer (CHME). Inclusion criteria at the level of patients were as follows: patients with OA of the hip and/or knee; in case of surgical interventions, data were only included when collected before surgical interventions (e.g., total knee or total hip replacement) or before other invasive interventions, because patients after surgery were considered a different patient population. At the level of instruments, inclusion criteria were as follows: self-assessment (self-reported or interview-based) questionnaires; questionnaires that contained ≥1 separate dimensions of either pain, physical function, or global perceived effect; both condition-specific and generic questionnaires were included; in case of different language versions of the same questionnaire, only the English version or the native version was included. Finally, at the level of the performed studies, inclusion criteria were as follows: the main focus of the article was the development, construction or psychometric evaluation of the instrument (only psychometric evaluations using classical test theory were included, evaluations based on item response theory (IRT) were excluded;

no checklist is currently available to rate psychometric evaluations based on IRT); data of patients with hip and/or knee OA were published separately in case of mixed populations (e.g., patients with rheumatoid arthritis and patients with OA); results had been published in English as a full report.

Data extraction and quality assessment

A checklist of specific criteria for quality assessment of instruments was used, consisting of a section with descriptive aspects of the instrument and a section with specific psychometric criteria (Appendix 1). The checklist was developed by Bot et al. (11), based on the work of Lohr et al. (12) and the checklist developed by Bombardier and Tugwell (13). This list of criteria has already been used in several systematic reviews (9;10). All qualities were rated as either positive, doubtful, or negative. In case no or insufficient information was available on an aspect, no rating was given. The psychometric qualities of each study were independently assessed by 2 reviewers (CV, GMD). Disagreements between the reviewers were resolved by discussion. Because the present review focused on pain, physical function, and patient global assessment, only information on these outcome dimensions was rated in case other dimensions were also included in the instrument (e.g., quality of life, mental functioning). When ≥2 studies were performed on the same psychometric qualities of the same instrument, involving the same population (e.g., hip/knee OA or outpatients/inpatients), the highest rating was taken.

Characteristics of the instruments

The descriptive data provide information about the target population, scales, and format of the instruments. Extracted data included target population, domains to which the scales could be classified (pain, physical functioning, emotional functioning, social functioning, general health and quality of life), number of scales, number of items, response options, range of score, mode of administration (self-administered or interview-based), ease of scoring method, and time needed to complete the questionnaire. Three types of self-reported instruments were distinguished: generic scales, which are designed for various populations of patients; condition-specific questionnaires, designed for a specific group of patients; and patient-specific questionnaires, designed for use with individual patients (14).

Psychometric qualities

Data on the characteristics of the study group (diagnosis and clinical features) were reported to reflect for which population the psychometric qualities (validity, reproducibility, responsiveness and interpretability) were assessed.

Validity

Validity is the degree to which an instrument measures the construct it is intended to measure (12). The content validity, internal consistency and construct validity of the instruments were evaluated. The content validity examines the extent to which the items adequately represent all the significant aspects of the construct being measured (15). The rating of content validity was positive if patient consultation was combined with either expert consultation or examination of literature during item selection in the construction phase of the instrument, and was doubtful if only patients were consulted during item selection.

The internal consistency, determined by calculating Cronbach's alpha, indicates the homogeneity of the items in a (sub)scale. To determine which selected items cluster together around one aspect (and thus form a separate (sub)scale), factor analyses has to be performed. Questionnaires were rated positive if factor analysis was conducted and Cronbach's alpha for each dimension separately was > 0.70 (16).

Construct validity refers to the extent to which scores on a particular instrument relate to other assessment tools in a manner that is consistent with theoretically derived hypotheses (17). A positive rating was achieved when hypotheses about the magnitude and direction of relationships of the questionnaire (sub)scales with reference instruments were specified, and when >75% of these hypotheses could be confirmed. If available, descriptive data on the distribution of scores, including information about the presence of floor or ceiling effect, were extracted. Floor and ceiling effects were considered present if >15% of the respondents achieved the highest or lowest possible score (18).

Reproducibility

Reproducibility is the extent to which an instrument yields stable scores over time among respondents who are assumed not to have changed on the

domains being assessed. Reproducibility was assessed by rating reliability and agreement. Reliability is the degree to which an instrument is free of measurement error. The intra class correlation coefficient (ICC) for each (sub)scale is preferred to calculate reliability. The test-retest reliability and interobserver reliability were rated positive if ICC was >0.70 and >0.60, respectively (12).

To quantify measurement error and detect systematic differences between 2 measurements, a measure of agreement is calculated. The 95% limits of agreement according to Bland and Altman (19) and the standard error of measurement (SEM) or smallest detectable (real) difference (SDD) (20) were considered to be adequate measures for agreement. Because it was not possible to define adequate cutoff points for the agreement, agreement was rated as positive if one of the adequate measures was presented.

Responsiveness

Responsiveness is the ability of an instrument to detect real or important change over time in the concept being measured (21). Predefined hypotheses about the relation of change in the instrument to corresponding changes in reference instruments have to be postulated. Responsiveness was rated positive if these hypotheses were presented and if >75% of these hypotheses could be confirmed.

Interpretability

Interpretability is defined as the degree to which (change) scores can be interpreted and a qualitative meaning can be assigned to quantitative scores (12). A minimum clinically important difference (MCID) should be defined to interpret change scores in the target population. Other information that improves interpretation of the scores includes, for instance, presentation of means and standard deviations of patients' scores before and after treatment, data on the distribution of scores in relevant subgroups, and relating changes in the instrument score to patients' global perceived change. A positive rating was achieved when at least 2 types of information were presented.

Overall quality

To obtain an overall score of the instruments, we counted the number of positive ratings for each instrument.

Results

Selection of the studies

The search identified a total of 1930 publications. After screening titles and abstracts 1777 studies were excluded. Of the remaining 153 publications, 37 publications were included after reading the full article. Reasons for exclusion were data collection after operation or other invasive interventions (n=54); no separate data presented for patients with hip or knee OA (n=25); no psychometric evaluation of the instrument (n=16); no self-assessment instrument (n=9); no outcome dimensions of pain, physical function, or patient global assessment (n=6); no English version of the instrument (n=3); no use of classical test theory (n=3). A total of 32 questionnaires were included in study, which were divided into 24 condition-specific instruments (22-56), 7 generic questionnaires (23;24;30;32;33;47;52;55;57-59), and 1 patient-specific instrument (57;58). The full names of the investigated questionnaires are presented in Table 1. Actually, the 24 condition-specific questionnaires included different versions of the same instrument. Five different versions of both the WOMAC and Lequesne Index were investigated, and 3 different versions of the Arthritis Impact Measurement Scales (AIMS) were included. Three versions of the WOMAC differed in response options (visual analog scale (VAS), Likert scale, and numeric scale), and 1 version identified the most important items specific for the individual patient (signal version). The last WOMAC version differed in number of items (modified WOMAC), which also applied to the Lequesne Index hip and knee (modified Lequense Index) and the AIMS (AIMS2, AIMS2 Short Form (AIMS2-SF)). Finally, the Lequense Index versions varied in the mode of administration (interview-based or self-reported). Because not all descriptive information on a specific instrument was published in the article describing the psychometric study of patients with OA, original articles about the development of the instruments involving other patient populations were consulted (60-68).

Table 1 Full names of the questionnaires included

Abbreviation	Full name
WOMAC VA3.0	Western Ontario and McMaster Universities Osteoarthritis Index, Visual Analog Scale
WOMAC LK	Western Ontario and McMaster Universities Osteoarthritis Index, Likert Scale
Lequesne Index	Algofunctional indices for the hip and knee or index of severity for hip / knee disease
KOOS	Knee Injury and Osteoarthritis Outcome Score
HOOS	Hip Osteoarthritis Outcome Score
SMFA	Short Musculoskeletal Function Assessment Questionnaire
J-MAP	Joint-Specific Multidimensional Assessment of Pain
SF-36	MOS Short Form 36
HAQ	Health Assessment Questionnaire
AIMS	Arthritis Impact Measurement Scales
AIMS2-SF	Arthritis Impact Measurement Scales-Short Form
IRGL	Influence of Rheumatic Disease on General Health and Lifestyle
QR&S	Questionnaire Rising and Sitting down
ADL difficulty scale	Activities of Daily Living difficulty scale
ADL pain scale	Activities of Daily Living pain scale
VAS	Visual Analog Scale
NHP	Nottingham Health Profile
SIP	Sickness Impact Profile

Description of questionnaires

All included questionnaires and the descriptive items that were rated are presented in Table 2. Most questionnaires were developed to assess pain (n=22) and/or physical function (n=26) in separate (sub)scales. Only the Lequesne Index and the Patient Based Measure combine these 2 aspects in 1 single index.

Some questionnaires have separate versions for hip OA and knee OA, such as the Hip Osteoarthritis Outcome Score (HOOS) and Knee Injury and Osteoarthritis Outcome Score (KOOS), and Lequesne Index Hip and Lequesne Index Knee. With the exception of the Patient Based Measure and the Knee Pain Scale, which were developed specifically for patients with knee OA, all other questionnaires can be used for both hip OA and knee OA.

Of the condition-specific questionnaires, the AIMS2 had the largest number of items (n=78), followed by the Influence of Rheumatic Disease on General Health and Lifestyle questionnaire (IRGL; n=68), AIMS (n=52) and Short Musculoskeletal Function Assessment Questionnaire (SMFA; n=46),

whereas the Knee Pain Scale (n=6) and Joint-Specific Multidimensional Assessment of Pain (J-MAP; n=9) had the smallest number of items. Within the generic questionnaires, the number of items varied even more because the single pain questions (VAS and Likert) consisted of only 1 item and the Sickness Impact Profile (SIP) consisted of 136 items. The majority of questionnaires can be completed within 10 minutes of time.

Most studies involved patients with knee OA (n=18) or both knee OA and hip OA (n=16); only 3 studies included patients with only hip OA. Concerning the setting of the studies, 25 studies included patients from an outpatient setting (e.g., through general practitioner, hospital mailing), 8 studies described inpatients (mostly from a hospital or rehabilitation center), and 4 studies included both outpatients and inpatients.

Table 2 Description of the hip/knee OA questionnaires.

Questionnaire (references)	Target population*	Domains†	Number of scales‡	Number of items	Number of response options	Range of scores	Time to administer (minutes)	Mode of administration
Condition-specific questionnaires								
WOMAC VA3.0 (26-28;33;34;37)	Hip/knee OA	pain, other symptoms, physical function	3	24	0-100	P: 0-500 S: 0-200 PF: 0-1700	< 10 (paper) 10-15 (computer)	self-administered on paper and computer
WOMAC LK3.0 (26;29;30;32;38)	Hip/knee OA	pain, other symptoms, physical function	3	24	5	P: 0-20 S: 0-8 PF: 0-68	< 10	self-administered, paper and telephone
WOMAC numeric scale (22-24;53;54)	Hip/knee OA	pain, other symptoms, physical function	3	24	11	P: 0-50; S: 0-20 PF: 0-170	< 10 (paper) 10-15 (computer)	self-administered on paper and computer
WOMAC signal (25)	Hip/knee OA	pain, other symptoms, physical function	3	3			< 10	self-administered
WOMAC VA modified (36)	Hip/knee OA	pain, physical function	2	14	0-100	P: 0-500 PF: 0-900	6	self-administered
Lequesne Index knee (34;42;56)	Knee OA	pain, physical function	1	11	2-6	0-24	3-4	interview-based
Lequesne Index hip (42;56)	Hip OA	pain, physical function	1	11	2-6	0-24	3-4	interview-based

Table 2 continued

Questionnaire (references)	Target population*	Domains†	Number of scales‡	Number of items	Number of response options	Range of scores	Time to administer (minutes)	Mode of administration
Lequesne-knee self-reported (53)	Knee OA	pain, physical function	1	11	2-6	0-24	3-4	self-administered
Lequesne hip self-reported (53)	Knee OA	pain, physical function	1	11	2-6	0-24	3-4	self-administered
Lequesne modified (35)	Hip/knee OA	pain, physical function	1	10	2-6	0-23	3.25	interview-based
HOOS (40;44)	Hip OA	pain, physical function, quality of life	5	40	5	P: 0-40; Sy: 0-20; A: 0-68; SP: 0-16; Q: 0-16	7-10	self-administered
KOOS (49-51)	Knee OA	pain, physical function, quality of life	5	42	5	P: 0-36; Sy: 0-28; A: 0-68; SP: 0-20; Q: 0-16	10	self-administered
A patient based measure (31)	Knee OA	pain, physical function	1	12	2-6	0-100	?	self-administered
Knee pain scale (46)	Knee OA	pain	4	6	5/6	1-5 or 1-6	?	self-administered
SMFA (39)	musculoskeletal extremity disorders	physical function	2	46	5	PF: 34-170; bother:12-60	?	self-administered

Table 2 continued

Questionnaire (references)	Target population*	Domains†	Number of scales‡	Number of items	Number of response options	Range of scores	Time to administer (minutes)	Mode of administration
J-MAP (45)	patients with pain in any joint	pain	2	9	Varies	0-100	?	self-administered
SF-36 disease specific for physical function (PF) and role limitations (RP) (47)	every specific disease	physical function	2	14	2 or 3	0-100	< 10	self-administered
HAQ (30;55)	arthritic conditions	physical function	8	20	2 or 4	0-3	5-8	self-administered
AIMS (55)	arthritic conditions	pain, physical function, mental function, social function,	9	52	Varies	0-10	15-20	self-administered
AIMS2 (43)	arthritic conditions	pain, physical function, menstal function, social function	12	78	Varies	0-10	23	self-administered
AIMS2-SF (48)	arthritic conditions	pain, physical function, mental function, social function	5	23	5	0-10	?	self-administered
IRGL (52)	rheumatoid arthritis	physical function, mental and social function	11	68	4	Varies	20	self-administered

Table 2 continued

Questionnaire (references)	Target population*	Domains†	Number of scales‡	Number of items	Number of response options	Range of scores	Time to administer (minutes)	Mode of administration
ADL difficultly scale (58)	arthritic conditions	physical function	1	8	4	1-4	5	self-administered
ADL pain scale (58)	arthritic conditions	pain	1	8	4	1-4	5	self-administered
Patient-specific questionnaires								
Patient global assessment (57;58)	general population	physical function	N/A	N/A	Varies	Varies	?	self-administered
Generic questionnaires								
Single question pain (VAS) (57;58)	general population	pain	N/A	N/A	0-100	0-100	?	self-administered
Single question pain (Likert) (57)	general population	pain	N/A	N/A	5	0-4	?	self-administered
NHP (52;59)	general population	physical function, mental function, social function	6	38	2	0-100	< 10	self-administered

Table 2 continued

Questionnaire (references)	Target population*	Domains†	Number of scales‡	Number of items	Number of response options	Range of scores	Time to administer (minutes)	Mode of administration
SF-36 (23;24;30;32;33;47)	general population	pain, physical function, mental function, social function, general health	8	total 36 items	Varies	0-100	10	self-administered
SIP (55)	general population	physical function, mental function, social function	12	136	2	0-100	20-30	self-administered
QR&S (52)	general population	physical function	2	32	2	0-10	?	self-administered

* Population for which the questionnaire has been developed; †domains: pain, other symptoms, physical functioning, emotional functioning, and social functioning;‡scales: a subscore within a questionnaire.
P: Subscale Pain; S: Subscale Stiffness; PF: Subscale Physical Function; Sy: Subscale Symptoms; A: Activity Limitations-Daily Living; SP: Activity Limitations-Sport and Recreation; Q: quality of life (hip or knee related); RP: role limitations physical; N/A: not available; ?: no information found

Psychometric qualities

The rating of the psychometric qualities of the hip/knee OA questionnaires is presented in Table 3, summarizing each aspect as good, doubtful, or poor quality. An empty spot indicates no or insufficient information about an aspect. Because most results of psychometric qualities are dependent on the population studied, the kind of population is presented in the table (a distinction is made between outpatients and inpatients, and between hip OA and knee OA). None of the questionnaires in this review have been adequately tested on all psychometric qualities of the checklist in patients with hip and/or knee OA.

Content validity

Almost all instruments scored positively on content validity, meaning that patients and investigators or experts were involved during the development of the questionnaire. Only one instrument, the Patient Based Measure, was scored negatively on content validity because consultation of patients in the development of the questionnaire was not reported.

Internal consistency

For a positive rating of the internal consistency, information was needed on the construct of the questionnaire (investigated by factor analysis) and on the Cronbach's alpha of each (sub)scale. Information on both aspects was available for 6 of the 32 questionnaires, of which only 4 questionnaires had a positive rating (Patient Based Measure, Knee Pain Scale, J-MAP, and AIMS2-SF). The other 2 questionnaires, WOMAC LK (Likert scale) and HOOS, were rated as doubtful. For the WOMAC LK the a priori dimensions could not be confirmed by factor analysis. The dimensions of the HOOS could be supported, with the exception of the subscale 'activity limitations-daily living' which loaded as 2 factors. Besides this, the activity subscale had a Cronbach's $\alpha > 0.95$.

The dimensionality of 3 other questionnaires was studied by factor analysis, but no information was available on the Cronbach's alpha of the subscales. The construct of the WOMAC VAS (3 subscales: pain, stiffness, and physical function), and modified WOMAC (pain and physical function subscales) could not be confirmed. A 2-factor solution for the Lequesne Index was found, while the Lequesne Index claims to measure a single

construct (69).

In 9 instruments, information on internal consistency was restricted to information on Cronbach's alpha only, which ranged from 0.70 to 0.96. The exceptions were the HAQ and the subscale 'role limitations' of the MOS Short Form 36 (SF-36) which had a Cronbach's α <0.70.

Construct validity

Only 7 of 26 studies (for knee OA: WOMAC VA3.0, modified WOMAC, modified Lequesne Index, KOOS, disease-specific SF-36 and SF-36; for hip OA: HOOS) that investigated construct validity presented hypotheses relating to the magnitude and direction of expected correlations with other instruments, which is a condition for a positive rating according to the criteria of the checklist.

The correlations between most (subscales of) questionnaires measuring pain were moderate (r = 0.40-0.70). Concerning physical function, the Lequesne Index, the WOMAC physical function subscale, the physical function subscale of the SF-36, and SMFA had high correlations (r>0.7) with each other. These results also apply to the HOOS and KOOS because the physical function subscales of the HOOS and KOOS are equal to the WOMAC physical function subscale. Of 10 questionnaires, floor and/or ceiling effects were investigated, mainly for outpatients with knee OA. No floor or ceiling effects were found, with the exception of some subscales (e.g., sports and recreation subscale of the KOOS, which showed a floor effect for outpatients with knee OA).

Reproducibility

Information on test-retest reliability was found for 13 questionnaires. Because of low ICC's (<0.70), low sample size (<50), or the use of other correlation measures than ICC, only 3out of the 13 questionnaires had a positive rating for reliability. The modified WOMAC and modified Lequesne Index appeared to be reliable questionnaires for patients with knee OA, whereas the HOOS was reliable for patients with hip OA (ICC between 0.78 and 0.95).

Table 3 Summary of the quality assessment of the included questionnaires

Questionnaire	Time to administer	Ease of scoring	Readability and compre-hension	Content validity	Internal consis-tency	Construct validity
Condition-specific questionnaires						
WOMAC VA3.0 (26-28;33;34;37)	+§ ; - $	o	+	+	o (a)	+ (a^1,b^1) o (a^2,b^2)
WOMAC LK3.0 (26;29;30;32;38)	+	+		+	- (b) o (a)	o (a,b)
WOMAC numeric scale (22-24;53;54)	+§ ; -$	+		+[3]	o (a)	o (a)
WOMAC signal (25)		o				
WOMAC VA modified (36)	+	o	+	+[3]		+ (a^1,b^1)
Lequesne Index knee (34;42;56)	+	+	+			o (a^1)
Lequesne Index hip (42;56)	+	+	+			o (a^2)
Lequesne-knee self-reported (53)	+	+			o (a^1)	o (a^1)
Lequesne hip self-reported (53)	+	+			o (a)	o (a^2)
Lequesne modified (35)	+	+	+			+ (a^1,b^1)
HOOS (40;44)	+	o		+	o (a^2)	+ (b^2)
KOOS (49-51)	+	o		+	o (a^1)	+ (a^1)
A patient based measure (31)		-		-	+ (a^1)	o (a^1)
Knee pain scale (46)		o		+ (a^1)	+ (a^1)	o (a^1)
SMFA (39)		o	+	+		o (b^1)
J-MAP (45)		-		+	+ (b^1)	o (b^1)

Questionnaire	Floor / ceiling effect	Reliability	Agree-ment	Respon-siveness	Interpre-tability	MCID	Positively rated qualities, no.
Condition-specific questionnaires							
WOMAC VA3.0 (26-28;33;34;37)	+ (a^1) ‡	o (a, b^1)	+ (a)	o (a,b)	+ (a)	+ (a)	8
WOMAC LK3.0 (26;29;30;32;38)	+ (a^1,b^1)	o (a)	+ (a,b)	o (a,b)	o (a^1,b^1)		5
WOMAC numeric scale (22-24;53;54)	+ (b‡)	o (a)	+ (b)	o (a,b)	+ (b)	+ (b)	7
WOMAC signal (25)				o (a^1)	o (a^1)		0
WOMAC VA modified (36)		+ (a^1,b^1)					5
Lequesne Index knee (34;42;56)		o (a^1)					3
Lequesne Index hip (42;56)		o (a^2)					3
Lequesne-knee self-reported (53)		o (a^1)					2
Lequesne hip self-reported (53)		o (a^2)					2
Lequesne modified (35)		+ (a^1,b^1)					5
HOOS (40;44)	+ (b^2)	+ (a^2)					5
KOOS (49-51)	+ (a^1); - (a^1)†	o (a^1)			+ (a^1)		5
A patient based measure (31)	+ (a^1)						2
Knee pain scale (46)		o (a^1)					2
SMFA (39)	+ (a)						3
J-MAP (45)							2

Table 3 continued

Questionnaire	Time to administer	Ease of scoring	Readability and comprehension	Content validity	Internal consistency	Construct validity
SF-36 disease specific for physical function (PF) and role limitations (47)	+				o (a^1)	+ (a^1)
HAQ (30;55)	+	o	+	+	o (a^1,b^1)	o (a^1,b^1)
AIMS (55)	-	-	+	+		o (a)
AIMS2 (43)	-	-	+	o	o (a)	
AIMS2-SF (48)		-		o	+ (a)	
IRGL (52)	-	+				
ADL difficultly scale (58)		o				o (a^1)
ADL pain scale (58)		o				o (a^1)
Patient-specific questionnaires						
Patient global assessment (57;58)	+	+				- (a^1)
Generic questionnaires						
Single question pain (VAS) (57;58)	+	o				o (a^1)
Single question pain (Likert) (57)	+	+				o (a^1)
NHP (52;59)	+	-		+		
SF-36 (23;24;30;32;33;47)	+	-			o (a^1)	+ (a^1); o (a^2)
SIP (55)	-	o		+		o (a)
QR&S (52)		o		+		

Ratings: + positive, o doubtful, - negative; (a) outpatients; (b) inpatients; [1] only knee OA; [2] only hip OA; [3] assessed from WOMAC VA3.0; § version on paper;$ version on computer; ‡ floor effect; ‡‡ ceiling effect; [9] all scales, except subscales reach and activities (floor effect: -) ; [8] all scales, except subscales physical function and role limitations (floor effect: -); ° subscale mobility; † floor effect for subscale Sports and Recreation

Questionnaire	Floor / ceiling effect	Reliability	Agree-ment	Respon-siveness	Interpre-tability	MCID	Positively rated qualities, no.
SF-36 disease specific for physical function and role limitations (47)							2
HAQ (30;55)	+ (a^1,b^1)[¶]			o (a^1,b^1)	o (a^1,b^1)		4
AIMS (55)				o (a)	o (a)		2
AIMS2 (43)					o (a)		2
AIMS2-SF (48)	+				+ (a)		3
IRGL (52)				o (a^o)	o (a^o)		1
ADL difficultly scale (58)				o (a^1)	o (a^1)		0
ADL painscale (58)				o (a^1)	o (a^1)		0
Patient-specific questionnaires							
Patient global assessment (57;58)				o (a^1)	+ (a^1)		3
Generic questionnaires							
Single question pain (VAS) (57;58)				o (a^1)	+ (a^1)		2
Single question pain (Likert) (57)				o (a^1)	+ (a^1)		3
NHP (52;59)		o (a^2)		o (a^o)	o (a^o)		2
SF-36 (23;24;30;32;33;47)	+ (a^1)[a]; + (b‡‡)		+ (b)	o (a, b)	+ (a, b),	+ (b)	6
SIP (55)				o (a)	o (a)		1
QR&S (52)				o (a)	o (a)		1

Information on agreement was available for 4 instruments (WOMAC VA3.0, WOMAC LK, WOMAC numeric scale, and SF-36). Either the standardized error of measurement (SEM) or smallest detectable difference (SDD) were presented.

Responsiveness

The responsiveness was investigated for 16 questionnaires. None of these studies presented hypotheses relating to the magnitude of change and/or relationships with change scores of other instruments. Therefore, all questionnaires were rated as doubtful on responsiveness. Responsiveness was quantified as either effect sizes, relative efficiency or standardized mean scores. Change scores were also calculated, and correlations with change scores of other instruments were presented. Some studies compared the responsiveness of ≥2 questionnaires. In general, the WOMAC appeared to be more responsive compared with the SF-36, in both patients with hip OA and those with knee OA(23;24;33;57). In patients with knee OA, the responsiveness of the SF-36 appeared to be comparable with the HAQ (30), just as the AIMS was as responsive as the SIP in patients with hip OA and knee OA (55).

Interpretability and MCID

Eight questionnaires were rated positive on interpretability by presenting at least 2 out of the 4 types of information. Only 1 study (that of the AIMS2-SF) intentionally paid attention to the interpretability of scores by comparing the scores on the AIMS2-SF in groups of patients that differed in duration of disease, number of comorbidities and general health perception (48). The MCID was calculated for 2 questionnaires, the WOMAC numeric scale and SF-36. Of 13 questionnaires, means and standard deviations of baseline and follow-up scores or scores of relevant subgroups were presented.

Overall score

After counting the total number of positive ratings for each instrument, the WOMAC VA3.0, WOMAC numeric scale, modified Lequesne Index, HOOS, and KOOS had the highest overall scores among the condition-specific instruments, with 8, 7, 5, 5, and 5 positive ratings, respectively. Concerning the generic questionnaires, the SF-36 obtained the highest overall

score, with 6 positive ratings.

Discussion

An extensive search strategy led to the identification of 32 self-assessment questionnaires for the evaluation of pain and physical functioning and patient global assessment in patients with OA of the hip and/or knee, for which descriptive and psychometric qualities had been investigated. Most questionnaires were condition-specific (n=24); the remainder were generic (n=7) and patient-specific (n=1) instruments. Twenty-two instruments were developed to rate pain; physical function was rated by 26 instruments. Concerning patient global assessment, only 1 instrument was found. Most studies included patients with knee OA (n=18) or both knee and hip OA (n=16); only 3 studies included patients with only hip OA. Many psychometric qualities were not properly tested for a large number of questionnaires, and none of the questionnaires were rated positive on all aspects of the checklist.

Overall, the condition-specific instruments (WOMAC, VAS version; Lequesne Index hip / knee; and HOOS / KOOS) had the best ratings for their descriptive and psychometric qualities for both pain and physical function. The WOMAC has been the most extensively studied instrument and received the best ratings for its descriptive and psychometric qualities. One should keep in mind that some instruments (such as the HOOS and KOOS) have not been studied extensively or have only been studied in other populations; rating of these instruments might improve when more studies have been conducted on their psychometric qualities. Concerning generic instruments, the SF-36 has been studied most often and demonstrated, overall, the highest ratings. The psychometric qualities of the patient-specific instrument on patient global assessment has been studied to a limited degree in patients with hip and/or knee OA. Therefore, only a small number of quality criteria could be rated. The same accounted for the single questions pain on VAS and Likert scale, which were investigated in a small number of studies.

To compare the results of trials and optimize the transparency of care, a core set of questionnaires in patients with hip and/or knee OA seems to be indicated. For example, a core set of qualified questionnaires will

facilitate the comparison and interpretation of the outcome of various treatment modalities in OA. At this moment, guidelines for outcome dimensions in OA trials, such as OMERACT, EULAR, FDA, and SADOA, differ in their recommendations of instruments or do not include recommendations at all (2). Our results suggest that, at this time, the most appropriate questionnaires to use in patients with hip and/ or knee OA seem to be the condition-specific questionnaire WOMAC and the generic questionnaire SF-36. Therefore, it is recommended that these questionnaires, completed with a patient-specific instrument on patient global assessment, are included in guidelines as a core set of instruments in patients with OA of hip and/or knee. However, more research is needed on the psychometric qualities of patient global assessment measures before making a choice of the most appropriate instrument in patients with hip and/or knee OA.

Nonetheless, which scale is most appropriate to use always depends on the particular purpose of the assessment. For example, for discriminative purposes, the instrument should have satisfactory ratings for reproducibility and agreement. The modified WOMAC, modified Lequesne Index, and HOOS were the only questionnaires with a positive rating for test-retest reliability. Alternatively, when the purpose is to evaluate changes over time, an instrument should have positive ratings for responsiveness and no floor or ceiling effects. Currently, all questionnaires were rated equally, namely, were rated doubtful, on responsiveness. In general, the condition-specific instrument WOMAC appeared to be more responsive compared with the generic SF-36. Both in daily practice and in clinical trials, changes over time are frequently evaluated: therefore, responsiveness of the instrument is an important condition in selecting a questionnaire. The low ratings on responsiveness are remarkable, and more solid research on responsiveness is needed.

The dimensionality of only 9 questionnaires was tested using factor analysis. When the dimensionality of a questionnaire has not been analyzed, the internal consistency as reflected by Cronbach's alpha might not be interpretable (69). The theoretical dimensional structure of only 4 questionnaires could be confirmed (namely, Patient Based Measure, Knee Pain Scale, J-MAP, and AIMS2-SF). The factor analysis of the other 5 instruments (WOMAC VA3.0, WOMAC LK, modified WOMAC, Lequesne Index, and HOOS) yielded either more dimensions or less dimensions than a

priori stated; therefore, internal consistency was rated as doubtful. It needs to be considered that only studies based on the classical test theory were included in the present review. Item response theory (IRT) also provides a model to evaluate health status questionnaires. In total, 3 studies on the WOMAC were excluded from this review because Rasch analyses were performed.

Some limitations of this study have to be mentioned. First, some consideration is recommended when generalizing the results. After all, we excluded studies of patients after operations (e.g, total hip replacement and total knee replacement) or other invasive interventions. Furthermore, because it is uncertain whether psychometric qualities of translated versions can be generalized to the original version, we only included the English version (or, in the absence of an English version, the native version) of the questionnaires. In total, 20 studies (concerning the WOMAC (n=12), KOOS (n=2), AIMS (n=2), NHP (n=2), Lequesne (n=1), and VAS (n=1)) were excluded because non-English version were evaluated. The results of this review are only applicable to the included populations and questionnaires. Second, the criteria we used to evaluate the quality of the instruments were helpful to provide information on the practical and psychometric properties in order to facilitate the choice between questionnaires. However, in our opinion, there is room for improvement of the checklist. First, no instructions are given on how to determine the overall best instrument. We counted the number of positive ratings to make an overall judgement of the instruments, which implies that all different qualities are equally important. Second, the criteria for construct validity and responsiveness to postulate specific hypotheses can be questioned. The absence of hypotheses in the publication might be due to a shortcoming of the author instead of a lower psychometric quality of the instrument. In constrast, the need for clearly defined objective cutoff points to rate construct validity and responsiveness is high. Strikingly, the authors of all studies that investigated construct validity and responsiveness of questionnaires concluded that the questionnaires were valid and responsive instruments for patients with hip or knee OA. The present criteria on postulating hypotheses are a first step towards clearly defining these objective cutoff points. Furthermore, as suggested by Bot et al. (9), authors can contribute to a good rating of questionnaires by clearly presenting the results of the studies they performed. The checklist, as used in

the present review, might be a good tool for authors to check whether their results are systematically and unambiguously presented.

In conclusion, although the final choice of a questionnaire depends on the purpose of the assessment, the WOMAC VA3.0 and SF-36 currently demonstrated the highest ratings overall for both descriptive and psychometric qualities. Therefore, these questionnaires are recommended for evaluating pain and physical function in patients with hip and/or knee OA. Completed with a measure on patient global assessment, these instruments could be recommended in guidelines concerning outcome measurement in OA trials.

References

1. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997; 24(4):799-802.
2. Bellamy N. Osteoarthritis clinical trials: candidate variables and clinimetric properties. J Rheumatol 1997; 24(4):768-778.
3. Guidelines for the clinical investigation of drugs used in rheumatic diseases, European drug guidelines, Series 5. World Health Organization, Regional Office for Europe. Copenhagen: European League Against Rheumatism. 1985.
4. Lequesne M, Brandt K, Bellamy N, Moskowitz R, Menkes CJ, Pelletier JP et al. Guidelines for testing slow acting drugs in osteoarthritis. J Rheumatol Suppl 1994; 41:65-71.
5. Bellamy N. Outcome measurement in osteoarthritis clinical trials. J Rheumatol 22[43S], 49-51. 1995.
6. Garratt AM, Brealey S, Gillespie WJ. Patient-assessed health instruments for the knee: a structured review. Rheumatology (Oxford) 2004; 43(11):1414-1423.
7. Rogers JC, Irrgang JJ. Measures of adult lower extremity function: The American Academy of Orthopedic Surgeons Lower Limb Questionnaire, the Activities of Daily Living Scale of the Knee Outcome Survey (ADLS), Foot Function Index (FFI), Functional Assessment System (FAS), Harris Hip Score (HHS), Index of Severity for Hip Osteoarthritis (ISH), Index of Severity of Knee Osteoarthritis (ISK), Knee Injury and Osteoarthritis Outcome Score (KOOS), and Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Arthr Care & Res 2003; 49(S5):S67-S84.
8. Sun Y, Sturmer T, Gunther KP, Brenner H. Reliability and validity of clinical outcome measurements of osteoarthritis of the hip and knee--a review of the literature. Clin Rheumatol 1997; 16(2):185-198.
9. Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. Ann Rheum Dis 2004; 63(4):335-341.
10. de Boer MR, Moll AC, de Vet HC, Terwee CB, Volker-Dieben HJ, van Rens GH. Psychometric properties of vision-related quality of life questionnaires: a systematic review. Ophthalmic Physiol Opt 2004; 24(4):257-273.
11. Bot SDM, Terwee CB, van der Windt DAWM, Bouter LM, Dekker J, de Vet HC. Psychometric evaluation of self-report questionnaires - the development of a checklist. In: Ader HJ, Mellenbergh GJ, editors. Proceedings of the second workshop on research methodology 25-27 June 2003. Amsterdam: VU University Amsterdam, 2003: 161-168.

12. Lohr KN, Aaronson NK, Alonso J, Burnam MA, Patrick DL, Perrin EB et al. Evaluating quality-of-life and health status instruments: development of scientific review criteria. Clin Ther 1996; 18(5):979-992.
13. Bombardier C, Tugwell P. Methodological considerations in functional assessment. J Rheumatol 1987; 14(S15):6-10.
14. Binkley J. Measurement of functional status, progress, and outcome in orthopaedic clinical practice. Orthopaedic Practice 1999; 11(1):14-21.
15. Guyatt G, Feeny D, Patrick D. Issues in quality-of-life measurement in clinical trials. Control Clin Trials 1991; 12(4 Suppl):81S-90S.
16. Nunnally J. Psychometric theory. 2nd ed. ed. New York: McGraw-Hill, 1978.
17. Kirshner B, Guyatt G. A methodological framework for assessing health indices. J Chronic Dis 1985; 38(1):27-36.
18. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? Qual Life Res 1995; 4(4):293-307.
19. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet 1986; 1(8476):307-310.
20. Beckerman H, Roebroeck ME, Lankhorst GJ, Becher JG, Bezemer PD, Verbeek AL. Smallest real difference, a link between reproducibility and responsiveness. Qual Life Res 2001; 10(7):571-578.
21. Terwee CB, Dekker FW, Wiersinga WM, Prummel MF, Bossuyt PM. On assessing responsiveness of health-related quality of life instruments: guidelines for instrument evaluation. Qual Life Res 2003; 12(4):349-362.
22. Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. J Rheumatol 2002; 29(1):131-138.
23. Angst F, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. Ann Rheum Dis 2001; 60(9):834-840.
24. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Care Res 2001; 45:384-391.
25. Barr S, Bellamy N, Buchanan WW, Chalmers A, Ford PM, Kean WF et al. A comparative study of signal versus aggregate methods of outcome measurement based on the WOMAC Osteoarthritis Index. Western Ontario and McMaster Universities Osteoarthritis Index. J Rheumatol 1994; 21(11):2106-2112.
26. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988; 15(12):1833-1840.

27. Bellamy N, Kean WF., Buchanan WW, Gerecz-Simon E, Campbell J. Double blind randomized controlled trial of sodium meclofenamate (Meclomen) and diclofenac sodium (Voltaren): post validation reapplication of the WOMAC Osteoarthritis Index. J Rheumatol 1992; 19(1):153-159.
28. Bellamy N, Campbell J, Stevens J, Pilch L, Stewart C, Mahmood Z. Validation study of a computerized version of the Western Ontario and McMaster Universities VA3.0 Osteoarthritis Index. J Rheumatol 1997; 24(12):2413-2415.
29. Bellamy N, Campbell J, Hill J, Band P. A comparative study of telephone versus onsite completion of the WOMAC 3.0 osteoarthritis index. J Rheumatol 2002; 29(4):783-786.
30. Brazier JE, Harper R, Munro J, Walters SJ, Snaith ML. Generic and condition-specific outcome measures for people with osteoarthritis of the knee. Rheumatology (Oxford) 1999; 38(9):870-877.
31. Clark JA, Spiro A, III, Fincke G, Miller DR, Kazis LE. Symptom severity of osteoarthritis of the knee: a patient-based measure developed in the veterans health study. J Gerontol A Biol Sci Med Sci 1998; 53(5):M351-M360.
32. Davey RC, Edwards SM, Cochrane T. Test-retest reliability of lower extremity functional and self-reported measures in elderly with osteoarthritis. Advances in Physiotherapy 2003; . 5(4):155-160.
33. Davies GM, Watson DJ, Bellamy N. Comparison of the responsiveness and relative effect size of the western Ontario and McMaster Universities Osteoarthritis Index and the short-form Medical Outcomes Study Survey in a randomized, clinical trial of osteoarthritis patients. Arthritis Care Res 1999; 12(3):172-179.
34. Faucher M, Poiraudeau S, Lefevre-Colau MM, Rannou F, Fermanian J, Revel M. Algo-functional assessment of knee osteoarthritis: comparison of the test-retest reliability and construct validity of the WOMAC and Lequesne indexes. Osteoarthritis Cartilage 2002; 10(8):602-610.
35. Faucher M, Poiraudeau S, Lefevre-Colau MM, annou F, ermanian J, evel M. Assessment of the test-retest reliability and construct validity of a modified Lequesne index in knee osteoarthritis. Joint Bone Spine 2003; 70(6):521-525.
36. Faucher M, Poiraudeau S, Lefevre-Colau MM, Rannou F, Fermanian J, Revel M. Assessment of the test-retest reliability and construct validity of a modified WOMAC index in knee osteoarthritis. Joint Bone Spine 2004; 71(2):121-127.
37. Feeny D, Blanchard CM, Mahon JL, Bourne R, Rorabeck C, Stitt L et al. The stability of utility scores: test-retest reliability and the interpretation of utility scores in elective total hip arthroplasty. Qual Life Res 2004; 13(1):15-22.
38. Kennedy D, Stratford PW, Pagura SMC, Wessel J, Gollish JD, Woodhouse LJ. Exploring the factorial validity and clinical interpretability of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Physiother Can 2003; 55(3):160-168.

39. Kirschner S, Walther M, Bohm D, Matzer M, Heesen T, Faller H et al. German short musculoskeletal function assessment questionnaire (SMFA-D): comparison with the SF-36 and WOMAC in a prospective evaluation in patients with primary osteoarthritis undergoing total knee arthroplasty. Rheumatol Int 2003; 23(1):15-20.
40. Klassbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. Scand J Rheumatol 2003; 32(1):46-51.
41. Lequesne M. Indices of severity and disease activity for osteoarthritis. Semin Arthritis Rheum 1991; 20(6 Suppl 2):48-54.
42. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee. Validation--value in comparison with other assessment tests. Scand J Rheumatol Suppl 1987; 65:85-89.
43. Meenan RF, Mason J.H., Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. Arthritis Rheum 1992; 35(1):1-10.
44. Nilsdotter AK, Lohmander LS, Klassbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)--validity and responsiveness in total hip replacement. BMC Musculoskelet Disord 2003; 4(1):10.
45. O'Malley KJ, Suarez-Almazor M, Aniol J, Richardson P, Kuykendall DH, Moseley JB, Jr. et al. Joint-specific multidimensional assessment of pain (J-MAP): factor structure, reliability, validity, and responsiveness in patients with knee osteoarthritis. J Rheumatol 2003; 30(3):534-543.
46. Rejeski WJ, Ettinger WH, Jr., Shumaker S, Heuser MD, James P, Monu J et al. The evaluation of pain in patients with knee osteoarthritis: the knee pain scale. J Rheumatol 1995; 22(6):1124-1129.
47. Ren XS, Kazis L, Lee A, Miller DR, Clark JA, Skinner K et al. Comparing generic and disease-specific measures of physical and role functioning: results from the Veterans Health Study. Med Care 1998; 36(2):155-166.
48. Ren XS, Kazis L, Meenan RF. Short-form Arthritis Impact Measurement Scales 2: tests of reliability and validity among patients with osteoarthritis. Arthritis Care Res 1999; 12(3):163-171.
49. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998; 28(2):88-96.
50. Roos EM, Roos HP, Lohmander LS. WOMAC Osteoarthritis Index--additional dimensions for use in subjects with post-traumatic osteoarthritis of the knee. Western Ontario and MacMaster Universities. Osteoarthritis Cartilage 1999; 7(2):216-221.
51. Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 2003; 1(1):17.

52. Steultjens MP, Roorda LD, Dekker J, Bijlsma JW. Responsiveness of observational and self-report methods for assessing disability in mobility in patients with osteoarthritis. Arthritis Rheum 2001; 45(1):56-61.
53. Stucki G, Sangha O, Stucki S, Michel BA, Tyndall A, Dick W et al. Comparison of the WOMAC (Western Ontario and McMaster Universities) osteoarthritis index and a self-report format of the self-administered Lequesne-Algofunctional index in patients with knee and hip osteoarthritis. Osteoarthritis Cartilage 1998; 6(2):79-86.
54. Theiler R, Spielberger J, Bischoff HA, Bellamy N, Huber J, Kroesen S. Clinical evaluation of the WOMAC 3.0 OA Index in numeric rating scale format using a computerized touch screen version. Osteoarthritis Cartilage 2002; 10(6):479-481.
55. Weinberger M, Samsa GP, Tierney WM, Belyea MJ, Hiner SL. Generic versus disease specific health status measures: comparing the sickness impact profile and the arthritis impact measurement scales. J Rheumatol 1992; 19(4):543-546.
56. Lequesne MG, Samson M. Indices of severity in osteoarthritis for weight bearing joints. Journal of Rheumatology 1991; . 18(SUPPL. 27):16-18.
57. Bolognese JA, Schnitzer TJ, Ehrich EW. Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. Osteoarthritis Cartilage 2003; 11(7):499-507.
58. Brooks R.H., Callahan L.F., Pincus T. Use of self-report activities of daily living questionnaires in osteoarthritis. Arthritis Care Res 1988; 1(1):23-32.
59. Hunt SM, McKenna SP, Williams J. Reliability of a population survey tool for measuring perceived health problems: a study of patients with osteoarthrosis. J Epidemiol Community Health 1981; 35(4):297-300.
60. Bergner M., Bobbitt R.A., Carter W.B., Gilson B.S. The Sickness Impact Profile: Development and final revision of a health status measure. Med.Care 19[8], 787-805. 1981.
61. Callahan L.F., Brooks R.H., Summey J.A., Pincus T. Quantitative pain assessment for routine care of rheumatoid arthritis patients, using a pain scale based on activities of daily living and a visual analog pain scale. Arthr & Rheum 30[6], 630-636. 1987.
62. de Bruin A.F., de Witte L.P., Stevens F., Diederiks J.P.M. Sickness Impact Profile: The state of the art of a generic functional status measure. Soc.Sci.Med. 35[8], 1003-1014. 1992.
63. Fries F. The assessment of disability: from first to future principles. Br J Rheum 22S, 48-58. 1983.
64. Huiskes C.J.A.E., Kraaimaat F.W., Bijlsma JW. Development of a self-report questionnaire to assess the impact of rheumatic diseases on health and lifestyle. J Rehab Sc 3[3], 65-70. 1990.
65. Huiskes CJAE, Kraaimaat FW, Bijlsma JW. De ontwikkeling van de IRGL: Een instrument om gezondheid te meten bij patiënten met reuma. Gedrag en Gezondh 18[2], 78-89. 1990.

66. Meenan R.F., Gertman P.M., Mason J.H. Measuring health status in arthritis: The Arthritis Impact Measurement Scales. Arthr & Rheum 23[2], 146-152. 1980.
67. Roorda LD, Roebroeck ME, Lankhorst GJ, van Tilburg T, Bouter LM. Measuring functional limitations in rising and sitting down: development of a questionnaire. Arch Phys Med Rehabil 77, 663-669. 1996.
68. Swiontkowski MF, Engelberg R, Martin DP, Agel J. Short musculoskeletal function assessment questionnaire: validity, reliability, and responsiveness. J Bone Joint Surg Am 1999; 81(9):1245-1260.
69. Cortina JM. What is coefficient alpha? An examination of theory and applications. J Appl Psychol 78, 98-104. 2004.

Appendix 1 Checklist for rating the psychometric quality of self-assessment questionnaires (9;11).

Psychometric quality	Definition	Criteria used to rate the psychometric quality
Time to administer	Time needed to complete the questionnaire	Rating: [+]less than 10 minutes [-]more than 10 minutes [ ] no information found on time to administer
Ease of scoring	Ease of method used to calculate the questionnaire's score	Rating: [+] easy: summing up of the items [o] moderate: visual analogue scale (VAS) or simple formula [-] difficult: VAS in combination with formula or complex formula [ ] no information found on calculation of score
Readability & comprehension	The questionnaire is understandable for all patients	Rating: [+] readability tested; result was good [-] inadequate readability [ ] no information found on readability and comprehension
Content validity	The extent to which the domain of interest is comprehensively sampled by the items in the questionnaire	1) patients were involved during item selection and/or item reduction 2) patients were consulted for reading and comprehension Rating: [+] patients and investigator /expert involved [o] patients only [-] no patient involvement [ ] no information found on content validity
Internal consistency	The extent to which items in a (sub)scale are intercorrelated; a measure of the homogeneity of a (sub)scale	1) Factor analysis was applied in order to provide empirical support for the dimensionality of the questionnaire. 2) Cronbach's alpha above 0.70 for each dimension/subscale. Rating: [+] adequate design & method; factor analysis supporting the dimension; alpha > 0.70. [o] doubtful method used or no factor analysis [-] inadequate internal consistency (alpha < 0.70) or dimensions not supported by factor analysis [ ] no information found on internal consistency

Appendix 1 continued

Psychometric quality	Definition	Criteria used to rate the psychometric quality
Construct validity	The extent to which scores on the questionnaire relate to other measures in a manner that is consistent with theoretically derived hypothesis concerning the domains that are measured	1) hypotheses were formulated 2) results were acceptable in accordance with ≥ 75% of hypotheses 3) an adequate measure was used Rating: [+] adequate design, method, and result [o] doubtful method used [-] adequate design & method and inadequate construct validity [ ] no information found on construct validity
Floor & ceiling effects	The questionnaire fails to demonstrate a worse score in patients clinically deteriorated and an improved score in patients who clinically improved	1) descriptive statistics of the distribution of scores were presented 2) ≤15% of respondents achieved the highest or lowest possible score Rating: [+] no floor / ceiling effects [-] >15% in extremities [ ] no information found on floor and ceiling effects
Test-retest reliability	The extent to which the same results are obtained on repeated administrations of the same questionnaire when no change in physical functioning has occurred	1) calculation of an intraclass correlation coefficient (ICC); ICC>0.70 2) time interval and confidence intervals (or n>50) were presented Rating: [+] adequate design, method, and ICC>0.70 [o] doubtful method [-] inadequate reliability, with adequate design and method [ ] no information found on test-retest reliability
Agreement	The ability to produce exactly the same scores with repeated measurements	1) for evaluative questionnaires reliability agreement should be assessed 2) limits of agreement, Kappa, or standard error of measurement (SEM) was presented Rating: [+] adequate design, method and result [o] doubtful method used [-] inadequate agreement, with adequate design and method [ ] no information found on agreement

Chapter

8

General Discussion

The main focus of the current thesis is to assess the effectiveness of a behavioral graded activity program in patients with osteoarthritis (OA) of the hip and/or knee. For this purpose, we performed a randomized controlled trial to compare behavioral graded activity (BGA) to usual care (UC: physiotherapy, mainly exercise therapy, according to the Dutch osteoarthritis guideline for physiotherapists (1)) in patients with OA. This final chapter describes and discusses the main results. First, the main results of the performed clinical trial will be presented and discussed, and attention is paid to some limitations of the study. Second, the main results of a systematic review on the psychometric qualities of osteoarthritis questionnaires will be described and discussed. Finally, the implications for clinical practice will be considered and recommendations are made for future research.

Randomized controlled trial on the effectiveness of BGA

Main findings

Effectiveness of behavioral graded activity

Beneficial effects were found for both treatments (BGA and UC), both in the short term and in the long term. The differences between BGA and UC in improvement for pain (VAS, WOMAC), physical function (WOMAC) and patient global assessment were small and not significant. This pattern was similar for the scores on the secondary outcome measures, with the exception of the patient oriented physical function (MACTAR) and 5 meter walking time test. For these outcome measures significant differences were found in favor of BGA at week 65. Furthermore, BGA patients reported to adhere significantly more to the home exercises and activities.

Cost-effectiveness

Concerning the costs of both treatments, no significant differences were found in both direct and indirect costs between BGA and UC. However, there is a slight indication that BGA may lead to less hospitalization and less work absenteeism than UC. Moreover, the incremental cost-effectiveness utility ratio indicated that BGA was not cost-effective.

On basis of the results of the randomized trial, it can be concluded that the superiority of BGA over UC, in general, has not been demonstrated.

Which patients benefit most?
By means of subgroup analyses it was investigated whether BGA has particular benefit in specific subgroups of patients. It appeared that patients with a relatively low level of physical functioning at the start of treatment and, to a lesser extent, patients with a low level of internal locus of control benefit more from BGA than UC. Also, limited evidence was found that patients with mild OA (no radiological evidence) have more benefit from UC and patients with obesity have more benefit from BGA. Additionally, on basis of our qualitative study, it appeared that patients with initial motives for long term goals seemed to be more adherent to their exercises compared to patients with initial motives for short term goals. Furthermore, an active involvement of patients during the treatment process seemed to relate to a higher adherence to perform activities in the long term.

*Recruitment strategie*s
In the randomized clinical trial 2 different methods were used to recruit patients: referrals by physiotherapists and invitations by newspaper articles. The influence of recruitment methods on the study population and the efficacy of interventions was evaluated. Patients recruited by physiotherapists reported more pain and tiredness, although less joints were affected with OA and the duration of complaints was lower compared to patients recruited by newspaper articles. Furthermore, patients recruited by physiotherapists needed more time to walk 5 meter and scored higher on the powerful-others scale of locus of control. After adjusting for baseline differences the effect of treatment was comparable for both recruitment groups, both in the short term and in the long term. In conclusion, a mix of recruitment strategies seems allowable, on the condition that corrections for baseline differences are included in the analyses.

Clinical relevance
The question remains if the significant effects, as found in our trial, are also clinical relevant for patients. Although both interventions in our trial were about equally effective, the overall improvement within groups was clinically relevant for all primary outcome measures. Angst et al. (2) concluded that, in patients with OA, changes over 12% compared to baseline, in pain and physical function, can be regarded as minimal clinically

important differences. Tubach et al. (3) stated that the minimal clinically important improvement for relative changes (changes compared to baseline) were between 32% and 40% for pain (VAS) and between 21% and 26% for physical function (WOMAC). In our study we found changes in pain after 65 weeks of 43% (BGA) and 37% (UC), and changes in physical function of 25% (both groups). Moreover, 56% of the BGA patients and 49% of the UC patients reported, after 65 weeks, to be improved on the patient global assessment. This measure is often used as indicator for clinically relevant differences.

Concerning the statistically significant differences between both interventions, it can be concluded that especially the differences as assessed on the outcome measure MACTAR seem to be clinically relevant. Unfortunately, no research has been done on the minimal clinical important difference / improvement of the MACTAR, but in our opinion a difference of 46% (6.02 for BGA compared to 3.27 for UC) can be interpreted as a clinically relevant difference. The clinical important difference of the walking time test is less convincing (an improvement of 9% compared to baseline in BGA patients, compared to 3% in UC patients).

When comparing the effects of our trial with other studies on the effectiveness of exercise therapy in patients with OA of the hip and/or knee, several conclusions can be drawn. First, it can be concluded that both treatments resulted in beneficial effects in the short term which even increased in long term, while in other trials the beneficial short term effects decreased over time (4-9). Although further comparison is biased by common problems (e.g., different outcome measures, different presentation of data, different groups of patients), within group effect sizes are calculated to compare the amount of effect for both short term and long term. Concerning short term effectiveness, the within group beneficial effects on pain seem to be relatively low in our trial compared to other trials. The effect sizes on pain vary from 0.30 (pain VAS) to 1.0 (pain, WOMAC) in both BGA and UC, compared to 1.2 (10), 0.6 and 2.1 (7) in other trials on exercise therapy. The within group beneficial short term effects on physical function, however, are relatively large in our trial (effect size of 0.7 in both groups) compared to effect sizes of 0.3 (10), and 0.6 (7) in other trials. Since the effect sizes of our trial on both pain and physical function increase in the long term (effect size pain: 0.5 (VAS) and 1.6 (WOMAC); effect size on

physical function: 0.9 (WOMAC)) and effect sizes of other trials decrease in time, the long term effectiveness, within-groups, of our trial is relatively large compared to other trials (effect sizes pain: 0.8, 0,2 and 1,6; effect sizes on physical function: 0.3, 0.0, 0.3) (7;9).

Several RCT's have been conducted on the effectiveness of behavioral graded activity in musculoskeletal patients. So far, limited evidence has been found on the effectiveness of BGA. More specifically, graded activity programmes have been shown to improve the level of daily activities, reduce disability, and reduce duration of sick leave in patients with back pain (11;12). On the other hand, other studies reported no significant effects on pain and functional status (13;14). Concerning patients with chronic shoulder complaints, BGA was more effective in restoring performance of daily activities, although the beneficial effects were small (15). Two remarks need to be made when these results are compared to the results of our study. First, up till now, the effectiveness of BGA has been investigated in patients with non-specific symptoms, without underlying pathologies, such as low back pain and shoulder complaints (11;13-16). However, OA is a chronic progressive musculoskeletal disorder, and caution is called for since overloading the joint might cause damage to the joint and exacerbation of the complaints. Such a progressive disorder demands specific adaptations of the treatment BGA. For example, the activities need to be increased gradually and skillfully, the physiotherapists need to be aware of active inflammatory processes and a stopping rule is included in the treatment. This stopping rule states that the gradual increase of activities has to be interrupted when an active inflammatory process is suspected or diagnosed. Hereafter, the increase of activities starts at a lower level. In case of recurrent inflammatory processes, the treatment goal needs to be changed and the rate of increasing activities needs to be decelerated. The present study was the first study investigating the effectiveness of BGA in patients with a progressive and specific chronic disease. Despite the adaptations of BGA, which diminish the risk of joint damage and/or exacerbation of complaints, the results within the BGA group were comparable to the results of other BGA studies.

The second remark on the results of our study compared to other BGA studies, is that most other studies made a comparison between BGA and usual care, which was defined as care given by general practitioners or

occupational physicians (14;15). In our study, just as in the study of Ostelo et al. (13), usual care was defined as usual care as given by physiotherapists. Considering that BGA was also given by physiotherapists, it can be concluded that the contrast between both treatments in our trial was lower compared to the other studies on the effectiveness of BGA.

Contrast between Behavioral Graded Activity and Usual Care

During the analysis of our trial we had to conclude that the contrast between the 2 interventions was less than expected. First, we expected that BGA patients would be treated at least 5 sessions more compared to the UC group, since BGA patients received additional boostersessions after the first 12 weeks of treatment. However, the average number of sessions was 14.1 of the BGA group versus 11.7 of the UC group (which was still a significant difference, $p<0.01$).

Second, the contrast between both interventions was smaller since not all BGA patients were treated according to the protocol. Specific elements of BGA were described in the registration forms, and baseline was set for only 70% of the BGA patients, and an exercise protocol was only made for 84% of the BGA patients. This might indicate that also other specific elements of BGA, such as reinforcing feedback and extinction of pain behavior (not registered) may not be adequately executed. The physiotherapists registered all treatments on registration forms. On basis of these forms it could be evaluated if patients were treated according to the BGA protocol. However, to get complete insight into the content of the treatment, recordings of sessions with audio- or videotapes are preferred. With these recordings it can be controlled to which extent physiotherapists give a treatment according to the BGA protocol. The fact that not all patients were treated according to the protocol can be explained by the novelty of BGA; physiotherapists had to get used to the behavioral approach. To increase their experience with BGA, all participating physiotherapists had to perform the BGA treatment on one “test” patient with chronic pain before including patients to the trial. After the treatment of the test patient, the treatment was evaluated with the physiotherapist on the basis of their experiences, the forms, charts and barriers of the physiotherapist. However, a variable number of, patients varying from 1 tot 15, were included by the physiotherapists. Therefore, it is understandable that some physiotherapists

did not completely master all aspects of a behavioral treatment like BGA. Moreover, as was suggested by King et al. (17), a 2 day-course is too short to completely master the skills which are necessary to treat patients according to a behavioral graded activity protocol.

Third, less contrast was caused by the education of the UC physiotherapists. In the recent trial BGA was compared with usual care, which was operationalized as treatment according to the Dutch OA guideline for physiotherapists. This guideline was developed in 1998 and mailed to all Dutch physiotherapists. No specific additional training was given to these physiotherapists with regard to this guideline but treatment according to the guideline was recommended. In the recent trial we did offer the physiotherapists in the usual care group a short training on the guideline. This short intervention could have caused that the usual care treatment as performed in the trial is more based on the guideline compared to the usual care in practice. In this case, the contrast between both interventions (UC and BGA) is lower in the trial compared to daily practice. The expectation is that in daily practice more passive techniques such as massage, and electrotherapy are used since these modalities were more common before the guideline OA was developed.

In conclusion, the contrast between BGA and UC was already relatively low, since we compared BGA with usual care of physiotherapists (which is mainly exercise therapy). This contrast even decreased because not all BGA patients were treated according to protocol and UC in the trial was more active and functional compared to daily practice. Since the aim of this trial was to investigate the superiority of BGA compared to usual care of physiotherapists, we did not include a group without treatment. Including such group would increase the contrast with BGA, but would not answer the question which physiotherapeutic treatment is most effective and which treatment should be used by physiotherapists.

As a consequence of the pragmatic design of the trial, both patients and physiotherapists could not be blinded for the allocated treatment. Placebo, or Hawthorne (i.e., a significant positive effect which is apparently due to the effect on the participants of knowing themselves to be studied in connection with the outcomes measured) effects can therefore not be refuted. However, the research assistants, assessing the outcome measures were unaware of treatment allocation.

Usual Care

On basis of literature, we only expected beneficial effects in the short term and not in the long term from UC (6;8). Surprisingly, beneficial long term effects were found for both BGA and UC. These unexpected beneficial effects of UC can be explained by a shift within the field of physiotherapy. First, the shift from a more passive approach (e.g., massage therapy and physiotherapy modalities) to an active approach (exercise therapy, education), as advocated in the guidelines, is gaining more and more support in the physiotherapy profession (1;18-20). Within this active approach, the exercises become more functional and task oriented recently (21-23). This shift towards a more functional and active approach was confirmed by the comparison of treatment goals and modalities as registered in the study of Van Baar (9;10) and our trial. Since the UC in our trial was based on the exercise therapy of Van Baar, comparable goals and modalities were expected. However, in van Baar's study, exercise therapy was mainly directed towards improvement of muscle strength (93% of treatments), improvement of range of motion (85%) and reduction of pain (80%). While in our trial improvement of activities / decrease of limitations in activities (84% of treatments) was the most frequently mentioned intervention in the UC group. Possibly, physiotherapists who are willing to participate in research are precursors of a shift within the field of physiotherapy, which might have increased the contrast with van Baar's study. To confirm the shift in approach and to investigate in which way the approach within the field of physiotherapy has changed in the last ten years, more research is needed.

Subgroups of patients

Since OA patients are heterogeneous in their complaints and restrictions of activities, it is unlikely that all patients will benefit from the same treatment. Therefore, it is important to match the applied intervention with the patients' clinical presentation (24;25). Evidence was found for beneficial effects of BGA in subgroups of patients. Furthermore, results of our qualitative study indicated that the patients' initial motive plays an important role in the adherence of patients, just as the involvement of patients during the treatment process. These findings confirm the disablement models, which recognize the influence of both extra-individual factors (health care, external supports) and intra-individual factors (coping, lifestyle changes) on

consequences of disease (26;27). Furthermore, the role of the initial motive of patients emphasizes the importance of including activities which are valuable for individual patients. Apparently, patients are willing to accept pain related sensations when performing activities which are valuable to them. This is in line with the developments within the behavioral approach (28;29). For readiness to change, in this case to successfully realize a behavioral change towards a more active lifestyle, 2 aspects are of importance: the patients' perceived importance to change and the patients' beliefs about one's ability to engage in this behavioral change. The patients' perceived importance to change concerns the costs and benefits associated with behavior change. For example, regular exercise can be seen as an important strategy to use, if patients believe that exercise will help them enough to outweigh the potential problems associated with performing these exercises (e.g., boredom, increased pain or discomfort, fear) (28). Therefore, concerning the treatment BGA, it is essential for the success of the treatment that patients choose activities which are valuable for them individually.

Cost-effectiveness

Although no differences were found between the cost-effectiveness of both interventions, one remark needs to be made. All cost-effectiveness analyses were based on the primary outcome measures. Since no differences between both interventions were found on the primary outcome measures, the results of this cost-effectiveness study were not unexpected. However, there was a slight indication that BGA may lead to less hospitalization and less work absenteeism. The question remains if significant differences between both treatments would be found if the cost-effectiveness analyses were performed with the secondary outcome measures in which we found significant differences between both treatments (MACTAR and walking time test). Considering the fact that the power of most cost-effectiveness studies is too low to detect significant difference, which is a general problem with studies on cost-effectiveness, we do not expect that differences in cost-effectiveness would be found when analyzing other outcome measures.

Recruitment of patients

It appeared that publishing articles in local newspapers was a more successful method for recruiting patients with OA. In a period of 6 months,

90 patients enrolled in the study through the articles in local newspapers, while the physiotherapists enrolled 110 patients in 18 months. However, to recruit 1 patient by newspaper took a lot more effort and time per patient from the investigators. This was due to the more extensive selection procedure for this group and the high number of ineligible persons who contacted the researchers. Since the majority of persons admitted through local newspapers did not meet the inclusion criteria, or decided not to participate after receiving information, only 23% of them participated in the study (against 81% of the patients who were recruited by physiotherapists).

In our study it was concluded that a mix of recruitment strategies does not seem to modify treatment outcome, on the condition that baseline differences are adjusted for. However, the use of 2 recruitment strategies did introduce a confounding factor, which was included in all analyses of the trial. This is partly in line with the conclusions of Geraets et al. (30) who recently compared patients with chronic shoulder complaints recruited by general practitioners (GP) with patients recruited by advertisement in a local newspaper. Both recruitment groups were similar in terms of most baseline characteristics, although improvements on outcome measures were greater in patients recruited by GP's. However, Geraets et al. concluded that clinical effectiveness was not modified by the recruitment strategy.

Based on these 2 studies, a mix of recruitment strategies should be considered as a solution for a delay in inclusion of patients, on the condition that such mix is considered as a putative confounder or modifying factor.

Outcome measures

The outcome measures of this study were chosen according to the core set of outcome measures, as defined by the OMERACT, namely pain, physical function and patient global assessment (31). According to our findings of the performed systematic review on outcome measures in patients with OA of the hip and/or knee, we assessed pain by VAS and WOMAC and physical function by WOMAC. However, we also included a more patient oriented measure on physical function, MACTAR (32).

Since the main objective of behavioral graded activity is to increase the patients' level of daily activities, an outcome measure assessing the performed level of patients' daily life activities should be included in studies on the effectiveness of BGA. However, most common measures on

physical function are too generic to record changes in activities which are different and specific for each individual patient. For example, the dimension physical function of the WOMAC consists of 17 various items, like walking, walking on stairs, gardening and household working. If only 1 or 2 out of the 17 WOMAC items are valuable activities for the individual patient, the responsiveness of the total dimension physical function will not be sufficient to record changes in these 2 activities. Therefore, in our opinion a more patient specific outcome measure should be included in studies on the effectiveness of behavioral graded activity. Such measure corresponds better with the main values of the individual patients. The patient oriented MACTAR (patients report up to 5 activities which bother them most in daily life) assesses the change over time in 5 activities which are valuable to the individual patient. Although, at this moment, no studies exist on the psychometric evaluation of the MACTAR in patients with OA of the hip and/or knee, we do recommend to include a measure like the MACTAR in studies on the effectiveness of BGA. Based on the literature, the validity and responsiveness of patient specific measures seem to be rather high, which seems to be achieved at the expense of a relatively low reliability (33).

Systematic review on osteoarthritis questionnaires

Main results

A systematic review was conducted to identify and describe available questionnaires designed to assess pain, physical function and patient global assessment in patient with OA of the hip and/or knee. Furthermore, evidence for the psychometric qualities of these questionnaires was evaluated. Most studies were found for the condition specific WOMAC and generic SF-36. These instruments demonstrated, overall, the highest ratings and seem to be the most appropriate instruments to measure pain and physical function.

Remarks

Besides the fact that an enormous amount of questionnaires were found in the domain of pain and physical function, several remarks need to be made. First, the highest ratings were found for instruments which were investigated most. Since recently developed instruments like HOOS and

KOOS were not yet thoroughly investigated, it is possible that the ratings of these instruments will increase in time. Second, in our review we have used an existing checklist (34) and we have chosen to count the number of positive ratings to make an overall judgment of the instruments, which implies that all qualities are equally important. Nevertheless, one can imagine that certain aspects (e.g., internal consistency, construct validity and reproducibility) are more important aspects of the psychometric quality compared to other aspects (e.g., interpretability, content validity). Finally, especially the concepts of validity and responsiveness were rated low. These low ratings were partly due to the absence of tested hypotheses. Therefore, higher ratings can be achieved by a proper and complete presentation of the methods and results by the authors. The checklist can be used as a guide for a proper presentation of psychometric research (34).

Performance based measures

Recently, Terwee et al. (35) performed a systematic review on performance based methods for measuring physical function of patients with OA of the hip or knee. The results of this review can be used in combination with our review on the OA questionnaires. These authors concluded that data on reproducibility and validity are mostly lacking and that many more well designed clinimetric or psychometric studies are needed. Furthermore, they concluded that consensus is needed on what activities should be included in a performance based test for patients with OA and which aspects of function should be measured (e.g., time, quality of movement) before a justified choice can be made of the most useful performance based instrument.

Implications for clinical practice

Based on the results of our trial, it can be concluded that both BGA and UC have beneficial and clinically relevant effects in the long term. At this moment, both treatments are equally beneficial for patients with OA of the hip and/or knee. The question arises what will be the role of BGA in the future management of OA.

Future of BGA

Although at this moment the superiority of BGA compared to usual care in general has not been demonstrated, we have the opinion that BGA is valuable for most patients with OA of the hip and/or knee and needs to be added as one of the options in the care of patients with OA. This point of view can be motivated by several findings.

First, to achieve optimum results of treatment the heterogeneity of the patients should be taken into account. Therefore, more attention should be paid in general to the matching of the treatment with the individual patient. In our study several subgroups of patients, which particularly benefit from BGA, were identified. Based on the results of our study, BGA might be the first treatment of choice for patients with a lower level of physical function and for patients with a lower level of internal locus of control. Since we are aware that more research needs to be done on this topic, we do not recommend to adapt the Dutch OA guideline for physiotherapists. Nevertheless, we do recommend to emphasize the beneficial effects of BGA more explicitly. Furthermore, we do advise to implement BGA for patients with OA, on the condition that physiotherapists have been educated in the required skills for delivering the treatment. However, considering the shift in approach (i.e. from a pain-contingent approach towards a time-contingent approach), it is questionable whether all physiotherapists can master the required specific skills and, subsequently, can deliver BGA properly.

Second, the success of a treatment depends on both the beneficial effects and on the opinions of the patients and physiotherapists. To investigate the opinions of the patients and physiotherapists about BGA, we conducted two qualitative studies. One qualitative study investigated the experiences (including benefits and barriers) of participating physiotherapists with the treatment BGA. The other qualitative study was described in chapter 5 and investigated, besides the factors influencing the success of treatment, the experiences of the patients with BGA. It appeared that almost all physiotherapists had positive experiences with the treatment. Especially the structural time-contingent way to gradually increase the level of activities, the coaching role of the physiotherapists and the responsibilities of the patients were appreciated most. On the other hand, BGA was reported to be labor intensive, mainly because of all the paperwork (e.g., graphs). These positive experiences were confirmed by the patients,

most patients were content about the treatment and would advise BGA to their family and/or friends. These findings confirm the value of a treatment like BGA.

Furthermore, based on the findings of our study, some improvement can be made to the BGA treatment, which possibly will lead to an increase in beneficial effects. For example, as was confirmed in our qualitative study, the initial motivation of patients is of great importance for the success of the treatment. In order to achieve a successful outcome of the intervention, physiotherapists are recommended to gain a clear understanding of the patients' initial motives for treatment and of the activities which are valuable for the individual patient. This is confirmed in the acceptance approach of McCracken et al. (29), assuming that the acceptance of chronic pain is an active willingness to engage in meaningful activities in life regardless of pain related sensations.

Education of the physiotherapists

Another important improvement is the education of the caregivers, i.e. physiotherapists, and based on the findings of our study, several recommendations can be made. Concerning the different approach of physiotherapists when applying BGA compared to UC (e.g., the shift from pain-contingent (UC) to time-contingent (BGA)), a 2-day education seems not sufficient to completely master all aspects of a behavioral treatment like BGA. On the other hand, physiotherapists themselves need to realize that BGA is a different approach compared to their usual care which needs some investment from their side. For example, physiotherapists need to be aware that they need to apply BGA to many patients to get to master all aspects of such behavioral approach.

Although the involvement of patients in the treatment process is already included in the education of physiotherapist, based on the finding of the qualitative study, it would be beneficial to emphasize the importance of an active patients' involvement right from the start of the treatment. This is in line with the literature on patient participation during goal setting and the influence of patients' involvement on the self-efficacy of patients (36;37).

Outcome measures

When performing studies on the effectiveness of behavioral treatments, researchers should include a primary outcome measure which is related to the main objective of the treatment. For this purpose, a patient oriented measure (e.g., the MACTAR) is, at this moment, one of the most appropriate questionnaires to use in behavioral graded activity studies. However, at this moment no information is available on the psychometric qualities of such measure in patients with OA. Future research should focus on this aspect.

Furthermore, as in other fields of research, the range of available measurements is very wide. A certain degree of standardization is needed for the comparison of results in research and in communication in clinical practice. Systematic reviews of the psychometric qualities of the available measurement instruments are a useful contribution to facilitate the selection of an adequate instrument (33). Additionally, researchers can contribute to a good rating of questionnaires by clearly presenting the results of their psychometric studies. The checklist, as used in our review, is a good tool for authors to check whether their results are systematically and unambiguously presented (34).

Recommendations for future research

Based on the findings of our study, several recommendations for future research can be made. First, to get a complete overview of the effectiveness of BGA compared to usual care, an extra follow up assessment would be useful. One of the explanations of the unexpected beneficial results of usual care after 65 weeks, was the small contrast between both treatments. This can be explained by the shift in the usual care of physiotherapists: an active and functional approach and an integration of behavioral aspects seem to gain more and more support in the usual care of the physiotherapists. However, the main difference between BGA and usual care remains the systematic and protocolized application of behavioral aspects in the BGA treatment, leading to an integration of a higher level of performed activities in the daily lives of patients. Therefore, we expect that this systematic application of behavioral aspects and time-contingent increase of the level of activities in patients' daily lives will lead to more beneficial effects in the

long term. Accordingly, we expect that the superiority of BGA (as was found on the MACTAR, 5 meter walking test and adherence rate, after 65 weeks), will be manifested even more after a longer time of follow up (i.e. 4,5 years after inclusion). Since BGA patients were motivated to gradually increase their level of activities, regardless of their pain, we mainly expect beneficial effects in outcome measures which are related to the increase in the level of activities, such as physical function, walking time test and the adherence to perform exercises/activities.

Second, concerning outcome measures, future research should focus on several aspects. To facilitate a certain degree of standardization, systematic reviews on the psychometric qualities of instruments need to be performed. Furthermore, to value the extent of the effectiveness of certain treatments, researchers should focus more on determining the minimal important clinical differences of instruments, i.e. the difference between both treatments which is valued as clinically relevant by the patients, of questionnaires. Finally, which is particularly important for studies involving patients with OA of the hip and/or knee, research should focus on the psychometric characteristics of the MACTAR.

Third, as mentioned before, one of the findings of our study was that UC, given by physiotherapists to patients with OA of the hip and/or knee, changed over time. In ten years time (between 1994 and 2004), a shift was perceived from more passive treatment towards a more active and more functional approach. This shift is probably not specific for patients with OA but can be perceived in the whole field of physiotherapy. Therefore, to confirm this shift, research should focus on the differences in treatment goals and interventions of patients within the physiotherapy.

Finally, during our trial it appeared that not all participating patients with OA tended to avoid physical activities. Besides this group, another group of patients could be distinguished (in the opinions of the participating physiotherapists), namely patients who tended to be too active and needed to be slowed down and regulated in their activities. It seems reasonable that both groups of patients need a different treatment approach. Therefore, in the future, research should not only focus on reactivating patients with OA, but also on finding the balance between performing not enough activities and performing too much activities.

References

1. Vogels EMHM, Hendriks HJM, van Baar ME, Dekker J, Hopman-Rock M, Oostendorp RA et al. Clinical practice guidelines for physical therapy in patients with osteoarthritis of the hip or knee. Amersfoort: KNGF, 2001.
2. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. Arthritis Rheum 2001; 45(4):384-391.
3. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N et al. Evaluation of clinically relevant changes in patient reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis 2005; 64(1):29-33.
4. Deyle GD, Henderson NE, Matekel RL, Ryder MG, Garber MB, Allison SC. Effectiveness of manual physical therapy and exercise in osteoarthritis of the knee. A randomized, controlled trial. Ann Intern Med JID - 0372351 2000; 132(3):173-181.
5. Ettinger WH, Burns R, Messier SP, Applegate W, Rejeski WJ, Morgan T et al. A randomized trial comparing aerobic exercise and resistance exercise with a health education program in older adults with knee osteoarthritis. The Fitness Arthritis and Seniors Trial (FAST). Journal of the American Medical Association 1997; 277(1):25-31.
6. Fransen M, McConnel S, Bell M. Therapeutic exercise for people with osteoarthritis of the hip or knee. A systematic review. J Rheumatol 2002; 29:1737-1745.
7. McCarthy CJ, Mills PM, Pullen R, Richardson G, Hawkins N, Roberts CR et al. Supplementation of a home-based exercise programme with a class-based programme for people with osteoarthritis of the knees: a randomised controlled trial and health economic analysis. Health Technol Assess 2004; 8(46):iii-61.
8. Van Baar ME, Assendelft WJ, Dekker J, Oostendorp RA, Bijlsma JW. Effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a systematic review of randomized clinical trials. Arthritis Rheum 1999; 42(7):1361-1369.
9. Van Baar ME, Dekker J, Oostendorp RA, Bijl D, Voorn TB, Bijlsma JW. Effectiveness of exercise in patients with osteoarthritis of hip or knee: nine months' follow up. Ann Rheum Dis 2001; 60(12):1123-1130.
10. Van Baar ME, Dekker J, Oostendorp RA, Bijl D, Voorn TB, Lemmens JA et al. The effectiveness of exercise therapy in patients with osteoarthritis of the hip or knee: a randomized clinical trial. J Rheumatol 1998; 25(12):2432-2439.
11. Lindstrom I, Ohlund C, Eek C, Wallin L, Peterson LE, Fordyce WE et al. The effect of graded activity on patients with subacute low back pain: a randomized prospective clinical study with an operant-conditioning behavioral approach. Phys Ther 1992; 72(4):279-290.

12. Lindstrom I, Ohlund C, Eek C, Wallin L, Peterson LE, Nachemson A. Mobility, strength, and fitness after a graded activity program for patients with subacute low back pain. A randomized prospective clinical study with a behavioral therapy approach. Spine 1992; 17(6):641-652.
13. Ostelo RW, de Vet HC, Berfelo MW, Kerckhoffs MR, Vlaeyen JW, Wolters PM et al. Effectiveness of behavioral graded activity after first-time lumbar disc surgery: short term results of a randomized controlled trial. Eur Spine J 2003; 12(6):637-644.
14. Staal JB, Hlobil H, Köke AJA, Twisk JWR, Smid T, van Mechelen W. The effects of graded activity for low back pain in occupational health on sick leave, functional status and pain: a randomised controlled trial. Ann Intern Med 2004; 140:77-84.
15. Geraets JJ, Goossens ME, de Groot IJ, de Bruijn CP, de Bie RA, Dinant GJ et al. Effectiveness of a graded exercise therapy program for patients with chronic shoulder complaints. Aust J Physiother 2005; 51(2):87-94.
16. Ostelo RW, de Vet HC, Vlaeyen JW, Kerckhoffs MR, Berfelo WM, Wolters PM et al. Behavioral graded activity following first-time lumbar disc surgery: 1-year results of a randomized clinical trial. Spine 2004; 28(16):1757-1765.
17. King M, Davidson O, Taylor F, Haines A, Sharp D, Turner R. Effectiveness of teaching general practitioners skills in brief cognitive behaviour therapy to treat patients with depression: randomised controlled trial. BMJ 2002; 324(7343):947-950.
18. Bekkering GE, Hendriks HJ, van Tulder MW, Knol DL, Hoeijenbos M, Oostendorp RA et al. Effect on the process of care of an active strategy to implement clinical guidelines on physiotherapy for low back pain: a cluster randomised controlled trial. Qual Saf Health Care 2005; 14(2):107-112.
19. Jette AM, Delitto A. Physical therapy treatment choices for musculoskeletal impairments. Phys Ther 1997; 77(2):145-154.
20. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P et al. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis. Ann Rheum Dis 2003; 62(12):1145-1155.
21. Alexander NB, Galecki AT, Grenier ML, Nyquist LV, Hofmeyer MR, Grunawalt JC et al. Task-specific resistance training to improve the ability of activities of daily living-impaired older adults to rise from a bed and from a chair. J Am Geriatr Soc 2001; 49(11):1418-1427.
22. De Vreede PL, Samson MM, van Meeteren NL, Duursma SA, Verhaar HJ. Functional-task exercise versus resistance strength exercise to improve daily function in older women: a randomized, controlled trial. J Am Geriatr Soc 2005; 53(1):2-10.
23. Ketelaar M, Vermeer A, Hart H, Petegem-van Beek E, Helders PJ. Effects of a functional therapy program on motor abilities of children with cerebral palsy. Phys Ther 2001; 81(9):1534-1545.
24. Childs JD, Fritz JM, Flynn TW, Irrgang JJ, Johnson KK, Majkowski GR et al. A clinical prediction rule to identify patients with low back pain most likely to benefit from spinal manipulation: a validation study. Annals of Internal Medicine 2004; 141(12):920-928.

25. George SZ, Delitto A. Clinical examination variables discriminate among treatment-based classification groups: a study of construct validity in patients with acute low back pain. Physical Therapy 2005; 85(4):306-314.
26. Guccione AA. Arthritis and the process of disablement. Phys Ther 1994; 74(5):408-414.
27. Verbrugge LM, Jette AM. The disablement process. Soc Sci Med 1994; 38(1):1-14.
28. Jensen MP, Nielson WR, Kerns RD. Toward the development of a motivational model of pain self-management. Journal of Pain 2003; 4(9):477-492.
29. McCracken LM, Carson JW, Eccleston C, Keefe FJ. Acceptance and change in the context of chronic pain. Pain 2004; 109(1-2):4-7.
30. Geraets JJ, de Groot IJ, Goossens ME, de Bruijn CP, de Bie RA, van den Heuvel WJ et al. Comparison of two recruitment strategies for patients with chronic shoulder complaints. Br J Gen Pract 2006; 56(523):127-133.
31. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997; 24(4):799-802.
32. 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(3):446-451.
33. Dekker J, Dallmeijer AJ, Lankhorst GJ. Clinimetrics in rehabilitation medicine: current issues in developing and applying measurement instruments 1. J Rehabil Med 2005; 37(4):193-201.
34. Bot SDM, Terwee CB, van der Windt DAWM, Bouter LM, Dekker J, de Vet HC. Psychometric evaluation of self-report questionnaires - the development of a checklist. In: Ader HJ, Mellenbergh GJ, editors. Proceedings of the second workshop on research methodology 25-27 June 2003. Amsterdam: VU University Amsterdam, 2003: 161-168.
35. Terwee CB, Mokkink LB, Steultjens MP, Dekker J. Performance-based methods for measuring the physical function of patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. Rheumatology (Oxford) 2006.
36. Baker SM, Marshak HH, Rice GT, Zimmerman GJ. Patient participation in physical therapy goal setting. Phys Ther 2001; 81(5):1118-1126.
37. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev 1977; 84(2):191-215.

Summary

The focus of this thesis is the effectiveness of behavioral graded activity (BGA) in patients with osteoarthritis (OA) of the hip and/or knee. As described in **Chapter 1**, OA is a slowly progressive musculoskeletal disorder, causing pain, joint stiffness, muscle weakness, and threatening mobility and an active lifestyle. Patients with OA tend to avoid certain activities because they anticipate that these activities increase pain and suffering. In the short term, pain can be reduced by avoiding activities. However, in the long term, avoidance of activities can have both physical (loss of mobility, muscle strength, and fitness) and psychological (loss of self-esteem, depression) consequences.

The principle objectives of managing OA are to control pain adequately, improve function and reduce disability. There is strong evidence that exercise therapy has a short term benefit for OA. However, these beneficial effects decrease over time and finally disappear. This decline is thought to be related to the difficulties people have in maintaining adherence to prescribed exercises. Therefore, to enhance long term benefit, adherence to exercise therapy is of utmost importance. Recently, the focus of attention within physiotherapy has shifted towards behaviorally oriented treatment, like BGA, which focuses less on pain and includes psychological and social factors in the treatment-process. Such intervention seems appropriate to increase the level of activities of patients with OA and to increase patients' adherence to these activities.

However, at the start of the present study the scientific evidence for the effectiveness of BGA in patients with a progressive and specific chronic disease, like OA of the hip and knee, was not available. Therefore, we performed a randomized controlled trial to study the effectiveness of BGA in patients with OA of the hip and/or knee. It was hypothesised that in the long term BGA results in less pain, less limitations in activities, and better patient global assessment (i.e. the effect of treatment perceived by patients themselves), compared to usual care of physiotherapists (UC). UC was operationalized as physiotherapeutic care according to the Dutch physiotherapy guideline for patients with hip and/or knee OA. It was also investigated whether specific subgroups of patients benefited more from BGA and which factors influenced the success of BGA-treatment. Because 2 methods of recruitment were used in the randomized clinical trial, the influence of recruitment strategy on patient characteristics and on the

efficacy of BGA/UC was studied. Finally, to facilitate the choice of the appropriate questionnaires to measure pain, physical function, and patient global assessment in patients with OA, an overview was given of the psychometric qualities of all questionnaires available for patients with OA. The following research questions were formulated in the present thesis:

1. What is the effectiveness of behavioral graded activity in patients with osteoarthritis of the hip and/or knee?
2. What is the cost-effectiveness of behavioral graded activity in patients with osteoarthritis of the hip and/or knee?
3. Which groups of patients with osteoarthritis of the hip and/or knee benefit particularly from behavioral graded activity?
4. Which factors influence the success of behavioral graded activity in patients with osteoarthritis of the hip and/or knee?
5. What is the influence of two recruitment strategies on the characteristics of patients and on the outcome of a randomized clinical trial involving patients with osteoarthritis of the hip and/or knee?
6. What is the evidence available on the psychometric qualities of osteoarthritis questionnaires?

Chapter 2 presents the results of a single blind randomized controlled trial on the effectiveness of BGA, compared to UC as provided by physiotherapists, in patients with OA of the hip and/or knee. The trial was conducted in primary care. Twohundred patients with OA of the hip and/or knee, according to the criteria of the American College of Rheumatology, were included. We randomized these 200 patients; 97 were allocated to BGA and 103 patients to UC.

The BGA program was based on operant conditioning with exercise therapy comprising boostersessions. The intervention was directed at increasing the level of activities in a time-contingent way, with the objective to integrate these activities in the patients' daily life. The treatment consisted of a 12-week period with a maximum of 18 sessions, followed by 5 pre-set boostermoments with a maximum of 7 sessions (in week 18, 25, 34, 42, and 55, respectively). UC was treatment according to the Dutch physiotherapy guideline for patients with OA, which consists of general recommendations, emphasizing provision of information and advice, exercise therapy, and encouragement of positive coping with the complaints. The treatment

consisted of a maximum of 18 sessions within a period of 12 weeks.

Primary outcome measures were pain (visual analog scale (VAS); Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) physical function (WOMAC) and patient global assessment (PGA). Secondary outcome measures were, among others, range of joint motion (ROM), muscle strength, patient-specific physical function (MACTAR), physical activity (5-meter walking time test), and quality of life (MOS Short Form 36 (SF-36)). Assessments were performed at week 0, 13, 39, and 65. At 65 weeks, 179 patients were assessed.

Both treatments showed beneficial within-group effects on the primary outcome measures, both at short term and in the long term. After 65 weeks, BGA patients improved respectively 42.8% and 25.6% compared to baseline on WOMAC pain and physical function. UC patients improved 36.8% and 25.1%, respectively. These within-group improvements were clinically relevant. On the other hand, the differences between the groups in improvement for pain, physical function and PGA were small and not significant. The same pattern, significant within-group effect and no significant differences between groups, was found for the secondary outcome measures, with the exception of patient-oriented physical function (MACTAR) and 5 meter walking time test. For these outcome measures significant differences were found in favor of BGA at week 65. Furthermore, BGA patients reported to adhere significantly more to the home exercises and activities (at 65 weeks: 56% of BGA patients compared to 33% of UC patients).

Since both interventions resulted in beneficial long term effects, the superiority of BGA over UC has not been demonstrated. It can be concluded that BGA seems an acceptable method to treat patients with OA of the hip and/or knee, with similar results compared to UC.

Chapter 3 evaluates the cost-effectiveness of the 2 interventions included in the current study from a societal perspective. To evaluate the consequences of the treatments, direct health care and non-health care costs were considered, as well as indirect costs. Costs were measured using cost diaries for the entire follow-up period of 65 weeks. The improvements on clinical outcomes (pain, physical function, PGA, and quality of life (Euroqol)) showed no relevant differences between the 2 treatment conditions. The

mean (95% CI) difference in total costs between the 2 treatment groups was -€773 (-€2360 ; €772), that is, BGA resulted in less costs but this difference was non-significant. The sensitivity analyses showed that these results were fairly robust. In conclusion, currently there is no evidence provided that BGA is either more or less effective or more or less costly than UC.

Since patients with OA are very heterogeneous in their complaints and restrictions of activities, it is likely that certain patients will benefit more from BGA than other patients. **Chapter 4** describes a study which investigated whether specific groups of patients with OA particularly benefit from BGA. Beneficial effects were expected in 1) patients with a relatively low level of physical functioning at the start of the treatment; 2) patients with passive coping styles to pain; and 3) patients with a relative low level of internal locus of control or a relative high level of powerful others locus of control.

Data were used from the randomized clinical trial as described in Chapter 3. To study whether differences existed in the effects of BGA and UC, in specific subgroups, the effect modification of treatment was tested by including interactions in the analysis of covariance (e.g., treatment X physical function). This analysis was based on the long term effectiveness (at 65 weeks) of both BGA and UC.

As expected, patients with a relatively low level of physical functioning showed larger beneficial effects of BGA compared to UC. Patients with more limitations in physical functioning, tend to avoid activities. Because of their inactivity, their physical condition deteriorates, resulting in more limitations. One of the primary goals of BGA is to gradually increase activity levels despite pain, and to educate patients that their disease can be self-managed, which might explain the success of BGA in patients with a low level of physical functioning.The other hypotheses could not be confirmed with the exception of the beneficial effects of BGA for patients with a low level of internal locus of control.

In addition, effect modification by patient characteristics were studied in an exploratory analysis. Characteristics of patients included demographic and clinical features. These analyses indicated beneficial effects of BGA in patients with obesity (BMI > 30) and patients with a relatively high score of pain at the start of treatment. Also, evidence was

found for beneficial effects of UC in patients without radiological evidence of OA.

On basis of these findings, it can be concluded that BGA (compared to UC) is the preferred treatment option for patients with a relatively low level of physical functioning at the start of the treatment and, to a lesser extent, patients with a low level of internal locus of control.

To explain the differences, after BGA treatment, between patients who successfully integrated activities in daily life and patients who did not succeed in integrating activities in daily life, further investigation is required. **Chapter 5** presents the results of a qualitative study on the question why some patients remained to have an active lifestyle, i.e. adhered to perform activities in the long term, after completion of BGA treatment and others did not.

Twelve patients were selected according to the model of deliberate sampling for heterogeneity. Selection was based on their success of treatment as assessed by PGA. Six patients were identified as being improved, and 6 patients as worsened on PGA. The study was based on open-ended-in-depth interviews centering on the patients' experiences of the treatment BGA. The interview departed from 3 themes: aspects related to experiences with the physiotherapist, aspects related to the content of BGA, and aspects related to the patient. Data collection was stopped after 12 interviews because no new relevant data seemed to emerge during the last interviews. The data were coded and analyzed using the methods developed in grounded theory.

During the interviews, patients were asked whether they integrated the activities in their daily life (in other words: whether they adhered to the exercises/activities). It appeared that PGA and adherence did not agree with each other. Because the main goal of BGA was to increase the patients' level of activities, we decided to focus on the adherence of the patients.

The findings from this study suggest that 2 factors influence the long term adherence to perform activities. First, the initial motivation of patients played an important role during the treatment process; some patients were motivated to reach short term goals (e.g., to decrease pain), while others were motivated to reach long term goals (e.g., to postpone an operation or to live independently as long as possible). It appeared that all patients adhering

to the activities, were initially motivated to reach long term goals with their visits to the physiotherapist. All patients who did not adhere to the activities, reported either a short term goal or had no specific goal to visit the physiotherapist. Therefore, there seems to be a relationship between the initial motivation of the patients to visit the physiotherapist and the adherence of patients. Second, the involvement of patients differed among patients. It appeared that all patients of the adherence group reported that they were actively involved in the whole treatment process. However, most patients of the non-adherence group reported that the physiotherapist made all decisions. Therefore, it seems that the active involvement of patients during the treatment process relates to a higher adherence of patients.

In conclusion, it would be beneficial to emphasize the importance of an active involvement of patients right from the start of the treatment. Furthermore, to increase the success of treatment, physiotherapists should gain a clear understanding of the patients' initial motives for treatment. However, it needs to be commented that the objective of this qualitative study was to generate hypotheses on factors influencing the success of BGA; more (quantitative) research is needed to test and possibly confirm these hypotheses.

To conduct a clinical trial in primary care, recruitment of a sufficient number of study participants is among the most challenging aspects. In the randomized controlled trial (as described in Chapter 3), patients were recruited by participating physiotherapists (n=110). Because the recruitment rate was rather slow, a second recruitment strategy was used, i.e. patients responded to articles about the benefit of exercise therapy and the performed study, published in local newspapers (n=90). **Chapter 6** evaluates the effect of 2 different recuitment methods on the characteristics of patients with OA of the hip and/or knee and on the efficacy of BGA and UC.

At baseline demographic, clinical, and psychosocial data were compared. Furthermore, the improvement in primary outcome measures (pain, physical function, and PGA) was compared after 13, 39, and 65 weeks.

Different recruitment methods appeared to attract different subjects. As expected, the newspaper-group reported signficantly less pain (VAS, WOMAC) and tiredness (VAS) at baseline, although more joints were

affected with OA. A possible explanation is that the newspaper group has learned to cope with the pain and limitations due to OA and are less inclined to consult general practitioners or physiotherapists for these complaints. This conclusion was confirmed by the higher score of the group recruited by physiotherapists on the powerful others scale of locus of control. Only small differences were found in the effectiveness of treatment (either BGA or UC), which disappeared after adjustment for baseline differences.

The findings of this study suggest that method of recruitment affects clinical characteristics of patients recruited for the study. However, a mix of recruitment strategies does not seem to affect treatment outcome, on the condition that baseline differences are adjusted for.

Both in clinical practice and research on patients with OA, outcome is evaluated using many different instruments. The Outcome Measures in Rheumatology Clinical Trials (OMERACT) group defined a core set of outcome dimensions for clinical studies in hip and knee OA, which are pain, physical function and patient global assessment. To facilitate the choice of the most appropriate questionnaires, a systematic comparison of the descriptive and psychometric qualities of the questionnaires is required. **Chapter 7** describes the results of a systematic review on the descriptive and psychometric qualities of self-assessment instruments on pain, physical function and PGA for patients with OA of the hip and/or knee. Instruments were identified through systematic literature searches. Information relating to the characteristics and validity, reproducibility, and responsiveness of questionnaires was systematically extracted from published papers and rated according to a checklist of criteria.

The 32 questionnaires that met the inclusion criteria could be divided in 24 condition specific instruments, 7 generic instruments and 1 patient-specific scale. Many psychometric qualities were not properly tested for a large number of questionnaires and none of the questionnaires rated positive on all aspects of the checklist. For example, all studies rated doubtful on responsiveness. The condition-specific WOMAC had been studied most extensively and received, overall, the best ratings for its descriptive and psychometric qualities. Concerning generic instruments, the SF-36 had been studied most often in patients with OA and demonstrated, overall, the highest rating. The psychometric qualities of the patient-specific

instrument on patient global assessment has been studied to a limited degree in patients with hip and/or knee OA.

In conclusion, at this moment the WOMAC and SF-36 seem to be the most appropriate instruments to measure pain and physical function in patients with hip and/or knee OA.

Chapter 8 discusses the main results of the randomized controlled trial in the effectiveness of BGA in patients with OA of the hip and/or knee and the systematic review on the osteoarthritis questionnaires.

Concerning the randomized controlled trial, the clinical relevancy of the (within-group) effects were critically examined and the results of the present trial were compared to other trials on the effectiveness of BGA.

Furthermore, the contrast between BGA and UC is discussed. The contrast between the 2 interventions was less than expected. Several explanations for this reduced contrast can be offered. First, although we expected that BGA patients would be treated at least 5 sessions more, the number of sessions of both interventions were comparable. Second, not all BGA patients were treated according to the protocol. Third, UC physiotherapists were educated in the guideline OA, which could have caused that UC in the trial was more based on the guideline compared to usual care in practice.

Also, the shift within the field of physiotherapy is reviewed. Recently, a shift from a passive approach (e.g., massage therapy and physiotherapy modalities) to an active approach (exercise therapy, education), as advocated in the guidelines, is gaining more and more support within the field of physiotherapy. This shift can explain the unexpected beneficial effects of UC in the long term.

Other topics which are discussed are the importance of matching interventions with subgroups of patients, the influences of recruitment methods, and the value of the patient-specific outcome measures like the MACTAR.

In addition, the main findings of the systematic review on the osteoarthritis questionnaires are discussed. The review was based on self-assessment questionnaires. In Chapter 8, attention is paid to the descriptive and psychometric qualities of performance-based measures. Furthermore, the used checklist is critically examined.

Finally, implications for clinical practice are considered. The future of BGA is discussed and the education of physiotherapists is advocated. Recommendations are given for future research on BGA, outcome measures and the treatment of patients with OA of the hip and/or knee.

Samenvatting

De belangrijkste doelstelling van dit proefschrift is om de effectiviteit van gedragsmatige oefentherapie ('behavioral graded activity'(BGA)) bij patiënten met artrose van heup en/of knie te onderzoeken. Zoals beschreven in **hoofdstuk 1**, wordt artrose gekenmerkt door het progressieve verlies van gewrichtskraakbeen in combinatie met een toegenomen activiteit van het onderliggende subchrondale bot. Dit kan klachten als pijn, gewrichtsstijfheid, spierzwakte en een verminderde mobiliteit veroorzaken en leiden tot een inactieve levensstijl. Patiënten met artrose hebben de neiging bepaalde activiteiten te vermijden omdat zij inschatten dat deze activiteiten tot een toename van pijn en andere klachten leiden. Dit vermijden van activiteiten kan op de korte termijn tot een afname van pijn leiden. Echter, op de langere termijn kan deze vermijding zowel fysieke (verminderde mobiliteit, spierkracht en fitheid) als psychologische (daling van eigenwaarde, depressiviteit) consequenties hebben.

De belangrijkste doelen bij het behandelen van artrose zijn het controleren van pijn, het verbeteren van het functioneren en het verminderen van beperkingen. Er is veel wetenschappelijk bewijs voor de effectiviteit van oefentherapie op de korte termijn, bij patiënten met artrose. Deze effecten nemen echter af in de tijd en verdwijnen op de lange termijn. Er lijkt een relatie te zijn tussen de afname van de gunstige effecten en de problemen die mensen ervaren om de voorgeschreven oefeningen te blijven uitoefenen. Therapietrouw lijkt dan ook een belangrijke rol te spelen bij het vasthouden van de effecten op de lange termijn. Recentelijk is er binnen het vakgebied fysiotherapie meer aandacht voor een gedragsmatig georiënteerde behandeling (zoals BGA). Deze behandeling richt zich minder op pijn van patiënten en neemt psychologische en sociale factoren mee. Een dergelijke interventie lijkt geschikt om het activiteitenniveau van patiënten met artrose te vergroten en om de therapietrouw ten aanzien van het uitvoeren van deze activiteiten van patiënten te verhogen.

Bij aanvang van de huidige studie was er echter geen wetenschappelijk bewijs voor de effectiviteit van BGA bij patiënten met een progressieve, chronische en specifieke aandoening, zoals artrose van heup of knie. Daarom hebben wij een gerandomiseerd klinisch experiment uitgevoerd om na te gaan wat het effect was van BGA bij patiënten met artrose van heup of knie. De verwachting was dat het toepassen van BGA op de lange termijn zou leiden tot minder pijn, minder beperkingen in

activiteiten en een beter zelf ervaren herstel (zoals beoordeeld door de patiënt zelf) in vergelijking met de standaard behandeling door de fysiotherapeut (UC). UC is geoperationaliseerd als fysiotherapeutische behandeling zoals omschreven in de Nederlandse richtlijn fysiotherapie voor patiënten met artrose van heup of knie. Verder is onderzocht of bepaalde subgroepen patiënten vooral baat hadden bij BGA en welke factoren het succes van BGA beïnvloeden. Aangezien twee methoden werden gehanteerd om patiënten voor de bovengenoemde studie te werven, is bekeken wat de invloed was van de wervingsmethoden op de kenmerken van de patiëntenpopulatie en op de effectiviteit van BGA/UC. Tenslotte is een overzicht gemaakt van de psychometrische kwaliteiten van alle beschikbare vragenlijsten voor patiënten met artrose op het gebied van pijn, fysiek functioneren en zelf ervaren herstel. Een dergelijk overzicht vergemakkelijkt de keuze van geschikte vragenlijsten.

De volgende onderzoeksvragen komen in dit proefschrift aan de orde:

1. Wat is de effectiviteit van gedragsmatige oefentherapie (BGA) bij patiënten met artrose van heup of knie?
2. Wat is de kosten-effectiviteit van gedragsmatige oefentherapie bij patiënten met artrose van heup of knie?
3. Welke patiënten met artrose van heup of knie hebben vooral baat bij gedragsmatige oefentherapie?"
4. Welke factoren beïnvloeden het succes van gedragsmatige oefentherapie bij patiënten met artrose van heup of knie?
5. Wat is de invloed van twee wervingsstrategieën op de kenmerken van de patiëntenpopulatie en op de uitkomst van een gerandomiseerd onderzoek bij patiënten met artrose van heup of knie?
6. Welk bewijs is beschikbaar over de psychometrische kwaliteiten van vragenlijsten, geschikt voor patiënten met artrose van heup of knie?

Hoofdstuk 2 presenteert de resultaten van een gerandomiseerd klinisch onderzoek naar de effectiviteit van BGA, in vergelijking tot UC, bij patiënten met artrose van heup of knie. Het onderzoek vond plaats in de eerstelijns gezondheidszorg. Tweehonderd patiënten met artrose van heup of knie, volgens de criteria van het Amerikaanse College of Rheumatology, werden ingesloten in het onderzoek. Deze 200 patiënten zijn vervolgens

gerandomiseerd in twee groepen: 97 patiënten werden toegewezen aan de BGA-groep en 103 patiënten aan de UC-groep.

De BGA-behandeling is gebaseerd op operante behandelprincipes en deze zijn gecombineerd met oefentherapie en terugkomsessies. Bij deze benadering is de opbouw niet afhankelijk van de pijn maar wordt de intensiteit stapsgewijs opgebouwd in de tijd. Dit wordt tijdcontingent genoemd. Het uiteindelijke doel van de behandeling is een hoger activiteitenniveau te bereiken en deze activiteiten in het dagelijks leven van patiënten te integreren. De behandeling bestond uit een periode van 12 weken met een maximum van 18 zittingen fysiotherapie, gevolgd door vijf vastgestelde terugkommomenten met een maximum van zeven zittingen. Deze terugkommomenten vonden in week 18, 25, 34, 42 en 55 plaats). UC was behandeling volgens de Nederlandse fysiotherapie richtlijn artrose van heup of knie. In deze richtlijn worden algemene aanbevelingen aan fysiotherapeuten gegeven, zoals het geven van informatie en advies aan patiënten, oefentherapie en het aanmoedigen van een positieve copingstijl ten aanzien van klachten. De behandeling bestond uit maximaal 18 zittingen in een periode van 12 weken.

Primaire uitkomstmaten waren pijn (gemeten met een visuele analoge schaal (VAS) en met de Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), fysiek functioneren (WOMAC) en zelf ervaren herstel (patient global assessment (PGA)). Secundaire uitkomstmaten waren onder andere gewrichtsmobiliteit, spierkracht, patiënt-specifiek fysiek functioneren (MACTAR), fysieke activiteit (5 meter looptest) en kwaliteit van leven (MOS Short Form 36 (SF-36)). De metingen vonden plaats voor de behandeling, direct na de behandeling (na 13 weken) en na 39 en 65 weken. Aan de laatste meting, na 65 weken, namen 179 patiënten deel.

Zowel op de korte als op de lange termijn vertoonden beide behandelingen gunstige effecten op de primaire uitkomstmaten. Na 65 weken, waren de BGA-patiënten respectievelijk 42.8% en 25.6%, ten opzichte van de aanvang van het onderzoek, vooruitgegaan op de uitkomstmaten pijn (WOMAC) en fysiek functioneren. UC-patiënten gingen respectievelijk 36.8% en 25.1% vooruit. Deze vooruitgang binnen de groepen was klinisch relevant. Tussen de groepen waren de verschillen in vooruitgang op pijn, fysiek functioneren en zelf ervaren herstel echter klein

en niet significant. Hetzelfde patroon, significante verschillen binnen de groepen en geen significante verschillen tussen de groepen, werd gevonden voor de meeste secundaire uitkomstmaten. De uitkomstmaten patiënt-specifiek fysiek functioneren (MACTAR) en de 5 meter looptest vormden hierop een uitzondering. Bij deze twee uitkomstmaten werden significante verschillen gevonden tussen BGA en UC, ten gunste van BGA. Verder bleken BGA patiënten vaker aan te geven dat ze trouw waren aan de oefeningen/activiteiten zoals afgesproken met de fysiotherapeut. Na 65 weken bleek 56% van de BGA patiënten nog oefeningen/activiteiten uit te voeren ten opzichte van 33% van de UC patiënten.

Aangezien beide behandelingen resulteerden in gunstige effecten op de lange termijn, kon de meerwaarde van BGA ten opzichte van UC niet worden aangetoond. De conclusie kan worden getrokken dat BGA een acceptabele methode is om patiënten met artrose van heup of knie te behandelen, met vergelijkbare resultaten als UC.

In **hoofdstuk 3** wordt de kosten-effectiviteit van beide interventies (BGA en UC) bekeken vanuit een maatschappelijk perspectief. Zowel directe kosten als indirecte kosten zijn vastgelegd aan de hand van kostendagboekjes. Alle deelnemende patiënten werden gedurende 65 weken gevolgd. Er waren tussen de groepen geen relevante verschillen in de verbetering op de klinische uitkomstmaten (pijn, fysiek functioneren, zelf ervaren herstel en kwaliteit van leven (Euroqol)). Wat betreft de totale kosten was het gemiddelde (met 95% betrouwbaarheidsinterval) verschil tussen de twee groepen -€773 (-€2360 ; €772). Dit betekent dat de totale kosten in de BGA groep €773 lager waren dan in de UC groep, hoewel dit niet significant verschilde. De sensitiviteitsanalyses toonden aan dat de resultaten robuust waren. Daarom is er op basis van dit kosteneffectiviteitsonderzoek geen bewijs geleverd dat BGA tot meer of minder kosten of effecten zal leiden.

Aangezien patiënten met artrose heterogeen zijn in klachten en beperkingen van activiteiten, ligt het voor de hand dat bepaalde patiënten meer baat hebben bij BGA dan andere patiënten. In **hoofdstuk 4** wordt een studie beschreven waarin onderzocht werd of er specifieke subgroepen patiënten met artrose zijn die vooral baat hebben van BGA. Gunstige effecten van BGA werden van te voren verwacht bij 1) patiënten met een relatief laag

niveau van fysiek functioneren bij aanvang van de behandeling; 2) patiënten die een passieve copingstijl hanteren bij pijn; 3) patiënten met een relatief lage score op interne locus of control of een relatief hoge score op 'powerful others locus of control'.

Voor deze studie werden data gebruikt van het gerandomiseerde onderzoek zoals beschreven in hoofdstuk 3. Om te onderzoeken of er tussen specifieke subgroepen verschillen waren in de effectiviteit van BGA en UC, werden interacties toegevoegd aan de covariantie analyses (bijvoorbeeld behandeling X fysiek functioneren). Deze analyse was gebaseerd op de effectiviteit van de behandelingen na 65 weken.

Zoals verwacht bleken patiënten met een relatief laag niveau van fysiek functioneren meer baat te hebben van BGA dan van UC. Een mogelijke verklaring hiervoor is dat patiënten met veel beperkingen in fysiek functioneren de neiging hebben om activiteiten te vermijden. Deze inactiviteit heeft tot gevolg dat hun fysieke conditie vermindert, wat vervolgens resulteert in meer beperkingen. Eén van de belangrijkste doelen van BGA is het niveau van activiteiten, ongeacht de pijn, stapsgewijs op te bouwen en patiënten te leren dat zij zelf met hun ziekte om kunnen gaan. De andere verwachtingen werden niet bevestigd met uitzondering van patiënten met een laag niveau van interne locus of control. Deze patiënten bleken meer baat te hebben van BGA.

Aanvullend is voor een aantal kenmerken uit de medische anamnese nagegaan of deze van invloed waren op het effect van BGA. Dit betroffen demografische en klinische kenmerken. Uit deze analyses kwam naar voren dat het hebben van obesitas (body mass index > 30) en een relatief hogere score op pijn bij aanvang van de behandeling tot een gunstig effect van BGA kunnen leiden. Ook werd er bewijs gevonden dat patiënten zonder radiologisch aangetoonde afwijkingen aan het gewricht baat hebben bij UC.

Op basis van deze resultaten kan geconcludeerd worden dat de behandeling BGA de voorkeur heeft boven UC bij patiënten met een relatief laag niveau van fysiek functioneren en, in mindere mate, bij patiënten met een laag niveau van interne locus of control.

Om te verklaren waarom sommige patiënten, na BGA behandeling, succesvol activiteiten in het dagelijks leven intergreerden en andere patiënten dit niet deden, is verder onderzoek nodig. In **hoofdstuk 5** worden

de resultaten van een kwalitatieve studie gepresenteerd. Deze kwalitatieve studie richtte zich op de vraag waarom sommige patiënten een actieve levensstijl aanhielden (en dus de activiteiten op de lange termijn bleven uitvoeren) na behandeling met BGA en anderen dit niet deden.

Voor deze studie zijn 12 patiënten geselecteerd volgens het 'model of deliberate sampling for heterogeneity' om de heterogeniteit van de patiënten te waarborgen. Patiënten werden geselecteerd op basis van het succes van de behandeling, zoals gescoord op het zelf ervaren herstel na 65 weken. Zes patiënten scoorden 'vooruitgegaan' op zelf ervaren herstel en zes patiënten scoorden 'verslechterd'. Het onderzoek was gebaseerd op open-eind diepte interviews waarin de ervaringen van patiënten met de behandeling BGA centraal stonden. Het interview besloeg drie thema's: aspecten gerelateerd aan ervaringen met de behandelend fysiotherapeut, aspecten gerelateerd aan de inhoud van BGA en aspecten gerelateerd aan de patiënten zelf. Omdat na 12 interviews geen nieuwe relevante informatie meer werd gevonden is op dat moment gestopt met het verzamelen van data. De data werden vervolgens gecodeerd en geanalyseerd volgens de principes van de 'grounded theory'.

Tijdens de interviews werd de patiënten gevraagd of ze de activiteiten in hun dagelijks leven integreerden (ofwel, of ze trouw waren aan de oefeningen/activiteiten). Het bleek echter dat het zelf ervaren herstel en deze therapietrouw niet overeen kwamen. Aangezien het verhogen van het activiteitenniveau van patiënten het voornaamste doel van BGA was, hebben we besloten de therapietrouw als uitgangspunt van dit onderzoek te nemen.

Uit de resultaten van deze studie bleek dat de therapietrouw om op de lange termijn activiteiten uit te voeren door twee factoren beïnvloed werd. Allereerst bleek dat de aanvankelijke motivatie van patiënten om naar de fysiotherapeut te gaan een belangrijke rol speelt tijdens het behandel proces; sommige patiënten waren gemotiveerd om korte termijn doelen te behalen (zoals het verminderen van pijn), terwijl andere patiënten gemotiveerd waren om lange termijn doelen te behalen (zoals het uitstellen van een operatie, het zelfstandig thuis blijven wonen). Het bleek dat alle patiënten die aangaven trouw te zijn aan de oefeningen/activiteiten, aanvankelijk een lange termijn doel voor ogen hadden. Daarentegen bleken alle patiënten, die niet meer oefenden of hun activiteiten niet hadden geïntegreerd in hun dagelijks leven,

aanvankelijk een korte termijn doel of geen specifiek doel te hebben. Daarom lijkt er een verband te zijn tussen de aanvankelijke motivatie van patiënten om naar de fysiotherapeut te gaan en de therapietrouw van patiënten (na 65 weken). Ten tweede bleek de betrokkenheid van patiënten bij het behandelproces te verschillen. Het bleek dat alle patiënten van de therapietrouwe groep aangaven dat zij actief betrokken waren bij het hele behandelproces. De meeste patiënten die niet therapietrouw waren gaven echter aan dat de fysiotherapeut de meeste beslissingen nam. Het lijkt dan ook dat een actieve betrokkenheid van patiënten gedurende het behandelproces verband houdt met een betere therapietrouw van patiënten.

Op basis van deze resultaten kan geconcludeerd worden dat het gunstig kan zijn om het belang van een actieve betrokkenheid van patiënten, al direct vanaf de aanvang van de behandeling, te benadrukken. Verder zouden fysiotherapeuten een duidelijk beeld moeten krijgen van de aanvankelijke motivatie van patiënten om naar de fysiotherapeut te komen. Er moet echter wel opgemerkt worden dat het doel van deze kwalitatieve studie is om hypothesen te genereren over factoren die het succes van BGA beïnvloeden; er is behoefte aan meer kwantitatief onderzoek zodat deze hypothesen verder getoetst kunnen worden.

Eén van de meest uitdagende aspecten bij het uitvoeren van een klinisch onderzoek in de eerstelijns gezondheidszorg is het werven van een voldoende aantal participanten. Voor het gerandomiseerde onderzoek (zoals omschreven in hoofdstuk 3) werden patiënten geworven via deelnemende fysiotherapeuten (n=110). Aangezien deze werving vrij traag verliep is een tweede wervingsmethode gebruikt, namelijk het publiceren van artikelen over het belang van oefentherapie bij artrose en de uitgevoerde studie. Deze artikelen werden geplaatst in lokale kranten en leverde uiteindelijk 90 deelnemers op. In **hoofdstuk 6** wordt het effect van twee verschillende wervingsmethodes op de kenmerken van de patiëntenpopulatie en de effectiviteit van BGA en UC geëvalueerd.

Demografische, klinische en psychologische data, gemeten bij aanvang van de studie, werden vergeleken. Verder werd de verandering in de primaire uitkomstmaten (pijn, fysiek functioneren en zelf ervaren herstel) vergeleken na 13, 39 en 65 weken.

Het bleek dat verschillende wervingsmethodes verschillende

personen aantrekt. Zoals verwacht gaf de groep patiënten die geworven waren via kranten significant minder pijn (VAS, WOMAC) en vermoeidheid (VAS) aan bij aanvang van de studie, hoewel meer gewrichten waren aangedaan met artrose. Een mogelijke verklaring is dat patiënten, die via kranten werden geworven, hebben leren omgaan met de pijn en beperkingen als gevolg van artrose en daardoor minder geneigd zijn hulp van artsen of fysiotherapeuten in te schakelen. Deze conclusie werd bevestigd door de hoger score op 'powerful others locus of control' van de groep patiënten die via de fysiotherapeuten werd geworven. Wat betreft de effectiviteit van BGA en UC bleek dat er slecht marginale verschillen werden gevonden. Deze verschillen verdwenen wanneer in de analyses werd gecorrigeerd voor verschillen in demografische, klinische of psychosociale kenmerken.

Op basis van de resultaten van deze studie kan geconcludeerd worden dat de wervingsmethode invloed heeft op de kenmerken van de patiëntenpopulatie. Een mix van wervingsmethodes lijkt de effectiviteit van behandelingen echter niet te beínvloeden, op voorwaarde dat er gecorrigeerd wordt voor eventuele verschillen tussen de groepen bij aanvang van de studie.

Zowel in de klinische praktijk als in het onderzoek bij patiënten met artrose worden verschillende instrumenten gebruikt om bijvoorbeeld de effectiviteit van behandelingen te evalueren. De Outcome Measures in Rheumatology Clinical Trials (OMERACT) groep heeft een core set van uitkomstmaten gedefinieerd, te gebruiken bij klinische studies bij patiënten met artrose van heup of knie. Deze core set omvat de uitkomstmaten pijn, fysiek functioneren en zelf ervaren herstel. Om de keuze van het meest geschikte instrument te vergemakkelijken, is een systematische vergelijking van de beschrijvende en psychometrische kwaliteiten van de instrumenten nodig. In **hoofdstuk 7** worden de resultaten gepresenteerd van een systematische literatuur studie naar de beschrijvende en psychometrische eigenschappen van vragenlijsten voor patiënten met artrose van heup of knie, op het gebied van pijn, fysiek functioneren en zelf ervaren herstel. De vragenlijsten werden gezocht middels een systematische literatuurstudie. Vervolgens werd informatie over de eigenschappen, validiteit, betrouwbaarheid en responsiviteit van de vragenlijsten systematisch geëxtraheerd uit de gepubliceerde artikelen en gescoord op een checklist met criteria.

In totaal voldeden 32 vragenlijsten aan de inclusiecriteria, welke verdeeld konden worden in 24 ziekte-specifieke vragenlijsten, zeven generieke vragenlijsten en één patiënt-specifieke schaal. Het bleek dat bij een groot aantal vragenlijsten veel psychometrische kwaliteiten niet goed onderzocht waren en dat geen van de vragenlijsten positief scoorde op alle criteria van de checklist. Zo scoorden alle vragenlijsten bijvoorbeeld dubieus op het item responsiviteit. De ziekte-specifieke vragenlijst WOMAC was het meest uitgebreid onderzocht en bleek, in het algemeen, de hoogste scores te krijgen voor de beschrijvende en psychometrische kwaliteiten. Wat betreft de generieke vragenlijsten bleek de SF-36 het meest onderzocht en in het algemeen de hoogste scores te krijgen. De psychometrische kwaliteiten van de patiënt-specifieke vragenlijst op het gebied van zelf ervaren herstel zijn slechts beperkt onderzocht bij patiënten met artrose van heup of knie.

Concluderend kan gesteld worden dat op dit moment de WOMAC en SF-36 de meest geschikte vragenlijsten lijken te zijn om pijn en fysiek functioneren te meten bij patiënten met artrose van heup of knie.

In **hoofdstuk 8** worden de belangrijkste resultaten bediscussieerd van zowel het gerandomiseerde klinische onderzoek naar de effectiviteit van BGA als de systematische literatuurstudie naar vragenlijsten voor patiënten met artrose.

Wat betreft het gerandomiseerde onderzoek wordt de klinische relevantie van de gunstige effecten binnen de groepen kritisch besproken. Ook worden de resultaten van de huidige studie vergeleken met andere studies naar de effectiviteit van BGA (bij andere patiëntenpopulaties).

Vervolgens wordt het contrast tussen BGA en UC bediscussieerd. Het contrast tussen beide interventies was lager dan verwacht. Hiervoor kunnen verschillende verklaringen worden gegeven. Allereerst bleek het aantal gegeven zittingen van beide interventies vergelijkbaar te zijn, terwijl verwacht werd dat patiënten uit de BGA groep minstens vijf zittingen meer zouden krijgen. Ten tweede bleken niet alle BGA patiënten te zijn behandeld volgens het behandelprotocol. Tenslotte werden UC fysiotherapeuten geschoold in het gebruik van de richtlijn artrose. Deze scholing kan tot gevolg hebben dat UC in het onderzoek is meer gebaseerd op de richtlijn dan UC uit de dagelijkse praktijk.

Verder is de verandering binnen de fysiotherapeutische

handelingspraktijk besproken. De laatste jaren is een verandering zichtbaar van een passieve benadering (met bijvoorbeeld massage en fysische technieken) naar een actieve benadering (oefentherapie, scholing) zoals omschreven wordt in de richtlijn artrose. Deze verandering kan de onverwachtse positieve lange termijn effecten van UC verklaren.

Andere onderwerpen die zijn beschreven zijn het belang van het afstemmen van de behandeling op specifieke subgroepen patiënten, de invloed van wervingsmethodes en het belang van patiënt-specifieke vragenlijsten zoals de MACTAR.

Bovendien zijn de belangrijkste resultaten van de systematische literatuurstudie over vragenlijsten voor patiënten met artrose kritisch besproken. In hoofdstuk 8 is aandacht geschonken aan de beschrijvende en psychometrische kwaliteiten van objectieve meetinstrumenten, of terwijl testen. Verder is de gebruikte checklist bediscussieerd.

Tot slot worden de implicaties van de resultaten van het proefschrift voor de dagelijkse praktijk besproken. De toekomst van BGA en de scholing van fysiotherapeuten zijn beschouwd. Ook zijn aanbevelingen gedaan voor toekomstig onderzoek naar de effectiviteit van BGA, uitkomstmaten en de behandeling van patiënten met artrose van heup of knie.

Dankwoord

In dit proefschrift staat het thema `stapsgewijs meer bewegen' centraal. Het schrijven van mijn proefschrift heb ik ervaren als het stapsgewijs beklimmen van een berg. Met dit dankwoord sta ik met een voldaan gevoel op de top. Gelukkig heb ik niet alle stappen alleen hoeven zetten; vele mensen hebben deze voor of met mij gezet. Hier door heb ik mijn gehele promotietijd als zeer prettig ervaren. Bij deze wil ik iedereen bedanken die er op de een of andere manier een bijdrage aan heeft geleverd. Daarnaast wil ik graag een aantal mensen in het bijzonder noemen.

Allereerst de patiënten, want zonder jullie zou er immers niets te onderzoeken zijn geweest. Bedankt voor het invullen van de vragenlijsten en alle bezoekjes aan de onderzoeksassistenten. Ook wil ik alle deelnemende fysiotherapeuten bedanken. Geweldig dat jullie, ondanks jullie werkdruk, bereid waren aan het onderzoek mee te werken en energie te steken in het aanleren van een nieuwe behandelvorm.

Els, mijn copromotor en dagelijks begeleider. Ik heb goede herinneringen aan de vele discussies die we hebben gehouden en aan onze gezellige kletspraatjes. Jij hebt er toe bijgedragen dat het promoveren goed te combineren was met mijn thuissituatie. Je ziekte aan het einde van mijn promotietraject leert alles wel weer relativeren.

Ik heb het geluk gehad begeleid te worden door twee deskundige promotoren. Joost, jij hebt een grote bijdrage geleverd aan de inhoudelijke kwaliteit van dit proefschrift. Ik vind het bijzonder dat je altijd je vinger op de zere plek weet te leggen maar dan ook direct meedenkt aan een goede oplossing. Ik wil je bedanken voor je grote betrokkenheid bij het onderzoek en bij mij, als onderzoeker. Ik heb in de afgelopen jaren veel van je geleerd en prettig met je samengewerkt; ik hoop dat deze samenwerking in stand blijft. Hans, bedankt voor je waardevolle opmerkingen tijdens onze bijeenkomsten, je inzicht in het functioneren van patiënten met artrose en je inzet tijdens het scoren van alle foto's!

Albère Köke, Maurits van Tulder, Rob Oostendorp en Veerle Coupé, als leden van de werkgroep hebben jullie het project inhoudelijk ondersteund, bedankt. Albère, met je enthousiasme en humor geef je me altijd veel energie. Jij hebt een belangrijke bijdrage geleverd aan het ontwikkelen van het behandelprotocol en het scholen van de fysiotherapeuten. Ik wil jou en de andere docenten, Rob Pelt, Rob Coenen, Mario Geilen, Carla Hoogervorst en Monique Hodiamont complimenteren

voor de wijze waarop jullie de fysiotherapeuten de aspecten van gedragsmatige oefentherapie hebben aangeleerd. Veerle, bedankt voor het mooie artikel dat je hebt geschreven over de kosten-effectiviteit van de interventies. De leden van de begeleidingscommissie, bestaande uit Rian Veldhuizen, Bart Staal, Annemiek de Crom, Vera Agterberg en Remco Reij wil ik langs deze weg bedanken voor jullie deskundige advies.

Martijn, Cami, Claudy en Sybille, bij 200 patiënten vier metingen afnemen was een hele klus. Martijn, gezien jouw grote bijdrage aan het project, je inzet en je enthousiasme, vind ik het erg leuk dat jij de kans krijgt een vervolg op het project uit te voeren. Een bijkomstig voordeel is dat dit onze plezierige samenwerking verlengt. Marian en Claire, als ik alleen al bedenk hoeveel patiënten jullie hebben gezien voor een screening (en hoeveel kilometers hiervoor zijn afgelegd…). De mijlpaal van 200 deelnemende patiënten was toch wel één van de hoogtepunten van de afgelopen periode, al werd dit gevoel versterkt doordat we daarmee het hoofdstuk bloedprikken konden afsluiten. Marinda, bedankt voor al je hand- en spandiensten met name tijdens het enorme gekkenhuis dat ontstond nadat artikeltjes waren geplaatst in lokale krantjes. Dit is dan ook direct de juiste plek om de medewerkers van de receptie van het NIVEL te bedanken. Het was ongelofelijk en totaal niet verwacht dat een artikel in een paar plaatselijke krantjes zo'n telefonische chaos bij ons teweeg kon brengen. Bedankt dat jullie alle patiënten geduldig te woord hebben gestaan. Timon, als stagiaire heb jij het kwalitatieve onderzoek uitgevoerd. Ik vond het heel bijzonder om, met jou samen, een aantal patiënten te bezoeken voor de diepte-interviews.

Marian Brouwer, bedankt voor al je werk om het proefschrift in zo'n korte tijd zo mooi te maken als het nu is. Ik zal alle toekomstige promovendi waarschuwen geen grote tabellen te maken!

De leden van de leescommissie, bestaande uit prof. dr. Francois Schellevis, prof. dr. Rob de Bie, dr. Raymond Ostelo, prof. dr. Jos Twisk, prof. dr. Trudy van de Bosch en prof. dr. Ben Dijkmans, bedankt voor het beoordelen van dit manuscript.

Zoals eerder vermeld heb ik de afgelopen jaren als zeer prettig ervaren. Dit heb ik grotendeels te danken aan de fijne collega's die ik bij het NIVEL heb of heb gehad. Het zijn er te veel om op te noemen, dus dat zegt al wel genoeg. Ik wil er toch een paar in het bijzonder noemen. Marieke, wij

zijn lang kamergenoten geweest en hebben alle lief en leed gedeeld. Ik vind het nog steeds heel jammer dat je niet meer bij ons werkt. Al mijn collega's van paramed, Gaby, Chantal, Ilse, Martijn, Wilma, Sybille en voor mijn gevoel ook nog Hannelore, bedankt voor jullie gezelligheid. Gaby, wij zijn de afgelopen jaren veel samen opgetrokken in artrose land en ik heb daar veel plezier van gehad. Fijn dat je mijn paranimf wil zijn! Wienke, Atie en Emmy, het is erg prettig dat jullie altijd zo belangstellend zijn.

Lieve vrienden, het was mijn doel dat mijn promotie mijn privé-leven niet al te zeer zou beïnvloeden. Dat is naar mijn idee goed gelukt. Dat kan ook niet anders met alle gezelligheid, etentjes en/of leuke weekendjes met elkaar. Laten we daar maar vooral mee doorgaan!

Mama, bedankt voor al je onvoorwaardelijke liefde, hulp en belangstelling. Bianca, mijn hele leven ben je als mijn grote zus een steun voor mij in de rug geweest. Het klopt voor mij dan ook helemaal dat je mij als paranimf tijdens mijn promotie bijstaat. Bedankt voor het maken van de voorkant; in dit schilderij heb je op een mooie manier jouw interpretatie van bewegen, knie en promoveren weergegeven. Verry en Inke, ook jullie bedankt voor jullie belangstelling en hulp. Michelle, wij kunnen samen hele leuke gesprekken voeren, zowel over ons privé-leven als inhoudelijk over het werk. Bedankt ook voor het doorspitten van het proefschrift; dat heeft in ieder geval veel kleine foutjes gescheeld!

Lieve, lieve Marc, ik geniet er elke dag van dat je lekker rustig en nuchter bent, maar ook zo ontzettend lief, zowel voor mij als voor onze kinderen. Het is prettig dat we samen zo van het leven en elkaar kunnen genieten! Lieve Noa, Mila en Kas. Wat relativeert er meer dan na een drukke dag thuis te komen en drie juichende kinderen op je af te zien komen? Heerlijk! Nu ik de 'top' van dit proefschrift bereikt heb, kan ik niet wachten tot ik met jullie samen de eerste bergen kan gaan beklimmen…

About the author

Cindy Veenhof was born on March 27, 1971 in Rockanje, the Netherlands. After completing secondary school in 1988 at the Rijksscholengemeenschap in Brielle, she participated for one year in a students' exchange programme and graduated at Jay Highschool in Arkansas, United States of America. From 1989 until 1994, she studied Human Movement Sciences at the Vrije Universiteit in Amsterdam. After her graduation she travelled in Asia, Australia en New Zealand. After returning, she worked as a researcher on the field of children with cerebral palsy at the Universiteit Utrecht and started her study Physiotherapy at the Hogeschool Utrecht in Utrecht. In 1999, she started working at the Netherlands Institute for Health Services Research (NIVEL), where she conducted a study on the effectiveness of a graded activity intervention in patients with osteoarthritis of the hip or knee. She was also engaged in other research projects. Currently, she is still working at the NIVEL institute as a senior researcher in the field of allied health care and physical activity.