Contributions to the management of osteoarthritis

Elien Mahler

Colofon

De uitsparing in de bergen vormt een silhouet dat samen met de weerspiegeling in het water een kniegewricht uitbeeldt. De bergen geven het gezonde deel weer, terwijl de rimpelige en rafelige weerspiegeling het artrotische gewricht met oneffenheden weergeeft. De ruwe keien en de witte bergen zijn een link naar de botstructuur. De stenen in het water staan voor de 'contributions': communicatie, wisselwerking, oversteken naar de andere zijde. Tenslotte illustreren de vogels de beweging en vooruitgang in het artrose landschap.

Cover artwork: Ilse Schrauwers (isontwerp)

Layout: burorub grafisch ontwerp, Nijmegen
Printing: Koninklijke Van der Most B.V., Heerde

ISBN: 978-90-39369-85-2

Publication of this thesis was financially supported by the Sint Maartenskliniek, Nijmegen, The Netherlands

Contributions to the management of osteoarthritis

Bijdragen aan de aanpak van artrose
(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof. dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op maandag 25 juni 2018 des middags te 2.30 uur

door

Elien Anna Maria Mahler

geboren op 26 december 1975 te Sint Anthonis

Promotoren:
Prof. dr. J.W.J. Bijlsma
Prof. dr. F.H.J. van den Hoogen

Copromotoren:
Dr. A.A. den Broeder
Dr. C.H.M. van den Ende

Table of contents

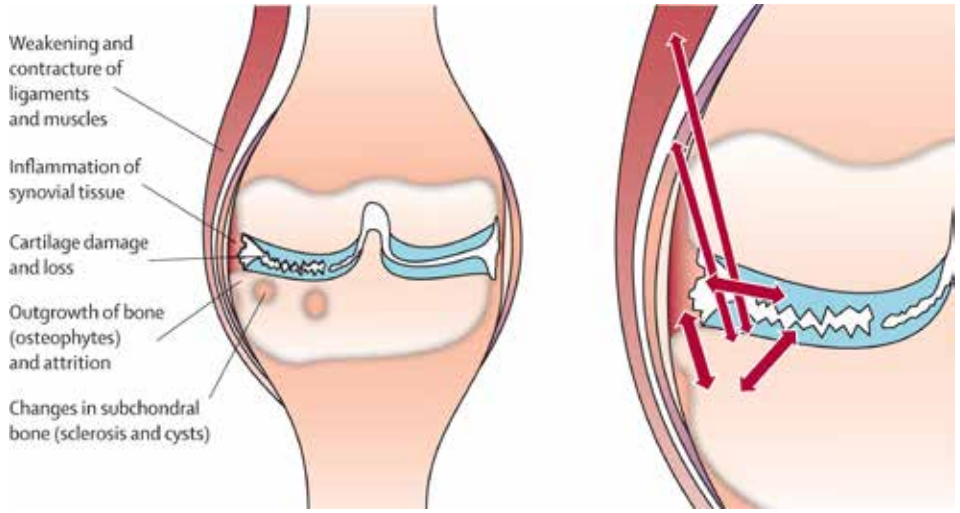
Chapter 1

General introduction

Introduction

Osteoarthritis (OA) is a multifactorial joint disease affecting all the tissues of the joint, and is characterized by cartilage breakdown, the formation of bony outgrowths at the joint margin (osteophytes), subchondral bone sclerosis, alterations to ligaments and muscles, and inflammation of the synovial membrane (figure 1).[1,2] OA causes pain and functional impairments, and reduced quality of life.[1,3,4] During physical examination a decreased range of motion, bony enlargements and deformities of the joint can be observed. OA can be classified according to different sets of classification criteria, focusing on either clinical or radiographic characteristics of OA, or on both. The most commonly used clinical classification criteria used are listed in Table 1.[5,6] Radiographic examination reveals structural abnormalities of the joint as osteophytes, joint space narrowing and sclerosis of subchondral bone.[7] However, only a weak association between symptomatic and radiographic OA has been established[8,9] which supports the need to focus on symptomatic outcomes.

Figure 1. Schematic drawing of an osteoarthritic joint.



Reproduced from Osteoarthritis: an update with relevance for clinical practice, JWJ Bijlsma, F Berenbaum, FPJG Lafeber, Lancet 2011; 377: 2115-26, with permission from Elsevier Ltd.
The different tissues involved in clinical and structural changes of the disease are shown on the left. Note that cartilage is the only tissue not innervated. On the right the bidirectional interplay between cartilage, bone, and synovial tissue involved in osteoarthritis is shown, and the two-way interaction between this interplay and the ligaments and muscles.

Table 1. American College of Rheumatology classification criteria for clinical knee and hip osteoarthritis.

Knee OA	Hip OA
Knee pain for most days of previous month *and* ≥ 3 of the 6 following:	Hip pain for most days of previous month *and*
Age > 50 years	Internal rotation of the hip < 15° *and*
Morning stiffness < 30 minutes	Erythrocyte sedimentation rate ≤ 45 mm/hour
Crepitus on active joint motion	*Or*
Bony tenderness	Hip pain for most days of previous month *and*
Bony enlargement	Internal rotation of the hip ≥ 15° *and*
No palpable warmth	Painful internal rotation of the hip *and*
	Morning stiffness ≤ 60 minutes *and*
	Age > 50 years

The relevance of osteoarthritis

OA is a serious health problem. OA is one of the leading causes of pain and disability, because the symptoms can be quite severe. In addition, it is the most prevalent chronic joint disease, with knee and hip being frequently involved joints.[1,10] The incidence and prevalence of symptomatic knee and hip OA have been rising substantially over the past several decades, mainly concurrent with an ageing population and the growing obesity epidemic.[10,11,12,13] The prevalence of OA varies and depends in particular on the definition used, the joint of interest, and the population measured. In the Netherlands in 2016, reported point prevalence rates in the general population were about 572.000 for knee OA and 397.000 for hip OA, with rates for women being almost twice as high as those for men.[14] In the USA, an estimated 10% of men and 13-18% of women above the age of 60 develop symptoms of knee OA.[11,13,15] The health future exploration 2018 for the Netherlands, shows that OA is the fastest growing disease with an increase from 1.2 million patients in 2015 to 2.2 million in 2040.[16]

OA has become the fastest growing cause of disability worldwide.[12] The pain and loss of function can be debilitating; in developed countries the resultant socioeconomic burden is large, costing between 1.0% and 2.5% of gross domestic product.[18] In the Netherlands, OA is the fourth cause in the elderly for loss of healthy life years.[19] Subsequently, OA will have a growing impact on health care (capacity) and future economic costs.[11,17] Considering the high clinical burden, the high and rising prevalence, and the increasing economic impact, there is a clear need to improve the management of OA. Management of OA could be improved by taking into consideration the following points: the measurement instruments of OA symptoms, identification of risk factors and optimizing treatment modalities.

Measuring symptoms of osteoarthritis

Measuring symptoms is of great importance for the management of OA. The Osteoarthritis Research Society International (OARSI) has recommended to use the three core clinical outcome domains: pain, physical function, and patient global assessment for assessing symptomatic outcomes.[20] Changes in these clinical outcome domains can be evaluated from the perspective of the patient by means of so-called patient reported outcome measures

(PROMs). PROMs allow patients to report how they feel, function and how their quality of life is impacted, without interpretation from healthcare professional or others.[20,21,22] Several measurement instruments are available to measure these PROMs. Importantly, these measurement instruments need to be methodologically sound. This means they need to be based on good quality criteria (i.e. clinimetric properties) which includes: content validity, criterion validity, construct validity, reliability, internal consistency, responsiveness, and interpretability.[23,24] Before its use in research and clinical practice, PROMs should have undergone rigorous testing and proven to be valid, reliable and responsive to change (i.e. be able to detect changes over time).

However, improvement of PROM scores does not necessarily translate into treatment success as perceived by the patient. In clinical research, it is considered to be of great importance to incorporate the patient's interpretation of outcomes in establishing the relevance of findings.[25] In fact, the interpretation of outcomes of clinical trials and the translation of data into daily practice can be conceptualized in two ways: 1) relevant amount of change (improvement of worsening), and 2) reaching an acceptable symptom state. With the first concept, emphasis is on whether or not an individual has changed or improved enough after an intervention to speak of a meaningful change or improvement (minimum clinically important change).[25] Whereas with the second, emphasis is on whether or not the achieved outcome is acceptable from the patient's perspective, often expressed as reaching an absolute value.[25,26] This value beyond which patients consider themselves well or consider their health state to be acceptable is called the patient acceptable symptom state (PASS).[26,27] In conclusion, it is important to take into account the patient's perspective of valid, reliable, and responsive important clinical outcome measures, by evaluating changes or assessing if an acceptable symptom state is reached.

Identifying subgroups

An unmet need for improving the management of OA is identifying risk factors for clinical progression. The natural course of pain and physical function in OA is highly variable and heterogeneous. Most patients appear to remain stable, while others will worsen and some even improve.[28,29,30,31] However, the individual course of OA is difficult to predict, because the underlying mechanisms and risk factors for clinical worsening are still largely unknown.[32] It is generally accepted that identification of risk factors for clinical worsening is important for both patients and clinicians, because these could be used to inform them about the prognosis.[32,33,34,35] Additionally, it is considered important that this identification will enable to improve and target treatments to specific subgroups of patients.[32,33,34,35] Recently two systematic reviews concluded that drawing conclusions on prognostic factors for clinical worsening of OA is hampered by the lack of an unambiguous and validated definition of clinical worsening in OA.[34,35] Therefore, little research data is available about the course and determinants associated with clinical deterioration of knee and hip OA.[32,34,35]

Identification of risk factors could also contribute to the identification of OA phenotypes i.e. distinguishing well-defined subgroups. In recent years it has been debated this would allow the identification of more efficient treatments for example for applying more advanced treatment options to specific subgroups of patients.[32,33,34,35,36] Knee OA is regarded as a heterogeneous disease with multiple etiologies. Because of this heterogeneity, it is considered very likely that there are different forms of OA.[36] This is probably also the reason why the treatment

principle one size fits all is not applicable for OA. Researchers are increasingly suggesting that the identification of phenotypes or subgroups could help us to better understand the driving factors in the development and progression of OA.[36,37,38] Identifying different phenotypes of OA is currently a subject of much research. However, a recent systematic review concluded that no generally evidence based classification system for OA phenotypes exists.[39] The authors proposed several phenotypes: chronic pain-associated OA (in which central mechanisms are prominent); inflammation-associated OA; metabolic syndrome-associated OA; OA associated with joint-localized bone and cartilage metabolism; mechanical load-associated OA; and OA with minimal disease and a low rate of progression.[39] As recently described by Biersma-Zeinstra and van Middelkoop, the challenge is first and foremost to agree on phenotypes that are relevant for diagnosis, prognosis and therapy before this concept of phenotypes can be applied.[37] Thereafter, the most discriminating factors that are easy to use and test in clinical practice could be identified for these phenotypes.

Treatment

Of course, improving treatment options for OA would be a of great value. Since no disease-modifying treatment exists for OA, current OA treatment focuses primarily on the reduction of symptoms as pain and loss of function.[1,40] Treatments can be categorized into non-surgical and surgical treatments. Several international consensus-based clinical guidelines for the management of knee and hip OA are available, emphasizing the importance and efficacy of non-surgical treatment modalities, which include exercise, analgesic use,[40,41] life style education and advice concerning physical activities focusing on improved muscle strength and aerobic capacity[42,43] and advice on weight loss in patients who are overweight.[11,44,45,46] However, in general, limited effect sizes for the non-surgical treatments and therapies of OA have been shown.[44]

When non-surgical treatments do not result in satisfactory reduction in symptoms, surgical options are often considered. These include corrective osteotomy, joint replacement, and joint distraction.[47,48] In general, total joint replacement and osteotomy are considered to be effective treatments in improving pain and function for symptomatic end-stage disease.[49,50] Total knee replacement is less effective in increasing functioning compared to total hip replacement.[47] However, functional outcomes can be poor and the lifespan of prostheses is not unlimited.[11,47,49] A systematic review of 17 prospective studies reported that about 20% of patients kept long-term pain after total knee replacement, and that approximately 9% of patients sustained long-term pain after total hip replacement.[51] There are several potential drawbacks of total knee replacement: the relatively high proportion (10-34%) of patients being dissatisfied after a knee replacement, the higher risks of complications, the limited lifespan of a prosthesis and poorer patient outcomes after revision arthroplasty.[52,53,54,55,56,57] Therefore, it is generally acknowledged that knee replacement should not be performed too early in the disease course.[49,51] Furthermore, in approximately 5% of the patients, surgery cannot be performed due to comorbidities or the patients is not willing to have total joint replacement.[58] Finally, studies regarding joint distraction are promising, but have small sample sizes and short follow-ups.[59] Therefore, more than for hip OA, there is a clear need for more effective non-surgical knee OA treatment options.

Low-dose radiation therapy (LDRT) may be a new non-surgical treatment option for OA patients in whom non-surgical interventions are insufficiently effective and for whom surgical treatments are not (yet) an option. In OA, low grade synovial inflammation is known to play a role. Previous in vitro and in vivo studies of OA in animal models have shown that LDRT exerts anti-inflammatory effects.[60] LDRT is widely used for benign disorders such as knee OA in Germany and Eastern European countries, although the evidence for its effect in clinical practice remains poor.[61,62] Thus, there is a need to improve knowledge about the effectiveness of LDRT as a treatment option for OA.

Aim and outline of this thesis

At the start of the clinical studies performed in the context of this thesis, there was a need to contribute to appropriate outcome measures, identification of risk factors, and evidence of additional non-surgical treatment options for established knee and hip OA. This thesis focuses on clinical outcomes of OA, since these are most important for clinical decision-making. This thesis mainly includes patients with established knee or hip OA and the vast majority were not yet deemed eligible for total joint replacement by their orthopaedic surgeon.

Aims of this thesis are

1. To contribute to the identification of appropriate clinical outcome measures in knee and hip OA that could be applied in future OA research (chapters 2 and 3).
2. To contribute to the identification of risk factors for clinical knee and hip OA that could be applied in OA health care and research, in order to enhance identifying subgroups (chapters 4,5 and 6).
3. To contribute to the knowledge about the effectiveness of low-dose radiation therapy as treatment for knee OA (chapters 7 and 8).

In **chapter 2** we assess the responsiveness (i.e. the ability to detect changes over time) of physical function of four PROMs in patients with knee OA by testing a priori defined hypotheses about expected correlations using the Consensus-based Standards for the selection of health status Measurements Instruments (COSMIN) methodology.
To incorporate the patients' perception in the interpretation of different PROMs assessing physical function, we estimated the patient acceptable symptom state (PASS) values in knee OA patients in **chapter 3**.
In **chapter 4**, we describe the development and validation of preliminary criteria for clinical worsening in knee and hip OA using a literature, expert opinion and data driven approach. Candidate worsening criteria were first tested in a derivation cohort, and finally examined in a validation cohort. These validated clinical worsening criteria are used in **chapter 5** in order to evaluate whether failure of optimal standardized non-surgical treatment is a risk factor for worsening over time.
In **chapter 6**, we describe the results of a cross-sectional study to determine whether the S100A8/A9 (also called calprotectin) serum levels are associated with clinical and structural characteristics in patients with established OA.
In **chapter 7** the results of a systematic literature review on the effectiveness of low dose radiation therapy on pain and functioning in patients with OA are described. In **chapter 8**, the results of a randomized, double-blinded, sham-controlled trial to evaluate the effectiveness of low-dose radiation therapy on symptoms in patients with knee OA are presented.
Finally, **chapter 9** summarizes the results of this thesis and discusses the main findings of this thesis. Furthermore, implications for clinical practice, gained insights and directions for future research are provided. A summary of this thesis in Dutch is provided in **chapter 10**.

References

1. Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice. Lancet 2011;377:2115-26.
2. Sellam J, Berenbaum F. The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. Nat Rev Rheumatol. 2010;6:625–635.
3. Davis MA, Ettinger WH, Neuhaus JM, Mallon KP. Knee osteoarthritis and physical functioning: evidence from the NHANES I epidemiologic followup study. J Rheumatol 1991;18:591–598.
4. Salaffi F, Carotti M, Stancati A, Grassi W. Health-related quality of life in older adults with symptomatic hip and knee osteoarthritis: a comparison with matched healthy controls. Aging Clin Exp Res 2005;17:255–63.
5. Altman R, Asch E, Bloch D, et al. Development of criteria for the classification and reporting of osteoarthritis. Classification of osteoarthritis of the knee. Arthritis Rheum 1986;29:1039–49.
6. Altman R, Alarcon G, Appelrouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum. 1991;34:505–14.
7. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.
8. Bedson J, Croft PR. The discordance between clinical and radiographic knee osteoarthritis: a systematic search and summary of the literature. BMC Musculoskelet Disord. 2008;9:116.
9. Kim C, Nevitt MC, Niu J, et al. Association of hip pain with radiographic evidence of hip osteoarthritis: diagnostic test study. BMJ 2015;351:h5983.
10. Deshpande BR, Katz JN, Solomon DH, Yelin EH, Hunter DJ, Messier SP, et al. Number of Persons With Symptomatic Knee Osteoarthritis in the US: Impact of Race and Ethnicity, Age, Sex, and Obesity. Arthritis Care Res (Hoboken) 2016;68:1743-50.
11. Glyn-Jones S, Palmer AJ, Agricola R, et al. Osteoarthritis. Lancet 2015;386:376-87.
12. Vos T, Flaxman AD, Naghavi M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380:2163-96.
13. Zhang Y, Jordan JM. Epidemiology of osteoarthritis. Clin Geriatr Med 2010;26:355-69.
14. Gommer AM, Poos MJJC, Hulshof T. Prevalentie en aantal nieuwe gevallen van artrose (in Dutch). 2017. Retrieved from www.volksgezondheidenzorg.info.
15. Helmick CG, Felson DT, Lawrence RC, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States: Part I. Arthritis Rheum 2008;58:15–25.
16. The public health future exploration 2018 of the RIVM the Netherlands, www. Vtv2018.nl.
17. Neogi T. The epidemiology and impact of pain in osteoarthritis. Osteoarthritis Cartilage 2013;21:1145-53.
18. Hiligsmann M, Cooper C, Arden N, et al. Health economics in the field of osteoarthritis: an expert's consensus paper from the European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO). Semin Arthritis Rheum 2013; 43:303–13.
19. nationaalkompas.nl. https://www.volksgezondheidenzorg.info/ranglijst/ranglijst-ziekten-op-basis-van-verlies-aan-gezonde-levensjaren-ziektejaarequivalenten-0#node-ranglijst-65-plussers-0.
20. McAlindon TE, Driban JB, Henrotin Y, et al. OARSI Clinical Trials Recommendations: Design, conduct, and reporting of clinical trials for knee osteoarthritis. Osteoarthritis Cartilage. 2015;23:747-60.
21. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61:102-9.
22. Valderas JM, Kotzeva A, Espallargues M, et al. The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. Qual Life Res. 2008;17:179-93.
23. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34-42.
24. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine. Cambridge University Press 2011.
25. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? Ann Rheum Dis 2007;66(Suppl III):iii40-iii41.
26. Tubach F, Ravaud P, Martin-Mola E, Awada H, et al. Minimum clinically important improvement and patient acceptable symptome state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. Arthritis Care Res (Hobroken) 2012;64:1699-707.
27. Tubach F, Wells GA, Ravaud P, Dougados M. Minimal clinically important difference, low disease activity state, and patient acceptable symptom state: methodological issues. J Rheumatol. 2005;32:2025-9.
28. Dekker J, van Dijk GM, Veenhof C. Risk factors for functional decline in osteoarthritis of the hip or knee. Curr Opin Rheumatol. 2009;21:520-4.
29. Holla JF, van der Leeden M, Heymans MW, et al. Three trajectories of activity limitations in early symptomatic knee osteoarthritis: a 5-year follow-up study. Ann Rheum Dis 2014;73:1369-75.
30. Wesseling J, Bierma-Zeinstra SM, Kloppenburg M, Meijer R, Bijlsma JW. Worsening of pain and function over 5 years in individuals with 'early' OA is related to structural damage: data from the Osteoarthritis Initiative and CHECK (Cohort Hip & Cohort Knee) study. Ann Rheum Dis 2015;74:347-53.
31. White DK, Keysor JJ, Lavalley MP, et al. Clinically important improvement in function is common in people with or at high risk of knee OA: the MOST study. J Rheumatol 2010;37:1244-51.
32. Conaghan PG, Kloppenburg M, Schett G, Bijlsma JWJ. EULAR osteoarthritis ad hoc committee. Ann Rheum Dis 2014;73:1442-5.
33. Belo JN, Berger MY, Reijman M, Koes BW, Bierma-Zeinstra SM. Prognostic factors of progression of osteoarthritis of the knee: a systematic review of observational studies. Arthritis Rheum. 2007;57:13-26.
34. de Rooij M, van der Leeden M, Heymans MW, et al. Prognosis of Pain and Physical Functioning in Patients With Knee Osteoarthritis: A Systematic Review and Meta-Analysis. Arthritis Care Res (Hoboken). 2016 ;68:481-92.
35. Bastick AN, Runhaar J, Belo JN, Bierma-Zeinstra SM. Prognostic factors for progression of clinical osteoarthritis of the knee: a systematic review of observational studies. Arthritis Res Ther 2015;17:152.
36. Karsdal MA, Michaelis M, Ladel C, et al. Disease-modifying treatments for osteoarthritis (DMOADs) of the knee and hip: lessons learned from failures and opportunities for the future. Osteoarthritis Cartilage. 2016;24:2013-21.
37. Bierma-Zeinstra SM, van Middelkoop M. Osteoarthritis: In search of phenotypes. Nat Rev Rheumatol. 2017;13:705-6.
38. Felson DT. Identifying different osteoarthritis phenotypes through epidemiology. Osteoarthritis Cartilage. 2010;18:601-4.
39. Deveza LA, Melo L, Yamato TP, Mills K, Ravi V, Hunter DJ. Knee osteoarthritis phenotypes and their relevance for outcomes: a systematic review. Osteoarthritis Cartilage. 2017;25:1926-41.
40. Hochberg MC, Altman RD, April KT, et al, and the American College of Rheumatology. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2012;64:465–74.
41. da Costa BR, Reichenbach S, Keller N, et al. Effectiveness of non-steroidal anti-inflammatory drugs for the treatment of pain in knee and hip osteoarthritis: a network meta-analysis. Lancet. 2017;390:e21-e33.
42. Uthman OA, van der Windt DA, Jordan JL, et al. Exercise for lower limb osteoarthritis: systematic review incorporating trial sequential analysis and network meta-analysis. BMJ 2013; 347: f5555.
43. Fransen M, McConnell S, Harmer AR, Van der Esch M, Simic M, Bennell KL. Exercise for osteoarthritis of the knee. Cochrane Database Syst Rev. 2015;1:CD004376.
44. McAlindon TE, Bannuru RR, Sullivan MC, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage. 2014;22:363-88.

1

45. Gudbergsen H, Boesen M, Lohmander LS, et al. Weight loss is effective for symptomatic relief in obese subjects with knee osteoarthritis independently of joint damage severity assessed by high-field MRI and radiography. Osteoarthritis Cartilage 2012;20:495–502.
46. Felson DT, Zhang Y, Anthony JM, Naimark A, Anderson JJ. Weight loss reduces the risk for symptomatic knee osteoarthritis in women. The Framingham Study. Ann Intern Med 1992;116:535–39.
47. Zhang W, Moskowitz RW, Nuki G, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008; 16: 137–62.
48. van der Woude JA, Wiegant K, van Heerwaarden RJ, et al. Knee joint distraction compared with total knee arthroplasty: a randomised controlled trial. Bone Joint J. 2017;99-B:51-58.
49. Carr AJ, Robertsson O, Graves S, et al. Knee replacement. Lancet 2012; 379:1331-40.
50. Brouwer RW, Huizinga MR, Duivenvoorden T, et al. Osteotomy for treating knee osteoarthritis. Cochrane Database Syst Rev. 2014;12:CD004019.
51. Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. BMJ Open 2012; 22; 2:e000435.
52. Hawker GA, Badley EM, Borkhoff CM et al. Which patients are most likely to benefit from total joint arthroplasty? Arthritis Rheum. 2013;65:1243-52.
53. Cushner F, Agnelli G, FitzGerald G, Warwick D. Complications and functional outcomes after total hip arthroplasty and total knee arthroplasty: results from the Global Orthopaedic Registry (GLORY). Am J Orthop (Belle Mead NJ) 2010;39:22–28.
54. Rissanen P, Aro S, Slätis P, et al. Health and quality of life before and after hip or knee arthroplasty. J Arthroplasty 1995;10:169–175.
55. Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. J Rheumatol 2000;27:1745–52.
56. Scott CE, Howie CR, MacDonald D, Biant LC. Predicting dissatisfaction following total knee replacement: a prospective study of 1217 patients. J Bone Joint Surg Br. 2010;92:1253-8.
57. Anakwe RE, Jenkins PJ, Moran M. Predicting dissatisfaction after total hip arthroplasty: a study of 850 patients. J Arthroplasty. 2011;26:209-13.
58. Frankel L, Sanmartin C, Conner-Spady B, et al. Osteoarthritis patients' perceptions of "appropriateness" for total joint replacement surgery. Osteoarthr Cartil 2012;20:967-73.
59. Flouzat-Lachaniette CH, Roubineau F, Heyberger C, Bouthors C. Distraction to treat knee osteoarthritis. Joint Bone Spine. 2017;84:141-44.
60. Arenas M, Sabater S, Hernández V, et al. Anti-inflammatory effects of low-dose radiotherapy: Indications, dose, and radiobiological mechanisms involved. Strahlentherapie und Onkol 2012; 188:975–81.
61. Leer JWH, Van Houtte P, Davelaar J. Indications and treatment schedules for irradiation of benign diseases: A survey. Radiother Oncol 1998;48:249–57.
62. Seegenschmiedt MH, Micke O, Willich N. Radiation therapy for nonmalignant diseases in Germany: Current concepts and future perspectives. Strahlentherapie und Onkol 2004;180:718–30.

Chapter 2

Responsiveness of four patient-reported outcome measures to assess physical function in patients with knee osteoarthritis

Elien A.M. Mahler
Nienke Cuperus
Johannes W.J. Bijlsma
Thea P.M. Vliet Vlieland
Frank H.J. van den Hoogen
Alfons A. den Broeder
Cornelia H.M. van den Ende

Published in Scandinavian Journal of Rheumatology 2016 Nov;45(6):518-527

Abstract

Objective

The aim of this study was to evaluate the responsiveness of four patient-reported outcome measures (PROMs) to measure change in physical function simultaneously in patients with knee osteoarthritis (OA) following currently recommended COSMIN (COnsensus-based Standards for the selection of health status Measurement INstruments) standards.

Methods

Patients with knee OA receiving conservative treatment following a stepped care approach were invited to complete a set of questionnaires at baseline and 3 months. Questionnaires included four widely used measures of physical function: the Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), the Lequesne algofunctional index (LAI), the Lower Extremity Functional Scale (LEFS), and the Western Ontario and McMaster Universities Osteoarthritis Index Physical Function subscale (WOMAC-PF). Responsiveness of physical function was investigated according to the COSMIN standard by testing 15 a priori defined hypotheses. Responsiveness was considered positive if > 75% of the hypotheses could be confirmed.

Results

A total of 161 patients participated (61% female, mean(SD) age 59(9) years and body mass index 29.7(5.0) kg/m^2. Baseline values of the four PROMs were, mean(SD): KOOS-PS 53.6 (16.8), LAI 11.0 (4.0), LEFS 40.6 (14.1), and WOMAC-PF 51.8 (19.4). We could confirm 12 out of 15 predefined hypotheses (80%) about expected correlations for the WOMAC-PF whereas for the KOOS-PS, LAI, and LEFS < 75% hypotheses could be confirmed (73, 67, and 73% respectively).

Conclusion

Our results suggest that the WOMAC-PF is able to detect changes over time in physical function and therefore should be the measure of first choice in clinical trials evaluating the effectiveness of an intervention on physical function in knee OA patients.

Introduction

Knee osteoarthritis (OA) is a common chronic disorder affecting all tissues of the knee joint and causing mainly pain and reduced physical function.[1-3] Physical function is therefore one of the core outcome dimensions in clinical practice and research in knee OA.[4] As no curative treatment is currently available for knee OA[2], treatment is usually aimed at improving symptoms (function, pain, stiffness) using a variety of pharmacological and non-pharmacological interventions including surgery.[4] The effectiveness of such interventions is frequently evaluated by clinicians in daily practice and researchers in clinical trials using core outcome measures that address the domains of pain and function as advocated.[5,6] Therefore the ability of a measurement instrument to detect changes over time (i.e. the responsiveness) regarding physical function, is of particular importance when selecting an instrument.[7-9]

Several patient-reported outcome measures (PROMs), either disease-specific or specific for musculoskeletal conditions affecting lower extremity function, are available for evaluating physical function in patients with knee OA. Among those, the Lequesne algofunctional index (LAI) and the physical function subscale of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC-PF) are recommended as measures of first choice in OA trials.[10,11] However, in two systematic reviews on measurement properties, it was concluded that the responsiveness of those measures was questionable as none of the included studies presented hypotheses relating to the magnitude of the change.[11,12] However, a new disease-specific measure and a measure specific for musculoskeletal conditions affecting lower extremity function have been developed to evaluate physical function in patients with knee OA: the short version of the Knee Injury and Osteoarthritis Outcome Score (KOOS-PS)[13] and the Lower Extremity Functional Scale (LEFS), respectively.[14] The selection of a measurement instrument is dependent on the measurement properties regarding reliability, validity, responsiveness, and interpretability.[8] Although the reliability and validity of these PROMs have been studied previously[13-20], so far no studies have examined the responsiveness of these latter instruments head-to-head in a single study.

A large number of definitions and methods have been proposed for assessing responsiveness, but the optimal method for evaluating responsiveness is still under debate.[21] There is growing consensus that responsiveness should be considered as a measure of longitudinal construct validity because responsiveness refers to measuring changes in the construct to be measured. Hence, responsiveness should, analogous to construct validity, preferably be evaluated by testing predefined hypotheses about expected correlations between changes in related measurements (convergent) or unrelated measures (discriminant) or expected between-group differences in changes (discriminative).[9,11] With this approach, the validity of the change scores can be assessed in contrast to the magnitude of the change scores as assessed by traditionally accepted methods such as the effect size (ES) or the standardized response mean (SRM).[21] The latter methods are known to have some disadvantages: (i) the effect is dependent on the patient group in which the measure is being calculated, (ii) the ES and the SRM are based on statistical considerations rather than on patients' judgement of what constitutes an important change and are also influenced by the sample size, and (iii) the ES is known to be related to the treatment effect rather than the quality of a measurement instrument.[21]

To our knowledge, no studies have evaluated the responsiveness of PROMs assessing physical function by postulating a priori hypotheses using the COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) methodology in patients with knee OA head-to-head.[9] In the current study we assessed the responsiveness of physical function of the four PROMs KOOS-PS, LAI, LEFS, and WOMAC-PF in patients with knee OA by testing predefined hypotheses about expected correlations between changes in measurements or expected differences in changes.

Methods

Participants and intervention

Recruitment to the study took place between July 2012 and January 2014 and consecutive patients, referred by orthopaedic surgeons, attending our specialty knee and hip rheumatology OA outpatient clinic were invited for this observational study. All patients fulfilled the American College of Rheumatology (ACR) clinical OA criteria (knee pain (> 15 days of last month) plus at least three of the following: age > 50 years, morning stiffness < 30 minutes, crepitus, bony tenderness, bony enlargement, or no palpable warmth) and were invited the day after their first outpatient visit to our department.[22,23] The exclusion criterion was: short-term indication (within 3 months) for knee replacement surgery. The local Medical Research Ethics Committee, region Arnhem-Nijmegen (The Netherlands) approved the study design (study number 2012/375). All patients gave their written informed consent to participate in the study.

Stepped care approach

Several international consensus-based clinical guidelines for the management of knee OA are available, emphasizing the importance and efficacy on non-surgical treatment modalities. Therefore, we selected patients who all received multimodal conservative treatment based on a published Dutch multidisciplinary stepped care approach for diagnosis and treatment of knee and hip OA.[24] At the rheumatology outpatient department, during a 90 min group visit (4-6 patients) led by a physician assistant and a specialized nurse and supervised by a rheumatologist (as described elsewhere[25]), patients received education, referral for physical therapy (prescription for both aerobic and strengthening exercises according to the graded activity principle, no instruction about group classes or individual instruction), step-up analgesics guided by a patient's pain level (change of policy regarding pain medication was considered in case of pain > 4 on a Numeric Rating Scale (NRS)), lifestyle advice concerning physical activity and advice on weight loss in patients with a BMI ≥28 kg/m^2. In this group visit, most components of the individual visits were retained such as private one-to-one conversations, while creating more time for patient education and discussion.[26] If a prescription for analgesics was considered necessary, we started with paracetamol (acetaminophen) at a fixed dose of 1000 mg three times a day. In case of lack of efficacy or recent consistent use of paracetamol, a non-steroidal anti-inflammatory drug (NSAID) was added. After 4 weeks, patients were contacted by telephone and if necessary the analgesics were switched.

Data acquisition

At baseline and at 3 months, patients completed a standardized set of questionnaires including sociodemographic information (age, sex, duration and localization of symptoms).

The set of postal questionnaires included four PROMs to measure physical function. Other outcome measures included in the set of questionnaires were derived from hypotheses that were a priori postulated by an expert group (see section on responsiveness). To prevent a learning effect as much as possible, we balanced the order of the questionnaires by varying questionnaires measuring physical function with questionnaires measuring other constructs. A reminder was sent to those patients who did not respond within 3 weeks. Patients were included in the analysis of the current study if they completed both the baseline and follow-up measurements.

PROMs

KOOS-PS. The seven-item short measure of physical function (KOOS-PS) is derived from the subscale "activities of daily living" (four items) and "sport and recreation" (three items) of the KOOS.14 With every item, patients rate the degree of difficulty they have experienced over the previous week due to their knee pain on a five-point Likert scale (0-4). Raw scores can range from 0 to 28 and in this study were converted to normalized scores ranging from 0 to100, with a score of 0 indicating no difficulty. The KOOS-PS has been shown to be valid and reliable in knee OA.[13,15-17]

LAI. The LAI is an 11-item questionnaire measuring the degree of functional disability, with four questions pertaining to activities of daily living.27 The total score can range from 0 to 24 points. The degree of functional disability corresponds with the following scores: a score ≥ 14 points indicates extremely severe disability, a score of 11-13 very severe disability, a score of 8-10 severe disability, a score of 5-7 moderate disability, and a score of 1-4 minimal disability. The LAI has demonstrated good reliability, whereas its validity has been questioned because of its different impacts in different interventions.[18]

LEFS. The LEFS is a 20-item condition-specific questionnaire designed for use in patients with musculoskeletal conditions affecting lower extremity function, including knee OA.[28] The instrument asks patients about their ability to perform general activities of daily living, general recreational activities, specific daily physical tasks, and specific recreational or occupational related tasks on the day of completing the questionnaire(0-4). The total score ranges from 0 to 80 points, with higher scores representing higher levels of functioning. Although the LEFS is not designed as a disease-specific questionnaire, it measures the same construct physical function and the Dutch LEFS questionnaire has been validated and shown to be reliable in knee OA patients.[14,19]

WOMAC-PF. The KOOS (www.koos.nu) includes the WOMAC OA index in its complete and original format (with permission). We used the 17-item subscale Activities of Daily Living (Likert scale version) from the KOOS to calculate the WOMAC physical function (WOMAC-PF) score (and thus four items are overlapping with the KOOS-PS). Standardized scores ranging from 0 to100 were used, where higher scores reflect better health status. The WOMAC has been the most extensively studied instrument in individuals with knee OA and has been shown to be valid and reliable.[20]

Patient's assessment of change. In the follow-up questionnaire, patients were asked to rate the extent to which their daily function and pain had been changed since the start of the

treatment on a seven-point Likert scale; that is, very much worsened, much worsened, slightly worsened, no change, slightly improved, much improved and very much improved.

Other PROMs. In addition to the above-mentioned PROMs, patients were asked to rate their functioning and pain in the preceding week on a 0-10-point Numeric Rating Scale (NRS) where 0 equals no symptoms. The patient's global assessment (PGA) of knee OA impact during the last week before visit was also assessed on this 0-10 point NRS scale. Moreover, physical function and mental health were measured with corresponding subscales of the 36-item Short Form Health Survey (SF-36), a widely used generic health status questionnaire comprising eight areas of health status, with higher scores indicating better health (range 0-100).[29,30] Fatigue was measured with the eight-item "Subjective Fatigue" subscale of the Checklist Individual Strength (CIS).[31] The total score can range from 8 to 56 points, with scores of ≥ 35 representing severe fatigue. Self-efficacy was assessed with the Dutch General Self-efficacy Scale (DGSS) and higher scores, ranging from 10 to 40, reflect higher self-efficacy.[32] Pain coping was assessed with the Pain Coping Inventory List (PCI)[33], which is a 33-item questionnaire that measures active and passive pain-coping strategies on a four-point Likert scale ranging from 'hardly ever' to 'very often'. Higher scores on the subscales active or passive coping indicate more use of an active or passive coping style. The Hospital Anxiety and Depression Scale (HADS) was used to assess anxiety and depression.[34-36] Both subscales consist of seven items with possible scores ranging from 0 to 21 for each subscale. Higher scores indicate higher levels of disorder.

Responsiveness

The COnsensus-based Standards for the selection of health status Measurement INstruments (COSMIN) and the quality criteria for measurement properties as proposed by Terwee and coworkers were followed for the assessment of responsiveness.[8,9] We defined responsiveness as the ability of an instrument to detect changes over time in the construct to be measured.[8,21] As we considered responsiveness as a measure of longitudinal validity and because of the lack of a gold standard, the basic approach we used to assess responsiveness was to postulate and test predefined specific hypotheses formulated by an expert group in analogy to construct validity.[8] These a priori defined hypotheses addressed expected correlations between changes in scores on the KOOS-PS, LAI, LEFS, and WOMAC-PF and changes in scores on other clinical (un) related measures (pain, mental health, fatigue, self-efficacy, coping, anxiety, and depression) or expected differences in correlation in changes between groups. In this way, responsiveness is independent of the magnitude of a change but measures changes in the concept being measured. However, in responsiveness studies, lower correlation coefficients are often found compared to those in construct validity studies, which can be explained by the fact that in the former, a correlation coefficient is obtained between change scores of two measurement instruments.[37] We installed an expert group consisting of researchers, epidemiologists, and physical therapists with published studies on OA and well versed with the current literature providing PROMs and a rheumatologist with clinical and research experience in the field of knee OA. Members of the expert group independently formulated hypotheses that were discussed in a group meeting until consensus was reached. In these hypotheses, the magnitude and direction of the correlation coefficients had to be clearly defined. In addition, the correlations had to address between-group changes in physical function and changes in scores on related measures (convergent), unrelated measures (discriminant) or differences between groups (discriminative).38 In addition, we added one hypothesis on the size of the area under of the receiver operating characteristic (ROC) curve (AUC), measuring the ability of a questionnaire to distinguish between patients who have and have not changed, according to an external anchor.[9] We calculated the ROC curve for the improved patients using the change in physical function scores of the PROMs and the patients' rating of change in physical function assessed by the transition question. Because a gold standard for change of physical function is lacking, we used a global rating scale as the gold standard for measuring change as currently recommended.[37] The rating of change was dichotomized to identify patients who were (very) much improved in physical function and remained stable (slightly improved, not changed, slightly worse). A correlation of $-0.3 \leq r \leq 0.3$ was considered weak whereas a correlation of $r > 0.3$ or $r < -0.3$ was considered moderate. An AUC of at least 0.70 was considered adequate.[10] Responsiveness was considered positive if >75% of the hypotheses were confirmed.[9]

Statistical analysis

The distributions of the study variables were inspected. Descriptive statistics are provided as mean and standard deviation (SD) or median and interquartile range (IQR) where appropriate, for continuous variables and numbers with percentages for categorical variables. Descriptive statistics were used to describe the study population. Paired t-tests were used to compare baseline with 3-month values. First, the scale scores of the KOOS-PS, LAI, LEFS, and WOMAC-PF were assessed for normality and missing data. We assessed floor and ceiling effects for each questionnaire at baseline; these were considered present if >15% of the patients scored the best or the worst possible score, respectively. Where appropriate, Pearson or Spearman rank correlation coefficients were computed to test the hypotheses. Evaluation of the hypotheses regarding differences in strength between different constructs was performed if the correlation coefficients were significantly different from each other by the Meng et al test. Assessment of the hypotheses regarding differences between groups (discriminative) was carried out by testing ROC curves which were estimated for the change in KOOS-PS, LAI, LEFS, and WOMAC-PF to distinguish between patients who indicated they were very much and much improved and those who indicated they were stable in physical function after 3 months (slightly worse, no change, slightly improved). All analyses were performed using STATA version 13.1.

Results

Table 1. Sociodemographic and disease-related characteristics of the 161 patients with knee OA.

Sociodemographic characteristics	
Female, n (%)	99 (61)
Age, years	59 (9)
Body Mass Index (BMI), kg/m2	29.7 (5.0)
Localization of symptoms, n (%):	
Left knee	39 (24))
Right knee	56 (35)
Both sides	66 (41)
Duration of complaints > 5 years, n (%)	59 (37)
PROMs regarding physical function	
KOOS-PS (range 0-100)	53.6 (16.8)
LAI (range 0-24)	11.0 (4.0)
LEFS (range 0-80)	40.6 (14.1)
WOMAC-PF (range 0-100)	51.8 (19.4)
Other PROMs	
NRS function (range 0-10)	5.5 (2.4)
NRS pain (range 0-10)	5.6 (2.1)
NRS PGA (range 0-10)	5.6 (2.6)
Physical function (SF-36) (range 0-100)	31.8 (9.1)
Mental health (SF-36) (range 0-100)	39.8 (4.7)
Fatigue (CIS) (range 8-56)	34.4 (11.8)
Self-efficacy (DGSS) (range 0-100)	32.3 (5.0)
Active coping (PCI) (range 0-4)	2.3 (0.5)
Passive coping (PCI) (range 0-4)	2.0 (0.4)
Anxiety (HADS) (range 0-21)	6.0 (4.4)
Depression (HADS) (range 0-21)	5.8 (3.8)
Changes in daily function rated on the transition question, n (%)	
Very much improved	4 (2.5)
Much improved	13 (8.1)
Slightly improved	28 (17.5)
No change	59 (36.9)
Slightly worsened	42 (26.3)
Much worsened	13 (8.1)
Very much worsened	1 (0.6)
Changes in pain rated on the transition question, n (%)	
Very much improved	4 (2.5)
Much improved	21 (13.1)
Slightly improved	27 (16.9)
No change	44 (27.5)
Slightly worsened	44 (27.5)
Much worsened	17 (10.6)
Very much worsened	3 (1.9)

OA: osteoarthritis; BMI: body mass index; PROMs: patient reported outcome measures; KOOS-PS: Knee Injury and Osteoarthritis Outcome Physical Function Short Form; LAI: Lequesne algofunctional index; LEFS: Lower Extremity Functional Scale; WOMAC-PF: Western Ontario and McMaster Universities Osteoarthritis Index Physical Function subscale; NRS: Numeric Rating Scale; PGA: Patient Global Assessment; SF-36: Short Form 36 Health Survey; CIS: Checklist Individual Strength; DGSS: Dutch General Self Efficacy; PCI: Pain Coping Inventory; HADS: Hospital Anxiety and Depression Scale.
Higher scores indicate worse physical function for the KOOS-PS and LAI and higher scores indicate better physical function for the LEFS and WOMAC-PF.
Higher scores indicate worse NRS function and more NRS pain, worse PGA, better physical function SF-36, better mental health SF-36, more fatigue, higher self-efficacy, more frequent use of coping strategies, more pronounced levels of anxiety and depression.
Data are shown as mean (SD) unless indicated otherwise.

Patient characteristics

In total, we invited 272 patients to participate in the current study, of whom 185 (68%) completed the baseline measurements. Of these patients, 161 (87%) completed the follow-up measurements and were included in the current analysis. No differences were found between the participants and non-participants with regard to sex, whereas the group of participants were significantly older than the non-participants (59 years vs. 56 years, p = 0.02). The majority of the cohort consisted of women (61%), with a mean age of 59 years (sd = 9.3) and a mean BMI of 29.7 kg/m^2 (sd = 5.0). The demographic and disease-related characteristics at baseline are presented in Table 1. Scores on the KOOS-PS, LAI, LEFS, and WOMAC-PF were normally distributed at both baseline and follow-up. At the scale level of these instruments, there were ≤ 5% missing values at both time points. Ceiling or floor effects were not present regarding the measures of KOOS-PS, LAI, LEFS, and WOMAC-PF.

The 3-month follow-up

For the whole group of patients, the majority of the PROMs remained stable after 3 months but significant improvements were observed in physical function measured with the KOOS-PS (p=0.03) and for fatigue, assessed by the CIS (p=0.03). The majority of patients (n=129, 80.6%) indicated on the transition question that they had remained stable in their daily function after 3 months (Table 1). Fourteen (8.8%) patients indicated (very) much worsening in their physical function. The 17 patients (10.6%) who indicated (very) much improvement showed significant improvements in physical function after 3 months in all four PROMs compared to baseline (p < 0.05) and in most of the other assessed PROMs (NRS function, NRS pain, NRS PGA, physical function SF-36, mental health SF-36, fatigue, passive coping and depression, data not shown). However, success on advice to exercise, adherence to analgesics, weight loss, and lifestyle changes were not evaluated.

Table 2. Correlation coefficients (r, 95% CI) between changes (Δ) in physical function in KOOS-PS, LAI, LEFS, and WOMAC-PF scores and changes (Δ) in scores on related and unrelated constructs.

	Δ KOOS-PS		Δ LAI		Δ LEFS		Δ WOMAC-PF	
	r	95% CI	r	95% CI	r	95% CI	r	95% CI
Δ NRS physical function	**0.31**	0.16 to 0.46	**0.49**	0.36 to 0.60	**-0.45**	-0.56 to -0.31	**-0.45**	-0.56 to -0.31
Δ NRS pain	**0.33**	0.17 to 0.46	**0.49**	0.36 to 0.60	**-0.45**	-0.57 to -0.32	**-0.45**	-0.57 to -0.31
Δ Physical function (SF-36)	-0.40	-0.53 to -0.25	**-0.44**	-0.56 to -0.31	**0.53**	0.40 to 0.63	**0.47**	0.34 to 0.59
Δ Mental health (SF-36)	-0.10	-0.26 to 0.06	-0.12	-0.27 to 0.04	**0.17**	0.01 to 0.32	0.15	-0.00 to 0.30
Δ Fatigue (CIS)	**0.22**	0.06 to 0.37	**0.21**	0.06 to 0.36	**-0.21**	-0.36 to -0.05	**-0.17**	-0.32 to -0.01
Δ Self-efficacy (DGSS)	-0.07	-0.23 to 0.09	0.11	-0.05 to 0.27	0.01	-0.15 to 0.16	-0.01	-0.17 to 0.14
Δ Active coping (PCI)	0.09	-0.07 to 0.25	0.13	-0.03 to 0.29	-0.07	-0.22 to 0.09	-0.09	-0.24 to 0.07
Δ Passive coping (PCI)	**0.24**	0.08 to 0.40	**0.32**	0.17 to 0.46	**-0.24**	-0.38 to -0.09	**-0.20**	-0.35 to -0.05
Δ Anxiety (HADS)	**0.25**	0.08 to 0.40	0.16	-0.00 to 0.31	-0.14	-0.29 to 0.02	**-0.16**	-0.31 to -0.00
Δ Depression (HADS)	**0.26**	0.10 to 0.41	**0.22**	0.06 to 0.37	**-0.18**	-0.33 to -0.02	**-0.20**	-0.35 to -0.05
Transition physical function	**-0.33**	-0.47 to -0.18	**-0.42**	-0.54 to -0.28	**0.53**	0.41 to 0.64	**0.41**	0.27 to 0.53
Transition pain	**-0.28**	-0.42 to -0.12	**-0.37**	-0.50 to -0.23	**0.48**	0.34 to 0.59	**0.34**	0.20 to 0.47

CI: confidence interval; KOOS-PS: Knee Injury and Osteoarthritis Outcome Physical Function Short Form; LAI: Lequesne algofunctional index; LEFS: Lower Extremity Functional Scale; WOMAC-PF: Western Ontario and McMaster Universities Osteoarthritis Index Physical Function subscale; NRS: Numeric Rating Scale; SF-36: Short Form 36 Health Survey; CIS: Checklist Individual Strength; DGSS: Dutch General Self Efficacy; PCI: Pain Coping Inventory; HADS: Hospital Anxiety and Depression Scale.
Higher scores indicate worse physical function for the KOOS-PS and LAI and higher scores indicate better physical function for the LEFS and WOMAC-PF.
Higher scores indicate worse NRS function and more NRS pain, better physical function SF-36, better mental health SF-36, more fatigue, higher self-efficacy, more frequent use of coping strategies, more pronounced levels of anxiety and depression.
Significant correlations($p < 0.05$) shown in bold.

Table 3. Results of predefined hypotheses to assess the responsiveness of the KOOS-PS, LAI, LEFS, and WOMAC-PF in knee OA.

Hypothesis		KOOS-PS	LAI	LEFS	WOMAC-PF
1.	There is at least a moderate correlation (r > 0.3) with Δ NRS function	yes	yes	yes	yes
2.	There is a weak correlation r ≤ 0.3 with Δ NRS pain	no	no	no	no
3.	There is at least a moderate correlation (r > 0.3) with Δ SF-36 subscale physical function	yes	yes	yes	yes
4.	There is a weak correlation r ≤ 0.3 with Δ SF-36 subscale mental health	yes	yes	yes	yes
5.	There is a weak correlation r ≤ 0.3 with Δ fatigue	yes	yes	yes	yes
6.	There is a weak correlation r ≤ 0.3 with Δ self-efficacy	yes	yes	yes	yes
7.	There is a weak correlation r ≤ 0.3 with Δ active coping	yes	yes	yes	yes
8.	There is a weak correlation r ≤ 0.3 with Δ passive coping	yes	no	yes	yes
9.	There is a weak correlation r ≤ 0.3 with Δ anxiety	yes	yes	yes	yes
10.	There is a weak correlation r ≤ 0.3 with Δ depression	yes	yes	yes	yes
11.	There is at least a moderate correlation (r > 0.3) with transition physical function	yes	yes	yes	yes
12.	There is a weak correlation r ≤ 0.3 with transition pain	yes	no	no	no
13.	The correlation with ΔNRS function is stronger than the correlation with ΔNRS pain	no	no	no	no
14.	The correlation with transition function is stronger than the correlation with transition pain	no	no	no	yes*
15.	The ability to distinguish patients who were (very) much improved and who were stable in physical function (AUC ≥ 0.7)	no	yes	yes	yes
	Confirmed, n/n (%)	11/15 (73)	10/15 (67)	11/15 (73)	12/15 (80)

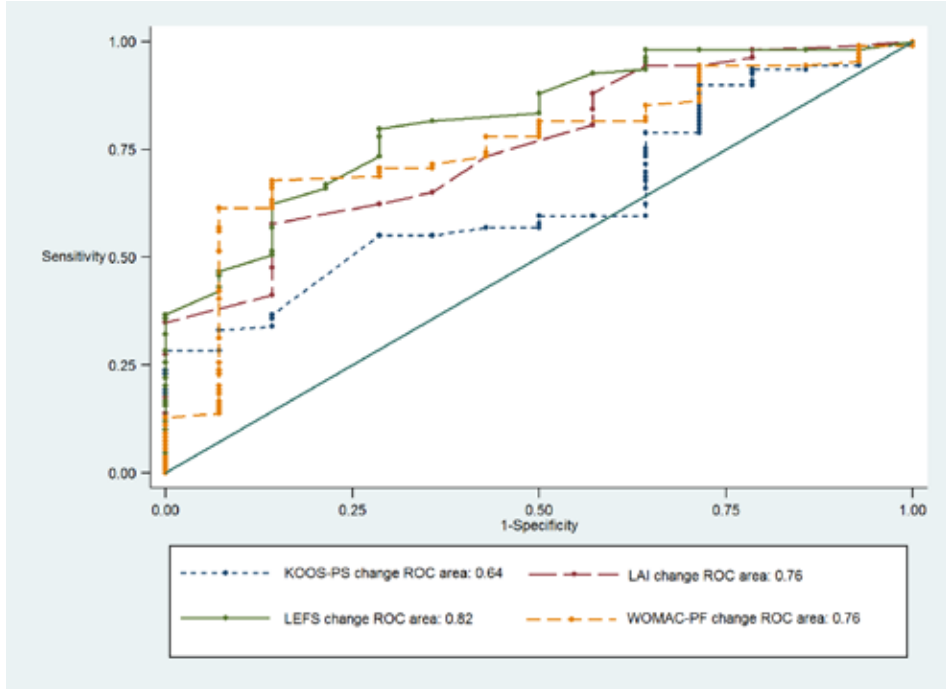
OA: osteoarthritis; KOOS-PS: Knee Injury and Osteoarthritis Outcome Physical Function Short Form; LAI: Lequesne algofunctional index; LEFS: Lower Extremity Functional Scale; WOMAC-PF: Western Ontario and McMaster Universities Osteoarthritis Index Physical Function subscale; Δ: change in; NRS: Numeric Rating Scale; SF-36: Short Form 36 Health Survey; AUC: area under the receiver operating characteristic curve.
Higher scores indicate worse physical function for the KOOS-PS and LAI and higher scores indicate better physical function for the LEFS and WOMAC-PF.
Higher scores indicate worse NRS function and more NRS pain, better physical function SF-36, better mental health SF-36, more fatigue, higher self-efficacy, more frequent use of coping strategies, more pronounced levels of anxiety and depression.
* Significantly stronger correlation, $p < 0.05$.

Responsiveness

Table 2 presents the correlation coefficients between changes in physical function on the four PROMs and changes in scores on related and unrelated constructs. We could confirm 12 out of 15 predefined hypotheses (80%) about expected (or absence of) correlations using the WOMAC-PF. The responsiveness of the WOMAC-PF could therefore be established as > 75% of the hypotheses were confirmed (Table 3). For the KOOS-PS, LAI and LEFS, respectively, 11 (73%), 10 (67%), and 11 (73%) hypotheses were confirmed and the 75% according to our definition of positive rating for responsiveness was not reached.

For all four PROMs, we were unable to confirm the hypothesized weak correlation between change in physical function and change in NRS pain since we found a moderate correlation ranging between 0.33 and 0.49 (hypothesis 2). Hypotheses 5 to 10 concerning unrelated measures (discriminant responsiveness) were almost all confirmed except for a weak correlation between changes in LAI and changes in passive coping, which could not be confirmed. Concerning hypothesis 12, for three of the four questionnaires we were unable to confirm the hypothesized weak correlation between change in physical function and transition in pain since we found a moderate correlation above 0.33, except for the KOOS-PS, where the weak correlation with transition in pain was confirmed (r = -0.28). In addition, for all questionnaires we were unable to confirm a higher strength of correlation between change in physical function and change in NRS function compared with the correlation between change in physical function and change in NRS pain (hypothesis 13). Only for the WOMAC-PF could we confirm the stronger correlation between change in physical function and transition function than the correlation with transition pain (hypothesis 14). The LAI, LEFS, and WOMAC-PF showed good ability to distinguish between patients who were (very) much improved and those who were stable (slightly worsened, no change, slightly improved) in physical function with an AUC of ≥ 0.7 (Figure 1).

Figure 1. ROC curves showing the sensitivity and 1-specificity of the change in four PROMs assessing physical function in patients who indicated they were (very) much improved compared with patients who indicated they were stable after 3 months.



KOOS-PS: Knee Injury and Osteoarthritis Outcome Physical Function Short Form; LAI: Lequesne algofunctional index; LEFS: Lower Extremity Functional Scale; WOMAC-PF: Western Ontario and McMaster Universities Osteoarthritis Index Physical Function subscale.

Discussion

This is the first study to comprehensively evaluate the responsiveness of the four PROMs KOOS-PS, LAI, LEFS, and WOMAC-PF to assess physical function in patients with clinical knee OA. According to the COSMIN standard, we tested predefined hypotheses about expected (or absence of) correlations between changes in constructs. Our results suggest that the WOMAC-PF is able to detect changes over time in physical function and should therefore be the first-choice measure in clinical trials evaluating the effectiveness of an intervention in knee OA patients.

An intriguing finding in our study that warrants closer inspection is that, contrary to our hypotheses, the strength of association of changes in physical function with changes in NRS pain and transition in pain was stronger than hypothesized, for all four PROMs. Whether physical function and pain can be assessed independently with PROMs is much debated.[14,19,38-41] Our findings suggest that the PROMs examined in this study suffer from construct contamination when assessing physical functioning. It has been suggested that performance-based physical functioning is less influenced by pain than self-reported physical functioning, and hence

that performance-based measures are probably better than questionnaires in capturing the construct of physical functioning.[38,42] However, several disadvantages of performance-based methods have been considered. They measure physical functioning in an artificial situation, are influenced by the subject's motivation to participate, and may provide little information about how a person copes in their own environment.[43] It has been argued that performance-based and self-reported questionnaires measure different aspects of function and offer complementary information.[38] So far, there is little insight into the measurement properties of performance-based measures in knee OA.[38,44,45] Further research addressing the added value of performance-based measures above questionnaires to measure physical function in knee OA is warranted.

Our study has several strengths. A stringent protocol consistent with the latest COSMIN standard was prepared and followed in which an anchor-based approach was used to assess responsiveness with 15 predefined hypotheses. We assessed responsiveness based on hypotheses regarding correlations (i) between changes in physical function and changes in scores on related measures (convergent), and (ii) between changes in physical function and unrelated measures (discriminant), and (iii) regarding differences between groups (discriminative). In addition, the responsiveness of the four PROMs was measured at the same time in the same population of patients with knee OA receiving the same treatment advice regarding conservative treatment, which increased the accuracy of the assessment of the four PROMs.[8,11,12] The results of previous studies assessing responsiveness were based on traditional methods(e.g. the ES and the SRM) that are known to be dependent of the type of intervention. This could explain the inconsistency of previous findings on the responsiveness of PROMs assessing physical function in knee OA. For example, the LAI has been found to be more responsive than the WOMAC-PF in patients with knee OA following exercise therapy and rehabilitation[18,27], whereas the WOMAC-PF subscale was found to be more responsive after hip or knee replacement[46], but not superior to the LEFS[39]. Because the responsiveness measured with the COSMIN standard is not expected to impact differently in different interventions or patient groups, further research is warranted to determine, preferably in a head-to-head study, whether our results are also valid for other, perhaps more effective, interventions. Importantly, all patients were included from daily clinical practice, which in our opinion provides external generalizability of our results.

Some limitations of our study and of studies assessing responsiveness in general should be acknowledged. First, the COSMIN standard was developed as a reporting tool for quality criteria of health status questionnaires, not for the purpose of rating the instruments. However, these criteria are also recommended for use in design validation studies of health questionnaires and the standard formulated an absolute cut-off point of > 75% of the hypotheses to be confirmed to consider responsiveness as positive.9 Furthermore, defining hypotheses remains arbitrary regarding the number of predefined hypotheses and the magnitude and direction of the correlation coefficients defined. To avoid this, we used a transparent method and clearly predefined hypotheses about the magnitude and direction of the correlation coefficients. It turned out that the majority addressed discriminant hypotheses regarding correlations with unrelated constructs. Probably more convergent hypotheses would have been more desirable with, for example, more hypotheses concerning the same construct measured with performance-based measures. All hypotheses were equally important and thus counted equally for the overall assessment that at least 75% of all hypotheses should be confirmed. For

both the KOOS-PS and the LEFS, 11 hypotheses were confirmed (73%). This is only one confirmed hypothesis less than the 12 confirmed hypotheses for the WOMAC-PF (80%). To date, there is no consensus or guideline about the nature and number of hypotheses that should be tested and confirmed or about weighted testing of hypotheses (i.e. giving convergent hypotheses greater weight). There is also no consensus about considering the use of the lower or upper bound of the 95% confidence interval of an association. It would be interesting to take this issue into account in future research or in an update of the COSMIN standard. Other aspects of the questionnaires were also not taken into account. Regarding hypothesis 15, the use of a single-item anchor may have been less robust than using a composite score of a multi-item change measure. However, this multi-item score for change of physical function is lacking. A response shift could also have influenced our results, although we deemed a time frame of 3 months long enough to detect improvement and short enough to minimize the risk of a response shift.[47,48] Furthermore, the questionnaires are validated for different time frames and it remains unclear if these different time frames might have influenced the correlation coefficients between the change scores (LEFS refers to function on the day and KOOS-PS and WOMAC-PF to function in the previous week). In addition, because of the homogeneous population, the results of our study seem to be confined to patients with symptomatic knee OA in secondary care. Nevertheless, our cohort is comparable to other cohorts, consisting mainly of overweight women with knee OA.[49,50] In addition, because we studied the responsiveness head-to-head, it is unlikely that our sample characteristics could have influenced the results. Finally, fewer patients than expected indicated (very) much improvement in their daily function after 3 months.[25] Although the COSMIN standard stresses that responsiveness is independent of the magnitude of the change, it is possible that the relatively small number of improved patients could have influenced the strength of the correlations we found.

In conclusion, based on the COSMIN standard, which stresses that responsiveness is independent of the magnitude of the change, our results suggest that the WOMAC-PF could be seen as the measure of first choice in clinical trials evaluating the effectiveness of an intervention to assess physical function in patients with knee OA. Further research is warranted to assess, preferably in a head-to-head study, whether our results are valid for other, perhaps more effective, interventions as well. Future research is necessary to determine whether the use of different predefined hypotheses, including hypotheses concerning worsening, yield the same results on responsiveness.

Acknowledgement
We thank all patients who participated in this study and the following members from the expert group (apart from the authors): TJ Hoogeboom and JE Vriezekolk.

2

References

1. Arden N, Nevitt MC. Osteoarthritis: epidemiology. Best Pract Res.Clin Rheumatol. 2006;20:3-25.
2. 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;133:726-37.
3. Jinks C, Jordan K, Croft P. Osteoarthritis as a public health problem: the impact of developing knee pain on physical function in adults living in the community: (KNEST 3). Rheumatology.(Oxford) 2007;46:877-81.
4. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines fot the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage. 2014;22:363-88.
5. Altman R, Brandt K, Hochberg M, Moskowitz R, Bellamy N, Bloch DA, et al. Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Results from a workshop. Osteoarthritis Cartilage. 1996;4:217-43.
6. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage. 2004;12:389-99.
7. Guyatt GH, Deyo RA, Charlson M, Levine MN, Mitchell A. Responsiveness and validity in health status measurement: a clarification. J Clin Epidemiol. 1989;42:403-8.
8. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63:737-45.
9. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34-42.
10. 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;16:185-98.
11. Veenhof C, Bijlsma JW, van den Ende CH, van Dijk GM, Pisters MF, Dekker J. Psychometric evaluation of osteoarthritis questionnaires: a systematic review of the literature. Arthritis Rheum. 2006;55:480-92.
12. Dobson F, Hinman RS, Hall M, Terwee CB, Roos EM, Bennell KL. Measurement properties of performance-based measures to assess physical function in hip and knee osteoarthritis: a systematic review. Osteoarthritis Cartilage. 2012;20:1548-62.
13. Perruccio AV, Stefan LL, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. Osteoarthritis Cartilage. 2008;16:542-50.
14. Hoogeboom TJ, de Bie RA, den Broeder AA, van den Ende CH. The Dutch Lower Extremity Functional Scale was highly reliable, valid and responsive in individuals with hip/knee osteoarthritis: a validation study. BMC Musculoskelet Disord. 2012;13:117.
15. Bond M, Davis A, Lohmander S, Hawker G. Responsiveness of the OARSI-OMERACT osteoarthritis pain and function measures. Osteoarthritis Cartilage. 2012;20:541-7.
16. Ruyssen-Witrand A, Fernandez-Lopez CJ, Gossec L, Anract P, Courpied JP, Dougados M. Psychometric properties of the OARSI/OMERACT osteoarthritis pain and functional impairment scales: ICOAP, KOOS-PS and HOOS-PS. Clin Exp Rheumatol. 2011;29:231-7.
17. Singh JA, Luo R, Landon GC, Suarez-Almazor M. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. J Rheumatol. 2014;41:509-15.
18. French HP, Fitzpatrick M, FitzGerald O. Responsiveness of physical function outcomes following physiotherapy intervention for osteoarthritis of the knee: an outcome comparison study. Physiotherapy. 2011;97:302-8.
19. Pua YH, Cowan SM, Wrigley TV, Bennell KL. The Lower Extremity Functional Scale could be an alternative to the Western Ontario and McMaster Universities Osteoarthritis Index physical function scale. J Clin Epidemiol. 2009;62:1103-11.
20. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Rheum. 2001;45:453-61.
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;12:349-62.
22. 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 Rheum. 1986;29:1039-49.
23. 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;34:505-14.
24. Smink AJ, van den Ende CH, Vliet Vlieland TP, Swierstra BA, Kortland JH, Bijlsma JW, et al. "Beating osteoARThritis": development of a stepped care strategy to optimize utilization and timing of non-surgical treatment modalities for patients with hip or knee osteoarthritis. Clin Rheumatol 2011;30:1623-9.
25. Snijders GF, den Broeder AA, van Riel PL, Straten VH, de Man FH, van den Hoogen FH, et al. Evidence-based tailored conservative treatment of knee and hip osteoarthritis: between knowing and doing. Scand J Rheumatol. 2011;40:225-31.
26. Jaber R, Braksmajer A, Trilling JS. Group visits: a qualitative review of current research. J Am Board Fam Med. 2006;19:276-90.
27. Lequesne MG. The algofunctional indices for hip and knee osteoarthritis. J Rheumatol. 1997;24:779-81.
28. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. Phys Ther. 1999;79:371-83.
29. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol. 1998;51:1055-68.
30. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
31. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res. 1994;38:383-92.
32. Teeuw B, Schwarzer R, Jerusalem M. Dutch adaptation of the General Self-Efficacy Scale. 1994. http://userpage.fu.berlin.de/~health/dutch.htm.
33. Kraaimaat FW, Evers AW. Pain-coping strategies in chronic pain patients: psychometric characteristics of the pain-coping inventory (PCI). Int J Behav Med. 2003;10:343-63.
34. Axford J, Butt A, Heron C, Hammond J, Morgan J, Alavi A, et al. Prevalence of anxiety and depression in osteoarthritis: use of the Hospital Anxiety and Depression Scale as a screening tool. Clin Rheumatol. 2010;29:1277-83.
35. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res. 2002;52:69-77.
36. Spinhoven P, Ormel J, Sloekers PP, Kempen GI, Speckens AE, Van Hemert AM. A validation study of the Hospital Anxiety and Depression Scale (HADS) in different groups of Dutch subjects. Psychol Med. 1997;27:363-70.
37. De Vet HCW, Terwee CB, Mokkink LD, Knol DL, eds. Measurement in medicine, 1st edn. Cambridge: Cambridge University Press, 2011.

38. Terwee CB, van der Slikke RM, van Lummel RC, Benink RJ, Meijers WG, de Vet HC. Self-reported physical functioning was more influenced by pain than performance-based physical functioning in knee-osteoarthritis patients. J Clin Epidemiol. 2006;59:724-31.
39. Stratford PW, Kennedy DM, Hanna SE. Condition-specific Western Ontario McMaster Osteoarthritis Index was not superior to region-specific Lower Extremity Functional Scale at detecting change. J Clin Epidemiol. 2004;57:1025-32.
40. Gandek B. Measurement properties of the Western Ontario and McMaster Universities Osteoarthritis Index: A systematic review. Arthritis Care Res.(Hoboken.) 2015;67:216-29.
41. Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G. A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. Arthritis Care Res. 1999;12:331-5.
42. 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;45:890-902.
43. Myers AM, Holliday PJ, Harvey KA, Hutchinson KS. Functional performance measures: are they superior to self-assessments? J Gerontol. 1993;48:M196-206.
44. Barten JA, Pisters MF, Huisman PA, Takken T, Veenhof C. Measurement properties of patient-specific instruments measuring physical function. J Clin Epidemiol. 2012;65:590-601.
45. Kroman SL, Roos EM, Bennell KL, Hinman RS, Dobson F. Measurement properties of performance-based outcome measures to assess physical function in young and middle-aged people known to be at high risk of hip and/or knee osteoarthritis: a systematic review. Osteoarthritis Cartilage. 2014;22:26-39.
46. Theiler R, Sangha O, Schaeren S, Michel BA, Tyndall A, Dick W, et al. Superior responsiveness of the pain and function sections of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) as compared to the Lequesne-Algofunctional Index in patients with osteoarthritis of the lower extremities. Osteoarthritis Cartilage. 1999;7:515-9.
47. Robling M, Hood K. Response shift, responsiveness or recall bias? Br J Gen Pract. 2002;52:585.
48. Sprangers MA, Schwartz CE. Integrating response shift into health-related quality of life research: a theoretical model. Soc Sci Med. 1999;48:1507-15.
49. van Dijk GM, Veenhof C, Spreeuwenberg P, Coene N, Burger BJ, van SD, et al. Prognosis of limitations in activities in osteoarthritis of the hip or knee: a 3-year cohort study. Arch Phys Med Rehabil. 2010;91:58-66.
50. Wesseling J, Dekker J, van den Berg WB, Bierma-Zeinstra SM, Boers M, Cats HA, et al. CHECK (Cohort Hip and Cohort Knee): similarities and differences with the Osteoarthritis Initiative. Ann Rheum Dis. 2009;68:1413-9.

Appendix

Authors' reply.

Response to comments by Stratford and Kennedy on: Responsiveness of four patient-reported outcome measures to assess physical function in patients with knee osteoarthritis

Elien A.M. Mahler
Nienke Cuperus
Johannes J.W. Bijlsma
Thea P.M. Vliet Vlieland
Frank H.J. van den Hoogen
Alfons A. den Broeder
Cornelia H.M. van den Ende

Published in Scandinavian Journal of Rheumatology 2017 Mar;46(2):169-170

We thank Stratford and Kennedy for their interest in our study in which we evaluated the responsiveness of four patient-reported outcome measures (PROMs) to measure change in physical function in patients with knee osteoarthritis (OA) following currently recommended COnsensus-based Standards for the selection of health Measurement INstrument (COSMIN) standards.[1] They have concerns about the conclusion because of three reasons, which will be discussed separately.

Their first point addresses the veracity of hypothesis number 14. There is growing consensus that responsiveness should be considered as a measure of longitudinal construct validity because responsiveness refers to measuring changes in the construct to be measured. Hence, responsiveness should preferably be evaluated by testing predefined hypotheses about expected correlations in different aspects of responsiveness, namely convergent, discriminant, and discriminative responsiveness. Hypotheses are about expected correlations between changes in related measurement (convergent; hypotheses 1, 3, 11, 13, and 14); unrelated measures (discriminative; hypotheses 2, 4-10, and 12) or expected between-group differences in changes (discriminative; hypothesis 15). Hypothesis 14 addresses convergent responsiveness in which expected correlations between changes in related measures (i.e. the construct physical function) are examined. Four PROMs on physical function were examined and the expert group a priori hypothesized that changes in the PROMs for physical function would be more strongly associated with answers on the transition question covering the same construct (i.e. physical function) than answers on the transition scale regarding a different construct (i.e. pain). Stratford and Kennedy state that for hypothesis 14 to have merit it must be applied in a context where function and pain are known to display different change trajectories. We agree that in this patient group changes in both pain and function are to be expected, but with the COSMIN approach, the validity of the change scores can be assessed in contrast to the magnitude of the change score as assessed by traditionally accepted methods such as the standardized response mean (SRM).

Their second concern considers the small difference in the number of confirmed hypotheses for the Western Ontario and McMaster Universities Osteoarthritis Index Physical Function subscale (WOMAC-PF) (12/15) and Lower Extremity Functional Scale (LEFS) (11/15). Of note, this regards the Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS) as well because 11 out of 15 hypotheses were confirmed. Our discussion has also pointed out this concern, and we agree with Stratford and Kennedy that, to date, there is no consensus or guideline about considering the use of the lower or upper bound of the 95% confidence interval of the cut-off point (i.e. > 75% of hypotheses). It would be interesting to take this into account in future research or in an update of the COSMIN standard. However, this update should also include guidelines about the nature and number of hypotheses, as defining more hypotheses would lead to a smaller confidence interval regarding this cut-off point. Moreover, the COSMIN standard is not developed to compare and rate different measurement instruments, but meant to confirm or reject the responsiveness of a particular measurement instrument. Therefore, we used the formulated absolute cut-off point of the standard of > 75% of the hypotheses to be confirmed to consider responsiveness as positive.[2,3] Moreover, for this reason we did not compare and rate the different measurements to assess physical function.

We do not agree on the third point, to interpret the point estimates for the four PROMs separately, because this was not part of the a priori defined hypotheses. Only hypotheses 13 and 14 incorporate stronger correlations and differences in correlations were tested as a priori agreed upon. Furthermore, in responsiveness studies, lower correlation coefficients are often found compared to those in construct validity studies, which can be explained by the fact that in the former, a correlation coefficient is obtained between change scores of two measurement instruments.[3] The COSMIN standard reports about using correlation coefficients, but we agree with Stratford and Kennedy that it would be interesting to take confidence intervals into account in an update. Furthermore, the areas under the curve (AUCs) reported in Figure 1 are presented to interpret hypothesis 15, which states that the AUC has to be ≥ 0.70. The 95% confidence intervals of AUCs of LEFS (0.82) and WOMAC-PF (0.76) do overlap (data not shown) and therefore it cannot be stated that this AUC of the LEFS might be higher or better than the WOMAC-PF.

In conclusion, given the strengths and limitations of the COSMIN standard, and considering the evidence in its entirety, we think that we assessed responsiveness of the four PROMs as recommended and that our conclusion is valid that the WOMAC-PF could be seen as the measure of first choice in clinical trials evaluating the effectiveness of an intervention to assess physical function in patients with knee OA. We agree that considering the use of the lower or upper bound of the 95% confidence interval of the absolute cut-off point (i.e. > 75% of hypotheses) would be interesting to take into account in future research or in an update of the standard.

References

1. Mahler EAM, Cuperus N, Bijlsma J, Vliet Vlieland T, van den Hoogen F, den Broeder AA, et al. Responsiveness of four patient-reported outcome measures to assess physical function in patients with knee osteoarthritis. Scand J Rheumatol 2016;45:518-27.
2. Terwee CB, Bot SD, de Boer MR, van der Windt DA, Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60:34-42.
3. De Vet HCW, Terwee CB, Mokkink LD, Knol DL, eds. Measurement in medicine, 1st edn. Cambridge: Cambridge University Press, 2011.

Chapter 3

Patient Acceptable Symptom State (PASS) in knee osteoarthritis patients succeeds across different patient-reported outcome measures assessing physical function, but fails across other dimensions and rheumatic diseases

Elien A.M. Mahler
Nadine Boers
Johannes W.J. Bijlsma
Frank H.J. van den Hoogen
Alfons A. den Broeder
Cornelia H.M. van den Ende

Published in Journal of Rheumatology 2018 Jan;45(1):122-127

Abstract

Objective

The aims of this study are: (1) to establish the Patient Acceptable Symptom State (PASS) cutoff values of different patient-reported outcome measures (PROMs) assessing physical function in patients with knee osteoarthritis (OA), and (2) to assess the influence of sex, age, duration of symptoms, and presence of depressive feelings on being in PASS.

Methods

Patients fulfilling the clinical American College of Rheumatology knee OA criteria received standardized nonsurgical treatment and completed different questionnaires at baseline and 3 months assessing physical function: Knee Injury and Osteoarthritis Outcome Score, Lequesne Algofunctional Index, Lower Extremity Functional Scale, numeric rating scale, and the physical function subscale of the Western Ontario and McMaster Universities Osteoarthritis Index. PASS values were defined as the 75th percentile of the score of questionnaires for those patients who consider their state acceptable.

Results

Of the 161 included patients, 62% were woman with a mean age of 59 years (SD 9) and body mass index of 30 kg/m^2 (SD 5). Standardized PASS values (95% CI) for different questionnaires for physical function varied between 48 (44-54) and 54 (50-56). Female patients and patients feeling depressed were found to have a lower probability to be in PASS for physical function, with odds ratios (95% CI) varying from 0.45 (0.23-0.91) to 0.50 (0.26-0.97) and from 0.27 (0.14-0.55) to 0.38 (0.19-0.77), respectively.

Conclusion

PASS cutoff values for physical function are robust across different PROMs in patients with knee OA. Our results indicate that PASS values are not consistent across dimensions and rheumatic diseases, and that the use of a generic PASS value for patients with OA or even patients with other rheumatic diseases might not be justifiable.

Introduction

Osteoarthritis (OA) is a joint disease that affects the entire joint and mainly causes pain, disability, and reduced quality of life.[1,2] Because there is no curative treatment available for OA, treatment aims to improve daily functioning and reduce symptoms. In clinical trials with patients, it is considered of great importance to incorporate the patient's interpretation of outcomes in establishing the relevance of findings.[3,4] Both the change in complaints (minimal clinically important improvement) and the absolute level of complaints (Patient Acceptable Symptom State (PASS)) are considered useful concepts for the interpretation of outcomes of clinical trials and the translation of data into daily practice.[5]

The PASS is considered a state and is defined as the highest level of symptoms that the majority of patients consider acceptable.[6,7] Although the PASS has shown to be a relevant concept in rheumatology, only a few studies have estimated and validated the PASS in knee OA.[8,9-11] However, different values were obtained, which might be explained by the selection of patients in a specific setting or country, or by the use of different followup periods across studies. In addition, different approaches have been used in the involvement of domains (i.e., pain, patient's global assessment (PGA), function) and rheumatic diseases for estimating PASS values. This has led to the estimation of generic PASS values incorporating different domains, as well as more specific PASS values for only 1 domain (i.e., pain, PGA, function) or for 1 rheumatic disease. As a result, the generalizability of PASS values to other patient settings, countries, languages, and cultures is speculative.[7,11] Hence, more insight into the variability of PASS values in different patient groups and/or settings is needed.

In addition, the extent to which PASS values reflecting a specific outcome domain, i.e., physical function, are robust across different questionnaires is unknown. Earlier studies examining the PASS value for physical function in OA have mostly used the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[12], while other reliable and valid patient-reported outcome measures (PROMs) for physical functioning are available. Frequently used validated questionnaires to asses physical functioning in OA are the short version of the Knee Injury and Osteoarthritis Outcome Score (KOOS-PS)[13], Lequesne Algofunctional Index (LAI)[14], the Lower Extremity Functional Scale (LEFS)[15], and the physical function subscale of the WOMAC (WOMAC-PF).[12] For the comparison of research findings using PASS cutoff values across studies, insight is needed in the variability of the PASS value across different questionnaires measuring the same construct; e.g., the PASS for self-reported physical function in OA.

Also, earlier research showed inconsistency in the influence of factors such as sex, age, and duration of symptoms on the PASS cutoff value.[6,7,11] Further, because the PASS is based on patients' opinion, it could be hypothesized that the presence of depressive symptoms influences patients' evaluation of their clinical status. So, it may be possible that depressive symptoms affect the acceptability of functioning; it has long been established that there is a strong relationship between the level of depression and the severity of pain.[16,17]

Therefore, we conducted this prospective study with the following aims: (1) to establish the PASS values across different PROMs assessing physical function in patients with knee OA, and (2) to assess the influence of sex, age, duration of symptoms, and comorbid depressive state on this estimated PASS value.

3

Materials and methods

Design, setting, and participants

Consecutive patients (≥ 18 years) with knee OA referred by orthopaedic surgeons to our specialty knee and hip rheumatology OA outpatient clinic between July 2012 and January 2014 were eligible for participation in this prospective observational cohort study, as described elsewhere.[18] All patients fulfilled the American College of Rheumatology clinical OA criteria: knee pain (over 15 days last month) and at least three of the following: age over 50 years, morning stiffness for at least 30 minutes, crepitus, bony tenderness, bony enlargement, or no palpable warmth.[19] The exclusion criterion was indication within 3 months for knee replacement surgery. The local Medical Research Ethics Committee, region Arnhem-Nijmegen (The Netherlands) approved the study design (study number 2012/375). All patients signed informed consent.

Stepped care approach

Non-surgical treatment modalities for the management of knee OA are recommended by several (inter)national consensus-based guidelines.[20-23] Therefore, all patients received multimodal conservative treatment based on a Dutch multidisciplinary stepped care approach for treatment of knee OA.[24] This includes education, referral for physical therapy (prescription for both aerobic and strengthening exercises according to the graded activity principle), step-up analgesics guided by a patient's pain level, lifestyle advice, and advice on weight reduction for patients with a body mass index (BMI) ≥28 kg/m^2. This approach recommends that more advanced options are considered only if the options listed previously failed to yield satisfactory results.[24] Patients attended a 90-minute during group visit (4-6 patients) to the rheumatology outpatient department, led by a physician assistant and a specialized nurse and supervised by a rheumatologist. When analgesics were found to be necessary, patients started with a fixed dose of 1000 mg paracetamol (acetaminophen) 3 times a day. In case of lack of efficacy of paracetamol, a nonsteroidal antiinflammatory drug (NSAID) was added. The patients were contacted after 4 weeks by telephone and if necessary, the analgesics were switched.

Outcome measures and data collection

Patients were asked to complete questionnaires at baseline and at 3 months, including sociodemographic information such as sex, age, and duration of symptoms. The postal questionnaires included the following 4 PROMs to measure physical function.

KOOS-PS. The KOOS-PF is a 7-item short questionnaire with 4 items on daily activities and 3 items on sport and recreation (5-point Likert scale version from 0 to 4).[13] Scores range from 0 to 28, and in our study the scores were converted to normalized scores ranging from 0 to 100, with higher scores defining higher levels of disability. The KOOS-PS has been shown to be valid and reliable in knee OA.[13,25-27]

LAI. The LAI is an 11-item questionnaire that measures pain, walking distance, and daily activities. The total score ranges from 0 to 24 points.[14] The degree of functional disability corresponds with the following scores: a score ≥ 14 points indicates extremely severe disability, a score of 11–13 very severe disability, a score of 8–10 severe disability, a score of 5–7 moderate disability, and a score of 1–4 minimal disability. The LAI has been shown to be reliable, although its validity has been questioned.[28]

LEFS. The LEFS is a 20-item condition-specific questionnaire on daily and recreational activities created for the use in patients with musculoskeletal conditions of the lower extremity, including knee OA (5-point Likert scale version from 0 to 4).[15] The total score ranges from 0 to 80 points, with higher scores defining higher levels of functioning. The LEFS has been validated (also in Dutch) and shown to be reliable in patients with knee OA.[29,30]

WOMAC-PF. The KOOS includes the WOMAC OA index in its complete and original format (with permission). We used the 17-item subscale with questions about activities of daily living from the KOOS to calculate WOMAC-PF score, originally developed for people with OA (5-point Likert scale version from 0 to 4) The score ranges from 0 to 68 points and in our study was converted to normalized scores ranging from 0 to 100, with higher scores defining higher levels of disability. The WOMAC is the most widespread studied and used instrument in individuals with knee OA and is shown to be valid and reliable.[12]

Other patient-related outcome measures. Next to the PROMs mentioned above, the questionnaire also included the depression subscale of the Hospital Anxiety and Depression Scale (HADS), which assesses depression and is validated and reliable and has been validated in patients with OA.[16,17,31] The depression subscale consists of 7 items with possible scores ranging from 0 to 21. A HADS score > 8 was considered as indicating depressive symptoms. Furthermore, patients were asked to rate their functioning and pain in the preceding week on a 0-10-point numeric rating scale (NRS) in which 0 equals no symptoms.[24] The PGA of knee OA effect was measured identically as well.

The PASS has been defined as the value below which the majority of patients consider themselves in an acceptable state of symptoms. At baseline and after 3 months, the PASS for physical function was defined using an external anchor question considering their condition of knee OA. This single question was asked to the patients: "Think about all consequences of the knee osteoarthritis in the last week. If you were to remain for the rest of your life as you were during the last week, would the current state be acceptable or unacceptable for you?" 32 Patients were included in the analysis of our current study if they completed both the baseline and followup measurements.

Statistical analysis

Descriptive statistics were used to describe the study population. All continuous outcomes are shown as means with SD when appropriate and dichotomous outcomes are shown in numbers with percentages. All scores of the PROMs were normalized (and inverted when necessary) to a range of 0 to 100, with 100 being maximal complaints. Scale scores of KOOS-PS, LAI, LEFS, and WOMAC-PF were assessed for normality and missing data. Floor and ceiling effects for each questionnaire at baseline were considered present if > 15% of the patients scored the best or worst possible score, respectively. The 75th percentile of the cumulative distribution score of the PROM scores at 3 months in patients who considered themselves at an acceptable state was used to determine the cutoff value of the PASS. This approach has been validated as a comparable alternative to the receiver-operation characteristic curve.[6,10,11,32] Thereafter, these cutoff values with their 95% CI of the 4 different PROMs assessing physical function were compared.

To examine the influence of covariates on the estimated absolute PASS value, a univariate and multivariate logistic regression was performed. As dependent variable, the absolute PASS value on group level of a particular PROM assessing physical function was used, separately for each PROM. The independent variables were sex, age, duration of symptoms, and having depressive feelings at baseline. To improve interpretation and clinical applicability, we dichotomized all independent variables: age ≥ 65 years (yes/no), duration of symptoms > 5 years (yes/no), and HADS > 8 (yes/no) as validated by Axford, et al.[17] Statistical analyses were performed using STATA version 13.1.

Results

Patient characteristics

In total, 272 eligible consenting patients were invited to participate, of whom 185 (68%) completed the baseline measurements. A total of 161 (87%) who completed the measurements at 3 months' followup were included in the analyses. Around two-thirds (62%) of the patients were female, the mean age was 59 years (SD 9), and the mean BMI was 29.7 kg/m² (SD 5.0; (Table 1). No differences were found between the participants and nonparticipants with regard to sex, although the participants were significantly older than the nonparticipants (59 years vs 56 years; p value = 0.02). The sociodemographic and disease-related characteristics of the patients are presented in Table 1. For each instrument, there were ≤ 5% missing values at both timepoints.

Table 1. Sociodemographic and disease-related characteristics of study sample (n=161). Values are given as mean (SD) unless indicated otherwise. The mean scores of the KOOS-PS, LAI, LEFS, and WOMAC-PF are presented using the usual score range and using a normalized score (0-100).

Characteristics	Baseline	3 Months
Female, n (%)	99 (62)	
Age, years	59 (9)	
BMI, kg/m2	29.7 (5)	
Duration of symptoms, n (%)		
> 5 years	59 (37)	
Localization of symptoms, n (%)		
Left knee	39 (24)	
Right knee	56 (35)	
Both sides	66 (41)	
Patients considering their state to be acceptable, n (%)	77 (49)	87 (56)
PROMs regarding physical function		
KOOS-PS		
Range 0-28	18.3 (5.2)	17.8 (5.3)
Range 0-100	53.6 (16.8)	51.7 (15.8)
LAI		
Range 0-24	11.0 (4.0)	10.9 (4.3)
Range 0-100	45.8 (16.7)	45.4 (18.1)

Characteristics	Baseline	3 Months
LEFS*		
Range 0-80	40.6 (14.1)	41.0 (15.3)
Range 0-100	49.2 (17.6)	48.8 (19.1)
WOMAC-PF		
Range 0-68	32.8 (13.2)	32.0 (13.2)
Range 0-100	48.2 (19.4)	47.0 (20.5)
Other PROMs		
NRS function (range 0-10)	5.5 (2.4)	5.4 (2.5)
NRS pain (range 0-10)	5.6 (2.1)	5.5 (2.2)
NRS PGA (range 0-10)	5.6 (2.6)	5.5 (2.2)
Depression (HADS; range 0-21)		
No depressive feelings	4.6 (3.0)	3.6 (1.9)
Depressive feelings	8.5 (3.9)	10.3 (2.9)

*For all PROMs except LEFS, higher scores reflect a higher level of disability.
BMI = Body Mass Index; WOMAC-PF = Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale; LAI = Lequesne Algofunctional Index; LEFS = Lower Extremity Functional Scale; KOOS-PS: Knee Injury and Osteoarthritis Outcome Score, short version; PROM: patient-reported outcome measure; NRS = Numerical Rating Scale; PGA = patient's global assessment; HADS = Hospital Anxiety and Depression Scale.

Table 2. The estimated PASS cutoff values at 3 months' followup for different PROMs assessing physical function.

PROM	PASS value at 3 months (95% CI)
KOOS-PS	
Range 0-28	19.5 (18.0-20.7)
Range 0-100	52.8 (48.5-56.8)
LEFS	
Range 0-80	37.0 (35.2-40.0)
Range 0-100	53.8 (50.0-56.1)
LAI	
Range 0-24	11.5 (10.5-13.0)
Range 0-100	47.9 (43.8-54.2)
WOMAC-PF	
Range 0-68	34.0 (31.0-38.0)
Range 0-100	50.0 (45.6-55.9)
NRS function	
Range 0-10	6.0 (5.0-6.0)
Range 0-100	60.0 (50.0-60.0)

All PROMs present scores with higher scores defining a higher level of disability with the exception of the nonstandardized LEFS score range, where higher scores define higher level of functioning. PASS: Patient Acceptable Symptom State; PROM: patient-reported outcome measure; WOMAC-PF: Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale; LAI: Lequesne Algofunctional Index; LEFS: Lower Extremity Functional Scale; KOOS-PS: Knee Injury and Osteoarthritis Outcome Score, short version; NRS: numeric rating scale.

3

PASS cutoff values

Table 2 displays that the PASS cutoff values for function determined by 4 questionnaires range from 48 for the standardized LAI (95%CI 44-54) to 54 for the standardized LEFS (95%CI 50-56). This table shows that the cutoff values with their 95% CI for the PASS for physical function are comparable across the 4 different standardized PROMs assessing physical function. The PASS values of NRS function, pain, and PGA turned out to be consistent as well with a cutoff value of 60 (50-60). The 75th percentile of the NRS for function gives a PASS value of 60 (data not shown).

At 3 months' followup, 56% (95% CI 48-64) of the patients considered their state to be acceptable. The proportion of patients with depressive symptoms remained stable at 3 months' followup (30% vs 31%). The univariate logistic regression analysis showed that age and duration of symptoms are not associated with reaching the estimated PASS value for physical function, whereas a significant association was found between sex and being in PASS in 3 out of 5 PROMs regarding function. Female patients have a smaller probability of reaching the estimated absolute PASS value for physical function than male patients with a significant OR varying from 0.27 (95%CI 0.14-0.55) to 0.38 (95%CI 0.19-0.77) for the different PROMs. Also, having depressive symptoms turned out to be associated with reaching the estimated absolute PASS value; patients having depressive symptoms have a smaller probability of reaching an acceptable state for physical function than patients without depressive symptoms, with a significant OR around 0.50 for all PROMs except the KOOS-PS (Table 3). The multivariate logistic regression analyses yielded similar results (data not shown).

Table 3. Significant OR of the univariate logistic regression for each PROM for reaching the absolute PASS value.

	PASS			
	NRS function	KOOS-PS	LAI	LEFS	WOMAC-PF
Dependent variables					
Female	0.45 (0.23-0.91)	0.47 (0.24-0.93)	0.50 (0.26-0.97)	0.27 (0.14-0.55)	0.32 (0.16-0.64)	0.36 (0.18-0.72)
Age ≥ 65 years						
Duration of symptoms > 5 years						
Depressive feelings	0.38 (0.19-0.77)					

The Odds Ratio (OR) is given for every independent variable in the logistic regression with 95% confidence intervals for each PROM for function. All ORs being significant at the 0.05 level are shown. PASS: Patient Acceptable Symptom State; PROM: patient-reported outcome measure; NRS: numeric rating scale; WOMAC-PF: Western Ontario and McMaster Universities Osteoarthritis Index physical function subscale; LAI: Lequesne Algofunctional Index; LEFS: Lower Extremity Functional Scale; KOOS-PS: Knee Injury and Osteoarthritis Outcome Score, short version.

Discussion

We documented the PASS cutoff values for physical function and its determinants across different PROMs in a cohort of patients with knee OA in the Netherlands. Our results show that these PASS cutoff values are relatively robust across different questionnaires measuring physical function. Also, in our knee OA cohort, and in line with previous results for OA, patients consider a higher level of symptoms acceptable than previously reported for other rheumatic diseases.[6,10] In addition, we observed that woman and depressive patients have a lower chance of reaching the estimated PASS value.

The consistency of the cutoff values of the PASS for physical function across different PROMs assessing the same construct physical function in a specific cohort represents the robustness of the PASS values for physical function across these 4 different PROMs measuring physical function. To our knowledge, the robustness of the PASS regarding 1 outcome domain, i.e., physical functioning, measured with 4 different PROMs, has never been studied before. Therefore, our findings suggest that different questionnaires may be used to determine PASS cutoff values for physical function of a certain population and setting. Future research is warranted to investigate this finding in other populations and settings, and to examine whether this robustness is also valid for other outcome domains.

We found a higher PASS value for physical function than the generic, multinational PASS value reported previously, applicable for 5 different rheumatic conditions (including hip and knee OA) and for different outcome domains (pain, PGA, physical function).[10] Our values are comparable with other studies reporting on the PASS value of physical function in the context of nonsurgical treatment of OA.[10,11] However, our findings support earlier findings that PASS values might be variable across different rheumatic diseases, countries, types of intervention, and outcome domains. In previous studies, higher PASS values for patients with OA than for other rheumatic diseases were found, which may be caused by not having high expectations from optimized nonsurgical treatment modalities in knee OA compared to, for example, expectations from rheumatoid arthritis treatment.[10] Bellamy et al questioned the generalizability of PASS values to other countries, languages, and cultures, because they found considerable variation in PASS values across countries.[11] In addition, the variability in PASS values across countries was confirmed in recent studies from France reporting on relatively low PASS values for physical function in patients with knee and hip OA.[6] The observation that lower PASS values after total joint replacement were estimated than after NSAID treatment in Spanish patients with OA suggests that the type of intervention could affect PASS values as well.[11,33] This is in line with previous suggestions that patients expect greater effects from surgery than from nonsurgical therapy.[34] Finally, several studies documented that PASS values for physical function are higher than those for other domains[10,11] Taking the above considerations into account, it is conceivable that PASS values are generalizable only to 1 outcome domain in a specific disease for a certain type of intervention, and thus that the use of a generic PASS value is not justifiable. An alternative could be to determine PASS values for each study separately, by including the standardized question in the data collection, rather than applying a generic (multinational) estimated value.

An intriguing finding of our study was that patients having depressive feelings are associated with a lower chance of reaching the estimated PASS value. In fact, this could be in line with the previous notion that acceptability of a certain disease state is not only dependent on the absolute level of complaints, but is also dependent on other factors, and in this particular case on the patient's mood. If future research does confirm this finding, this may create a new point of view when treating a patient with OA who does not reach an acceptable symptom state.

Potential limitations of our study include the quite small cohort used in our study compared to the cohorts used in earlier studies to determine the PASS. However, because we used a homogeneous cohort, our findings seem generalizable to Dutch patients who are not yet deemed eligible for surgery. Another limitation for studies examining PASS values in general could be that a response shift took place, in which perception of the disease state changes during the assessments.[35] In addition, a general limitation for studies using PASS values is that there is no uniform approach to establishing a PASS value; the question asked to the patient varied across earlier studies and the time extent was different.[6] If the PASS is to become a universal concept for defining interventional success, a standard anchor question for meaningful comparison of results across groups should be established, in particular with regard to the duration of an acceptable state. We would suggest that PASS implies a state without change, that is, the time spent in the state as "rest of your life" as recommended by the Outcome Measures in Rheumatology [8,34,36]

In conclusion, PASS cutoff values are robust for physical function across different PROMs in patients with knee OA. However, our results indicate that PASS values are not consistent across dimensions and rheumatic diseases, and that the use of a generic PASS value for patients with OA or even patients with other rheumatic disorders might not be justifiable.

Acknowledgement
We thank all patients who participated.

3

References

1. Cooper C, Adachi JD, Bardin T, Berenbaum F, Flamion B, Jonsson H et al. How to define responders in osteoarthritis. Curr Med Res Opin 2013;29:719-29.
2. Zhang W, Doherty M, Peat G, Bierma-Zeinstra MA, Arden NK, Bresnihan B et al. EULAR evidence-based recommendations for the diagnosis of knee osteoarthritis. Ann Rheum Dis 2010;69:483-9.
3. Altman R, Brandt K, Hochberg M, Moskowitz R, Bellamy N, Bloch DA et al. Design and conduct of clinical trials in patients with osteoarthritis: recommendations from a task force of the Osteoarthritis Research Society. Results from a workshop. Osteoarthritis Cartilage 1996;4:217-43.
4. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage 2004;12:389-99.
5. Wells G, Boers M, Shea B, Anderson J, Felson D, Johnson K et al. MCID/Low Disease Activity State Workshop: low disease activity state in rheumatoid arthritis. J Rheumatol 2003;30:1110-1.
6. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N et al. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. Ann Rheum Dis 2005;64:34-7.
7. Perrot S, Bertin P. "Feeling better" or "feeling well" in usual care of hip and knee osteoarthritis pain: determination of cutoff points for patient acceptable symptom state (PASS) and minimal clinically important improvement (MCII) at rest and on movement in a national multicenter cohort study of 2414 patients with painful osteoarthritis. Pain 2013;154:248-56.
8. Hochberg MC, Wohlreich M, Gaynor P, Hanna S, Risser R. Clinically relevant outcomes based on analysis of pooled data from 2 trials of duloxetine in patients with knee osteoarthritis. J Rheumatol 2012;39:352-8.
9. Dougados M, Moore A, Yu S, Gitton X. Evaluation of the patient acceptable symptom state in a pooled analysis of two multicentre, randomised, double-blind, placebo-controlled studies evaluating lumiracoxib and celecoxib in patients with osteoarthritis. Arthritis Res Ther 2007;9:R11.
10. Tubach F, Ravaud P, Martin-Mola E, Awada H, Bellamy N, Bombardier C et al. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. Arthritis Care Res (Hoboken) 2012;64:1699-707.
11. Bellamy N, Hochberg M, Tubach F, Martin-Mola E, Awada H, Bombardier C et al. Development of multinational definitions of minimal clinically important improvement and patient acceptable symptomatic state in osteoarthritis. Arthritis Care Res (Hoboken) 2015;67:972-80.
12. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Rheum 2001;45:453-61.
13. Perruccio AV, Stefan LL, Canizares M, Tennant A, Hawker GA, Conaghan PG et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS) - an OARSI/OMERACT initiative. Osteoarthritis Cartilage 2008;16:542-50.
14. Lequesne MG. The algofunctional indices for hip and knee osteoarthritis. J Rheumatol 1997;24:779-81.
15. Binkley JM, Stratford PW, Lott SA, Riddle DL. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. North American Orthopaedic Rehabilitation Research Network. Phys Ther 1999;79:371-83.
16. Axford J, Heron C, Ross F, Victor CR. Management of knee osteoarthritis in primary care: pain and depression are the major obstacles. J Psychosom Res 2008;64:461-7.
17. Axford J, Butt A, Heron C, Hammond J, Morgan J, Alavi A et al. Prevalence of anxiety and depression in osteoarthritis: use of the Hospital Anxiety and Depression Scale as a screening tool. Clin Rheumatol 2010;29:1277-83.
18. Mahler EAM, Cuperus NC, JWJ Bijlsma, Vliet Vlieland TPM, van den Hoogen FHJ, den Broeder AA, et al. Responsiveness of four patient-reported outcome measures to assess physical function in patients with knee osteoarthritis. Scand J Rheumatol 2016;45:518-27.
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. Diagnostic and Therapeutic Criteria Committee of the American Rheumatism Association. Arthritis Rheum 1986;29:1039-49.
20. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage 2010;18:476-99.
21. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. Arthritis Care Res (Hoboken) 2012;64:465-74.
22. Conaghan PG, Dickson J, Grant RL. Care and management of osteoarthritis in adults: summary of NICE guidance. BMJ 2008;336:502-3.
23. Fernandes L, Hagen KB, Bijlsma JW, Andreassen O, Christensen P, Conaghan PG et al. EULAR recommendations for the non-pharmacological core management of hip and knee osteoarthritis. Ann Rheum Dis 2013;72:1125-35.
24. Smink AJ, van den Ende CH, Vliet Vlieland TP, Swierstra BA, Kortland JH, Bijlsma JW et al. "Beating osteoARThritis": development of a stepped care strategy to optimize utilization and timing of non-surgical treatment modalities for patients with hip or knee osteoarthritis. Clin Rheumatol 2011;30:1623-9.
25. Bond M, Davis A, Lohmander S, Hawker G. Responsiveness of the OARSI-OMERACT osteoarthritis pain and function measures. Osteoarthritis Cartilage 2012;20:541-7.
26. Ruyssen-Witrand A, Fernandez-Lopez CJ, Gossec L, Anract P, Courpied JP, Dougados M. Psychometric properties of the OARSI/OMERACT osteoarthritis pain and functional impairment scales: ICOAP, KOOS-PS and HOOS-PS. Clin Exp Rheumatol 2011;29:231-7.
27. Singh JA, Luo R, Landon GC, Suarez-Almazor M. Reliability and clinically important improvement thresholds for osteoarthritis pain and function scales: a multicenter study. J Rheumatol 2014;41:509-15.
28. French HP, Fitzpatrick M, FitzGerald O. Responsiveness of physical function outcomes following physiotherapy intervention for osteoarthritis of the knee: an outcome comparison study. Physiotherapy 2011;97:302-8.
29. Hoogeboom TJ, de Bie RA, den Broeder AA, van den Ende CH. The Dutch Lower Extremity Functional Scale was highly reliable, valid and responsive in individuals with hip/knee osteoarthritis: a validation study. BMC Musculoskelet Disord 2012;13:117.
30. Pua YH, Cowan SM, Wrigley TV, Bennell KL. The Lower Extremity Functional Scale could be an alternative to the Western Ontario and McMaster Universities Osteoarthritis Index physical function scale. J Clin Epidemiol 2009;62:1103-11.
31. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. An updated literature review. J Psychosom Res 2002;52:69-77.
32. Kviatkovsky MJ, Ramiro S, Landewe R, Dougados M, Tubach F, Bellamy N et al. The Minimum Clinically Important Improvement and Patient-acceptable Symptom State in the BASDAI and BASFI for Patients with Ankylosing Spondylitis. J Rheumatol 2016;43:1680-6.
33. Escobar A, Gonzalez M, Quintana JM, Vrotsou K, Bilbao A, Herrera-Espineira C et al. Patient acceptable symptom state and OMERACT-OARSI set of responder criteria in joint replacement. Identification of cut-off values. Osteoarthritis Cartilage 2012;20:87-92.
34. Tubach F, Ravaud P, Beaton D, Boers M, Bombardier C, Felson DT et al. Minimal clinically important improvement and patient acceptable symptom state for subjective outcome measures in rheumatic disorders. J Rheumatol 2007;34:1188-93.

35. McPhail S, Haines T. Response shift, recall bias and their effect on measuring change in health-related quality of life amongst older hospital patients. Health Qual Life Outcomes 2010;8:65.
36. Putrik P, Ramiro S, Hifinger M, Keszei AP, Hmamouchi I, Dougados M et al. In wealthier countries, patients perceive worse impact of the disease although they have lower objectively assessed disease activity: results from the cross-sectional COMORA study. Ann Rheum Dis 2016;75:715-20.

Chapter 4

How should worsening in osteoarthritis be defined? Development and initial validation of preliminary criteria for clinical worsening in knee and hip osteoarthritis

Elien A.M. Mahler
Alfons A. den Broeder
Thasia G. Woodworth
Vincent J.J.F. Busch
Frank H.J. van den Hoogen
Johannes W.J. Bijlsma
Cornelia H.M. van den Ende

Published in Scandinavian Journal of Rheumatology 2017 Sep;46(5):396-406

Abstract

Objective

There is a need to define and validate measures of clinical worsening in knee and hip osteoarthritis (OA). The objectives of this exploratory project were: (1) to characterize worsening criteria in knee and hip OA using psychometric methods; (2) to estimate their sensitivity and specificity; (3) to validate and compare these criteria with worsening criteria previously described in the literature.

Methods

An Expert Group reached consensus on 10 sets of worsening criteria to be tested in observational data sets of patients with knee or hip OA who received multimodal conservative treatment. These sets included 219 patients (derivation cohort) and 296 patients (validation cohort). We estimated minimal clinically important worsening (MCIW) values for pain, function, stiffness, and patient global assessment, and tested candidate worsening criteria in the derivation cohort. Finally, using patient judgement, we examined sensitivity and specificity of literature-based as well as candidate worsening criteria in the validation cohort.

Results

Literature-based worsening criteria were found to have high specificity (range 60-92%) but low sensitivity (range 22-59%). Two out of 10 candidate worsening criteria constructed by the Expert Group showed an acceptable combination of sensitivity and specificity in the derivation cohort, which was confirmed in the validation cohort (ranging from 54% to 65% and 67% to 74%, respectively).

Conclusion

This is the first study to describe symptomatic worsening criteria based on expert consensus after examining the performance of candidate criteria derived from the literature applied to data in an observational study. The newly proposed worsening criteria show an acceptable combination of sensitivity and specificity.

Introduction

Over the past few years, increasing attention has been directed towards the development and use of outcome measures on an individual level in patients with osteoarthritis (OA). In OA, multiple sets of criteria are available to distinguish between patients who clinically improve and those who do not.[1-3] However, not all patients improve with treatment, and thus, determining which patients have deteriorated is important, both in research and in clinical care.[4] Consequently, many studies have been and are being performed to determine prognostic factors for OA, but mainly focus on radiographic progression, while a clear discordance between radiographic and symptomatic OA has formerly been described.[5] In prediction studies that focus on symptomatic progression of OA, validated worsening criteria could be used as an outcome measure to identify variables which may predict deterioration. In future longitudinal clinical care, worsening criteria could also assist in monitoring patients in clinical practice and selecting patients in whom treatment should be modified. In contrast to clinically important improvement criteria, validated worsening criteria to define clinically important (symptomatic) worsening for outpatients with knee and hip OA have not been described, although several studies used self-defined worsening criteria that have not been validated.[1,4]

Indeed, there is no consensus on how clinical worsening in knee or hip OA should be defined and, as a consequence, a large variety in definitions of clinical knee OA remains.[4] Several definitions of worsening on a group level, using mean changes, have been used, in particular to describe the progress over time in cohorts of patients.[4,6-9] However, mean changes with a continuous measure are not useful to measure worsening on an individual level. Only a few studies have used individual (dichotomized) clinical worsening criteria. These non-validated criteria vary considerably with respect to the following aspects: measured domain(s) (pain, function, stiffness, and/or patient global assessment (PGA)), the measurement instrument used, the type of change (absolute or relative, or a combination of absolute and relative) and the amount of absolute and/or relative change.[10-20] This implicates that worsening can be defined in numerous ways depending on the choices per aspect.[11,14,17,18]

An important aspect of worsening criteria is the amount of change in relevant domains. The amount of change (relative and/or absolute) can be based on consensus among experts, and/or based on patient self-report to examine minimal clinically important worsening (MCIW) values for identified domains. The MCIW can be defined as the smallest difference in domain score, associated with patients' perception of worsening, with a request to change their healthcare management.[21] The patients' perception of change regarding longer time spans in, for example, prediction or observational studies, needs to be transformed into changes in measurement instruments to minimize the risk of a response shift, as was recognized when developing responder criteria for OA.[22] Because the amount of worsening that patients consider important is different from the amount of improvement[18,23-26], arbitrarily defining worsening using the inverse of proposed minimal clinically important improvement values is likely to be inconsistent with the patient perspective.[1,11,27,28] However, MCIW thresholds for the domains pain, function, stiffness, and PGA have not yet been identified.

To address the need for worsening criteria and MCIW values applicable to evaluating individual patients in longitudinal studies, the aims of this exploratory project were: (1) to characterize

worsening criteria in knee and hip OA using state-of-the-art psychometric methods; (2) to estimate their sensitivity and specificity; (3) to validate and compare these criteria with worsening criteria previously described in the literature.

Patients and methods

Expert group process

We followed a six-step approach to develop and validate individual clinical worsening criteria in this exploratory project (Table 1). In steps 2 and 3, an Expert Group (n = 9), comprising researchers and clinicians experienced in the field of knee and hip OA and well versed with the current literature providing patient-reported outcome measures (PROMs) (epidemiologists, orthopaedic surgeon, psychologist, researchers, rheumatologists), decided on the sets of previously used and newly proposed worsening criteria to be evaluated.

Step 1: review of the literature. Previously used worsening criteria up to 2014 were identified and reviewed by means of a literature search in MEDLINE using the terms: worsening or deterioration or flare or progression, and osteoarthritis knee or osteoarthritis hip. Additional references were identified from references of identified publications and abstracts of American College of Rheumatology (ACR), European League Against Rheumatism (EULAR), and Osteoarthritis Research Society International (OARSI) meetings. From this review, only worsening criteria that were dichotomized on an individual level were selected.

In addition, to prepare for the expert meeting, the Outcome Measures in Rheumatology (OMERACT)-OARSI study regarding the development of responder criteria for OA and the literature upon which this is based were thoroughly studied so that the Expert Group was well informed on aspects considered to be important when developing clinically improvement criteria for OA (i.e. outcome domains, type (absolute and/or relative), amount of change, and measurement instruments).[1]

Step 2: selection of previously described worsening criteria. The face validity of previously used worsening criteria was scored by each member of the Expert Group, before the expert meeting, on a scale ranging from 1 to 3, with 3 being the best score. In the expert meeting, the Expert Group selected, based on the presented total scores for each previously used worsening criteria and on the basis of consensus, the criteria that should be further validated in our study (agreement among ≥80% of the members of the Expert Group).

Step 3: development of candidate worsening criteria. In a consensus meeting, the Expert Group used a nominal group process to reach agreement by discussion and voting regarding the most appropriate outcome domains, type (absolute and/or relative changes), amount of clinically important change for each domain (what amount and/or using our estimated MCIW values), and measurement instruments of first choice that should comprise a set of worsening criteria.

Step 4: construction of newly proposed sets of worsening criteria. Using the transition scale as external anchor, we estimated MCIW values with an anchor-based approach for the four domains: numeric rating scale (NRS) and Western Ontario and McMaster Universities (WOMAC) pain, function, stiffness; and PGA in the derivation cohort. The MCIW was estimated

for both the absolute difference (3 month value minus baseline value) and the relative difference (3 month value minus baseline value, divided by baseline value). We deemed a time frame of 3 months long enough to allow for worsening and brief enough to minimize the risk of a response shift.[29] The mean score differences between the "equal" and the "slightly worse" group defined the MCIW values.[21,30] The MCIW values were presented as mean or median, as appropriate.

Thereafter, based on prerequisites agreed upon by the expert meeting in step 3 and the results of the MCIW in step 4, we constructed 10 new candidate worsening criteria sets (see Table 6). A priori, it was decided to restrict the maximal number of candidate worsening criteria sets to 10.

Step 5: estimation of sensitivity and specificity of newly proposed worsening criteria in derivation cohort. We examined the sensitivity and specificity of the newly proposed worsening criteria in the derivation cohort using the transition scale. We classified the responses on the transition scale into two categories: importantly worsened (slightly worse, worse and much worse) and not importantly changed (equal). The five sets that performed best according to the highest Youden index (J: sensitivity + specificity - 1) and receiver operating characteristic (ROC) point closest to the maximum (0,1) in the derivation cohort were selected to be validated in the validation cohort (step 6).[31] This ROC point is similar to the concordance-statistic (c-statistic).

Step 6: validation of previously used and newly proposed worsening criteria. Using the transition scale, we examined the sensitivity and specificity of the literature-based worsening criteria with acceptable face validity, the MCIW values of pain, function, stiffness, and PGA, together with the five sets of newly proposed worsening criteria in the validation cohort. We determined the three best performing worsening sets according to the Youden index (J) and c-statistic.

Study design and patients

We included 515 consecutive consenting patients attending our specialty knee and hip rheumatology OA outpatient clinic and fulfilling the ACR clinical criteria for knee or hip OA.[32,33] All patients in this observational study received multimodal treatment for 3 months, which comprised education, physical therapy, step-up analgesics, and advice on weight reduction if indicated, as described elsewhere.[34] Visits were planned at baseline and 3 months at the outpatient clinic and at weeks 4 and 8 by telephone. Exclusion criteria were: other rheumatic or severe orthopaedic diseases leading to inflammatory arthritis or secondary OA, comorbidity exceeding the complaints or limitations of the knee or hip OA, orthopaedic procedures planned within the next 3 months, or cognitive or sensorimotor problems interfering with questionnaire completion. The index joint was the most symptomatic knee or hip at baseline. All patients signed informed consent according to the Declaration of Helsinki and the study was approved by the local medical research ethics committee, region Arnhem-Nijmegen (The Netherlands) (study number 2009/095).

Data acquisition

At inclusion/baseline, demographic and clinical data were collected and all patients completed PROMs (see below) for four domains: pain, function, stiffness, and PGA of index joint symptom impact (PGA).[34] Patient data were included in the present analyses when the

transition question at 3 months was completed. In total, 219 patients, who were included between 1 July 2007 and 31 July 2009 and followed for 3 months, were used as the derivation cohort. The validation cohort consisted of 296 patients, recruited between 1 August 2009 and 1 May 2011, who completed PROMs at a 3 month visit.

Patient Reported Outcome Measures (PROMs)

Pain intensity and PGA of disease impact during the week before the visit were assessed at baseline and the 3 month visit and measured on a 0-10 point NRS, where 0 equals no symptoms. Patients also completed the Dutch Knee/Hip injury and Osteoarthritis Outcome Score (KOOS/HOOS) questionnaire (Likert-scale version) at baseline and the 3 month visit.[35] From the KOOS/HOOS, the WOMAC score can be derived, with WOMAC pain, function, and stiffness subscales presented as transformed scores ranging from 0 to 100, where 100 equals no symptoms.

Because a gold standard to determine clinically important change in index joint symptoms is lacking, we used as an external anchor the "transition" method based on the patient's perception of change in index joint symptoms, as currently recommended.[21,36,37] The transition questionnaire assessed the change in health related to the index OA joint at 3 month follow-up compared to baseline on a seven-point Likert scale (much worse, worse, slightly worse, no change, slightly improved, improved, much improved).

Statistical analysis

Checks for assumptions of normality were performed for all continuous baseline data, including pain, function, stiffness, and PGA. Descriptive statistics were provided as mean ± standard deviation (SD) or median and interquartile range (IQR) for continuous variables and percentages for categorical variables. To test the robustness of our results, a sensitivity analysis was performed on an imputed dataset - in which missing values of PROMs were imputed using multiple imputation to create 20 data sets - and results were combined using Ruben's rules.[38] Only the results without multiple imputation are shown, as multiple imputation did not change the results. All statistical analyses were performed using STATA 13.1.

Results

Expert group process

The review of the literature yielded 11 different sets of worsening criteria.[1,2,11,12,14,17,18,20,39-41] Based on examination of these sets, the nominal group process led to four main aspects to be taken into account when developing new criteria for clinically important worsening in knee and hip OA: (1) stiffness as a domain for PROMs is less validated than pain, function, and PGA; (2) newly defined worsening criteria should contain the domains pain and/or function, and optional PGA; (3) criteria for worsening should reflect both absolute and relative change[1,2,42-45]; and (4) criteria for worsening could either contain one domain (pain, function, or PGA) or a combination of domains. The magnitude of absolute change for worsening criteria should be 1-2 for NRS pain and PGA and 10-20 for WOMAC subscales, whereas the magnitude of relative change should range between 10% and 25%, and above 50% if one domain is tested. In addition, the magnitude of change from the results of the MCIW should preferably be taken into account. Finally, it was decided that different measurements per domain can be used; for instance, either NRS or WOMAC pain for the domain pain.

The Expert Group's review of the literature yielded five studies presenting different worsening criteria with acceptable face validity that were included in the validation round.[1,11,14,17,18] In addition, the Expert Group constructed 10 sets of newly proposed worsening criteria.

Table 1. Overview of the six-step approach - based on literature, expert opinion, and data-driven approach - for the development and validation of clinical worsening criteria.

	Expert group (n=9)	Derivation cohort (n=219)	Validation cohort (n=296)
Step 1	• Review of literature and identification of worsening criteria • Review of literature and important characteristics for developing improvement criteria in OA		
Step 2	• Selection of previously used worsening criteria		
Step 3	• Development of candidate worsening criteria		
Step 4	• Construction of new sets of worsening criteria based on expert consensus	• Estimation of MCIW values for pain, function, stiffness, and PGA	
Step 5		• Estimation of sensitivity and specificity of newly proposed worsening criteria	
Step 6			• Validation of previously used worsening criteria • Validation of MCIW values of pain, PGA, function, and stiffness • Validation of sensitivity and specificity of newly proposed worsening criteria

OA: osteoarthritis; MCIW: minimal clinically important worsening; PGA: patient global assessment.

Table 2. Baseline characteristics of patients with knee or hip osteoarthritis.

	Derivation cohort (n = 219)	Validation cohort (n = 296)
Age, years	55.2 (10.1)	55.2 (10.4)
Women	145 (66.2)	196 (66.2)
Body mass index, kg/m², median (IQR)	28.7 (25.5 – 32.5)	27.9 (25.4 – 33.2)
Duration of symptoms, years, median (IQR)	4.1 (1.9 – 9.6)	3.6 (1.5 – 9.6)
Index joint, knee	172 (81.9)	246 (83.1)
Kellgren and Lawrence ≥ 2	136 (63.0)	201 (72.0)
NRS pain (0-10)	6.1 (2.0)	6.0 (1.8)
PGA (0-10)	6.7 (2.1)	6.1 (2.2)
WOMAC pain (0-100)	48.7 (20.8)	48.1 (19.6)
Function (0-100)	46.5 (21.2)	47.0 (19.9)
Stiffness (0-100)	43.1 (23.2)	45.0 (23.7)

Data are shown as mean ± SD or N (%) unless stated otherwise, IQR: interquartile range.
NRS: numeric rating scale and PGA: patient global assessment (both scales 0-10, where 0 equals no symptoms); WOMAC: Western Ontario and McMaster Universities pain, function, and stiffness (scale 0-100, where 100 equals no symptoms).

Table 3. Mean difference (T 3 months - T 0) in outcome scores by transition reply at the 3 month follow-up for the total group of patients with knee or hip OA and baseline values.

Transition reply	No. of patients (n=515) (%)		Change in NRS pain		Change in WOMAC function		Change in WOMAC stiffness	
Much worsened	14	(2.7)	1.1	(1.9)	-5.2	(9.3)	-5.1	(13.8)
Worsened	34	(6.6)	0.4	(1.5)	-3.1	(16.6)	-5.6	(27.7)
Slightly worsened	64	(12.4)	0.4	(1.9)	0.6	(13.9)	-3.3	(23.4)
No change	232	(45.1)	-0.9	(2.0)	3.6	(16.3)	0.5	(20.3)
Slightly improved	96	(18.6)	-1.8	(1.7)	10.1	(14.6)	3.1	(18.9)
Improved	57	(11.1)	-2.8	(2.3)	17.1	(17.5)	14.5	(20.5)
Much improved	18	(3.5)	-4.2	(1.8)	17.7	(23.5)	13.4	(22.7)
Baseline values			6.0	(1.9)	46.7	(20.4)	44.1	(23.5)

Transition reply	No. of patients (n=515) (%)		Change in PGA		Change in WOMAC pain		Change in WOMAC function		Change in WOMAC stiffness	
Much worsened	14	(2.7)	1.5	(2.3)	-2.5	(13.7)	-5.2	(9.3)	-5.1	(13.8)
Worsened	34	(6.6)	4.1	(2.0)	-5.2	(13.8)	-3.1	(16.6)	-5.6	(27.7)
Slightly worsened	64	(12.4)	-0.2	(2.0)	1.4	(18.9)	0.6	(13.9)	-3.3	(23.4)
No change	232	(45.1)	-1.4	(2.2)	3.7	(19.3)	3.6	(16.3)	0.5	(20.3)
Slightly improved	96	(18.6)	-1.9	(2.2)	9.3	(14.1)	10.1	(14.6)	3.1	(18.9)
Improved	57	(11.1)	-2.9	(2.4)	17.5	(17.1)	17.1	(17.5)	14.5	(20.5)
Much improved	18	(3.5)	-2.5	(3.3)	24.3	(23.1)	17.7	(23.5)	13.4	(22.7)
Baseline values			6.3	(2.1)	48.3	(20.1)	46.7	(20.4)	44.1	(23.5)

Data are shown as N (%) or meanchange ± SDchange. Baseline values are presented as mean ± SD.
OA: osteoarthritis; NRS: numeric rating scale pain; PGA: patient global assessment (both scales 0-10, where 0 equals no symptoms); WOMAC: Western Ontario and McMaster Universities pain, function, and stiffness (scale 0-100, where 100 equals no symptoms).

Analyses: patient characteristics

The longitudinal observational study in which data were collected for derivation and validation of worsening criteria included a total of 515 patients, of whom 66.2% were female, with a mean age of 55.2 ± 10.3 years and a median body mass index of 28.4 kg/m² (IQR 25.4-32.9 kg/m²). At baseline, the mean NRS pain and PGA were indicating moderate to high pain levels and disease impact as well as moderate to high levels of WOMAC pain, function, and stiffness, as shown in Table 2. Most patients (82.6%) had primarily knee OA, with a median duration of symptoms at baseline of 4.0 years (IQR 1.6-9.6 years). At baseline, 68.1% of all patients had a Kellgren & Lawrence score ≥ 2 and this percentage was somewhat lower in the derivation cohort than the in validation cohort (63.0% vs 72.0%, p = 0.03); no other differences were found between the derivation and validation cohorts (Table 2).

Analyses: Patient Reported Outcome Measures (PROMs)

There were few missing data: for the 515 included patients, missing data for the PROMs at baseline were 4.1%, and at 3 months, 13.1%. After 3 months, 112 patients (21.7%) worsened (slightly worse, worse, much worse), 232 patients (45.0%) did not change, and 171 patients (33.2%) improved, as indicated by the transition question. In the derivation cohort, 40 patients (18.3%) worsened, compared to 72 patients (24.3%) in the validation cohort (slightly worse, worse, much worse) (not significant). Table 3 shows the mean difference in outcome scores by transition reply at the 3 month follow-up. Patients who reported no change after 3 months showed a statistically significant improvement in all outcome measures (except for stiffness) with relatively higher changes in NRS pain and PGA. In addition, a clear asymmetry between the magnitude of change in the direction of worsening versus improvement can be seen in all outcome measures, with the amount of worsening in complaints in those patients who reported being worsened being lower than the amount of improvement in complaints in patients who reported being improved. The MCIW values are shown in Table 4.

Table 4. Minimal clinically important worsening (MCIW) values after outpatient conservative treatment in patients with knee or hip osteoarthritis in the derivation cohort (n=219).

	MCIW	
	Absolute	Relative % (compared to baseline score)
NRS pain	1.8	18.1
WOMAC pain	-4.1	-7.7
WOMAC function	-3.1	-3.9
WOMAC stiffness	-7.7	-25.0
PGA	0.6	8.7

Data are shown as mean for absolute MCIW and median for relative MCIW.
NRS numeric rating scale pain, PGA patient global assessment (both scales 0-10, where 0 equals no symptoms); WOMAC: Western Ontario and McMaster Universities pain, function, and stiffness (scale 0-100, where 100 equals no symptoms).

The previously used worsening criteria, including the inverse of the responder criteria, were shown to be moderately to highly specific (range 60-92%), but lacked sensitivity (range 22-59%) (Table 5).[1] The absolute and relative change based on estimated MCIW values showed the

same pattern, with a moderate to good specificity (range 66-92%), but low sensitivity (range 26-48%).

The sensitivity and specificity of worsening criteria proposed by the Expert group ranged from 26% to 65% and 67-92%, respectively, in the validation cohort (Table 6 for NRS pain and Table 7 for WOMAC pain). The three newly proposed worsening sets that performed best incorporated smaller absolute and relative changes compared with improvement criteria. Selecting the best performing worsening sets using the Youden Index and c-statistic yielded similar results, which confirms the robustness of the selection. Set 7 performed best, with a sensitivity and specificity of 59% and 74%, respectively, when using NRS pain, and a sensitivity and specificity of 54% and 74%, respectively, when using WOMAC pain. This set of worsening criteria that performed best in terms of sensitivity and specificity addressed: worsening in (1) pain ≥ 20% and absolute change ≥ 20 [in the case of using NRS pain (scale 0-10): absolute change ≥ 2], *or* (2) function ≥ 10% and absolute change ≥ 10, *or* (3) PGA ≥ 10% and absolute change ≥ 1 (Figure 1). The sensitivity analysis using the imputed dataset yielded similar results (data not shown).

Figure 1. Worsening criteria that performed best.

Worsening in:

- **pain ≥ 20% and absolute change ≥ 20 *or***
- **function ≥ 10% and absolute change ≥ 10 *or***
- **PGA ≥ 10% and absolute change ≥ 1**

Pain and function: WOMAC scale 0-100; PGA: patient global assessment, scale 0-10
*: in case of using NRS pain (scale 0-10): absolute change ≥ 2.

Table 5. Sensitivity and specificity of previously used worsening criteria and minimal clinically important worsening (MCIW) values in the validation cohort.

		Validation cohort (n=296)			
		Sensitivity (95% CI)	Specificity (95% CI)	*J*	ROC point (95% CI)
Previously used worsening criteria					
1	Alschuler 2013: change in pain or function ≥50% or change in pain and function ≥20%				
	NRS pain	0.48 (0.36-0.60)	0.79 (0.70-0.86)	0.27	0.63 (0.56-0.70)
	WOMAC pain	0.42 (0.30-0.56)	0.75 (0.65-0.83)	0.17	0.59 (0.51-0.66)
2	Angst 2002 (scale WOMAC 0-10)				
	Change in absolute WOMAC pain ≥0.64	0.32 (0.21-0.46)	0.74 (0.64-0.82)	0.06	0.53 (0.45-0.60)
	Change in relative WOMAC pain ≥14%	0.39 (0.27-0.53)	0.75 (0.65-0.83)	0.14	0.57 (0.49-0.64)
	Change in absolute WOMAC function ≥1.03	0.29 (0.18-0.42)	0.86 (0.77-0.92)	0.14	0.57 (0.50-0.64)
	Change in relative WOMAC function ≥ 22%	0.34 (0.22-0.47)	0.88 (0.79-0.93)	0.22	0.61 (0.54-0.68)
	Change in absolute WOMAC stiffness ≥0.29	0.41 (0.28-0.54)	0.69 (0.59-0.78)	0.09	0.55 (0.47-0.63)
	Change in relative WOMAC stiffness ≥6%	0.41 (0.28-0.54)	0.69 (0.59-0.78)	0.09	0.55 (0.47-0.63)
3	Pham 2004: change in pain or function ≥ 50% and absolute change ≥ 20 or improvement in at least 2 of the 3 following: pain ≥ 20% and absolute change ≥ 10, function ≥ 20% and absolute change ≥ 10, PGA ≥ 20% and absolute change ≥ 10				
	Inverse responder criteria with NRS pain	0.23 (0.14-0.34)	0.92 (0.86-0.96)	0.15	0.57 (0.52-0.63)
	Inverse responder criteria with WOMAC pain	0.22 (0.13-0.34)	0.87 (0.80-0.93)	0.09	0.55 (0.49-0.60)
4	White 2010: absolute score WOMAC function at 3 months ≥36.1 (range 0-68)	0.59 (0.46-0.72)	0.60 (0.50-0.70)	0.20	0.60 (0.52-0.68)
5	Yusuf 2011: change in WOMAC pain absolute ≥9.7 or WOMAC function ≥9.3 (scale 0-100)	0.42 (0.30-0.56)	0.71 (0.61-0.79)	0.13	0.57 (0.49-0.64)
	MCIW values				
1	Change in NRS pain absolute ≥ 1.8 and relative ≥ 18.1%	0.26 (0.16-0.39)	0.92 (0.85-0.97)	0.18	0.59 (0.53-0.65)
2	Change in WOMAC pain absolute ≤ -4.1 and relative ≤ -7.7%	0.48 (0.34-0.61)	0.66 (0.56-0.76)	0.14	0.57 (0.49-0.65)
3	Change in WOMAC function absolute ≤ -3.1 and relative ≤ -3.9%	0.46 (0.33-0.59)	0.73 (0.63-0.82)	0.19	0.60 (0.52-0.67)
4	Change in WOMAC stiffness absolute ≤ -7.7 and relative ≤ -25.0%	0.31 (0.19-0.44)	0.80 (0.71-0.87)	0.10	0.55 (0.48-0.62)
5	Change in PGA absolute ≥ 0.6 and relative ≥ 8.7%	0.46 (0.32-0.59)	0.83 (0.73-0.89)	0.28	0.64 (0.57-0.72)

NRS: numeric rating scale pain; PGA: patient global assessment (both scales 0-10, where 0 equals no symptoms); WOMAC: Western Ontario and McMaster Universities pain, function, and stiffness (scale 0-100, where 100 equals no symptoms); CI: confidence interval; J = Youden index; ROC: receiver operating characteristic.

Table 6. Sensitivity and specificity of newly developed worsening criteria using numeric rating scale (NRS) pain, in the derivation and validation cohorts.

		Derivation cohort (n=219)				Validation cohort (n=296)			
		Sensitivity (95% CI)	Specificity (95% CI)	*J*	ROC point (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	*J*	ROC point
Newly proposed worsening criteria									
Set 1	Worsening in pain ≥50% and absolute change ≥2	0.18 (0.08-0.34)	0.95 (0.89-0.90)	0.13	0.57 (0.50-0.63)				
Set 2	Worsening in: • pain ≥50% and absolute change ≥2 *or* • function ≥20% and absolute change ≥10 *and* • PGA ≥20% and absolute change ≥10	0.18 (0.07-0.33)	0.95 (0.89-0.99)	0.13	0.56 (0.50-0.63)				
Set 3	Worsening in: • pain ≥50% and absolute change ≥2 *or* • function ≥20% and absolute change ≥10	0.33 (0.19-0.49)	0.80 (0.71-0.87)	0.13	0.56 (0.48-0.65)				
Set 4	Worsening in pain ≥20% and absolute change ≥2	0.28 (0.15-0.45)	0.91 (0.84-0.96)	0.20	0.60 (0.52-0.67)	0.26 (0.16-0.39)	0.92 (0.85-0.97)	0.18	0.59 (0.53-0.65)
Set 5	Worsening in • pain ≥20% and absolute change ≥2 *and* • function ≥20% and absolute change ≥20	0.08 (0.02-0.20)	0.98 (0.93-1.00)	0.06	0.53 (0.49-0.57)				
Set 6	Worsening in: • pain ≥20% and absolute change ≥2 *or* • function and PGA ≥10% and absolute change ≥ 1 (PGA) or ≥10 (function)	0.30 (0.17-0.47)	0.90 (0.82-0.95)	0.20	0.60 (0.52-0.68)	0.27 (0.17-0.39)	0.92 (0.86-0.96)	0.19	0.60 (0.54-0.65)
Set 7	Worsening in: • pain ≥20% and absolute change ≥2 *or* • function ≥10% and absolute change ≥10 *or* • PGA ≥10% and absolute change ≥1	0.53 (0.36-0.69)	0.69 (0.60-0.78)	0.22	0.61 (0.52-0.70)	0.59 (0.47-0.71)	0.74 (0.66-0.82)	0.34	0.67 (0.60-0.74)
Set 8	Worsening in: • pain ≥10% and absolute change ≥1 *or* • function ≥10% and absolute change ≥10	0.55 (0.39-0.71)	0.65 (0.55-0.74)	0.20	0.60 (0.51-0.69)	0.52 (0.40-0.64)	0.74 (0.65-0.81)	0.28	0.63 (0.56-0.70)
Set 9	Worsening in at least 2 of the 3 following: • pain ≥10% and absolute change ≥1 • function ≥10% and absolute change ≥10 • PGA ≥10% and absolute change ≥1	0.29 (0.15-0.46)	0.85 (0.77-0.92)	0.14	0.57 (0.49-0.65)				
Set 10	Worsening in • pain ≥10 % and absolute change ≥ 10 *or* • function ≥10 % and absolute change ≥ 10 *or* • PGA ≥10 % and absolute change ≥ 1	0.65 (0.48-0.79)	0.59 (0.49-0.68)	0.24	0.62 (0.53-0.71)	0.65 (0.53-0.76)	0.68 (0.58-0.76)	0.32	0.66 (0.49-0.65)

NRS: numeric rating scale pain; PGA: patient global assessment (both scales 0-10, where 0 equals no symptoms); WOMAC: Western Ontario and McMaster Universities function, scale 0-100, where 100 equals no symptoms);
CI: confidence interval; J = Youden index; ROC: receiver operating characteristic.

Table 7. Sensitivity and specificity of newly developed worsening criteria using Western Ontario and McMaster Universities (WOMAC) pain, in the derivation and validation cohorts.

		Derivation cohort (n=219)				Validation cohort (n=296)			
		Sensitivity (95% CI)	Specificity (95% CI)	*J*	ROC point (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	*J*	ROC point (95% CI)
Newly proposed worsening criteria									
Set 1	Worsening in pain ≥50% and absolute change ≥ 20	0.06 (0.01-0.20)	0.91 (0.84-0.96)	-0.03	0.49 (0.44-0.54)				
Set 2	Worsening in: • pain ≥50% and absolute change ≥ 20 *or* • function ≥20% and absolute change ≥10 *and* • PGA ≥20% and absolute change ≥10	0.05 (0.01-0.17)	0.92 (0.86-0.97)	-0.02	0.49 (0.44-0.53)				
Set 3	Worsening in: • pain ≥50% and absolute change ≥ 20 *or* • function ≥20% and absolute change ≥10	0.27 (0.13-0.44)	0.82 (0.73-0.89)	0.09	0.54 (0.46-0.63)	0.34 (0.22-0.47)	0.85 (0.76-0.91)	0.19	0.60 (0.52-0.66)
Set 4	Worsening in pain ≥20% and absolute change ≥ 20	0.15 (0.05-0.31)	0.87 (0.78-0.93)	0.02	0.51 (0.44-0.58)				
Set 5	Worsening in • pain ≥20% and absolute change ≥ 20 *and* • function ≥20% and absolute change ≥20	0.03 (0.00-0.15)	0.90 (0.82-0.95)	-0.08	0.46 (0.42-0.50)				
Set 6	Worsening in: • pain ≥20% and absolute change ≥ 20 *or* • function and PGA ≥10% and absolute change ≥ 1 (PGA) or ≥10 (function)	0.13 (0.04-0.27)	0.87 (0.81-0.94)	0.01	0.51 (0.45-0.57)				
Set 7	Worsening in: • pain ≥20% and absolute change ≥ 20 *or* • function ≥10% and absolute change ≥10 *or* • PGA ≥10% and absolute change ≥1	0.51 (0.35-0.68)	0.70 (0.60-0.78)	0.21	0.60 (0.51-0.70)	0.54 (0.41-0.66)	0.74 (0.65-0.81)	0.27	0.64 (0.56-0.71)
Set 8	Worsening in: • pain ≥10% and absolute change ≥ 10 *or* • function ≥10% and absolute change ≥10	0.44 (0.27-0.62)	0.76 (0.66-0.84)	0.20	0.60 (0.50-0.70)	0.41 (0.28-0.54)	0.75 (0.65-0.83)	0.15	0.58 (0.50-0.65)
Set 9	Worsening in at least 2 of the 3 following: • pain ≥10% and absolute change ≥ 10 • function ≥10% and absolute change ≥10 • PGA ≥10% and absolute change ≥1	0.27 (0.13-0.44)	0.83 (0.74-0.90)	0.10	0.55 (0.46-0.63)	0.32 (0.21-0.46)	0.85 (0.76-0.91)	0.17	0.59 (0.52-0.66)
Set 10	Worsening in • pain ≥10 % and absolute change ≥ 10 *or* • function ≥10 % and absolute change ≥ 10 *or* • PGA ≥10 % and absolute change ≥ 1	0.56 (0.40-0.72)	0.67 (0.57-0.76)	0.23	0.62 (0.52-0.71)	0.57 (0.44-0.68)	0.67 (0.57-0.75)	0.23	0.62 (0.54-0.69)

PGA: patient global assessment (scale 0-10, where 0 equals no symptoms); WOMAC: Western Ontario and McMaster Universities pain and function (scale 0-100, where 100 equals no symptoms); CI: confidence interval; J = Youden index; ROC: receiver operating characteristic.

Discussion

To our knowledge, this exploratory project is the first to propose a validated set of clinical worsening criteria for patients with knee or hip OA. These clinical worsening criteria could be a starting point to help patients, doctors, and researchers in distinguishing patients who have clinically worsened over time from those who have not. Evaluation of study results using these criteria as an outcome measure in longitudinal studies could also facilitate research to predict factors associated with symptomatic progression in knee or hip OA; this is a recently identified research priority area for EULAR.[46]

How would we recommend the application (and further validation) of these preliminary criteria? Dichotomized clinical worsening criteria increase the interpretability, making it more meaningful and likely to be applied both in clinical practice and in the research setting. Because of the elaborated insight into the performance of the different clinical worsening criteria, a selection could be made depending on the goals for use, and also the measurement instrument used for pain (NRS or WOMAC pain). Our newly proposed worsening criteria are to be preferred to literature-based criteria, as they show an acceptable combination of sensitivity and specificity. These preliminary validated worsening criteria could be used as an outcome measure in longitudinal studies to predict factors associated with clinical worsening. Future research should focus on further refinement of these criteria to improve sensitivity and specificity. However, our proposed sets of worsening criteria show similar values of sensitivity but better specificity compared to the OMERACT-OARSI responder criteria.[1] Future research should also take into account growing evidence that OA is a heterogeneous condition that probably requires differentiation in phenotypes.[47] Heterogeneity may explain, in part, the relatively low sensitivity that we found and suggests that different PROM domains may apply for different OA phenotypes.

Some findings of our study warrant closer inspection. An intriguing finding was that the stable patient group, assessed by the transition scale, reported significant improvements in PROMs on pain, PGA, and function after 3 months. This phenomenon could be due to a response shift, influenced by the expectations of these patients. However, we deemed a period up to 3 months long enough to allow for worsening and brief enough to minimize the risk of a response shift.[29] This issue warrants further research. An additional observation is that the proposed clinical worsening criteria incorporate relatively small absolute and relative changes compared with responder criteria, which confirms previous literature indicating that the amount of change that patients consider important is different with worsening versus improvement. This supports using separate values for improvement and deterioration and may indicate that MCIW values are group specific.[23-26] This confirms the need to develop separate worsening criteria for OA. Caution is needed when interpreting and using published MCIW values, because these values have a number of disadvantages, such as its dependency on the baseline scores, on characteristics of the population, and on the method used for assessment. Because of growing consensus that a single MCIW value cannot be applied across all populations, we did not use previously published MCIW values.[48,49] Furthermore, we used both absolute and relative change scores, and the latter method is frequently used to adjust for baseline covariates. Recent research suggests that researchers and clinicians should consider both the final state as well as change. However, in our study, adding the final state to the sets of clinical worsening criteria using an absolute cut-off point of the involved domains did not influence the results in terms of optimizing the Youden index or c-statistic (data not shown).

Our study has several strengths. A stringent protocol consistent with recent recommendations for PROM validation was prepared and followed. All PROMs were targeted on the index joint, being the most symptomatic knee or hip at baseline. We believe that our literature review identified all relevant articles that have previously used worsening criteria, and that our expert process rigorously developed consensus regarding criteria that should be examined to characterize clinically important worsening of a signal joint for individual patients with knee or hip OA. We first examined all of the expert-identified criteria for sensitivity and specificity in a derivation cohort of more than 200 OA patients, and then tested those that performed the best in a larger cohort of almost 300 patients. Importantly, all patients were included from daily clinical practice, which, in our opinion, provides external generalizability of our results. The last point is especially important, because previous studies have shown different amounts of change in PROM values in daily practice compared to those in clinical trials.[3]

Some limitations of our study and of validation studies in general should be acknowledged. Most of the patients in our reference study had knee OA, so generalizability for hip OA is not assured. In assessing validity based on sensitivity and specificity, transition questions completed by patients were used, often considered as a reasonable proxy for a gold standard, which is consistent with consensus obtained at OMERACT 8.[28] This method is in line with current thinking of adding patients' perspectives and was also used when proposing improvement criteria.[1,28] However, the anchor-based method has three problems with regard to validity of the anchor.[48] First, a single item is assumed to be less reliable, and thus less valid, then a score of a multi-item instrument. Secondly, patients' ratings on an anchor are more highly correlated with the follow-up score than with the baseline score, and this may lead to recall bias or response shift and could have influenced our results. Although we deemed a time frame of 3 months long enough to detect worsening and brief enough to minimize the risk of a response shift.[29], we cannot rule out a possible differential response shift between the transition question and the criteria tested. A third validity problem of the anchor is that we cannot be sure that "improved" means an important change to patients.[48] For now, this external anchor seems to be the best, taking into account the literature regarding this topic and the available data in our study. Finally, we did not examine duration of worsening, and future studies should take this key factor into account. If duration is also included, with, for example, a second measurement after 1-2 weeks, the reported worsening criteria may increase in specificity, but lose sensitivity.

Importantly, not all issues considered relevant previously in OMERACT filter 2.0 could be taken into account; for example, the patients' perspective, cross-cultural differences, and contextual factors.[50] However, the process for developing clinical worsening criteria proved challenging, and thus, we consider our results as an initial step towards assuring the validity of ultimately reliable and useful clinical worsening criteria. This exploratory project could be seen as starting point for future research that includes patient focus groups to complement expert consensus recommendations, ideally conducted internationally, enabling identification of cross-cultural and contextual factors to ensure a comprehensive and consensus-based list of key domains to be measured to develop a sensitive, responsive, and reliable instrument to measure OA worsening. Methods described by rheumatoid arthritis flare working groups may be useful for patients to address which domains based on PROMs should be included, followed

by an inventory of adding more objective features, such as laboratory values.[51] However, adding imaging modalities, for example, for this purpose will be difficult, because these are known to change very slowly.[9]

Because the patients' perception of change regarding longer time spans (> 3 months) in prediction or observational studies needs to be transformed into changes in measurement instruments to minimize the risk of a response shift, we described newly symptomatic worsening criteria.[22]

In conclusion, the results from this first exploratory project proposing preliminary worsening criteria in OA using a literature, expert opinion, and data-driven approach, showed that previously used criteria for clinical worsening showed a large variety in definitions of clinical worsening. Previously used criteria for clinical worsening as an outcome measure are specific but lack sensitivity, and we showed that newly clinical worsening criteria should incorporate relatively small absolute and relative changes compared with improvement criteria. These criteria, indicating symptomatic worsening over time, could be used as an outcome measure to facilitate research on prognostic factors for the symptomatic progression of OA.

Acknowledgement
We thank all patients who participated and the members of the Expert Group (apart from the authors: K. Bevers, N. Cuperus, C. Kersten and J. Vriezekolk).

References

1. Pham T, van der Heijde D, Altman RD, Anderson JJ, Bellamy N, Hochberg M, et al. OMERACT-OARSI initiative: osteoarthritis research society international set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage 2004;12:389-99.
2. Tubach F, Ravaud P, Martin-Mola E, Awada H, Bellamy N, Bombardier C, et al. Minimum clinically important improvement and patient acceptable symptome state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. Arthritis Care Res (Hobroken) 2012;64:1699-707.
3. Perrot S and Bertin P. "Feeling better" or "feeling well" in usual care of hip and knee osteoarthritis pain: determination of cutoff points fot patient acceptable symptom state (PASS) and mimimal clinically important improvement (MCII) at rest and on movement in a national multicenter cohort study of 2414 patients with painful osteoarthritis. Pain 2013;154:248-56.
4. Bastick AN, Runhaar J, Belo JN, Bierma-Zeinstra SMA. Prognostic factors for progression of clinical osteoarthritis of the knee: a systematic review of observational studies. Arthritis Res Ther 2015;17:152.
5. Bedson J, Croft PR. The discordance between clinical and radiographic knee osteoarthritis: a systematic search and summary of the literature. BMC Musculoskelet Disord. 2008;9:116.
6. Holla JFM, Steultjens MPM, Roorda LD, Heymans MW, ten Wolde S, Dekker J, et al. Prognostic factors for the two-year course of activity limitations in early osteoarthritis of the hip and/or knee. Arthritis Care Res (Hobroken) 2010;62:1415-25.
7. Riddle DL, Moxley G, Dumenci L. Associations between statin use and changes in pain, function and structural progression: a longitudinal study of persons with knee osteoarthritis. Ann Rheum Dis 2013;72:196-203.
8. Sharma L, Cahue S, Song J, Hayes K, Pai Y, Dunlop D. Physical functioning over three years in knee osteoarthritis. Arthritis Rheum 2003;48:3359-70.
9. Wesseling J, Bierma-Zeinstra SMA, Kloppenburg M, Meijer R, Bijlsma JWJ. Worsening of pain and function over 5 years in individuals with "early" OA is related to structural damage: data from the osteoarthritis initiative and CHECK (Cohort Hip & Cohort Knee) study. Ann Rheum Dis 2013;0:1-7.
10. Kastelein M, Luijsterburg PAJ, Belo J, Verhaar JAN, Koes BW, Bierma-Zeinstra SMA. Six-year course and prognosis of nontraumatic knee symptoms in adults in general practise: a prospective cohort study. Arthritis Care Res (Hobroken) 2011;63:1287-94.
11. Yusuf E, Bijsterbosch J, Slagboom PE, Kroon HM, Rosendaal FR, Huizinga TWJ, et al. Association between several clinical and radiological determinants with long-term clinical progression and good prognosis of lower limb osteoarthritis. Plos one 2011;6:e25426.
12. Jinks C, Jordan P, Blagojevic M, Croft P. Predictors of onset and progression of knee pain in adults living in the community. A prospective study. Rheumatology 2008;47:368-74.
13. Pisters MF, Veenhof C, van Dijk GM, Heymans MW, Twisk JWR, Dekker J. The course of limitations in activities over 5 years in patients with knee and hip osteoarthritis with moderate functional limitations: risk factors for future for functional decline. Osteoarthritis Cartilage 2012;20:503-10.
14. Alschuler KN, Molton IR, Jensen MP, Riddle DL. Prognostic value of coping strategies in a community-based sample of persons with chronic symptomatic knee osteoarthritis. Pain 2013;154:2775-81.
15. Glass NA, Torner JC, Frey Law LA, Wang K, Yang T, Nevitt MC, et al. The relationship between quadriceps muscle weakness and worsening of knee pain in the MOST cohort: a 5-year longitudinal study. Osteoarthritis Cartilage 2013;21:1154-9.
16. Urish KL, Keffalas MG, Durkin JR, Miller DJ, Chu CR, Mosher TJ. T2 texture index of cartilage can predict early symptomatic OA progression: data from the osteoarthritis initiative. Osteoarthritis Cartilage 2013;21:1550-7.

17. White DK, Zhang Y, Niu J, Keysor JJ, Nevitt M, Lewis C, et al. Do worsening knee radiographs mean greater chances of severe functional limitation? Arthritis Care Res (Hobroken) 2010;62:1433-9.
18. Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. J Rheumatol 2002;29:131-9.
19. Brandt K, Heilman DK, Slemenda C, Katz BP, Mazzuca SA, Braunstein EM, et al. Quadriceps strength in women with radiographically progressive osteoarthritis of the knee and those with stable radiographic changes. J Rheumatol 1999;26:2431-7.
20. Sands GH, Brown PB, Essex MN. The efficacy of continuous versus intermittent celecoxib treatment in osteoarthritis patients with body mass index ≥ 30 and < 30 kg/m2. Open Rheumatol J 2013;7:32-7.
21. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the mimimal clinically important difference. Controlled Clin Trials 1989;10:407-15.
22. Kamper SJ, Ostelo RWJG, Knol DL, Maher CG, de Vet HCW, Hancock MJ. Global perceived effect scales provided reliable assessmetns of health transition in people with musculoskeletal disorders, but ratings are strongly influenced by current status. J Clin Epidemiol 2010;63:760-6.
23. Cella D, Hahn EA, Dineen K. Meaningful change in cancer specific quality of life scores : differences between improvement and worsening. Qual Life Res 2002;11:207-21.
24. Guyatt GH, Jaeschke RJ. Reassessing quality-of-life instruments in the evaluation of new drugs. Pharmacoeconomics 1997;12:621-6.
25. Schwartz AL, Meek PM, Nail LM, Fargo J, Lundquist M, Donofrio, et al. Measurement of fatigue: determining minimally important clinical differences. J Clin Epidemiol 2002;55:239-44.
26. Rouquette A, Blanchin M, Sébille V, Guillemin F, Côte S, Fallissard B, et al. The minimal clinically important difference determined using item response theory models: an attempt to solve the issue of the association with baseline scores. J Clin Epidemiol 2014;67:433-40.
27. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? Ann Rheum Dis 2007;66(Suppl III):iii40-iii41.
28. Tubach F, Ravaud P, Beaton D, Boers M, Bombardier C, Felson DT, et al. Minimal clinically important improvement and patient acceptable symptom state for subjective outcome measures in rheumatic disorders. J Rheumatol 2007;34:1188-93.
29. Robling M, Hood K. Response shift, responsiveness or recall bias. Br J Gen Pract 2002;52:585.
30. Redelmaier DA, Lorig K. Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. Arch Intern Med 1993;153:1337-42.
31. Perkins NJ, Schisterman EF. The inconsistency of 'optimal' cutpoints obtained using two criteria based on the receiver operating characteristic curve. Am J Epidemiol 2006;163:670-5.
32. 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 Rheum 1986;29:1039-49.
33. 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;34:505-14.
34. Snijders GF, van den Ende CHM, Fransen J, van Riel PLCM, Stukstette MJPM, Defoort KC, et al. Fatigue in knee and hip osteoarthritis: the role of pain and physical function. Rheumatology (Oxford) 2011;50:1894-900.
35. De Groot IB, Favejee MM, Reijman M, Verhaar JAN, Terwee CB. The Dutch version of the knee injury and osteoarthritis outcome score: a validation study. Health Qual Life Outcomes 2008;6:16.

36. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinicaly relevant changes in patient-reported outcomes in knee and hip osteoarthritis: the minimal clinically important improvement. Ann Rheum Dis 2005;64:29-33.
37. De Vet HCW, Terwee CB, Mokkink LD, Knol DL. Measurement in Medicine. New York: Cambridge University Press, 2011.
38. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clincal research: potential and pitfalls. BMJ 2009;338:b2393.
39. Williams VJ, Piva SR, Irrgang JJ, Crossley C, Fitzgerald K. Comparison of reliability and responsiveness of patient-reported clinical outcome measures in knee osteoarthritis rehabilitation. J Orthop Sports Phys Ther 2012;42:716-23.
40. Stannus OP, Jones G, Blizzard L, Cicuttini FM, Ding C. Associations between serum levels of inflammatory markers and change in knee pain over 5 years in older adults: a prospective cohort study. Ann Rheum Dis 2013;72:535-40.
41. Peat G, Thomas E. When knee pain becomes severe: a nested case-control analysis in community-dwelling older adults. J Pain 2009;10:798-808.
42. Altman R, Brandt T, Hochberg M, Moskovitz R. Design and conduct of clinical trials of patients with osteoarthritis. Recommendations from a task force of the Osteoarthritis Research Society. Osteoarthritis Cartilage 1996;4:217-43.
43. Bellamy N, Kirwan J, Altman R, Boers M, Brandt K, Brooks P, et al. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip and hand osteoarthritis. Results of consensus development at OMERACT III. J Rheumatol 1997;24:799-802.
44. Group for the respect of ethics and excellence in science (GREES): osteoarthritis section. Recommendations for the registration of drugs used in the treatment of osteoarthritis. Ann Rheum Dis 1996;55:552-7.
45. Dougados M, Leclaire P, van der Heijde D, Bloch DA, Bellamy N, Altman RD. Response criteria for clinical trials on osteoarthritis of the knee and hip: a report of the Osteoarthritis Research Society International Standing Committee for Clinical Trials respons criteria initiative. Osteoarthritis Cartilage 2000;8:395-403.
46. Conaghan PG, Kloppenburg M, Schett G, Bijlsma JWJ. Osteoarthritis research priorities: a report from a EULAR ad hoc expert committee. Ann Rheum Dis 2014;73:1142-5.
47. Bijlsma JWJ, Berenbaum F, Lafeber FPJG. Osteoarthritis: an update with relevance for clinical practice. Lancet 2011;377:2115-26.
48. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol 2010;63:524-34.
49. Mills KAG, Naylor JM, Eyles JP, Roos EM and Hunter DJ. Examining the minimal important difference of patient-reported outcome measures for individuals with knee osteoarthritis : a model using the knee injury and osteoarthritis score. J Rheum. 2016;43:395-404.
50. Boers M, Kirwan JR, Wells G, Beaton D, Gossec L, d'Agostino MA, et al. Developing core outcome measurement sets for clinical trials : OMERACT Filter 2.0 J Clin Epidemiol 2014;67:745-53.
51. Bartlett SJ, Bykerk VP, Cooksey R, Choy EH, Alten R, Christensen R, et al. Feasibility and domain validation of rheumatoid arthritis flare cor domain set: report of the OMERACT 2014 RA flare group. J Rheumatol 2015;42:2185-9.

Chapter 5

Short term clinical worsening is a clear predictor for worsening at 2 years in established knee and hip osteoarthritis

Elien A.M. Mahler
Alfons A. den Broeder
Nathan den Broeder
Johannes W.J. Bijlsma
Gijs F. Snijders
Frank H.J. van den Hoogen
Cornelia H.M. van den Ende

Submitted

Abstract

Objective

To estimate (1) the proportion of knee and hip OA patients showing worsening at 2 years, and (2) to examine the additional predictive value of failure of optimised nonsurgical treatment for worsening at 2 years.

Methods

Data of patients participating in the longitudinal CONTROL-PRO study (patients fulfilling American College of Rheumatology clinical criteria for knee or hip OA) were used. Measurements of pain, functioning and patient global assessments were performed at baseline, 3 months and 2 years. Worsening at 2 years was defined as fulfilling the recently validated clinical worsening criteria for knee and hip OA, or total joint replacement (TJR). Logistic regression was performed with worsening at 2 years as dependent variable.

Results

The 297 included patients were predominantly women (66%) with a mean age of 55 years. At 2 years, 61% showed worsening (knee 59%; hip 71%) and 24% had undergone a TJR (knee 19%; hip 51%). Clinically worsening at 3 months appeared to be a clear independent predictor for worsening at 2 years (odds ratio (OR) 2.8 95% CI 1.5–5.2) with a moderate discriminative ability (area under the receiver operating characteristic curve (AUC) 0.68 95% CI 0.57–0.70). Similar results were obtained when only TJR at 2 years was used as outcome measure (OR 4.1 95% CI 2.0–8.4) with good AUC (0.82 95% CI 0.76–0.87).

Conclusion

Our findings suggest that re-assessment of symptoms after optimised nonsurgical treatment could be meaningful in clinical decision making for TJR. Furthermore, this information could be used to identify subgroups of patients potentially eligible for novel and advanced treatment options.

Introduction

Osteoarthritis (OA) is considered to be the most prevalent chronic joint disease and is one of the leading causes of pain and disability worldwide, with knee and hip being frequently involved joints.[1-4] Meanwhile, the incidence and prevalence of knee and hip OA is rising substantially due to the ageing population and the epidemic of obesity which portends the associated future economic burden.[3-5,6] The natural course of pain and physical functioning in knee and hip OA is highly variable: most patients have been found to remain stable, while a subset will gradually worsen.[7-10] Both the involvement of a high socioeconomic burden as well as the variability on the natural course, mandate that identification of risk factors for clinical decline are important.[11] This could be used to inform both patients and healthcare professionals, to identify patients at risk for deterioration in order to adapt treatment or to select individuals potentially eligible for novel therapies.

Longitudinal studies on validated clinical outcomes in knee and hip OA are lacking and therefore, little is known about the course and determinants associated with clinical deterioration of knee and hip OA.[12] In contrast, many studies have been performed to determine prognostic factors for radiographic progression of knee and hip OA.[8,13] However, a clear discordance between radiographic and symptomatic knee OA has been well established. This highlights the need to also focus on symptomatic rather than radiological outcomes.[7,12,14,15] Symptomatic progression of knee OA is most relevant for both patient and healthcare professionals. Therefore, an understanding of the risk factors that predict clinical worsening in knee and hip OA would be useful to give insight in daily clinical practice.

Validated clinical worsening criteria have not been available up to recently. This is corroborated by two recently published systematic reviews of prognostic factors for symptomatic progression of knee OA, concluding that it was impossible to properly summarize the evidence due to different ways of measuring clinical progression.[12,14] Recently, we validated clinical worsening criteria that have been proposed to identify patients who have been deteriorated, enabling longitudinal outcome studies on determinants for clinical worsening over time of knee and hip OA[16], which corresponds to the current opinion to use symptom progression as outcome measure.[12,13]

Several international consensus-based clinical guidelines for the management of knee and hip OA are available, emphasizing the importance and efficacy on nonsurgical treatment modalities, which include education, exercise, step up analgesics, life style advice concerning physical activity and advice on weight loss in patients that are overweight.[17] An important issue in clinical practice would be to evaluate whether failure of optimal standardized nonsurgical treatment, is an additional risk factor for worsening over time, beyond history taking and physical examination. Therefore, the aims of this study are to estimate 1) the proportion of knee and hip OA patients showing worsening at 2 years, and 2) to examine the additional predictive value of failure of optimized standardized nonsurgical treatment during 3 months for worsening at 2 years.

Methods

Design, setting and participants

This study is part of the longitudinal study CONTROL-PRO (Cohort Of Non-invasively Treated Osteoarthritis of Lower Extremities – Pain, function and Radiological Outcome).[18] Consecutive patients at the rheumatology specialized outpatient clinic, carried out in the framework of a specialized knee and hip OA outpatient clinic, were invited to participate. All patients fulfilled the clinical American College of Rheumatology (ACR) criteria for knee or hip OA and were at inclusion deemed ineligible for total joint replacement (TJR) by their orthopaedic surgeon. The most symptomatic knee or hip at baseline was considered the index joint.

All patients received standardized nonsurgical treatment during the first 3 months which included education, referral for physical therapy (aerobic and strengthening exercises), step-up analgesics using acetaminophen based on the NRS pain (patients were contacted every 4 weeks; next step only if NRS pain > 4 unless contraindicated), followed by a first NSAID, substitution of NSAID and tramadol thereafter), and advice on weight reduction if indicated (goal 5% weight loss when BMI ≥ 28 kg/m^2), as described elsewhere.[18] Exclusion criteria were: other rheumatic or severe orthopaedic diseases leading to inflammatory arthritis or secondary OA, co-morbidity exceeding the complaints or limitations of the knee or hip OA, orthopaedic procedures planned within the next three months, or cognitive or sensorimotor problems interfering with questionnaire completion. For the current study, patients were invited for a follow-up visit after 2 years if they completed both baseline and 3 months follow-up visits, and were included when they indeed completed the 2 years follow-up visit. The local Medical Research Ethics Committee, region Arnhem-Nijmegen (The Netherlands) approved the study design (study number 2009/095). All patients signed informed consent.

Data acquisition

Visits were scheduled at baseline, at 3 months, and at 2 years. Two year visits were scheduled only for those patients who completed both baseline and 3 month questionnaires, and had not undergone a TJR. At inclusion, demographic and OA-related characteristics were collected, using a standardized interview and physical examination as described elsewhere.[18] The number of comorbidities was assessed using the long version of the Dutch Arthritis Impact measurement Scales.[19] At baseline, erythrocyte sedimentation rate (ESR) and C-reactive protein (CRP) were measured and radiographs were assessed. To examine structural abnormalities, knee (weight-bearing posterior-anterior fixed flexion) or hip radiographs (both anterior-posterior supine position) were obtained. Scoring of the index joint was done blinded for clinical data, using Kellgren and Lawrence (K&L) grading system, atlas based, by an experienced rheumatologist.[20,21] Previous intraobserver reliability (kappa) for K&L score ranged from 0.68 to 0.89 (re-scored in 20 participants.[18] At all visits, patients completed a standardized set of patient reported outcomes measures.

Patient Reported Outcome Measures (PROMs)

Pain intensity and the patient global assessment (PGA) of OA impact during the last week were measured on a 0-10 point numeric rating scale (NRS) where 0 equals no symptoms. Patients also completed the Dutch Knee/Hip injury and Osteoarthritis Outcome Score (KOOS/HOOS) questionnaire (Likert scale version).[22] From the KOOS/HOOS, the Western Ontario and McMaster Universities (WOMAC) scores can be derived, with WOMAC pain, function, and stiffness subscales presented as standardized scores ranging from 0 to 100, where 100 equals no symptoms.[23] Fatigue was measured at baseline and 3 months with the 8-itemed "Subjective Fatigue" subscale of the Checklist Individual Strength (CIS).[24] The total score can range from 8-56 points where scores of ≥ 35 represent severe fatigue. Fear of movement was measured with the Tampa Scale for Kinesiophobia (TSK)[25], where scores > 37 represent excess fear of movement. Mental health was measured with the mental component score and calculated with corresponding subscales of the 36-item Short Form Health Survey (SF-36), a widely used generic health status questionnaire comprising eight areas of health status, with higher scores indicating better health (range 0-100).[26,27]

Primary outcome

Worsening at 2 years was operationalized as TJR in the index joint, or fulfilling recently validated clinical worsening criteria for knee and hip OA: worsening in: pain ≥ 20% and absolute change ≥ 20 or function ≥ 10% and absolute change ≥ 10 or PGA ≥10% and absolute change ≥1 (scale 0-10) compared to baseline values.[16] We used two different pain outcome measures -NRS and WOMAC- and consequently two distinct sets of worsening criteria i.e. worsening using NRS pain and worsening using WOMAC pain, respectively. As described elsewhere, these literature- and expert-group-based worsening criteria were first tested in a derivation cohort (n=219) and confirmed in a validation cohort (n=296). Both datasets incorporated observational data of patients with knee and hip OA who received multimodal conservative treatment. This set performed best regarding sensitivity (59%) and specificity (74%).[16] Clinical worsening at 3 months was dichotomized similarly, but with different time points i.e. change between 3 months and baseline values.

Statistical analysis

Patient characteristics, follow-up at 3 months and 2 years

Descriptive statistics were provided as mean and standard deviation (SD) or median and interquartile range (IQR) or numbers with percentages when appropriate. T-tests or chi-squared tests were used to compare baseline and 3 month values between patients who were included in the current analysis and patients who were lost to follow-up and to compute differences in knee and hip OA patients for worsening at 2 years.

Additional predictive value of failure of optimized nonsurgical treatment during 3 months

Multivariate logistic regression analyses were performed with clinical worsening at 2 years as dependent variable. According to the TRIPOD statement, backward logistic regression analyses - guided by the Akaike information criterion (p=0.157) - was used to build the full model (which included only baseline independent variables) for worsening at 2 years (dependent variable).[28] Based on the literature and clinical relevance, the following independent variables were selected: age, gender, BMI, affected joint, comorbidities, pain, PGA, function, stiffness, CRP, fatigue, mental component scale of SF-36 and K&L score and used in developing the full multivariable model. Separate models were run for NRS pain and for WOMAC pain, where both the independent variable for pain and the outcome measure for pain differed. Results are presented as odds ratio (OR) with 95% confidence interval (95% CI). Secondly, clinical worsening at 3 months despite optimized nonsurgical treatment was added as independent variable to the model. For including variables, we used a rule of thumb as recommended by various authors[29], that a minimum of 10 events per variable is required to obtain a reliable and concise prediction model. To reduce the impact of missing data, data at baseline and 3 months

5

was imputed using multiple imputations to create 20 datasets and results were combined using Rubin's rules.[30,31] The discriminatory ability of the final model was estimated using the area under (AUC) the receiver operating characteristic (ROC) curve, which is similar to the concordance-statistic (c-statistic). An AUC of 1 indicates perfect discrimination, while an AUC of 0.5 indicates discrimination no better than chance. Moreover, the positive and negative predictive values (PPV and NPV) as well as sensitivity and specificity of fulfilling clinical worsening criteria at 3 months and worsening at 2 years were estimated. Finally, the pre- and post-test probability was calculated and considered clinically relevant when the increase was above 15%.[32,33] Furthermore, we performed two sensitivity analyses; one with TJR at 2 years as dependent variable and one on the subgroup of patients with knee OA. All analyses were performed using STATA 13.1.

Results

Patient characteristics

No relevant and significant differences were found between the patients included in the analyses (n=297) and the patients who did not reply to the invitation for the 2 years assessment (Figure 1, n=142, 32.4%) with regard to all baseline values presented in Table 1, except for the proportion of patients with baseline K&L ≥ 2 (included 74% versus not-replying 60%, p=0.06). Three months data were available for 54 out of 142 not-replying patients and we found no significant difference in proportion with clinical worsening between patients not-replying and patients included in the current analysis (24%; 95% CI 22–27% versus 28%; 95% CI 25–31% respectively). The cohort consisted predominantly of women (66%), with a mean age of 55 years and median BMI of 28 kg/m² who are moderate to severely disabled by their disease considering the relatively high scores for pain and PGA, and the high proportion of patients showing fear of movement and severe fatigue.

Figure 1. Flowchart of patient recruitment and dropout.

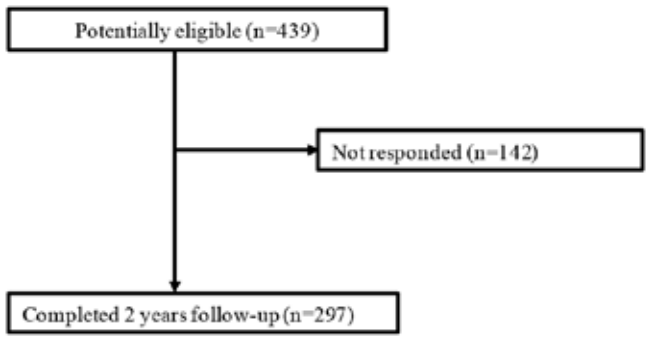


Table 1. Baseline characteristics of 297 patients with knee or hip osteoarthritis.

Sociodemographic characteristics	
Age, years	55.0 (9.6)
Women, n (%)	195 (65.6)
Body mass index, kg/m², median (IQR)	27.9 (25.3 – 32.9)
Duration of symptoms , years, median (IQR)	3.8 (1.6 – 10.4)
Index joint knee, n (%)	252 (84.8)
Education, low/middle, n (%)	214 (72.1)
Comorbidities, > 1, n (%)	119 (40.1)
Clinical parameters	
NRS pain (0-10)	5.9 (1.8)
NRS PGA (0-10)	6.1 (2.1)
WOMAC pain (0-100)	48.8 (19.2)
Function (0-100)	47.9 (19.7)
Stiffness (0-100)	45.9 (23.4)
ESR (mm/h), above upper limit, n (%)	30 (12.1)
C-reactive protein above upper limit, n (%)	16 (6.5)
Severe fatigue (CIS ≥ 35), n (%)	114 (43.9)
Fear of movement (TSK > 37), n (%)	147 (49.5)
SF-36 mental component score (range 2-71)	51.2 (11.6)
Kellgren and Lawrence ≥ 2, n (%)	209 (74.1)#

Values are mean (SD) unless stated otherwise; IQR: interquartile range
NRS: numeric rating scale; PGA: patient global assessment; WOMAC pain, function and stiffness, Western Ontario and McMaster Universities Osteoarthritis Index scale; ESR: Erythrocyte sedimentation rate; CIS: Checklist Individual Strength; TSK: Tampa Scale for Kinesiophobia; SF-36: Short Form 36 Health Survey;
Higher scores indicate more NRS pain, worse PGA, better scores for WOMAC pain, function and stiffness, better mental and physical health SF-36
15 missing values

Follow-up at 3 months and 2 years

Of the 297 patients in the cohort, a total of 79 (28%) and 181 patients (61%) clinically worsened at 3 months and 2 years respectively. A total of 71 out of 181 patients (39%) who worsened at 2 years - i.e. 24% of the whole group of patients - had undergone a TJR in the index joint on average 1.1 years (SD 0.5) after inclusion. As shown in figure 2, a higher proportion of knee OA patients showed clinical worsening at 2 years compared to hip OA patients (40% versus 20%, p<0.01). However, the proportion of patients who underwent a TJR was lower for knee than hip OA (19% versus 51% respectively, p<0.0001). We found no difference in the proportion of patients not having worsened at 2 years between knee and hip OA patients (41% and 29%, respectively, p = 0.13). Out of the 79 patients who showed clinically worsening at 3 months 27 (34%) had TJR, and 34 (43%) maintained to be clinically worsened at 2 years. Median BMI did not change from baseline to 3 months and 2 years follow-up.

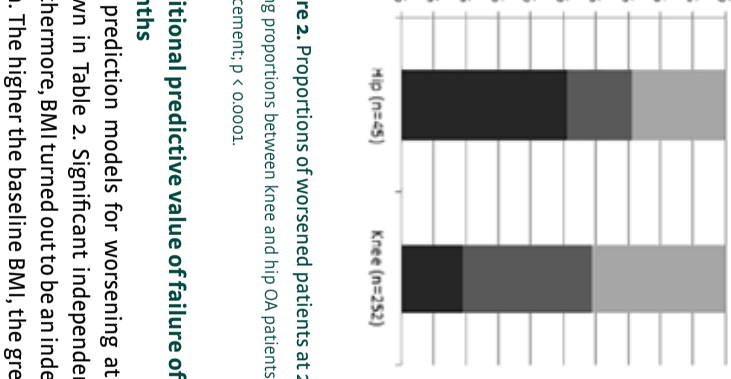


Figure 2. Proportions of worsened patients at 2 years per index joint.
Testing proportions between knee and hip OA patients: no worsening; p = 0.13, clinically worsened; p < 0.01, and total joint replacement; p < 0.0001.

Additional predictive value of failure of optimized nonsurgical treatment during 3 months

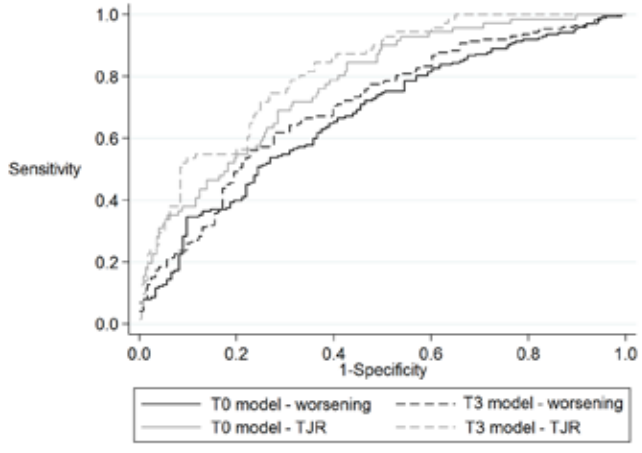
The prediction models for worsening at 2 years (defined as clinical worsening or TJR) are shown in Table 2. Significant independent baseline predictors are: PGA and K&L score ≥ 2. Furthermore, BMI turned out to be an independent predictor when using WOMAC, but not NRS pain. The higher the baseline BMI, the greater the risk of worsening at 2 years. Adding clinical worsening at 3 months (yes/no) as independent variable to the baseline model, resulted in an additional predictor for worsening at 2 years with an adjusted OR of 2.8 (95% CI 1.5–5.2 in NRS pain model). Overall, the discriminative ability of the model with clinical worsening at 3 months added to baseline variables, was fair with an AUC of 0.68 (95% CI 0.62–0.74), indicating moderate ability to discriminate between patients with and without (clinical) worsening at 2 years (figure 3). Table 3 shows a high positive predictive value, low negative predictive value, low sensitivity, and high specificity for clinical worsening at 3 months and outcome at 2 years. The positive likelihood ratio of 2.2 (95% CI 1.4–3.5) suggests that taking clinical worsening at 3 months despite optimized nonsurgical treatment into account increases the pre-test probability of 61% for worsening at 2 years to a post-test probability of 78% (Table 3).

Table 2. Multivariate logistic regression model with determinants presented as OR (95% CI) for worsening at 2 years in knee and hip OA patients.

	Worsening using NRS pain (N_{yes}=181; N_{no}=116)			Worsening using WOMACpain (N_{yes}=173; N_{no}=123)*		
	OR	95% CI	AUC	OR	95% CI	AUC
Baseline model			0.63 (0.57-0.70)		0.67 (0.61-0.73)	
Patient global assessment (range 0-10)	0.81	0.71-0.93		0.85	0.75-0.96	
Kellgren & Lawrence score ≥ 2 (range 0-4)	1.62	0.94-2.81		2.13	1.20-3.78	
WOMAC stiffness (0-100)	0.99	0.98-1.00				
BMI (kg/m²)				1.06	1.01-1.10	
Hip index joint				1.98	0.97-4.06	
Baseline model with clinical worsening at 3 months			0.68 (0.62-0.74)			0.70 (0.64-0.76)
Clinical worsening at 3 months	2.84	1.54-5.22		2.47	1.34-4.55	
Patient global assessment baseline (range 0-10)	0.83	0.72-0.96)		0.87	0.77-0.99	
Kellgren & Lawrence scale ≥ 2 (range 0-4)	1.69	0.96-2.99		2.26	1.26-4.06	
WOMAC stiffness baseline (0-100)	0.99	0.98-1.00				
BMI baseline (kg/m²)					1.05	1.01-1.10
Hip index joint					1.96	0.95-4.07

OR: odds ratio, CI: confidence interval; AUC: area under the ROC curve/c-statistic; OA: osteoarthritis; n: number;
NRS: numeric rating scale pain; WOMAC: Western Ontario McMaster Universities Index of osteoarthritis; PGA: patient global assessment, BMI: body mass index.
* Total number of patients with complete WOMAC pain = 296, due to missing values.
Higher scores indicate worse PGA, more NRS pain, and better scores for WOMAC stiffness.

Figure 3. ROC curves showing the sensitivity and 1-specificity for both the baseline model (T0) and the baseline model with clinical worsening at 3 months added (T3), for 2 different dependent outcome measures at 2 years: (1) worsening (TJR or fulfilling clinical worsening criteria, Table 2), and (2) TJR (Table 4) in patients with knee or hip OA.



Pain was calculated using WOMAC pain.
WOMAC: Western Ontario and McMaster Universities Osteoarthritis Index scale; TJR: total joint replacement.

Table 3. 2 x 2 table of clinical worsening at 3 months and worsening at 2 years.

		Worsening at 2 years		
		Yes	No	Total
Clinical worsening at 3 months	Yes	66	19	85
	No	115	97	212
	Total	181	116	297

Data shown as n (%)
Positive predictive value 77.6% (95% CI 67.1–85.7); negative predictive value 47.8% (95% CI 39.0–52.7); sensitivity: 36.5% (95% CI 29.5–44.0); specificity: 83.6% (95% CI 75.389.6) Positive likelihoodratio 2.23 (95% CI 1.41–3.51); Negative likelihoodratio 0.76 (95% CI 0.68–0.85).

Sensitivity analyses using TJR as dependent outcome, yielded similar independent clear predictors for TJR: clinical worsening at 3 months, K&L score ≥ 2, and affected joint (Table 4). The adjusted OR for clinical worsening at 3 months for having a TJR at 2 years was 4.1 (95% CI 2.0–8.4 in NRS pain model). Overall, the discriminative ability of this model, showed a good AUC of 0.82 (95% CI 0.77–0.87), indicating good ability to discriminate between patients with and without TJR. Similar conclusions could be drawn for the sensitivity analyses using only knee OA patients (OR clinically worsened at 3 months 5.2 95% CI 2.2–11.9 and OR K&L score increased to 8.1 95% CI 2.2–29.3).

Table 4. Multivariate logistic regression model with determinants presented as OR (95% CI) for total joint replacement in knee and hip OA patients.

	Total joint replacement (using NRS pain) (N_{yes}=181; N_{no}=116)			Total joint replacement (using WOMAC pain) (N_{yes}=173; N_{no}=123)*		
	OR	95% CI	AUC	OR	95% CI	AUC
Baseline model			0.77 0.71-0.83			0.77 0.72-0.83
Patient global assessment (range 0-10)	1.20	1.01-1.42		1.15	0.98-1.36	
Kellgren & Lawrence score ≥ 2 (range 0-4)	4.32	1.66-11.26		4.38	1.20-3.78	
WOMAC stiffness (0-100)	0.98	0.97-1.00		0.99	1.72-11.14	
Hip index joint	6.61	3.03-14.42		5.60	2.62-11.95	
Comorbidities >1				1.90	0.88-4.09	
Duration of symptoms (years)	1.03	1.00-1.06		1.03	0.99-1.06	
BMI (kg/m²)	1.04	0.99-1.09				
Age (years)	1.02	0.99-1.06				
Baseline model with clinical worsening at 3 months			0.82 0.76-0.87			0.82 0.77-0.87
Clinical worsening at 3 months	4.11	2.00-8.48		4.06	1.98-8.34	
Patient global assessment baseline (range 0-10)	1.30	1.08-1.55		1.28	1.07-1.54	
Kellgren & Lawrence score ≥ 2 (range 0-4)	4.72	1.79-12.43		3.24	1.84-11.95	
WOMAC stiffness baseline (0-100)	0.98	0.97-1.00		0.99	0.97-1.00	
Hip index joint	7.38	3.22-16.94		6.21	2.78-13.89	
Comorbidities > 1				1.78	0.81-3.92	
Duration of symptoms (years)	1.03	1.00-1.07		1.03	0.99-1.07	
BMI (kg/m²)	1.03	0.98-1.08				
Age (years)	1.03	1.00-1.06				

OA: osteoarthritis; n: number; OR: odds ratio; CI: confidence interval; AUC: area under the ROC curve/c-statistic; NRS: numeric rating scale pain; WOMAC: Western Ontario and McMaster University Index of osteoarthritis; PGA: patient global assessment; BMI: body mass index.
* Total number of patients due to missing values WOMAC pain = 296
Higher scores indicate worse PGA, more NRS pain, and better scores for WOMAC stiffness

Discussion

To our knowledge, this is the first longitudinal study on the additional predictive value of failure of optimized nonsurgical treatment during 3 months for worsening at 2 years in knee and hip OA. Our results show that more than half of patients with established knee and hip OA in secondary care showed worsening at 2 years, despite optimized nonsurgical treatment. We also found that clinical worsening at 3 months is a clear independent predictor for worsening at 2 years.

How could our results be used in clinical practice? Our results could be used for patient information and to guide both patients and (orthopaedic) surgeons in decision making about the appropriate timing of TJR. This study adds that patients who are clinically worsened at 3 months despite optimized multimodal nonsurgical treatment, have an almost threefold increased odds ratio for having worsened at 2 years. This corresponds to an increase of pre- and post-test probability from 61 to 78%. This increase is above the considered clinically relevant cut-off of 15% improvement in probability of response after a positive test.[32,33] In addition, our worsening criteria are easy to assess in clinical practice. These advantages favour the use of clinical worsening criteria to monitor the symptoms of patients with established OA and suggest that patients with persisting symptoms after optimized nonsurgical treatment, should be referred back to the (orthopaedic) surgeon to reconsider the TJR indication. This predictor could be used to identify a more severely affected subgroup of patients that would be eligible for TJR.[34] Lastly, using clinical worsening criteria could support the identification of subgroups of patients potentially eligible for novel and advanced treatment options.[35-38]

A remarkable finding is the relatively high proportion of worsened patients on the short term, compared with previous OA cohorts. This is not surprising, since most of the well-known OA cohorts focus on early OA patients.[7,8,10,14,39] This difference is most likely explained by the selection of patients who were not yet deemed eligible for TJR by their surgeon. This homogenous population might have led to a selection of patients with a relatively high clinical burden and may hamper the generalizability of our results. Therefore, our study population is not representative for the general OA population, but generalizable to this more established OA population. Therefore, future research is warranted and should aim to investigate other OA populations and settings, for example OA populations from primary care who are referred to secondary care.

An interesting finding is the lower proportion of TJR in knee OA patients compared to hip OA patients, whilst the total proportion of worsened patients between knee and hip OA patients was similar. This could be explained by the better longterm outcomes of a TJR of the hip than the knee (for example limited lifespan, and higher risk for serious adverse events for TJR of the knee) [38,40-46], whereby clinicians might be more reluctant to decide for TJR in knee than hip OA. Furthermore, as expected, K&L score turned out to be a strong predictor, especially for predicting TJR at 2 years, which might be explained by the influence the K&L score has on the decision of an orthopaedic surgeon to propose a TJR. Moreover, a remarkable finding is the direction of the association between baseline PGA and the probability of worsening at 2 years. Where one might expect a worse PGA to correspond to a higher chance of worsening at 2 years, the opposite was found. This finding might be explained by regression to the mean effect.

Several strengths of this study should be considered. Overall, our study was well-powered and we chose a validated dichotomous measure for worsening combining arthroplasty with clinical worsening at 2 years incorporating the domains pain, function and PGA, the outcome measures advised according to the current opinion to use symptom progression as outcome measure.[12,13] Considering our homogenous population, the results of our study seem to be generalizable to patients with established knee and hip OA for whom decision making about TJR is forthcoming.

Some limitations that we faced should be reflected on. First, estimated risks from prediction models have the tendency to be overestimated and thus further validation is required. In addition, most of the patients in our study had knee OA, so generalizability for hip OA is not assured. Furthermore, the substantial proportion of patients not replying to our invitation for the 2-years assessment could have influenced the results, although this seems unlikely, considering the lack of relevant differences in most relevant baseline variables and the similarity in proportion of replying and non-replying patients showing clinical worsening at 3 months. In addition, while the non-replying rate was quite high, it was comparable with other OA studies in which patients are not remaining under medical treatment.[39] Lastly, adherence to treatment in clinical practice is quite challenging[47,48] and subsequently, non-adherence to treatment could have influenced our results. However, we can only speculate about the potential direction. Nevertheless, given the challenge of adherence in both our study and clinical practice, our results are more likely to be representative for daily clinical practice, which strengthens the external generalizability of our results.

In conclusion, in light of our findings, we suggest that re-assessment of OA symptoms after optimized nonsurgical treatment could be meaningful for both patients and surgeons in clinical decision making for TJR. Furthermore, this information could be used to identify subgroups of patients potentially eligible for novel and advanced treatment options.

Acknowledgement
We thank all patients who participated.

References

1. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am 2007;89:780-5.
2. Deshpande BR, Katz JN, Solomon DH, Yelin EH, Hunter DJ, Messier SP, et al. Number of Persons With Symptomatic Knee Osteoarthritis in the US: Impact of Race and Ethnicity, Age, Sex, and Obesity. Arthritis Care Res (Hoboken) 2016;68:1743-50.
3. Bijlsma JW, Berenbaum F, Lafeber FP. Osteoarthritis: an update with relevance for clinical practice. Lancet 2011;377:2115-26.
4. Singh JA, Vessely MB, Harmsen WS, Schleck CD, Melton LJ, III, Kurland RL, et al. A population-based study of trends in the use of total hip and total knee arthroplasty, 1969-2008. Mayo Clin Proc 2010;85:898-904.
5. Murphy L, Schwartz TA, Helmick CG, Renner JB, Tudor G, Koch G, et al. Lifetime risk of symptomatic knee osteoarthritis. Arthritis Rheum 2008;59:1207-13.
6. Ravi B, Croxford R, Reichmann WM, Losina E, Katz JN, Hawker GA. The changing demographics of total joint arthroplasty recipients in the United States and Ontario from 2001 to 2007. Best Pract Res Clin Rheumatol 2012;26:637-47.
7. van Dijk GM, Dekker J, Veenhof C, van den Ende CH. Course of functional status and pain in osteoarthritis of the hip or knee: a systematic review of the literature. Arthritis Rheum 2006;55:779-85.
8. Wesseling J, Bierma-Zeinstra SM, Kloppenburg M, Meijer R, Bijlsma JW. Worsening of pain and function over 5 years in individuals with 'early' OA is related to structural damage: data from the Osteoarthritis Initiative and CHECK (Cohort Hip & Cohort Knee) study. Ann Rheum Dis 2015;74:347-53.
9. Holla JF, van der Leeden M, Heymans MW, Roorda LD, Bierma-Zeinstra SM, Boers M, et al. Three trajectories of activity limitations in early symptomatic knee osteoarthritis: a 5-year follow-up study. Ann Rheum Dis 2014;73:1369-75.
10. White DK, Keysor JJ, Lavalley MP, Lewis CE, Torner JC, Nevitt MC, et al. Clinically important improvement in function is common in people with or at high risk of knee OA: the MOST study. J Rheumatol 2010;37:1244-51.
11. Conaghan PG, Kloppenburg M, Schett G, Bijlsma JWJ. EULAR osteoarthritis ad hoc committee. Ann Rheum Dis 2014;73:1442-5.
12. Bastick AN, Runhaar J, Belo JN, Bierma-Zeinstra SM. Prognostic factors for progression of clinical osteoarthritis of the knee: a systematic review of observational studies. Arthritis Res Ther 2015;17:152.
13. Belo JN, Berger MY, Reijman M, Koes BW, Bierma-Zeinstra SM. Prognostic factors of progression of osteoarthritis of the knee: a systematic review of observational studies. Arthritis Rheum 2007;57:13-26.
14. de Rooij M, van der Leeden M, Heymans MW, Holla JF, Hakkinen A, Lems WF, et al. Prognosis of Pain and Physical Functioning in Patients With Knee Osteoarthritis: A Systematic Review and Meta-Analysis. Arthritis Care Res (Hoboken) 2016;68:481-92.
15. Kim C, Nevitt MC, Niu J, Clancy MM, Lane NE, Link TM, et al. Association of hip pain with radiographic evidence of hip osteoarthritis: diagnostic test study. BMJ 2015;351:5983.
16. Mahler E, den Broeder AA, Woodworth TG, Busch V, van den Hoogen FH, Bijlsma J, et al. How should worsening in osteoarthritis be defined? Development and initial validation of preliminary criteria for clinical worsening in knee and hip osteoarthritis. Scand J Rheumatol 2017;1-11.
17. McAlindon TE, Bannuru RR, Sullivan MC, Arden NK, Berenbaum F, Bierma-Zeinstra SM, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014;22:363-88.
18. Snijders GF, den Broeder AA, van Riel PL, Straten VH, de Man FH, van den Hoogen FH, et al. Evidence-based tailored conservative treatment of knee and hip osteoarthritis: between knowing and doing. Scand J Rheumatol 2011;40:225-31.

19. Riemsa RP, Taal E, Rasker JJ, Houtman PM, Van Paassen HC, Wiegman O. Evaluation of a Dutch version of the AIMS2 for patients with rheumatoid arthritis. Br J Rheumatol. 1996;35:755-60.
20. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007;15 Suppl A:A1-56.
21. KELLGREN JH, LAWRENCE JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494-502.
22. de Groot IB, Favejee MM, Reijman M, Verhaar JA, Terwee CB. The Dutch version of the Knee Injury and Osteoarthritis Outcome Score: a validation study. Health Qual Life Outcomes 2008;6:16.
23. McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Rheum 2001;45:453-61.
24. Vercoulen JH, Swanink CM, Fennis JF, Galama JM, van der Meer JW, Bleijenberg G. Dimensional assessment of chronic fatigue syndrome. J Psychosom Res 1994;38:383-92.
25. Kori SH, Miller RP, Todd DD. Kinesiophobia: A new view of chronic pain behavior. Pain Management. 1990;3:35–43.
26. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol 1998;51:1055-68.
27. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992;30:473-83.
28. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med 2015;162:1-73.
29. Steyerberg EW. Clinical prediction models, New York: Springer-Verlag, 2009.
30. Sterne JA, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ 2009;338:b2393.
31. Von Hippel PT. Regression with missing ys: an improved strategy for analyzing multiply imputed data. Sociological Methodology 2007;37:83-117.
32. Cuppen BV, Welsing PM, Sprengers JJ, Bijlsma JW, Marijnissen AC, van Laar JM, et al. Personalized biological treatment for rheumatoid arthritis: a systematic review with a focus on clinical applicability. Rheumatology (Oxford) 2016;55:826-39.
33. de Vet HCW, Terwee CB, Mokkink LB and Knol DL. Measurement in medicine. Cambridge University Press 2011.
34. Ferket BS, Feldman Z, Zhou J, Oei EH, Bierma-Zeinstra SM, Mazumdar M. Impact of total knee replacement practice: cost effectiveness analysis of data from the Osteoarthritis Initiative. BMJ 2017;356:1131.
35. French SD, Bennell KL, Nicolson PJ, Hodges PW, Dobson FL, Hinman RS. What do people with knee or hip osteoarthritis need to know? An international consensus list of essential statements for osteoarthritis. Arthritis Care Res (Hoboken). 2015;67:809-16.
36. Hagen KB, Smedslund G, Osteras N, Jamtvedt G. Quality of Community-Based Osteoarthritis Care: A Systematic Review and Meta-Analysis. Arthritis Care Res (Hoboken) 2016;68:1443-52.
37. Felson DT. Clinical practice. Osteoarthritis of the knee. N Engl J Med 2006;354:841-8.
38. Skou ST, Roos EM, Laursen MB, Rathleff MS, Arendt-Nielsen L, Simonsen O, et al. A Randomized, Controlled Trial of Total Knee Replacement. N Engl J Med 2015;373:1597-606.
39. Eyles JP, Mills K, Lucas BR, Williams MJ, Makovey J, Teoh L, et al. Can We Predict Those With Osteoarthritis Who Will Worsen Following a Chronic Disease Management Program? Arthritis Care Res (Hoboken) 2016;68:1268-77.
40. Hawker GA, Badley EM, Borkhoff CM, Croxford R, Davis AM, Dunn S, et al. Which patients are most likely to benefit from total joint arthroplasty? Arthritis Rheum 2013;65:1243-52.

41. Cushner F, Agnelli G, FitzGerald G, Warwick D. Complications and functional outcomes after total hip arthroplasty and total knee arthroplasty: results from the Global Orthopaedic Registry (GLORY). Am J Orthop (Belle Mead NJ) 2010;39:22-8.
42. Bourne RB, Chesworth BM, Davis AM, Mahomed NN, Charron KD. Patient satisfaction after total knee arthroplasty: who is satisfied and who is not? Clin Orthop Relat Res 2010;468:57-63.
43. Rissanen P, Aro S, Slatis P, Sintonen H, Paavolainen P. Health and quality of life before and after hip or knee arthroplasty. J Arthroplasty 1995;10:169-75.
44. Bullens PH, van Loon CJ, de Waal Malefijt MC, Laan RF, Veth RP. Patient satisfaction after total knee arthroplasty: a comparison between subjective and objective outcome assessments. J Arthroplasty 2001;16:740-7.
45. Katz JN. Parachutes and Preferences--A Trial of Knee Replacement. N Engl J Med 2015;373:1668-9.
46. Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. BMJ Open 2012;2:e000435.
47. Pisters MF, Veenhof C, de Bakker DH, et al. Behavioural graded activity results in better exercise adherence and more physical activity than usual care in people with osteoarthritis: a clusterrandomised trial. J Physiother 2010;56:41–7.
48. Beinart NA, Goodchild CE, Weinman JA, et al. Individual and intervention-related factors associated with adherence to home exercise in chronic low back pain: a systematic review. Spine J 2013;13:1940–50.

Chapter 6

Association between serum levels of the proinflammatory protein S100A8/A9 and clinical and structural characteristics of patients with established knee, hip, and hand osteoarthritis

Elien A.M. Mahler
Manon C. Zweers
Peter L. van Lent
Arjen B. Blom
Frank H.J. van den Hoogen
Wim B. van den Berg
Johannes Roth
Thomas Vogl
Johannes W.J. Bijlsma
Cornelia H.M. van den Ende
Alfons A. den Broeder

Published in Scandinavian Journal of Rheumatology 2015;44(1):56-60

Abstract

Objective

To explore the association between S100A8/A9 serum levels with clinical and structural characteristics of patients with established knee, hip, or hand osteoarthritis (OA).

Methods

A cross-sectional exploratory study was conducted with 162 OA patients. Measures for pain, stiffness, and function included the Western Ontario and McMaster Universities Osteoarthritis (WOMAC) questionnaires or the Australian Canadian Osteoarthritis Hand (AUSCAN) Index and for structural abnormalities, osteophytes and joint space narrowing grades. The association between S100A8/A9 and clinical or structural characteristics was analyzed using linear regression or logistic regression when appropriate.

Results

The mean age of the OA patients was 56 years, 71% were female, and 61% had a Kellgren and Lawrence (K&L) score ≥ 2. The serum S100A8/A9 level did not differ between knee, hip, and hand OA patients and no association was found between serum S100A8/A9 and clinical characteristics. The serum S100A8/A9 level was negatively associated with the sum score of osteophytes after adjusting for sex and body mass index (BMI) (adjusted β −0.015, (95% confidence interval (CI) −0.030 to 0.001, p = 0.062) and positively associated with erythrocyte sedimentation rate (ESR) > 12 mm/hour (adjusted OR 1.002, 95% CI 1.000–1.004, p = 0.049) for each increase of S100A8/A9 of 1 ng/mL. For hand OA patients, a negative association of S100A8/A9 with sum score of joint space narrowing was found (adjusted β −0.007, 95% CI −0.016 to 0.001, p = 0.099).

Conclusion

The results from this cross-sectional exploratory study do not support an important role for serum S100A8/A9 levels as a biomarker for clinical and structural characteristics in established knee, hip, and hand OA patients. The inverse association with structural abnormalities and the positive association with ESR may reflect inflammatory synovial processes in patients with OA before structural abnormalities occur.

Introduction

Osteoarthritis (OA) is a chronic disease of the joint in which three tissues play a pivotal role: articular cartilage, subchondral bone and synovial tissue. Although the pathophysiology of OA has long been thought to be primarily cartilage driven, recent evidence shows an additional and integrated role of bone and synovial tissue.[1] Synovial inflammation corresponds to clinical symptoms such as joint swelling and inflammatory pain, and it is thought to be secondary to cartilage debris and catabolic mediators entering the synovial cavity.[1] As a result, biochemical markers (biomarkers) of synovial inflammation could be indicators for both clinical signs and cartilage destruction. Biomarkers are considered, among others, to aid in diagnosing OA in an earlier stage when structural damage is limited and could still be modulated and used for in identifying targets for disease-modifying therapies and in defining phenotypes. Despite the increasing evidence for a role of synovitis in the pathophysiology of OA, biomarkers of synovial inflammation have not yet been extensively studied in human OA.

The biomarker S100A8/A9 is a proinflammatory protein of the S100 family (a major leucocyte protein also called calprotectin, myeloid-related proteins 8 and 14 heterocomplex) that is highly expressed by synovial lining macrophages of inflamed tissue in (juvenile) RA.[2,3] Thus, S100A8/A9 is released locally by infiltrating phagocytes at the site of inflammation and may diffuse from inflamed joints into the circulation, where it can be measured in plasma. Furthermore, S100A8/A9 is strongly expressed by synovial tissue in experimental OA and a positive correlation with development of human early OA has been found.[4]

The S100A8/A9 heterodimer has been shown to be a reliable indicator of disease activity and joint inflammation in other inflammatory rheumatic diseases, including RA, juvenile RA, psoriatic arthritis, and spondyloarthropathy, and recently with radiographic progression in RA.[2,3,5] An elevation of this biomarker has been demonstrated in erosive hand OA and decreasing levels were observed after joint replacement of knee and hip, indicating the possible role of this biomarker in OA joints.[6,7] Thus, although S100A8/A9 has been shown to correlate well with laboratory, clinical, and radiological assessments in several inflammatory rheumatic diseases, with regard to OA that shows low-grade synovitis, this has yet to be established.

Therefore, we conducted a cross-sectional exploratory study to determine whether S100A8/A9 serum levels are associated with clinical and structural characteristics in patients with established knee, hip, or hand OA.

Methods

Patient characteristics

This study included 162 patients attending our outpatient clinic and fulfilling the American College of Rheumatology (ACR) clinical criteria for knee, hip, or hand OA. Fifty-seven patients with knee OA and 47 patients with hip OA were randomly selected from our observational cohort study for patients with knee or hip OA and 58 patients were randomly selected from a controlled trial on the effect of a multidisciplinary self-management program in people with hand OA.[8,9] The patients gave written consent according to the Declaration of Helsinki and the study was approved by the local Medical Research Ethics Committee, region

Arnhem-Nijmegen (The Netherlands, study numbers 2009/095 & 2007/260). Demographic and OA-related characteristics were obtained using a standardized interview and physical examination as described previously.[8] On the same day, clinical data and radiographs were assessed. Exclusion criteria were: inflammatory rheumatic diseases, cognitive or sensorimotor problems interfering with the use of questionnaires, previous joint replacement surgery in one of the hand joints or in the index joint regarding knee and hip, planned orthopaedic procedures within the next 12 weeks, not able to write and/or understand the Dutch language.

Clinical characteristics

Clinical variables consisted of pain and global disease activity measured on numerical rating scales (range 1–10), erythrocyte sedimentation rate (ESR), and C-reactive protein (CRP). Pain, stiffness, and function were assessed using the Knee Injury and Osteoarthritis Outcome Score (KOOS) or the Hip disability and Osteoarthritis Outcome Score (HOOS) (Likert-scale version) questionnaires for knee and hip OA, respectively, which include the Western Ontario and McMaster Universities Osteoarthritis (WOMAC) index in its complete and original format (with permission, www.koos.nu) and the Australian Canadian Osteoarthritis Hand (AUSCAN) index for hand OA. WOMAC pain, stiffness, and function subscales are presented as normalized scores (0–100, where 100 equals extreme symptoms) and AUSCAN normalized subscores were transformed to scores between 0 and 100 by multiplying the score by 100/maximum value of the subscore (20, 36, and 4 for pain, function, and stiffness, respectively).[10] Laboratory measurement of ESR and CRP was performed on the day of examination (both analyzed by standard in-house methodology, with upper normal levels of 20 mm/h for women and 15 mm/h for men for ESR and 10 mg/L for CRP). Serum samples were frozen at −80°C for assessment of S100A8/A9 after completion of the clinical study. The complex S100A8/A9 was measured using an enzyme-linked immunosorbent assay (ELISA), as described previously[11], with normal levels of 440 ng/mL (range 393–487), and coefficients of variation of < 11% within and 6% between assays.

Structural abnormalities

For structural abnormalities, knee (weight-bearing posterior-anterior fixed flexion), hip, and hand radiographs (both anterior–posterior) were obtained. Scoring of the index joint (the left or right side, whichever gave the most complaints) was done blinded for clinical data, using the Kellgren and Lawrence (K&L) grading system, atlas based, in five (0–4) grades by an experienced rheumatologist. For hand OA, the mean score was used. In addition, radiographs were scored according to the new OARSI atlas of radiographic features for hip, knee, and hand OA[12] by a research physician. Sum scores of osteophyte grades and joint space narrowing at different joint sites were obtained. Sum score of osteophytes ranged from 0 to 12 for knee OA (osteophytes (grade 0–3) at medial and lateral tibial plateaus and at medial and lateral femoral condyles), from 0 to 9 for hip OA (osteophytes (grade 0–3) at superior and inferior femoral head and superior acetabulum), and from 0 to 29 for hand OA (osteophytes (grade 0–3) at proximal and distal interphalangeal joints (PIP and DIP), first carpometacarpal (CMC) joint, and the presence or absence of osteophytes at the interphalangeal (IP) joint of thumb and naviculotrapezial joint (NTJ)). Finally, the sum score of osteophytes was normalized by dividing the sum score by the maximum score of the joint (12 for knee, 9 for hip, and 29 for hand OA). The sum score of joint space narrowing ranged from 0 to 6 for knee OA (grade 0–3 at medial and lateral tibiofemoral compartment), from 0 to 6 for hip OA (grade 0–3 at superior and medial femoroacetabular joint), and from 0 to 29 for hand OA (grade 0–3 at PIP and DIP and first CMC and the presence or absence of joint space narrowing at the IP and NTJ). The sum score of joint space narrowing was normalized by dividing the sum score by the maximum score of the joint (6 for knee and hip and 29 for hand OA). Intra-observer reliability (kappa) ranged from 0.68 to 0.89 for osteophyte grading and from 0.78 to 0.94 for measurement of joint space narrowing at the different joint sites (rescored in 20 participants).

Statistical analysis

Statistical analysis was performed using the statistical software package Stata10 (StataCorp, College Station, TX, USA). Differences between participant characteristics within knee, hip, or hand OA were determined by analysis of variance (ANOVA) or Kruskal–Wallis one-way ANOVA when appropriate. In case of non-normal distribution, outcome measures were dichotomized using the median split method. The association between serum levels of S100A8/A9 (independent) and clinical data (dependent) and structural abnormalities (dependent) was analyzed using linear regression or logistic regression when appropriate and presented as the β-coefficient or odds ratio (OR). Potential confounders (such as body mass index (BMI), gender, age, duration of symptoms) were checked by calculating the correlation with the dependent and independent variable and hereafter remaining potential confounders were added to the association model (forward selection). Analysis of the total group and of each subgroup was performed. For multivariate analysis, we set $\alpha < 0.10$.

Results

Patient characteristics

The mean age of the patients was 56 years, 71% were female, and 61% had a K&L score ≥ 2. All baseline characteristics are shown in Table 1. Differences in clinical and disease characteristics were observed between the patients in the hand group vs. patients in the knee and hip group: hand OA patients were older, were more likely to be female, reported a longer duration of symptoms, less pain and lower global disease activity, but scored highest on the normalized sum score of osteophytes. No difference between patients in the knee, hip, and hand OA group was observed regarding BMI, pain (WOMAC or AUSCAN), stiffness, function, ESR, and CRP. The serum levels of S100A8/A9 were comparable among the knee, hip, and hand patients.

Association between serum S100A8/A9 and clinical data and structural abnormalities

No association between the serum level S100A8/A9 and normalized pain, stiffness, or function was found. Furthermore, in the total group, serum level of S100A8/A9 was positively associated with ESR (Table 2), in that with every increase in serum level of S100A8/A9 by 1 ng/mL, the odds for ESR > 12 mm/hour increased by 1.002. Sex and BMI were identified as possible confounders. Analysis adjusted for sex and BMI showed that S100A8/A9 was negatively associated with the normalized sum score of osteophytes, with a β-coefficient of −0.015 for each increase in S100A8/A9 of 1 ng/mL (Table 2). Only in the hand OA group was a negative association of serum level of S100A8/A9 and sum score of joint space narrowing found while adjusting for sex and BMI (β-coefficient −0.007, 95% confidence interval (CI) −0.016 to 0.001, p = 0.099). This association was not found in the knee and hip OA group.

6

Table 1. Baseline characteristics of 162 patients with knee, hip, or hand OA.

Age (years), mean (SD)	56.0 (8.8)
Women, n (%)	115 (71)
Body mass index (kg/m2)	27.5 (24.7-31.2)
Duration of symptoms more than 5 years , n (%)	36 (22)
Clinical characteristics	
Pain on numerical rating scale (0-10)	5 (3-7)
Global assessment of disease activity on numerical rating scale (0-10)	7 (5-8)
Pain, WOMAC or AUSCAN normalized score (0-100 (extreme symptoms))	55 (30-70)
Stiffness, WOMAC or AUSCAN normalized (0-100 (extreme symptoms))	63 (38-75)
Function, WOMAC or AUSCAN normalized (0-100 (extreme symptoms))	56 (32-75)
Erythrocyte sedimentation rate (mm/h)	12 (7-18)
C-reactive protein < 5 mg/L, n (%)	132 (89)
Serum level S100A8/A9 (ng/mL)	335 (240-460)
Knee OA patients (n = 57)	330 (250-470)
Hip OA patients (n = 47)	330 (230-470)
Hand OA patients (n = 58)	350 (230-450)
Kellgren and Lawrence < 2 (n = 63)	360 (240-480)
Kellgren and Lawrence ≥ 2 (n = 97)	330 (240-420)
Structural abnormalities	
Kellgren and Lawrence ≥ 2, n (%)	97 (61)
Sum score of osteophytes, normalized score (range 0-100)	22.2 (8.3-33.3)
Sum score of joint space width, normalized score (range 0-100)	33.3 (16.7-50.9)

Data are shown as median (interquartile range) unless stated otherwise.
BMI, Body mass index; K&L, Kellgren and Lawrence; ESR, erythrocyte sedimentation rate; CRP, C-reactive protein.
Sum scores were normalized by dividing the sum score by the maximum score of the joint (osteophytes knee 12, hip 9, hand 29 and joint space width knee 6, hip 6, and hand 29).

Table 2. Association of serum levels of S100A8/A9 in ng/mL with clinical characteristics or structural abnormalities expressed as β-coefficient or odds ratio (OR) for each increase in S100A8/A9 with 1 ng/mL with 95% CIs and p-values adjusted for sex and BMI in patients with knee, hip, or hand OA.

Dependent variable	β-coefficient or *OR*	95% CI	p-value
Clinical characteristics			
Pain	0.007	−0.118 to 0.026	0.52
Stiffness	0.005	−0.014 to 0.023	0.62
Function	0.001	−0.018 to 0.020	0.94
ESR (> 12mm/h)	*1.002*	*1.000–1.004*	0.049
CRP	0.003	−0.001 to 0.007	0.18
Structural abnormalities			
Sum score of osteophytes*	−0.015	−0.030 to 0.001	0.06
Sum score of joint space width*	−0.000	−0.022 to 0.022	1.00

BMI: body mass index; C:, confidence interval; ES:, erythrocyte sedimentation rate; CRP:, C-reactive protein.
* X-rays of 14 patients were missing for logistic reasons or because patients declined to have an X-ray.

Discussion

To our knowledge, this cross-sectional study is the first to show that S100A8/A9 serum levels in patients with established knee, hip, and hand OA are not associated with clinical characteristics regarding pain, stiffness, and function, but are positively associated with ESR and negatively associated with the sum score of osteophytes. Furthermore, in the hand OA group, the S100A8/A9 serum level showed an inverse relationship with the sum score of joint space narrowing. In our OA population, serum levels of S100A8/A9 were comparable among the knee, hip, and hand patients. Our results do not support an important role for serum S100A8/A9 levels as a biomarker for clinical and structural characteristics.

We did not observe an association between S100A8/A9 serum levels and pain, stiffness, and function in knee, hip, or hand OA despite increased expression of S100A8/A9 by synovial tissue in experimental OA4 and in synovitis in RA.[2] However, the positive association found between serum level of S100A8/A9 and ESR underlines the finding that S100A8/A9 may reflect inflammatory synovial processes in OA.[4] Furthermore, the inverse association between serum S100A8/A9 levels with structural abnormalities in this established OA population may reflect an inverse temporal relationship, with active inflammatory synovial processes in patients with OA before structural abnormalities occur. This finding is in line with previous studies showing that high levels S100A8/A9 may predict joint destruction in patients with early symptomatic OA4 and that S100A8/A9 might only have a role in early cartilage damage, as S100A8/A9 does not have a sustained role in cartilage degradation in experimental OA.[13]

Our results seem to be in contrast with the previously described positive association between S100A8/A9 serum level and CRP and between S100A8/A9 serum level and joint damage in patients with RA and psoriatic arthritis.[5,14] As the inflammation in RA is more severe than in OA, as indicated by 89% of our OA population having a CRP within the normal range, the lack of

association between S100A8/A9 serum level and CRP might well be a floor effect. As the ESR is less specific for inflammation compared to CRP and may be influenced by other factors such as obesity, further research is necessary to examine the association between S100A8/A9 serum levels and acute phase reagents in OA. Moreover, in animal experimental studies it has been shown that S100A8 and S100A9 both have strong cartilage-degrading properties.[15] There could be several explanations for this apparent discrepancy. First, we measured S100A8/A9 only in serum and not in synovial fluid, which might have given a more accurate representation of the local expression of S100A8/A9. However, measuring the level of S100A8/A9 in synovial fluid as a biomarker would not be easy to perform in daily clinical practice. S100A8/A9 serum levels probably reflect the synovial inflammation in OA, which is rarely as severe as in RA.[1] Although S100A8/A9 serum levels in OA are much lower than, for example, in RA, it is conceivable that serum levels of S100A8/A9 in OA do not necessarily increase if local concentrations in the synovial fluid are mildly elevated.[2] However, a strong correlation was found previously between the levels in synovial fluid and plasma in juvenile idiopathic arthritis.[3] Finally, it could be that cartilage and bone destruction in OA occur through a process independent of S100A8/A9.

Our study has several limitations. First, differences in patient and disease characteristics could have masked associations between S100A8/A9 serum levels and clinical or structural characteristics. Furthermore, interpretation of subgroup analyses was hampered by limited power. Lack of precision could therefore also be an explanation for absence of an association between serum levels of S100A8/A9 and clinical characteristics. However, power calculations show that, given our sample size, a correlation of 0.30 (and 0.40 for each subgroup) could be detected with a power of 80%. Therefore, we consider that clinically important associations should have been detected. Second, because this is a population with established OA, the associations in patients with very early OA might be different, especially because the results may reflect an inverse temporal relationship with an active inflammatory synovial process in patients with OA before structural abnormalities occur. Third, we included OA patients with main complaints of knee, hip, or hand, but the number of joints involved and the possible presence of generalized OA were unknown in these patients. Recently, more than half of patients with knee OA were shown to have generalized OA[16] and patients with hand OA reported considerable levels of disease impact across localizations[17]; hence the possible presence of generalized OA could have decreased the precision of our results as well. However, this possible involvement of generalized OA, along with the known occurrence of hand OA with OA at other sites[18], indicates that systemic factors may play a role and therefore supports combining knee, hip. and hand OA together.

The results from this cross-sectional exploratory study do not support a role for serum S100A8/A9 levels as a biomarker for severity of clinical symptoms and structural abnormalities in established knee, hip, and hand OA patients. The inverse association with the sum score of osteophytes and joint space narrowing and the positive association with ESR support the previous finding that S100A89/A9 may reflect inflammatory synovial processes in patients with OA before structural abnormalities occur, and warrant further exploration in longitudinal studies.

Acknowledgement
We thank all patients for their collaboration in this study.

References

1. Bijsma JWJ, Berenbaum F, Lafeber FPJG. Osteoarthritis: an update with relevance for clinical practise. Lancet 2011;377:2115-26.
2. Sunahori K, Yamamura M, Yamana J, Takasugi K, Kawashima M, Yamamoto H, et al. The S100A8/A9 heterodimer amplifies proinflammatory cytokine production by macrophages via activation of nuclear factor kappa B and p38 mitogen-activated protein kinase in rheumatoid arthritis. Arthritis Res Ther 2006;8:R69.
3. Frosch M, Ahlmann M, Vogl T, Wittkowski H, Wulffraat N, Foell D, et al. The myeloid-related proteins 8 and 14 complex, a novel ligand of toll-like receptor 4, and interleukin-1beta form a positive feedback mechanism in systemic-onset juvenile idiopathic arthritis. Arthritis Rheum 2009;60:883-91.
4. van Lent PLEM, Blom AB, Schelbergen RFP, Slöetjes A, Lafeber FPJG, Lems WF, et al. Active involvement of alarmins S100A8 and S100A9 in the regulation of synovial activation and joint destruction during mouse and human osteoarthritis. Arthritis Rheum 2012;64:1466-76.
5. Hammer HB, Odegard S, Fagerhol MK, Landewé R, van der Heijde D, Uhlig T, et al. Calprotectin (a major leucocyte protein) is strongly and independently correlated with joint inflammation and damage in rheumatoid arthritis. Ann Rheum Dis 2007;66:1093-7.
6. Punzi L, Ramonda R, Deberg M, Frallonardo P, Campana C, Musacchio E, et al. Coll2-1, Coll2-1NO2 and myeloperoxidase serum levels in erosive and non-erosive osteoarthritis of the hands. Osteoarthritis and Cartilage 2012;20:557-61.
7. Deberg M, Dubuc J-E, Labasse A, Sanchez C, Quettier E, Bosseloir A, et al. One-year follow-up of Coll2-1, Coll201NO and myeloperoxydase serum levels in osteoarthritis patients after hip or knee replacement. Ann Rheum Dis 2008;67:168-74.
8. Snijders GF, den Broeder AA, van Riel PL, Straten VH, de Man FH, van den Hoogen FH, et al; NOAC study group. Evidence-based tailored conservative treatment of knee and hip osteoarthritis; between knowing and doing. Scand J Rheumatol 2011;40:225-31.
9. Stukstette MJ, Dekker J, den Broeder AA, Westeneng JM, Bijlsma JWJ, van den Ende CHM. No evidence for the effectiveness of a multidisciplinary group based treatment program in patients with osteoarthritis of hands on the short term; results of a randomized controlled trial. Osteoarthritis Cartilage 2013;21:901-10.
10. Bellamy N, Campbell J, Haraoui B, Buchbinder R, Hobby K, Roth JH, et al. Clinimetric properties of the AUSCAN Osteoarthritis Hand Index: an evaluation of reliability, validity and responsiveness. Osteoarthritis Cartilage 2002;10:863-9.
11. Roth J, Goebeler M, van den Bos C and Sorg C. Expression of calcium-binding proteins MRP8 and MRP14 is associated with distinct monocyte differentiation pathways in HL-60 cells. Biochem Biophys Res Commun 1993;191:565-70.
12. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage 2007;15(Suppl A):A1-56.
13. Zreiqat H, Belluoccio D, Smith MM, Wilson R, Rowley LA, Jones K, et al. S100A8 and S100A9 in experimental osteoarthritis. Arthritis Res Ther 2010;12:R16.
14. Hammer HB, Odegard S, Syversen SW, Landewé R, van der Heijde D, Uhlig T, et al. Calprotectin (a major S100 leucocyte protein) predicts 10-year radiographic progression in patients with rheumatoid arthritis. Ann Rheum Dis 2010;69:150-4.
15. van Lent PL, Grevers L, Blom AB, Arntz OJ, van de Loo FAJ, van der Kraan, et al. Myeloid-related proteins S100A8/S100A9 regulate joint inflammation and cartilage destruction during antigen-induced arthritis. Ann Rheum Dis 2008;67:1750-8.
16. Forestier R, Francon A, Briole V, Genty C, Chevalier X, Richette P. Prevalence of generalized osteoarthritis in a population with knee osteoarthritis. Joint Bone Spine 2010;78:275-8.

17. Moe RH, Grotle M, Kjeken I, Hagen KB, Kvien TK, Uhlig T. Disease impact of hand OA compared with hip, knee and generalized disease in specialist rheumatology health care. Rheumatology(Oxford) 2013;52:189-196.
18. Dahaghin S, Bierma-Zeinstra SM, Reijman M, Pols HA, Hazes JM, Koes BW. Does hand osteoarthritis predict future hip or knee osteoarthritis? Arthritis and Rheumatism 2005;11:3520-7.

Chapter 7

The efficacy and safety of low-dose radiotherapy on pain and functioning in patients with osteoarthritis: a systematic review

Michiel J.M. Minten
Elien A.M. Mahler
Alfons A. den Broeder
Jan Willem H. Leer
Cornelia H.M. van den Ende

Published in Rheumatology International 2016 Jan;36(1):133-42

Abstract

Objective
Low-dose radiotherapy (LDRT) has been widely used for treatment of non-malignant disorders since its introduction and animal studies show anti-inflammatory effects in osteoarthritis (OA). However, the evidence for its effect in clinical practice remains unclear. Therefore, the aim of this study is to systematically summarise the literature on effectiveness of LDRT on pain and functioning in patients with OA and its safety.

Methods
Broad search terms were used to search PubMed, EMBASE, and Web of Science. Primary inclusion criteria were: osteoarthritis as indication, radiotherapy as intervention, written in English, German or Dutch, and published since 1980. Study quality was assessed using the EPHPP Quality Assessment Tool for Quantitative Studies (scale: strong, moderate, weak).

Results
Seven studies were suitable for inclusion, all with retrospective uncontrolled observational design. Methodological quality of all studies was judged as weak. Most studies used 2-3 RT-sessions per week for 2 weeks, some with booster session after 6 weeks. Generally, non-validated single-item measurement instruments were used to evaluate the effect of LDRT on pain and function. Across the studies, in 25–90% and 29–71% of the patients pain and functioning improved, respectively. Side effects were described in one study, none were reported.

Conclusion
Our results show that there is insufficient evidence for efficacy or to confirm the safety of LDRT in treatment of OA, due to absence of high-quality studies. Therefore, a well-designed, sham-controlled and blinded randomised trial, using validated outcome measures is warranted to demonstrate the value of LDRT for OA in clinical practice.

Introduction

Osteoarthritis (OA) is the most prevalent form of joint disease. For example, it is estimated that up to 20% of the population of the United States is affected.[1] In OA, the volume and the quality of cartilage is decreased due to an imbalance in breakdown and synthesis of cartilage. Although the exact aetiology of OA is unknown, it is now clear that ligaments and subchondral bone are affected and that inflammation is involved.[2–4] OA is characterised by pain and stiffness in affected joints, resulting in limitations in physical functioning and a loss of health-related quality of life.[5,6] The disease develops progressively, but symptoms might remain stable or even improve.[7] Known risk factors for the development of OA are old age, genetic predisposition and increased body mass index.[8–10]

Since there is no disease-modifying treatment available, current OA treatment is symptomatic.[11] Treatments can be categorised into non-surgical (e.g. education, exercise, insoles, braces and acetaminophen or NSAIDs use) and surgical treatments. However, effect sizes of non-surgical treatment options are small to moderate.[12] When non-surgical treatments do not result in satisfactory reduction in complaints, surgical options are often considered, with total joint replacement (TJR) as the most common option. Although TJR is effective in improving pain and function, there are several drawbacks in surgical treatments: it is associated with risk for complications, it is a costly procedure, and implants have a limited life time.[13–16] Furthermore, in approximately 5% of the patients, surgery cannot be performed due to comorbidities and the absence of patients' willingness to have TJR.[17] Considering this, the need for additional and more effective non-surgical treatments is evident.

A potential target for OA treatment may be synovial inflammation. Recent research suggests a crucial role of inflammation in OA pathogenesis. Inflamed synovium is thought to produce catabolic and pro-inflammatory mediators which, consequently, alter the balance of cartilage matrix degradation and repair, leading to excess production of enzymes responsible for cartilage breakdown.[4] This cartilage alteration, in turn, amplifies synovial inflammation, resulting in a vicious circle.[4] Ultrasound research has shown that inflammatory features as effusion, synovial thickening, and positive power Doppler signal are present in 96–100% of the patients with hand OA[18] and in 40–65% in patients with knee OA[19, 20] It could be hypothesised that when these inflammatory factors are reduced, the vicious cycle of inflammation and cartilage breakdown in OA is broken, and that pain and functioning improve. Recent evidence from both in vitro and in vivo studies shows that treatment with external beam radiotherapy (RT) in a low-dose (LD) has an anti-inflammatory effect in OA animal models.[21] Therefore, low-dose radiotherapy (LDRT) may be a promising non-surgical treatment option for OA. Although no definition of LDRT exists, in this paper, "low-dose" is added to "radiotherapy" to distinguish the dosage used for treatment of OA from the dosage used for treatment in oncology, which is typically 5–10 times higher.

RT has been used in various dosages for treating non-malignant disorders, including OA, since its introduction in the late nineteenth century.[22–24] A survey of radiotherapy institutes across the world has shown that LDRT is more commonly used as treatment for OA in eastern Europe, compared to western Europe and North America (85% vs. 23%, respectively).[25] As the technique has improved over the course of time (e.g. smaller radiation fields, new radiation equipment, and lower doses), the treatment of symptomatic OA with radiotherapy has become more

acknowledged within the professional community of radiotherapists in the West, resulting in increased use of this modality.[26, 27] However, despite the high acceptance in certain parts of the world, the use of LDRT as treatment for OA is not well documented. Nonetheless, a survey among RT departments in Germany reported that over 9000 patients with OA are treated with LDRT every year, in order to relieve their pain, and that this number almost tripled over 5–8 years.[28]

However, despite widespread use of LDRT in the treatment of OA, the evidence on the effects of LDRT on pain and functioning in patients with OA remains unclear. Therefore, the aim of the current study is to provide a systematic review of the literature on the effectiveness and safety of LDRT on pain and functioning in patients with OA.

Materials and methods

All studies investigating the effect of external beam radiotherapy on pain and/or functioning in patients with OA were collected for this study. No distinction was made in OA location or duration of complaints. Studies published prior to 1980 were excluded to improve extrapolation of results to the current clinical practice, as relatively recent technological advances have improved target volume definition and accuracy of radiation dose delivery.[26] Recommendations for reporting systematic reviews and meta-analyses (PRISMA) were followed in the current study.[29]

Eligibility criteria: inclusion and exclusion

Studies were eligible if they aimed to treat OA, the intervention was external beam radiotherapy, the effect of the radiotherapy intervention on pain or functioning was assessed and if publication type was "journal article". Studies were excluded if no primary data were presented, if written in a language other than English, German or Dutch, or published prior to 1980. Selection of studies was unrestrained with regard to radiation source, radiotherapy protocol, study design, presence of control group, and outcome measures used to assess pain and functioning.

Literature search

A computerised search, with broad search terms was performed in the PubMed, EMBASE, and Web of Science databases on 20 April 2015. The search was based on (Medical) Subject Headings (PubMed, EMBASE), Topics (Web of Science) and matches in title, abstract and article text for the terms osteoarthritis, degenerative joint disease, radiotherapy and irradiation or synonyms. Broad search terms were used to minimise the chance of missing relevant publications. The full search strategy is available in supplementary data. Additionally, the reference lists of eligible publications were scanned for potentially eligible publications, missed by the computerised search.

Selection procedure

First, titles of the publications found by the literature search were screened for possible eligibility by MM and CvdE. Second, if titles were inconclusive, abstracts – and when still inconclusive full text versions – were retrieved for further analysis. In case of disagreement, the eligibility of the study was discussed until full agreement was reached.

Data extraction

A predefined, self-designed data extraction form was used to support data extraction from included papers. This form included items on study characteristics (year of publication, study design), intervention (dose, frequency, irradiated area and size), patient characteristics (OA location, number, age, gender, duration of complaints), outcome measures, results and safety. When results for different musculoskeletal disorders were reported, only results of patients with OA were extracted and are presented in this study.

Study quality

Methodological quality of included publications was assessed independently by MM and CvdE, using the EPHPP Quality Assessment Tool for Quantitative Studies.[30] This tool assesses eight different domains: selection bias, study design, confounders, blinding, data collection methods, withdrawals and drop-outs, intervention integrity and analyses. Combining all domains, every publication was rated on a three-level scale: strong (no weak ratings in all domains), moderate (one weak rating in all domains) or weak (two or more weak ratings in all domains). In case of disagreement on quality domains, the study was discussed until consensus was reached.

Best-evidence synthesis

To merge results of the included studies, a best-evidence synthesis (BES) was performed for both outcome measures pain and functioning. With use of this synthesis, the conclusion of this review will be based on results of the studies with the highest quality, as previously described.[31] In this BES, not only randomised controlled trials (RCT's), but also non-controlled designs as observational studies can be enclosed. The different possible levels of evidence are shown in Table 1.

Table 1. Best evidence synthesis, possible levels.

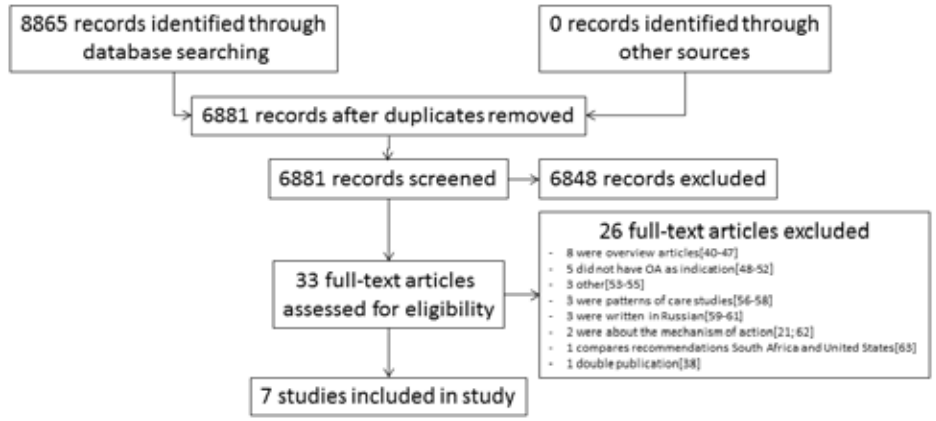
Strong evidence	Provided by consistent, statistically significant findings in outcome measures in at least two high quality RCT's
Moderate evidence	Provided by consistent, statistically significant findings in outcome measures in at least one high quality RCT and at least one low quality RCT or high quality CCT
Limited evidence	Provided by statistically significant findings in outcome measures in at least one high quality RCT
	Or
	Provided by consistent, statistically significant findings in outcome measures in at least two high quality CCT's (in absence of high quality RCT's)
	Provided by statistically significant findings in outcome and/or process measures in at least one high quality CCT or low quality RCT (in absence of high quality RCT's)
	Or
	Provided by consistent, statistically significant findings in outcome and/or process measures in at least two OD's with sufficient quality (in absence of RCT's and CCT's)
Indicative findings	Provided by statistically significant findings in outcome measures in at least one high quality CCT or low quality RCT (in absence of high quality RCT's)
	Or
	Provided by consistent, statistically significant findings in outcome and/or process measures in at least two OD's with sufficient quality (in absence of RCT's and CCT's)
No or insufficient evidence	In the case that results of eligible studies do not meet the criteria for one of the above stated levels of evidence
	Or
	In the case of conflicting (statistically significant positive and statistically significant negative) results among RCT's and CCT's
	Or
	In the case of no eligible studies

RCT: randomised controlled trial; CCT: controlled clinical trial; OD: design other than controlled designs

Results

The computerised search generated 8865 hits; 6881 publications remained after duplicates removal. Screening of titles, abstracts and full text articles resulted in eight eligible publications.[32–39] However, two publications reported the same results from the same study.[37, 38] Therefore, the most recent publication was excluded[38], resulting in seven included publications (Figure 1).

Figure 1. Flow diagram of study selection.



Study characteristics and results are shown in Table 2. All studies had a retrospective observational design, without a control group. Included patient populations varied between a homogenous group of patients with OA at one location [33,39], a heterogeneous group of patients with OA at different locations [32,37], and study samples of patients with different musculoskeletal disorders.[34-36] However, none of the studies described the criteria for diagnosis. In total, 2164 OA patients received LDRT in the included studies (i.e. 1867 knee, 133 shoulder, 96 hip, 49 spine, and 19 thumb). In general, patients were between 50 and 70 years of age, about 2/3 was female and duration of complaints varied from less than half a year[36] to 15 years[37]. Two studies reported about radiological OA severity of included patients.[37,39] Both used OA grade according to Kellgren and Lawrence.[37,39,64] The population's baseline level of pain was not reported in any of the included studies; the baseline level of functioning was reported in one study.[32] Eighty-five percent of their population had movement restrictions (14, 62, and 9% for minor, moderate and complete movement restriction, respectively). As intervention, four studies used two [32,34–36] and one study used three [37] LDRT sessions per week for 2 weeks. In one study, the frequency of sessions varied between one and five times per week [39] and one study did not report on the frequency of LDRT sessions.[33] One study used a booster repeat after 6 weeks, when the initial effect was not satisfactory.[37] Overall, a dose of 0.5–1.0 Gray (Gy; 1 Gy = 1 joule of absorbed energy per kilogram) per session was applied, with a total dose of 3.5–6.0 Gy. Five studies presented short-term follow-up results (≤ 3 months) [33-36,39], including assessments directly after intervention [33–35,39]; while three studies presented long-term results (> 3 months). [32,36,37] The quality of all included studies was considered weak. In particular, the domains of

confounding, blinding and data collection were considered weak over studies. The rating of study quality per study per domain is available in supplementary data.

Pain

In general, non-validated single-item measurement instruments were used to evaluate the effect of LDRT on pain. The rating score according to Von Pannewitz[65] was used in three studies.[34,37,39] With this rating score, improvement of pain was rated on a five-point scale: now painless, markedly improved, improved, stable and worse. Comparable non-validated scores (e.g. three-item transition scales: (1) complaint free, improved, no change; (2) no improvement, good effect, excellent effect) were used in the other four studies.[33–36] On the short term, decrease in pain, as indicated by transition scale, was observed in 13–90% of the patients, whereas long-term analgesic effects were observed in 44–87% of the patients. However, no studies with sufficient quality were retrieved. Thus, according to the BES, there is insufficient evidence for a positive effect of LDRT on pain.

Functioning

The effect of LDRT on functioning in OA was assessed by three studies.[36,37,39] In the study of Keilholz et al.[37], joint specific scores were used as outcome measures: Harris hip score for hip OA, Constant score for shoulder OA, Japanese Knee-Score, Tegner-Lysholm-score and Insall-Knee score for knee OA and a self-developed score for thumb OA. Depending on the OA location, improvement in functioning was reported in 55 (thumb) to 71% (hip) of the patients treated with LDRT. Sautter-Bihl et al. reported results on improvement of mobility.[36] No further specification was given. Their results suggest that 81 and 72% of the patients improved in mobility, for shoulder and knee OA, respectively. Keller et al. asked included patients about ability to move in a three-level classification in a survey, 2–14 years after intervention.[39] They report improvement in 39.8% of the patients; 56.5% remained stable, and 3.7% reported worse ability to move. However, no studies with sufficient quality were retrieved. Therefore, according to the BES, there is insufficient evidence for a positive effect of LDRT on functioning.

Side effects

Of the included studies, two report on the possibility of side effects on the short term[36,37]; one of those explicitly reported that no side effects were observed on the short-term [36], and the other did not present data on short-term side effects.[37] Potential long-term side effects were discussed by four of the included studies, but no results were presented. [33,36,37,39] According to the BES, there is insufficient evidence for the safety of LDRT as treatment for OA.

Table 2. Study design, patient characteristics, outcome measures and results

Publication	Design	Intervention	Patients (OA, n)	Gender (% male/female)	Age ± Sd (years)	Radiological OA severity	Duration of complaints	Follow up	Pain		Functioning	
									Outcome measures	Results	Outcome measures	Results
Lindner 1982 [32]	Retrospective observational; no control group	1.0 or 2.0 Gy 2x/week Total dose: 4 to 10 Gy	Degenerative joint illness: Knee: 105 Hip: 53 Spine: 49 Shoulder: 42	Knee: 21/79 Hip: 25/75 Spine: 39/61 Shoulder: 58/42	87% between 50 and 80	-	<8 weeks: 10% 8wks-6mnts: 27% 0.5-2.0yrs: 28% 2.0-5.0yrs: 17% >5.0yrs: 12%	average of 4.2 years 4 - 114 months	Average score, according to following categories: Free of complaints: 100 Significantly improved: 60 Improved: 30 Not influenced: 0	Knee score: 29.1/100; 57% improved Hip score: 22.5/100; 44% improved Spine score: 38.6/100; 74% improved Shoulder score: 43.1/100; 76% improved Total score: 31.9/100; 61% improved	-	-
Keinert 1982 [33]	Retrospective observational; no control group	0.5 or 1.0 Gy Total of 6 to 8 times Total dose: 3.0 to 8.0 Gy	OA of Knee: 290	28/72	63 (61 to 70)	-	<1year: 47% >1year: 44% unknown: 9%	Direct after therapy; 6 weeks after therapy	1) Free of complaints; 2) Better; 3) No change	Direct: 1) 8%; 2) 56%; 3) 36% 6 weeks: 1) 38%; 2) 43%; 3) 19%	-	-
Yaneva 1986 [34]	Retrospective observational; no control group	1, 2 or 4 fields, depending on localisation 387x10-4 C/kg per field; 2x/week Total of 6 to 8 times Total dose: ≤2064x10-4 C/kg per field	OA of Knee: 341 OA of Hip: 32	32/68 (not specified for OA location)	64±12.4 (not specified for OA location	-	-	Immediately after treatment	1) excellent effect 2) good effect 3) effect absent	Knee OA: 1) 3.8% 2) 87.2% 3) 10.0% Hip OA: 1) 12.5% 2) 68.8% 3) 18.8%	-	-

Table 2 continued

Publication	Design	Intervention	Patients (OA, n)	Gender (% male/female)	Age ± Sd (years)	Radiological OA severity	Duration of complaints	Follow up	Pain		Functioning	
									Outcome measures	Results	Outcome measures	Results
Gärtner 1988[35]	Retrospective observational; no control group	0.75 or 1 Gy 2x/week Total dose: 3.75 to 5 Gy	OA of Hip: 5 OA of Knee: 21	-	Knee: 58.4 Hip: 58.0	-	Knee: 9 years Hip: 9 year	End of treatment; 3 months after treatment	1) Free of complaints 2) Improved 3) No change 4) Deteriorated 5) Significantly deteriorated	End of treatment: ≤2) 37.5%(hip); 24.5%(knee) 3) 62.5%(hip); 65.8%(knee) ≥4) 0%(hip); 18.7%(knee) 3 months after treatment: ≤2): 12.5%(hip); 34.8%(knee) 3): 62.5%(hip); 51.7%(knee) ≥4): 25.0%(hip); 13.5%(knee)	-	-
Sautter-Bihl 1993[36]	Retrospective observational; no control group	0.5 to 1 Gy 2x/week Total dose: 2.5 to 6 Gy	OA of Shoulder and Periarthritis humeroscapularis: 74 OA of Knee: 42	Overall: 53/47	Overall: 54 year (29-81)	-	Shoulder: <0.5 year: 24% 0.5- <1year: 35% 1- <2 year: 5% ≥2 year: 35% Knee: <0.5 year: 14% 0.5- <1year: 26% 1- <2 year: 45% ≥2 year: 14%	End of treatment (pain); Long term (undescribed; pain and mobility)	Therapy effect on pain: 1) Very good 2) Good 3) Satisfying 4) Little effect 5) No effect	End of treatment: ≤3) 57%(shoulder); 62%(knee) 4) 23%(shoulder); 24%(knee) 5) 20%(shoulder); 14%(knee) Long term: ≤3) 70%(shoulder); 71%(knee) 4) 17%(shoulder); 10%(knee) 5) 13%(shoulder); 19%(knee)	Therapy effect on mobility: 1) Very good 2) Good 3) Satisfying 4) Little effect 5) No effect	Long term: ≤3) 74%(shoulder); 62%(knee) 4) 7%(shoulder); 10%(knee) 5) 20%(shoulder); 29%(knee)

Table 2 continued

Keilholz 1998[37]	Retrospective observational; no control group	0.5 or 1.0 Gy 3x/week for 4 weeks Total dose: 6 or 12 Gy	Degenerative osteoarthritis: Thumb: 19 Knee: 31 Hip: 6 Shoulder: 17	Overall: 37/63	Overall: 62±14	Thumb: 9 patients classified; K&L > 2: 67% Shoulder: 25 patients classified; K&L > 2: 68% Knee: 36 patients classified; K&L > 2: 56% Hip: 7 patients classified; K&L > 2: 72%	average 4 years (median 3 years; 0.5 - 15 years)	Mean of 4 years (1-10 years)	Pain: 1) Free of complaints 2) Significant improvement 3) Improvement 4) No effect 5) Deteriorated	≤3) 50%(thumb); 65%(shoulder); 64%(knee); 72%(hip) 4) 45%(thumb); 29%(shoulder); 26%(knee); 14%(hip) 5) 0%(thumb); 6%(shoulder); 10%(knee); 0%(hip)	Stable or improved according to: Harris Hip Score (hip) Constant score (shoulder) Japanese Knee score (knee) Self-developed score (thumb)	Stable: 45%(thumb); 41%(shoulder); 33%(knee); 29%(hip) Improved: 55%(thumb); 59%(shoulder); 67%(knee); 71%(hip)
Keller 2013[39]	Retrospective observational; no control group	0.5 to 1.5 Gy 1, 2 or 5x/week Total dose: 0.5 to 10Gy	Painful OA of Knee: 1037	30.5/69.5	≥60: 30.8% <60: 69.2%	Minimal: 18.3% Moderate: 35.0% Severe: 46.7%	<1year: 24.6% 1-3years: 24.8% >3years: 50.6%	Immediately after therapy or 2 months after therapy Survey 2 to 14 years after therapy	Pain: 1) Painless 2) Markedly improved 3) improved 4) stable 5) worse	Immediately or 2 months after therapy: ≤3) 79.3 4) 12.0% 5) 8.7% Survey: ≤3) 49.1% 4) 46.3% 5) 4.6%	Ability to move in a 3-level classification: Worse, Stable and Improved During survey	Worse: 3.7% Stable: 56.5% Improved: 39.8%

-: not applicable

Discussion

This review is the first to systematically summarise the literature on the effects of radiotherapy on pain and functioning in patients with OA. Our results show that there is insufficient evidence for a beneficial effect on pain or function, or to confirm the safety of LDRT in treatment of OA, due to the absence of studies with acceptable quality. Only studies with a retrospective, observational design and without a control group were retrieved from our search.

It is striking that, despite insufficient scientific evidence, LDRT is broadly applied as therapy for OA in large parts of the world, because although it may be effective, LDRT is not without possible disadvantages. First, there is patient exposure to radiation. So far, no data have been reported on the long-term side effects of LDRT. However, the risk has been estimated, based on mathematical models. RT dosed at 6 Gy for OA of the knee joint appears to relate to an effective dose of 13 mSv (comparable to a routine abdomen-pelvis CT-scan[66]) and an average attributable lifetime risk for an induced fatal tumour about 0.7 in thousand patients treated at the age of 50 (0.3 in thousand patients treated at the age of 75).[67] When more proximal joints (e.g. hip or shoulder) are irradiated, exposure to more susceptible tissues (e.g. bone marrow or intestine mucosa), and thus the risk for tumours, will increase. Secondly, LDRT is relatively costly and complex, as multiple medical professionals are involved (e.g. radiotherapist, physicist, technologist) and radiotherapy equipment in a safe environment is required. Considering these disadvantages, it could be considered unethical to expose patients to LDRT when clear scientific evidence for a positive effect in clinical practice is absent. We believe that treatment with LDRT only has a place in clinical practice if its effect size is proportional to the disadvantages and compares favourably to other treatment modalities. Furthermore, LDRT should at first only be used as treatment in patients not responding to generally-accepted non-surgical treatments (e.g. pharmacological treatment and physical therapy). When these requirements are met, the number of patients needed to treat, to relieve complaints in one patient, will be low and unnecessary exposure to radiation will be minimised.

Despite the low scientific evidence for efficacy in clinical practice, multiple studies in animal models, investigating the underlying working mechanism of LDRT on OA, have been performed. The results have been summarized by Arenas et al.[21] They conclude that LDRT is able to decrease both clinical inflammatory parameters and improve histological markers. Overall, it is suggested that LDRT is able to reduce the inflammatory factors associated with OA and break the vicious cycle of cartilage break down. Paradoxically, this effect of LDRT is opposite to the effect of high-dose RT, as applied in treatment of malignant diseases, where DNA is damaged, resulting in apoptosis.[68]

In light of the above, a high-quality RCT of LDRT in OA patients is warranted to overcome the following issues with the current data. Firstly, the included studies had no blinding and control group. Therefore, their positive results could be the result of placebo effect and regression to the mean, which are known to be substantial in OA patients.[69] Additionally, known factors that are associated with higher placebo effects are present here, including more frequent intervention sessions and increased invasiveness.[70] Thus, it is conceivable that the positive effects of multiple sessions of LDRT, reported by studies included in this review, could (to a large extent) consist of placebo effect and regression to the mean. Secondly, included studies did not describe the patient population clearly. For example, it is unclear what definition of

OA was used and what the baseline levels of pain were in the include studies. Therefore, it is difficult to extrapolate the results to well-defined populations as OA according to the American College of Rheumatology classification criteria for hand, hip, or knee OA.[71–73] Thirdly, no validated outcome measures were used. All studies used transition scales for pain, either the rating score according to Von Pannewitz or comparable transition scales, ignoring the levels of pain before the intervention. Therefore, the interpretation of the results is problematic.

There are some limitations to this review. First, our search excluded studies published prior to 1980. Therefore, there is a possibility that early high-quality studies exist and are not included in this review. Nonetheless, this restriction was chosen to improve generalisability of the results from this review to the current clinical practice. Radiation techniques, doses and field sizes have changed over time, and studies from the early days of radiotherapy research are likely not be comparable from later ones. For instance, two RCT's investigating the effect of LDRT on patients with OA were published in the 1970s.[74,75] In both studies, the therapy consisted of three RT sessions, with a dose per session ranging between the equivalent of 1–2.5 Gy. When comparing this to the recent German guidelines for radiation therapy of painful degenerative skeletal disorders[76], only half the number of currently advised sessions were applied. Furthermore, the dose per session was up to five times higher than recommended nowadays. Overall, early studies tend to use higher dosed radiotherapy, while the anti-inflammatory effects are seen due to low dosages. Second, the Quality Assessment Tool for Quantitative Studies, that was used as a tool for scoring methodological quality, is a non-precise tool with only three possible scores (strong, moderate and weak).[30] Nevertheless, it is highly unlikely that the conclusion of this review would be different if another quality assessment tool was used.

Conclusion

Although widely used and based on pathophysiological work, there is insufficient evidence for a positive effect of LDRT on pain and functioning in OA patients, due to the absence of high-quality studies. Therefore, a well-designed, sham-controlled, blinded, randomised trial using validated outcome measures is warranted to justify the use of LDRT as treatment for OA in clinical practice.

Acknowledgement

We thank Joke Vriezekolk for her help and advice regarding the search strategy of this review.

References

1. Helmick CG, Felson DT, Lawrence RC, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. Arthritis Rheum 2008;58:15–25.
2. Grynpas MD, Alpert B, Katz I, et al. Subchondral bone in osteoarthritis. Calcif Tissue Int 1991;49:20–26.
3. Tan AL, Grainger AJ, Tanner SF, et al. High-resolution magnetic resonance imaging for the assessment of hand osteoarthritis. Arthritis Rheum 2005;52:2355–2365.
4. Sellam J, Berenbaum F. The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. Nat Rev Rheumatol 2010;6:625–635.
5. Davis M a., Ettinger WH, Neuhaus JM, Mallon KP. Knee osteoarthritis and physical functioning: Evidence from the NHANES I epidemiologic followup study. J Rheumatol 1991;18:591–598.
6. Salaffi F, Carotti M, Stancati A, Grassi W. Health-related quality of life in older adults with symptomatic hip and knee osteoarthritis: a comparison with matched healthy controls. Aging Clin Exp Res 2005;17:255–263.
7. Van Dijk GM, Dekker J, Veenhof C, Van Den Ende CHM. Course of functional status and pain in osteoarthritis of the hip or knee: A systematic review of the literature. Arthritis Care Res 2006;55:779–785.
8. Dahaghin S, Bierma-Zeinstra SM a, Ginai a Z, et al. Prevalence and pattern of radiographic hand osteoarthritis and association with pain and disability (the Rotterdam study). Ann Rheum Dis 2005;64:682–687.
9. Spector TD, MacGregor AJ. Risk factors for osteoarthritis: Genetics. Osteoarthritis Cartilage 2004;12:S39–44.
10. Oliveria S a, Felson DT, Cirillo P a, et al. Body weight, body mass index, and incident symptomatic osteoarthritis of the hand, hip, and knee. Epidemiology 1999;10:161–166.
11. Bijlsma JWJ, Berenbaum F, Lafeber FPJG (2011) Osteoarthritis: An update with relevance for clinical practice. Lancet 377:2115–2126.
12. Zhang W, Nuki G, Moskowitz RW, et al. OARSI recommendations for the management of hip and knee osteoarthritis. Part III: Changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage 2010;18:476–499.
13. Hawker G a., Badley EM, Borkhoff CM, et al. Which patients are most likely to benefit from total joint arthroplasty? Arthritis Rheum 2013;65:1243–1252.
14. Cushner F, Agnelli G, FitzGerald G, Warwick D. Complications and functional outcomes after total hip arthroplasty and total knee arthroplasty: results from the Global Orthopaedic Registry (GLORY). Am J Orthop (Belle Mead NJ) 2010;39:22–28.
15. Rissanen P, Aro S, Slätis P, et al. Health and quality of life before and after hip or knee arthroplasty. J Arthroplasty 1995;10:169–175.
16. Jones CA, Voaklander DC, Johnston DW, Suarez-Almazor ME. Health related quality of life outcomes after total hip and knee arthroplasties in a community based population. J Rheumatol 2000;27:1745–52.
17. Frankel L, Sanmartin C, Conner-Spady B, et al. Osteoarthritis patients' perceptions of "appropriateness" for total joint replacement surgery. Osteoarthritis Cartilage 2012;20:967–973.
18. Kortekaas MC, Kwok WY, Reijnierse M, et al. Follow-up study of inflammatory ultrasound features in hand osteoarthritis over a period of 3 months: Variable as well as constant. Osteoarthritis Cartilage 2014;22:40–43.
19. Bevers K, Bijlsma JWJ, Vriezekolk JE, et al. The course of ultrasonographic abnormalities in knee osteoarthritis: 1 year follow up. Osteoarthritis Cartilage 2014;22:1651–1656.
20. Chan KKW, Sit RWS, Wu RWK, Ngai AHY. Clinical, Radiological and Ultrasonographic Findings Related to Knee Pain in Osteoarthritis. PLoS One 2014;9:e92901.
21. Arenas M, Sabater S, Hernández V, et al. Entzündungshemmende Effekte von niedrigdosierter Strahlentherapie. Strahlentherapie und Onkol 2012;188:975–981.

22. Sokoloff N (1898) [X-rays against joint rheumatism]. Fortschr Röntgenstr 1:209–13.
23. Schmid-Monnard C. [About cure of joint rheumatism by X-rays in children]. Fortschr Röntgenstr 1898;1:209.
24. Anders JM, Daland J, Pfahler GF. The Treatment of Arthritis Deformans With the Roentgen Rays. a Preliminary Report. JAMA J Am Med Assoc 1906;XLVI:1512–1514.
25. Leer JWH, Van Houtte P, Davelaar J. Indications and treatment schedules for irradiation of benign diseases: A survey. Radiother Oncol 1998;48:249–257.
26. Freeman CR, Podgorsak EB. Radiation oncology: times of practice change. Curr Oncol 2007;14:165–166.
27. Leer JW, van Houtte P, Seegenschmiedt H. Radiotherapy of non-malignant disorders: Where do we stand?. Radiother Oncol 2007;83:175–177.
28. Seegenschmiedt MH, Micke O, Willich N. Radiation therapy for nonmalignant diseases in Germany: Current concepts and future perspectives. Strahlentherapie und Onkol 2004;180:718–730.
29. Moher D, Liberati A, Tetzlaff J, Altman DG. Reprint--preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Phys Ther 2009;89:873–880.
30. Thomas BH, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: Providing the research evidence for public health nursing interventions. Worldviews Evidence-Based Nurs 2004;1:176–184.
31. Steultjens EMJ, Dekker J, Bouter LM, et al. Occupational therapy for stroke patients: A systematic review. Stroke 2003;34:676–686.
32. Lindner H, Freislederer R. Long term results of radiotherapy of degenerative joint diseases (author's transl). Strahlentherapie 1982;158:217–223.
33. Keinert K, Schalldach U, Schumann E. Results of irradiation therapy of arthrosis deformans of the knee-joint. 1982;37:445–447.
34. Yaneva M, Vlaikova M, Dimitrova M, Grigorov L. X-ray therapy in degenerative-distrophic diseases of the bones and joints. Folia Med (Plovdiv) 1986;28:26–28.
35. Gärtner C, Schüttauf M, Below M, et al. The radiotherapeutic treatment of chronically relapsing, degenerative skeletal changes at the Oncology Clinic (Charité). Radiobiol Radiother (Berl) 1988;29:689–698.
36. Sautter-Bihl M-L, Liebermeister E, Scheurig H, Heinze H-G. Analgetische Bestrahlung degenerativentzündlicher Skeletterkrankungen: Nutzen und Risiko. DMW - Dtsch Medizinische Wochenschrift 1993;118:493–498.
37. Keilholz L, Seegenschmiedt H, Sauer R. Radiotherapy for painful degenerative joint disorders. Indications, technique and clinical results. Strahlenther Onkol 1998;174:243–250.
38. Ruppert R, Seegenschmiedt MH, Sauer R. Radiotherapy of osteoarthritis. Indication, technique and clinical results. Orthopade 2004;33:56–62.
39. Keller S, Müller K, Kortmann R-D, et al. Efficacy of low-dose radiotherapy in painful gonarthritis: experiences from a retrospective East German bicenter study. Radiat Oncol 2013;8:29.
40. Buchholz J, Ludewig E. How sensible is radiation therapy in the treatment of degenerative joint diseases? Prakt Tierarzt 2009;90:416-422.
41. Calin A. X-radiation in the management of rheumatoid disease. Br J Hosp Med 1985;33:261–265.
42. Hassenstein E. Radiation treatment of benign diseases--indications, results and technic. Rontgenblatter 1986;39:21–23.
43. Hensel E. Degenerative and inflammatory diseases and their radiotherapy. Wien Med Wochenschr 1980;130:152–153.
44. Merte H, Pfab R. Radiotherapy of degenerative inflammatory joint diseases in the elderly. A worthwhile therapeutic measure. Fortschr Med 1995;113:323–324.

45. Puett DW, Griffin MR. Published trials of nonmedicinal and noninvasive therapies for hip and knee osteoarthritis. Ann Intern Med 1994;121:133–140.
46. Seegenschmiedt MH, Micke O. Strahlentherapie nichtmaligner Erkrankungen. Strahlentherapie und Onkol 2012;188:272–290.
47. Nissanke M, Sindzingre A. Institutional foundations for shared growth in sub-saharan Africa. African Dev Rev 2006;18:353–391.
48. Hackenberg L, Schäfer U, Micke O, Liljenqvist U. Die schmerzbestrahlung des chronischen, degenerativen lumbalsyndroms - Ergebnisse einer prospektiven randomisierten studie. Z Orthop Ihre Grenzgeb 2001;139:294–297.
49. Öner-Dinçbas F, Akyazici E, Koca S, et al. Radiotherapy for benign diseases: A retrospective analysis. J BU ON 2004;9:57–61.
50. Ott OJ, Hertel S, Gaipl US, et al. Benign painful shoulder syndrome: Initial results of a single-center prospective randomized radiotherapy dose-optimization trial. Strahlentherapie und Onkol 2012;188:1108–1113.
51. Schäfer U, Micke O, Willich N. Radiotherapy of pain in degenerative bone diseases. Rontgenpraxis 1996;49:251–254.
52. Wannenmacher M. The importance of radiotherapy in backache. Radiologe 1983;23:174–178.
53. Feldmann HJ. [Painful degenerative and inflammatory diseases of joints and soft tissue: x-ray instead of pharmaceutics?]. MMW Fortschr Med 2013;155:42.
54. Grosu a L. Radiotherapy outside of oncology. When is it indicated? (interview by Waltraud Paukstadt). MMW Fortschr Med 2001;143:16.
55. Merte H, Proske H, Pfab R. Radiotherapy of degenerative-inflammatory joint diseases - allows telethermography to verify the indication. 1993;44:273–275.
56. Mücke R, Seegenschmiedt MH, Heyd R, et al. Strahlentherapie bei schmerzhafter Kniegelenkarthrose (Gonarthrose): Ergebnisse einer deutschen Patterns-of-Care-Studie. Strahlentherapie und Onkol 2010;186:7–17.
57. Schultze J, Eilf K. Perspektiven der strahlentherapie gutartiger erkrankungen. Strahlentherapie und Onkol 2006;182:259–262.
58. Seegenschmiedt MH, Katalinic a, Makoski HB, et al. Radiotherapy of benign diseases: a pattern of care study in Germany. Strahlenther Onkol 1999;175:541–547.
59. Bogdashevskaia VG, Nadtochiĭ a G, Shestakov IN. X-ray study in secondary osteoarthrosis deformans of the temporomandibular joints in children. Stomatologiia (Mosk) 1983;62:27–29.
60. Pereslegin I a, Zolotkov a G, Aleĭ nikov GE. Radiotherapy in degenerative-dystrophic and inflammatory diseases of the osteoarticular system. Med Radiol (Mosk) 1984;29:38–41.
61. Pereslegin I a, Podliashchuk EL, Ustinova VF. Problems and perspectives of radiotherapy of non-tumor diseases. Vestn Rentgenol Radiol 1990;54–58.
62. Rödel F, Kamprad F, Sauer R, Hildebrandt G. Functional and molecular aspects of anti-inflammatory effects of low-dose radiotherapy. Strahlenther Onkol 2002;178:1–9.
63. Pretorius FJ, Alberts a S, Klopper ER. Radiation of benign disease--a South African perspective. S Afr J Surg 1995;33:115–116.
64. Kellgren JH, Lawrence JS. Radiological Assessment of Osteo-Arthrosis. Ann Rheum Dis 1957;16:494–502.
65. Von Pannewitz G. [The X-ray therapy of arthritis deformans-Clinical and experimental studies]. In: Holfelder H (ed) Ergebnisse der Medizinischen Strahlenforsch. Thieme, Leipzig, 1933;pp 61–126
66. Smith-Bindman R. Radiation Dose Associated With Common Computed Tomography Examinations and the Associated Lifetime Attributable Risk of Cancer. Arch Intern Med 2009;169:2078.
67. Jansen JTM, Broerse JJ, Zoetelief J, et al. Estimation of the carcinogenic risk of radiotherapy of benign diseases from shoulder to heel. Radiother Oncol 2005;76:270–277.

68. Dewey WC, Ling CC, Meyn RE. Radiation-induced apoptosis: Relevance to radiotherapy. Int J Radiat Oncol Biol Phys 1995;33:781–796.
69. Zhang W, Robertson J, Jones a C, et al. The placebo effect and its determinants in osteoarthritis: meta-analysis of randomised controlled trials. Ann Rheum Dis 2008;67:1716–1723.
70. Abhishek a., Doherty M. Mechanisms of the placebo response in pain in osteoarthritis. Osteoarthritis Cartilage 2013;21:1229–1235.
71. Altman R, Asch E, Bloch D, 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 Rheum 1986;29:1039–1049.
72. Altman R, Alarcón G, Appelrouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hand. Arthritis Rheum 1990;33:1601–1610.
73. Altman R, Alarcón G, Appelrouth D, et al. The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505–514.
74. Goldie I, Rosengren B, Moberg E, Hedelin E. Evaluation of Radiation Treatment of Painful Conditions of the Locomotor System: A Double Blind Study. Acta Oncol (Madr) 1970;9:311–322.
75. Valtonen EJ, Lilius HG, Malmio K. The value of roentgen irradiation in the treatment of painful degenerative and inflammatory musculoskeletal conditions. A double-blind study. Scand J Rheumatol 1975;4:247–249.
76. Ott OJ, Niewald M, Weitmann H-D, et al. DEGRO guidelines for the radiotherapy of non-malignant disorders. Strahlentherapie und Onkol 2014;191:1–6.

Supplementary file 1. Search strategy.

EMBASE

exp osteoarthritis/
OR
*arthrosis deformans/
OR
exp arthritis/
OR
osteoarthritis.af.
OR
degenerative arthritis.af.
OR
degenerative joint disease*".af.
OR
"*arthrosis deformans".af.
OR
"*arthritis*".af.

AND

exp radiotherapy/
OR
exp radiation/
OR
exp irradiation/
OR
radiotherapy.af.
OR
radiation therapy.af.
OR
"*radiation".af.
OR
irradiation.af.
OR
"Radiotherap*".af.
OR
X-ray therapy.af.

AND

Journal: Article.pt.

PubMed

Osteoarthritis

OR
degenerative arthritis
OR
degenerative joint disease*
OR
*arthrosis deformans
OR
arthritis

AND

radiotherapy
OR
radiation therapy
OR
*radiation
OR irradiation
OR
Radiotherap*
OR
"x-ray therapy"

Filters activated: Journal Article, Publication date from 1980/01/01
Web of Science

TS=osteoarthritis
OR
TS="degenerative arthritis"
OR
TS="degenerative joint disease*"
OR
TS=*arthrosis deformans
OR
TS=*arthr*is
OR
TS=arthritis
OR
TS=*arthritis
OR
TS=arthr*is

AND

TS=radiotherapy
OR
TS="radiation therapy"
OR

TS=*radiation*
OR
TS=radiotherapy*
OR
TS= "X-ray therapy"

AND

Document type: Article

Databases=SCI-EXPANDED, SSCI, A&HCI Timespan=1980-2014

Supplementary file 2. Study quality as scored using the EPHPP Quality Assessment Tool for Quantitative Studies.[21]

Study		A		Selection bias	B	Design	C		Confounders	D		Blinding	E		Data collection
		Q1	Q2	Section rating	Q1	Section rating	Q1	Q2	Section rating	Q1	Q2	Section rating	Q1	Q2	Section rating
Lindner[23]	1982	4	5	Weak	5	Moderate	3	4	Weak	1	1	Weak	2	3	Weak
Keinert[24]	1982	4	5	Weak	5	Moderate	3	3	Weak	1	1	Weak	2	2	Weak
Yaneva[25]	1986	4	5	Weak	5	Moderate	3	4	Weak	1	1	Weak	2	2	Weak
Gärtner[26]	1988	4	5	Weak	5	Moderate	3	4	Weak	1	1	Weak	2	2	Weak
Sautter-Bihl[27]	1993	2	5	Moderate	5	Moderate	3	4	Weak	1	1	Weak	2	2	Weak
Keilholoz[28]	1998	4	5	Weak	5	Moderate	3	4	Weak	1	1	Weak	1	1	Strong
Keller[30]	2013	4	5	Weak	5	Moderate	3	1	Weak	1	1	Weak	3	3	Weak

Continued; NA: not applicable

Study		F		Dropouts	G			Intervention	H				Analyses	Global rating
		Q1	Q2	Section rating	Q1	Q2	Q3	Section rating	Q1	Q2	Q3	Q4	Section rating	
Lindner[23]	1982	3	3	Weak	4	3	6	Weak	NA	Individual	2	2	Weak	Weak
Keinert[24]	1982	3	3	Weak	4	3	6	Weak	NA	Individual	2	2	Weak	Weak
Yaneva[25]	1986	3	4	Weak	4	3	6	Weak	NA	Individual	2	2	Weak	Weak
Gärtner[26]	1988	3	1	Strong	4	3	6	Weak	NA	Individual	2	2	Weak	Weak
Sautter-Bihl[27]	1993	1	3	Weak	1	3	6	Moderate	NA	Individual	2	2	Weak	Weak
Keilholoz[28]	1998	1	1	Strong	4	3	6	Moderate	NA	Individual	3	3	Weak	Weak
Keller[30]	2013	1	3	Weak	4	4	6	Weak	NA	Individual	2	2	Weak	Weak

Chapter 8

Effectiveness of low-dose radiation therapy on symptoms in patients with knee osteoarthritis: a randomised, double-blinded, sham-controlled trial

Elien A.M. Mahler*
Michiel J.M. Minten*
Mathilde M. Leseman-Hoogenboom
Philip M.P. Poortmans
Jan Willem H. Leer
Simone S. Boks
Frank H.J. van den Hoogen
Alfons A. den Broeder
Cornelia H.M. van den Ende

*both authors contributed equally

Submitted

Research in context

Evidence before this study

Osteoarthritis (OA) is a serious health problem, being the most prevalent chronic joint disease and one of the leading causes of pain and disability. Low-dose radiation therapy (LDRT) is widely used as OA treatment in some parts of the world. To identify the evidence for this treatment in clinical practice, we performed a systematic literature review, which was published in 2015. For this review, broad search terms were used to identify publications from 1980 and later, written in English, German, or Dutch, indexed in PubMed, EMBASE, and Web of Science. Primary inclusion criteria for studies were osteoarthritis as indication and radiotherapy as intervention, in title, abstract, and/or article text. Search terms used were osteoarthritis, degenerative joint disease, radiotherapy, and irradiation or synonyms. Seven clinical studies were suitable for inclusion, showing a reduction in pain in 25–90% and an improvement in functioning in 29–71% of the patients. The methodological quality of included publications was assessed independently by two researchers, using the EPHPP Quality Assessment Tool for Quantitative Studies. The methodological quality of all studies was judged as weak (no blinding, retrospective design, uncontrolled studies, and non-validated single-item outcome measures). Therefore, we concluded that there is insufficient high-level evidence available to indisputably demonstrate the effectiveness of LDRT in OA patients. In addition, two low-quality randomised controlled trials in OA patients were published in the 1970s and showed no effect of a higher dose radiation therapy than recommended in current guidelines. Therefore, high-quality research is needed, and we performed this randomised, double-blinded, sham-controlled trial using validated outcome measures.

Added value of this study

This is the first high-quality study. We found that treatment with LDRT does not lead to a substantial reduction of symptoms when compared with sham treatment. In addition, no changes in subclinical inflammatory signs assessed by ultrasound, magnetic resonance imaging, and serum inflammatory markers were observed.

Implications of all the available evidence

To the best of our knowledge, this is the first high-quality randomised controlled trial on the effectiveness of LDRT on symptoms and inflammatory signs in knee OA. We found no substantial reduction in symptoms after treatment with LDRT in knee OA. In view of the absence of other high-level quality evidence in favour of LDRT, we advise against its use as treatment for knee OA. Because LDRT is still widely used in some countries, future efforts should focus on de-implementation of LDRT, by changing beliefs of both patients and involved clinicians about the efficacy of LDRT.

Abstract

Objective

Low-dose radiation therapy (LDRT) for benign disorders such as knee osteoarthritis (OA) is widely used in some parts of the world, despite absence of controlled studies. We evaluated the effect of LDRT on symptoms and inflammation in knee OA patients.

Methods

In this randomised, double-blinded, sham-controlled trial (RCT), we recruited knee OA patients from the community and one outpatient clinic in the Netherlands. Key eligibility criteria were: age ≥ 50 years, fulfil clinical ACR knee OA criteria, pain score ≥ 5/10, and non-response to analgesics and exercise therapy. We randomly assigned patients 1:1 to receive LDRT (1 Gray per fraction) or sham intervention six times in 2 weeks, using a web-response system (block size 2,4,6 stratified by pain [>8 versus ≥8/10]. The primary outcome was the proportion of responders, according to the OMERACT-OARSI criteria, 3 months post-intervention. Secondary outcomes included inflammatory signs assessed by ultrasound, magnetic resonance imaging, and serum inflammatory markers. Patients and investigators, including those assessing outcomes, were masked to treatment assignment. Intention-to-treat analyses were performed. This trial is registered with the Dutch Trial Register, number NTR4574.

Results

From October 2015, through February 2017, we randomly assigned 55 patients: 27 (49%) to LDRT, and 28 (51%) to sham. At 3 months post-intervention, 12/27 patients (44%; 95% CI 26–63%) in the LDRT versus 12/28 patients (43%; 95% CI 25–61%) in the sham group responded; difference 2% (95% CI -25–28%, p=0·9), odds ratio adjusted for the stratifying variable was 1·1 (95% CI 0·4–3·2, p=0·9). No differences in any of the inflammatory signs were observed.

Conclusion

We found no substantial benefit of LDRT in knee OA patients, neither for the clinical symptoms, nor for the inflammatory signs. Therefore, based on this RCT and the absence of other high-quality evidence, we advise against the use of LDRT as treatment for knee OA.

Introduction

Osteoarthritis (OA) is considered to be the most prevalent chronic joint disease and is one of the leading causes of pain and disability worldwide, with the knee being the most frequently affected joint.[1,2] Since there is no disease-modifying treatment available, current knee OA management is symptomatic. However, in general, limited effect sizes for the non-surgical treatments and therapies of knee OA have been shown.3 When non-surgical treatments do not result in satisfactory reduction in symptoms, surgical options are often considered, but for many knee OA patients, total knee replacement (TKR) is not (yet) an option, considering the balance between the potential benefits and drawbacks. In general, TKR has good clinical outcomes.4 However, given the potential drawbacks with regard to the proportion of patients being dissatisfied after a TKR, the risks of complications, the limited lifespan of a prosthesis and poorer patient outcomes after revision arthroplasty, it is generally acknowledged that TKR should not be performed too early in the disease course.[4,5] Furthermore, it was recently shown that the current practice of total joint replacement in the U.S.A. had minimal effects on quality of life and quality-adjusted life years at the group level.[6] Therefore, there is a clear need for more effective non-surgical knee OA treatment options.

Although OA is considered a clinically non-inflammatory condition in general, subclinical synovial inflammation is prevalent, and it has been suggested to play an important role in the pathophysiology of the disease.[2,7,8] Furthermore, recent studies have suggested that synovitis in knee OA might play a more significant role than previously thought, since it is associated with pain and structural damage.[9-12] Thus, synovial inflammation may be a potential target for therapeutic approaches.

Previous in vitro and in vivo studies of OA in animal models have shown that low-dose radiation therapy (LDRT) exerts anti-inflammatory effects.[13] LDRT may be an additional non-surgical treatment option for knee OA patients in whom non-surgical interventions are insufficiently effective and for whom surgical treatments are not (yet) an option. LDRT is indeed widely used for benign disorders such as knee OA in some parts of the world, but this procedure is relatively unknown in other parts.[14,15] However, our recent systematic literature review showed that there is currently insufficient evidence available to demonstrate indisputably the effectiveness of LDRT in clinical practice, due to the absence of high-quality studies with a randomised design.[16] For that reason, we conducted a controlled trial primarily to evaluate the effectiveness of LDRT on symptoms in knee OA patients and, secondly, to examine the effects of LDRT on inflammatory signs.

Methods

Study design

This randomised, double-blinded, sham-controlled superiority trial (RCT) was performed in two centres in Nijmegen, the Netherlands. All visits and data collection, except the intervention, took place at the rheumatology and radiology outpatient clinics of the Sint Maartenskliniek, a hospital specialised in rheumatology, orthopaedics, and rehabilitation. The intervention was performed at the Department of Radiation Oncology of the Radboud university medical center. This study adhered to the Declaration of Helsinki, and national and international law and was approved Arnhem–Nijmegen ethics committee (study number 2014-275). The study was registered in the Dutch Trial Register (trial number NTR4574).

Patients

We enrolled patients from the rheumatology outpatient clinic of the Sint Maartenskliniek, and through advertisements in local newspapers (MM). Patients were eligible if they fulfilled the American College of Rheumatology (ACR) clinical knee OA criteria (knee pain >15 days of the previous month plus at least three of the following: morning stiffness <30 minutes, crepitus, bony tenderness, bony enlargement, or no palpable warmth). Other inclusion criteria were: (1) age ≥50 years; (2) a numeric rating scale (NRS) pain score ≥5 (scale 0–10) in the index knee during 15 of the previous 30 days; and (3) insufficient response to both analgesics and exercise therapy. Key exclusion criteria were: (1) treatment by a physical therapist in the previous 6 months; (2) previous surgical treatment of the knee or scheduled for surgical treatment in the following 12 months; (3) NRS pain score >2/10 in the contralateral knee or in one or both hips; (4) intramuscular or intra-articular corticosteroid injections received in the previous four weeks; (5) fibromyalgia according to the 2010 ACR diagnostic criteria or other rheumatic diseases; (6) Kellgren & Lawrence (K&L) score > 3; and (7) secondary knee OA due to trauma. Potential participants were initially screened using a standardized telephone questionnaire. Thereafter, screening for final eligibility was accomplished at the outpatient clinic by a rheumatologist (EM), or a physician assistant (VS), and researcher (MM). All patients gave written informed consent before study entry.

Patients were allowed to use analgesics, but they were encouraged not to change use during the study period. Patients were discouraged using corticosteroid injections or receiving active treatment by a physical therapist during the study period. However, if there was need for these treatments, they were allowed and the use was monitored.

Randomisation and masking

Included patients were randomly assigned 1:1, at the Department of Radiation Oncology, to the LDRT or sham intervention using a web-response system. An independent physicist, who was also involved with the dose calculation, generated the sequence and assigned participants to the interventions. To ensure balance in the possible confounder level of pain, assignment was stratified for intensity of pain score (NRS pain <8 versus ≥8/10) using stratified block randomisation (random block size of 2,4, or 6). The total process, including patient instructions, procedures, and marking of the target location, was accomplished before randomisation and was therefore performed blinded for patients and study personnel, including those assessing outcomes. After the randomization procedure, the unblinded radiotherapy technologist and radiotherapist were not involved in direct patient contact anymore. In the sham arm, the radiation therapy device was not activated, and these patients were exposed to a recording of a sound from an irradiating treatment machine.

Procedures

Patients visited the Radboud university medical center, department of Radiation Oncology, the Netherlands, seven times: once for preparations, and six times for the intervention. Before the first (sham) fraction was applied, the exact target locations of the index knee were marked by a trained radiotherapy technologist. The experimental intervention consisted of external beam LDRT with a total dose of 6 Gray, applied in six fractions of 1 Gray over 2 weeks according

to the 2015 consensus guidelines for radiation therapy of benign diseases of the German Society for Radio-oncology.[17] Fractions were scheduled every other day on weekdays. The sham intervention consisted of six fractions of 0 Gray. The intervention (LDRT or sham) was scheduled to start within 2 weeks after baseline measurement. Except for the fraction dose, all instructions and proceedings were identical for both groups.

All patients visited the Sint Maartenskliniek twice; a combined screening and baseline visit (T0), and a follow-up visit at 3 months post-intervention (T3). Before these visits, and at 1 and 2 months post-intervention (T1 and T2), a set of postal questionnaires was sent. Clinical parameters assessed by patient-reported outcome measures (PROMs) and inflammatory signs assessed by serum inflammatory markers, ultrasound (US), and magnetic resonance imaging (MRI) are described in detail below. In order to reduce random measurement error, questionnaires were administered twice at baseline and T3 within 7 to 14 days in between. Mean scores were used for analyses.

Demographic and OA-related characteristics were collected, using a standardized interview and physical examination. A radiograph of the index knee was taken (weight-bearing posterior-anterior fixed flexion) and scored blinded for clinical data, using the atlas-based K&L grading system, by a rheumatologist with ample experience in K&L scoring (EM).

Patient-reported outcome measures

Patients completed a standardized set of PROMs at baseline, T1, T2, and T3. Patients completed the Dutch Knee injury and Osteoarthritis Outcome Score (KOOS) questionnaire (Likert scale version). From this questionnaire, the Western Ontario and McMaster Universities (WOMAC) scores were derived, with WOMAC pain, function, and stiffness subscales presented as standardized scores ranging from 0 to 100, where higher scores reflect better health status. In addition, pain intensity and the patient global assessment (PGA) of knee OA impact during the previous week were measured on a 0–10 point numeric rating scale (NRS), where 0 equals no symptoms. Quality of life was measured by the physical and mental component scores (PCS, MCS) using corresponding subscales of the 36-item Short Form Health Survey (SF-36), standardized for general population (mean 50). The number of comorbidities according to the long version of the Dutch Arthritis Impact Measurement Scales was calculated (range 0–15). In addition, to take the patient acceptable symptom state (PASS) into account, patients were asked at T3 to think about the consequences of the knee OA during the previous week and to respond to the question: "If you were to remain for the rest of your life as you were during the last week, would the current state be acceptable or unacceptable for you?". Lastly, patients were asked about analgesic use and intra-articular corticosteroid injections during the previous month at baseline, T1, T2, and T3. Finally, patients filled out their presumption about the assigned treatment at T3, to estimate the quality of study blinding.

Inflammatory signs

Inflammatory signs were assessed at baseline and T3 by ultrasonography, MRI, and serum inflammatory markers i.e. the erythrocyte sedimentation rate (ESR, upper level women: 20 mm/h, men: 15 mm/h) and C-reactive protein (CRP, upper level 5 mg/l).

Ultrasound (Philips iU22 with a 50 mm linear transducer [frequency 5–12 MHz]) of the knee was performed by a researcher trained in musculoskeletal US (MM). The items measured were absolute synovial effusion (mm) and synovial thickness (mm) measured at suprapatellar and both medial and lateral parapatellar recesses (mean scores, based on protocol as described elsewhere. (BruynAR[18]).

Non-contrast-enhanced MRI of the index knee was performed at baseline and T3, using a 3.0 Tesla whole-body scanner (Philips Ingenia) utilising a dedicated phased–array 16ch dStream knee coil. Patients were in a supine position with the knee flexed. Sagittal and coronal PD SPAIR (repetition time [TR] 4130 ms and 5307 ms, respectively; echo time [TE] 30 ms; slice thickness [SL] 2•5 mm; slice gap 0•3 mm; turbofactor 17; field of view [FOV] 180x180x110 mm and 180x150x99 mm, respectively; matrix 580x421 and 580x373, respectively) sequences were performed, followed by axial PD (TR 4931 ms; TE 30 ms; SL 2•5 mm; slice gap 0•3 mm; turbofactor 17; FOV 160x160x154 mm; matrix 532x377) and axial T1 (TR 664 ms; TE 15 ms; SL 2•5 mm; slice gap 0•8 mm; turbofactor 6; FOV 160x160x154 mm; matrix 384x316) sequences. Synovial membrane and Hoffa's fat pad imaging was performed using the sagittal PD SPAIR and axial PD sequences. Three inflammatory signs were assessed semiquantitatively (all graded 0–3, with 0 representing normal situation), using validated scoring systems: (1) synovitis assessed as synovial membrane thickness at four regions: medial and lateral recesses, and medial and lateral suprapatellar bursa. Scores from four locations were summed to a maximum of 12[19], (2) effusion-synovitis, for a maximum of 3[20], and (3) Hoffa's fat pad synovitis, for a maximum of 3 in line with the MRI Osteoarthritis Knee Score scoring system.20 The reading was performed blinded by one musculoskeletal radiologist (SB) with ample experience in standardized semiquantitative assessment of knee OA.

Outcomes

The primary outcome was the proportion of responders at T3, according to the OMERACT-OARSI responder criteria (either relative improvement in pain or function ≥50% and an absolute improvement of ≥20/100 points, or two of the following: pain, function or patient's global assessment [relative improvement ≥20% and ≥10/100 points absolute for pain and function or ≥1/10 point absolute for PGA]).[21]
Secondary outcomes were clinical parameters (WOMAC pain, function, and stiffness, PGA, NRS pain, quality of life) and inflammatory signs assessed by imaging and inflammatory markers. The proportion of responders at T1 and T2 was also assessed.

Adverse events (AEs) and serious adverse events (SAEs) were identified by the researcher at the follow-up visits and were registered according to the Common Terminology Criteria for Adverse Events (CTCAE) version 4.0. Special attention was paid to skin and subcutaneous tissue disorders.

Statistical analysis

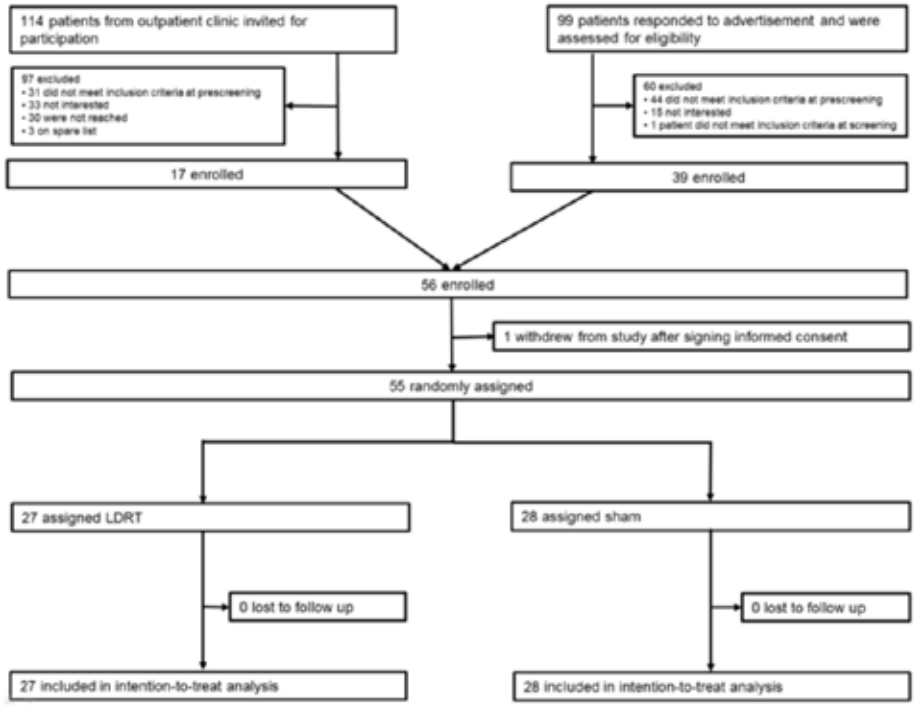
A potential role for LDRT in clinical practice is only warranted if this treatment modality results in large improvements in the relief of symptoms, with low numbers needed to treat. Therefore, this superiority study was powered to detect a large, short- to-medium effect of LDRT in patients with knee OA. The following assumptions were made: (1) a difference of 40% in the proportion of responders, according to the OMERACT-OARSI responder criteria between the LDRT group and the sham group, (2) the proportion of responders in the sham group being 40%, and (3) a power of 80% and a one-sided alpha 5% level. Allowing for dropouts we aimed to include 27 persons per group.

8

The analyses were performed blinded for assignment. Intention-to-treat primary analysis was performed to estimate the difference in proportion of responders between baseline and T3 between the two treatment groups, and 95% confidence intervals (CIs) were calculated. To adjust for the stratifying variable (NRS pain <8 versus ≥8/10), we performed logistic regression for the primary analysis and linear or logistic regression when applicable for secondary analyses, corrected for stratum. In addition, we performed sensitivity analysis adjusting for potential imbalanced confounding variables. All analyses were performed using STATA 13.1.

Results

Figure 1. Trial profile.



LDRT = low-dose radiation therapy

From October 2015 through February 2017, 55 of the 56 eligible patients were enrolled: 27 in the LDRT group and 28 in the sham group (see figure 1); one withdrew from the study after inclusion but before randomisation because an abnormality was detected during the nationwide breast cancer screening. We included 55 patients, instead of the 54 patients as planned before, because the last two patients were planned for screening on the same day and both were found to be eligible. Twenty-eight patients (51%) were female, mean age 65 years (standard deviation (SD) 9), median body mass index 27 (interquartile range (IQR) 24–31) kg/m², and the patients were moderately to severely disabled by their disease considering the scores for pain, function, and PGA (table 1). Baseline characteristics were similar between the two groups, except for a slightly lower mean age, higher median BMI, worse PGA, and higher

proportion of patients with ESR above upper limit in the LDRT group (table 1). During the month prior to baseline, 19 (70%) and 16 (57%) patients, respectively, used analgesics in the LDRT and sham groups. Three patients (14%) in the LDRT group and four patients (11%) in the sham group were included in the stratum with an NRS pain ≥8/10. The majority (n=39; 71%) was recruited by advertisement, and these patients showed a slightly better WOMAC pain, but comparable NRS pain, a slightly worse PCS, comparable K&L scores, and were more often male (54% versus 38%) than the patients recruited from the outpatient clinic (difference mean WOMAC pain 11; 95% CI 2·5–20, and normalised PCS 6·0; 95% CI 1·8–10).

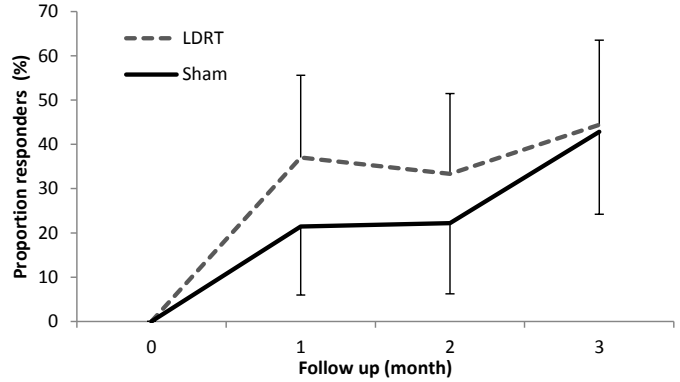
Table 1. Baseline characteristics of 55 knee patients with osteoarthritis.

	LDRT group (n=27)	Sham group (n=28)
Sociodemographic characteristics		
Age (years), mean (SD)	62 (9)	68 (9)
Sex		
Female, n (%)	15 (56%)	13 (46%)
Male, n (%)	12 (44%)	15 (54%)
BMI (kg/m²), median (IQR)	29 (25–30)	26 (24-31)
Duration of symptoms ≤ 5 years, n (%)	16 (59)	14 (52)#
Kellgren and Lawrence ≥ 2, n (%)	15 (56)	17 (61)
Comorbidities ≤ 1, n (%)	19 (70)	19 (68)
Clinical parameters included in the primary outcome		
WOMAC pain (0-100), mean (SD)	59 (14)	61 (17)
WOMAC function (0-100), mean (SD)	60 (17)	62 (19)
NRS PGA (0-10), mean (SD)	5·6 (2·2)	4·6 (2·3)
Other clinical parameters		
WOMAC stiffness (0-100), mean (SD)	47 (13)	55 (20)
NRS pain (0-10), mean (SD)	5·8 (1·6)	5·4 (1·6)
SF-36 mental component scale, mean (SD)	53 (10) ‡	52 (10)
SF-36 physical component scale, mean (SD)	39 (7)	39 (8)
Inflammatory signs		
Ultrasound		
Sum of synovial thickness (mm), median (IQR)	2 (1-3)	2 (1-3)
Sum of synovial effusion (mm), median (IQR)	6 (4-8)	5 (4-7)
MRI		
Sum of effusion-synovitis (0-12), median (IQR)	5 (3-8)	4 (2-7)
Synovial effusion-synovitis (0-3), median (IQR)	2·0 (1·0-2·0)	1·0 (1·0-2·0)
Hoffa's fat pad synovitis (0-3), median (IQR)	1·0 (1·0-2·0)	1·0 (1·0-2·0)
Serum inflammatory markers		
ESR, above upper limit, n (%)	10 (37)	4 (14)
CRP, above upper limit, n (%)	5 (19)	7 (25)

Data are shown as mean (SD), median (IQR), or n (%) Higher scores indicate better scores for WOMAC pain and function and stiffness, worse NRS PGA, worse NRS pain, better mental and physical component scales (reference U.S.A. population = 50), more synovitis (ultrasound and MRI). #1 missing value, ‡2 missing values LDRT=low-dose radiation therapy. SD=standard deviation. IQR=interquartile range. BMI=body mass index. IQR=interquartile range. WOMAC=pain, function and stiffness, Western Ontario and McMaster Universities Osteoarthritis Index scale. NRS=numeric rating scale. PGA=patient global assessment. SF-36=36-item short Form Health Survey. MRI=magnetic resonance imaging. ESR=Erythrocyte sedimentation rate. CRP=C-reactive protein.

All 55 randomised patients completed the study with very good adherence to (sham) treatment; one patient in the LDRT group discontinued prematurely because of severe back pain after a collapse hampering further LDRT after two fractions. Five percent of the data were re-entered, and differences with the original database were checked (<0•1%). Almost no data were missing (mean 1•1% per item, range 0•5–1•8%, and 0•5% of the planned visits; (imaging and laboratory markers are lacking from one patients at T3 but not the questionnaires); thus data imputation was deemed unnecessary. There were no differences in missing levels between the groups.

Figure 2. Proportion of responders over 3 months, with its 95% confidence interval, for LDRT and sham intervention.



LDRT=low-dose radiation therapy

At 3 months post-intervention, 12/27 (44%, 95% CI 26–63%) patients in the LDRT group and 12/28 (43%, 95% CI 25–61%) in the sham group met the primary outcome, i.e. OMERACT-OARSI responder criteria; this resulted in a difference of 2% (95% CI -25 to 28%, p=0•9, table 2, figure 2). Logistic regression for response adjusting for the stratified variable (NRS ≥8/10) yielded similar results with an odds ratio (OR) of 1•1 (95% CI 0•4–3•2, p=0•9) for the LDRT group compared with the sham group. Subsequently, we cannot reject our null hypothesis of no effect of the treatment, and the CIs around the between group difference show that a difference in effectiveness of LDRT versus sham treatment of over 28% is highly unlikely. Sensitivity analyses adjusting for potential confounders (age, BMI, PGA) yielded similar results (OR 1.3; 95% CI 0•4–4•2 for the LDRT group compared with the sham group).

Table 2. Proportion of responders of the two groups, the difference in proportion between groups and the odds ratio (with 95% confidence intervals) according to the OMERACT-OARSI criteria.[21]

Time after intervention	LDRT group	Sham group	Difference in proportion	Odds ratio for LDRT
1 month	37 (19-55)	21 (6-37)	16 (-8-39)	2.3 (0•7-7•5)
2 months	33 (16-51)	22 (9-42)	11 (-13-35)	1.8 (0•5-6•3)
3 months	44 (26-63)	43 (25-61)	2 (-25-28)	1.1 (0•4-3•2)

LDRT=low-dose radiation therapy

No significant differences were found between the two groups regarding WOMAC pain, function, and PGA (the domains of the primary outcome) at months 1,2, and 3 post-intervention (T3 data are shown in table 3). No differences in any other secondary outcome including the inflammatory signs measured by imaging or inflammatory markers were observed (table 3, figures 3 and 4).

The number of responders at T1 and T2 are shown in table 2 and figure 2. Analgesics during the third follow-up month were used in 15 (56%) and 13 (43%) patients in the LDRT group and the sham group, respectively. One patient received an intra-articular injection during follow-up (sham group second month post-intervention). This patient was a non-responder at 3 months post-intervention. Seventeen patients (63%, 95% CI 44–78%) in the LDRT group versus 23 patients (82%, 95% CI 64–92%) in the sham group reported being in PASS at T3. The number of patients who thought they had received and actually had received LDRT was comparable: 10/25 (40%, 2 missing values) and 12/26 (46%, 2 missing values) for the LDRT group and sham group, respectively, demonstrating adequate blinding.

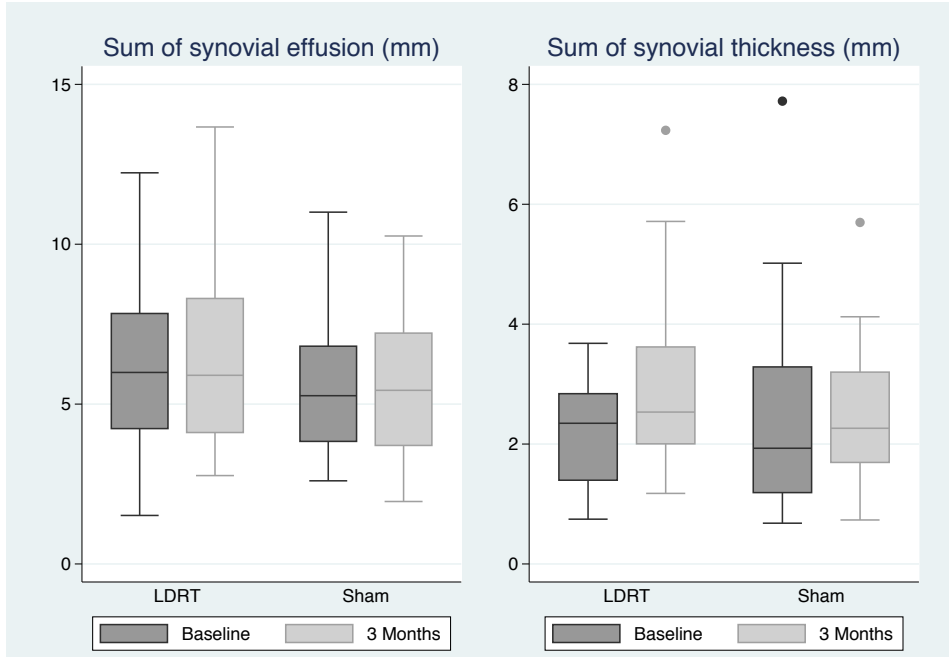
The occurrence of both AEs and serious adverse events SAEs was comparable between the two groups. Two SAEs in the sham group were observed: colon carcinoma was diagnosed post-intervention in two patients, which we expect not to be related to the intervention. Three AEs occurred: one LDRT patient suffered from a collapse as mentioned above, one patient in the sham group experienced severe knee pain during and after the intervention, and one patient in the sham group experienced cold sensations in the lower index leg. Local reactions were comparable in both groups. Fatigue was recorded in six (22%) patients versus three patients (11%) in the LDRT group and sham group, respectively.

Table 3. Absolute changes (SD or IQR where appropriate) of clinical parameters and inflammatory signs per group and differences between changes expressed as odds ratio (OR) or ß-coefficient (95% CI), where appropriate for the LDRT group.

	Change (baseline – 3 months)		OR or ß-coefficient (95% CI) for
	LDRT group	**sham group**	**LDRT group**
Clinical parameters included in the primary outcome			
WOMAC pain (range 0-100), mean (SD)	8 (13)	11 (14)	-3 (-10-4)
WOMAC function (range 0-100), mean (SD)	9·7 (8)#	6.3 (14)	4 (-3-10)
PGA (range 0-10), mean (SD)	-1·0 (2)	-0·9 (3)	0 (-1-1)
Clinical parameters not included in the primary outcome			
WOMAC stiffness (0-100), mean (SD)	-11 (19)	9 (21)	2 (-8 to 13)
NRS pain (0-10), mean (SD)	-1·1 (1·6)	-1·3 (2·4)	0·1 (-0·9-1·2)
SF36 mental component scale (normalised to 50), mean (SD)	0·9 (8·4)‡	-4·2 (10)#	5 (0 to 10)
SF36 physical component scale (normalised to 50), mean (SD)	0·1 (7·0)‡	2·4 (6·9)#	-2 (-6-2)
Inflammatory signs			
Ultrasound			
Sum of synovial thickness (mm), median (IQR)	0·7 (1·3)	0·1 (1·2)#	0·6 (-0·1-1·3)
Sum of synovial effusion (mm), median (IQR)	0·4 (1·8)	0·1 (1·1)#	0·3 (-0·5-1·1)
MRI			
Sum of effusion-synovitis (0-12), median (IQR)	-0·1 (0·7)	0·0 (0·7)#	-0·0 (-0·4-0·4)
Synovial effusion-synovitis (0-3), median (IQR)	0·0 (0·4)	-0·1 (0·4)#	0·0 (-0·2-0·3)
Hoffa's fat pat synovitis (0-3), median (IQR)	-0·1 (0·6)	0·0 (0·4)#	-0·1 (-0·4-0·2)
Serum inflammatory markers			
ESR, above upper limit, n (%, 95%CI)	-3 (-11 (-24-2))	5 (18 (3-33))	1 (0-2)†
CRP, above upper limit, n (%)	0 (0 (-11-11))	0 (0 (-22-22))#	1 (0 to 2)†

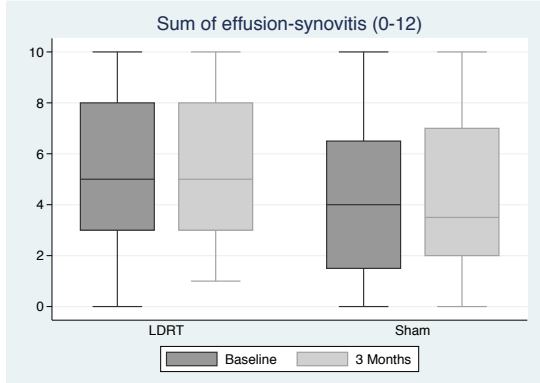
Change values between baseline and 3 months are presented as mean changes (SD), median (IQR), or n (%)
Pain and function are presented as WOMAC, range 0-100 where 100 equals no symptoms;
PGA is presented as NRS, scale 0-10 where 0 equals no symptoms
Positive changes (columns 1 and 2) indicate less WOMAC pain, better function, worse PGA, less stiffness, more NRS pain, better MCS and PCS, increase in ultrasound synovial thickness, and effusion, increase in MRI synovitis, increase in number of patients above upper limit (ESR and CRP).
Positive differences (column 3) indicate less pain, better function, worse PGA, increase in synovial thickness, effusion, ESR, CRP of LDRT group compared to sham group.
Differences between groups is estimated by linear or logistic regression when appropriate, adjusted for stratification of NRS ≥8/10
1 missing value
‡ 2 missing values
† OR for ESR or CRP above upper limit (yes/no) at 3 months post-intervention
SD=standard deviation. IQR=interquartile range. OR=odds ratio. CI=confidence interval. LDRT=low-dose radiation therapy. WOMAC=Western Ontario and McMaster Universities Osteoarthritis Index scale. PGA=patient global assessment NRS= numeric rating scale. SF-36=36-item Short Form Health Survey. MRI=magnetic resonance imaging. ESR=Erythrocyte sedimentation rate. CRP=C-reactive protein. MCS=mental component score. PCS=physical component score.

Figure 3. Median (IQR) sum scores of synovial effusion and thickness (mm) at baseline and 3 months post-intervention.



IQR=interquartile range. LDRT=low-dose radiation therapy.

Figure 4. Median (IQR) sum scores of effusion-synovitis (0-12) at baseline and 3 months post-intervention.



IQR=interquartile range. LDRT=low-dose radiation therapy.

Discussion

To our knowledge, we performed the first RCT to evaluate the effectiveness of LDRT in knee OA patients with the radiation dose as recommended in current guidelines.[17] We showed that treatment with LDRT, compared with sham, does not lead to a substantial reduction of symptoms. Considering the limits of the 95% CIs, a difference exceeding 28% of responders between groups seems unlikely. We also found no differences in changes of pain, function, PGA between the LDRT and sham groups. In addition, we found no substantial impact on imaging or laboratory inflammatory signs.

Our study has a number of strengths. These include the randomised, sham-controlled design, blinding of both patients and study assessors, use of validated outcome measures, and use of a well-defined patient population. Furthermore, the number of patients needed according to our sample size calculation was met, and both lost to follow-up and missing data were low.

Some methodological choices can be challenged. Firstly, the pre-specified 40% difference margin can be considered relatively large. However, this seems clinically justifiable, because we felt that LDRT could have a place in clinical practice only when its effect would outweigh the time investment, patients' burden, radiation exposure, and costs. As a result, we cannot rule out the existence of a small effect of LDRT in knee OA. However, considering the limits of the 95% CI of our results, a difference exceeding 28% of responders between groups seems unlikely. This limited sample size could, despite randomisation, also have resulted in imbalance of potential confounders between the two groups. However, adjusted analyses for confounding yielded similar results.

It can be argued that the absence of effect of LDRT in our study reflects a poor choice of patient population, treatment, or outcome measures. However, we did include the relevant patient population, being patients with established knee OA, considering the quite severe baseline symptoms, the K&L scores, and physically impaired but not mentally impaired quality of life. In addition, the dose of the LDRT is comparable with that used in previous studies and that currently recommended in clinical practice.[16,17] Also, a follow-up of 3 months seems adequate, as short- to-medium term effects were to be expected. A valid point of criticism could be the low-to-moderate sensitivity and specificity of the OMERACT-OARSI responder criteria[21] as this could have led to misclassification and, subsequently, underestimation of the effectiveness of LDRT. However, these are the best and most often used clinical outcome measures in knee OA intervention studies. Considering also the lack of effect of LDRT in all other outcome measures, we consider our results very robust.

How should our results be interpreted in the view of existing evidence on LDRT in knee OA treatment? There have been several studies on LDRT in knee OA showing improvement of pain and/or function; however all of them suffered from methodological shortcomings, ie, uncontrolled and/or retrospective design without blinding, and non-validated single-item outcome measures.[16] Therefore, we concluded recently in a systematic literature review that there is insufficient high-level evidence for a positive effect of LDRT on pain and functioning in OA patients.[16] In addition, two low-quality RCTs published in the 1970s, relating to patients suffering from a range of painful locomotor ailments including knee OA patients, showed no effect of radiation therapy as used at that time with a relatively high dose.[22,23] Of note, we found

a substantial 3 months response of 40% in both groups, illustrating the substantial effect of mainly a placebo effect and regression to the mean. In view of the absence of other high-level quality evidence in favour of LDRT, we hypothesise that these two effects are also responsible for the previously reported improvements of LDRT on symptoms in studies suffering from several methodological shortcomings. This is in accordance with previous research, and in particular for rather invasive interventions such as LDRT, which are associated with higher placebo effects.[24] In conclusion, we consider our results as valid, in contrast to previous clinical studies.

The external generalisability of our findings is strengthened by the similarity of baseline characteristics between current patients and patients previously included in previous OA research. However, selectivity of the sample could have influenced this external generalisability, given the relatively high proportion of men, as well as better baseline pain and physical quality of life of patients recruited by newspaper advertisement compared with outpatients. Nevertheless, this mixture of recruitment strategies attracts a more heterogeneous group of OA patients and has previously shown not to influence the efficacy of the intervention and even increases the external generalisability.[25,26]

In addition to the absence of clinical response, we also found no substantial impact on inflammatory signs assessed by ultrasound, MRI and serum inflammatory markers. We used validated MRI and US scores, which strengthens the internal validity of these findings.[18-20] However, several weaknesses regarding these secondary outcomes should be mentioned. Firstly, because our study was primarily not powered to detect substantial differences in inflammatory signs, we can only state that LDRT did not result in large differences between the two groups. Secondly, it can be debated whether we should have selected patients with a minimal threshold of inflammation at baseline. We decided not to, because an inflammatory phenotype is not well defined or validated, and because synovitis is known to fluctuate during the disease course[11], and the majority of knee OA patients have been shown to have inflammatory signs anyway.[27,28] The last was also seen in our patients. In addition, our baseline values of inflammatory signs are similar to those of previous studies that selected patients with signs of synovial inflammation.[19,20] Also, additional analyses comparing knee OA patients with and without inflammatory signs at baseline yielded similar results (data not shown). Thirdly, in general, the gold standard method for detecting synovitis is histological analysis of samples obtained by biopsy. However, non-invasive imaging techniques, including US and MRI, are reported to perform well when correlating them with histological observations of inflammation in OA.[7,19,20,29]

In conclusion, we were not able to show an effect of LDRT treatment on symptoms and inflammatory signs in knee OA. In view of the absence of other high-level quality evidence in favour of LDRT, we advise against its use as treatment for knee OA. Because this treatment is still widely used in some countries, future efforts should focus on de-implementation of LDRT, by changing the beliefs of involved clinicians and health professionals about the efficacy of LDRT that are not based on scientific grounds. Additionally, it is important that future research should also focus on the quality of the scientific evidence of LDRT treatment for other benign (musculoskeletal) disorders, for which high-quality studies are also lacking.[30]

8

Acknowledgement
We are indebted to all patients who participated in the study, and Vincent HHP Straten, physician assistant, for helping to screen patients.

References

1 Deshpande BR, Katz JN, Solomon DH, et al. Number of persons with symptomatic knee osteoarthritis in the US: impact of race and ethnicity, age, sex, and obesity. Arthritis Care Res (Hoboken) 2016; 68: 1743–50.
2 Glyn-Jones S, Palmer AJ, Agricola R, et al. Osteoarthritis. Lancet 2015; 386: 376–87.
3 McAlindon TE, Bannuru RR, Sullivan MC, et al. OARSI guidelines for the non-surgical management of knee osteoarthritis. Osteoarthritis Cartilage 2014; 22: 363–88.
4 Carr AJ, Robertsson O, Graves S, et al. Knee replacement. Lancet 2012; 379:1331–40.
5 Beswick AD, Wylde V, Gooberman-Hill R, Blom A, Dieppe P. What proportion of patients report long-term pain after total hip or knee replacement for osteoarthritis? A systematic review of prospective studies in unselected patients. BMJ Open 2012; 22; 2:e000435.
6 Ferket BS, Feldman Z, Zhou J, Oei EH, Bierma-Zeinstra SM, Mazumdar M. Impact of total knee replacement practice: cost effectiveness analysis of data from the Osteoarthritis Initiative. BMJ 2017; 356: j1131.
7 Sellam J, Berenbaum F. The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. Nat Rev Rheumatol 2010; 6: 625–35.
8 Robinson WH, Lepus CM, Wang Q, et al. Low-grade inflammation as a key mediator of the pathogenesis of osteoarthritis. Nat Rev Rheumatol 2016; 12: 580–92.
9 Yusuf E, Kortekaas MC, Watt I, Huizinga TW, Kloppenburg M. Do knee abnormalities visualised on MRI explain knee pain in knee osteoarthritis? A systematic review. Ann Rheum Dis 2011; 70: 60-7.
10 Atukorala I, Kwoh CK, Guermazi A, et al. Synovitis in knee osteoarthritis: a precursor of disease? Ann Rheum Dis 2016; 75: 390-5.
11 de Lange-Brokaar BJ, Ioan-Facsinay A, Yusuf E, et al. Evolution of synovitis in osteoarthritic knees and its association with clinical features. Osteoarthritis Cartilage 2016; 24: 1867–74.
12 Krasnokutsky S, Belitskaya-Lévy I, Bencardino J, et al. Quantitative magnetic resonance imaging evidence of synovial proliferation is associated with radiographic severity of knee osteoarthritis. Arthritis Rheum 2011; 63: 2983–91.
13 Arenas M, Sabater S, Hernández V, et al. Anti-inflammatory effects of low-dose radiotherapy. indications, dose, and radiobiological mechanisms involved. Strahlenther und Onkol 2012; 188: 975–81.
14 Leer JWH, Van Houtte P, Davelaar J. Indications and treatment schedules for irradiation of benign diseases: a survey. Radiother Oncol 1998; 48: 249–57.
15 Seegenschmiedt MH, Micke O, Willich N. Radiation therapy for nonmalignant diseases in Germany: current concepts and future perspectives. Strahlenther und Onkol 2004; 180: 718–30.
16 Minten MJM, Mahler E, den Broeder AA, et al. The efficacy and safety of low-dose radiotherapy on pain and functioning in patients with osteoarthritis: a systematic review. Rheumatol Int 2016; 36: 133–42.
17 Ott OJ, Niewald M, Weitmann H-D, et al. DEGRO guidelines for the radiotherapy of non-malignant disorders. Strahlenther und Onkol 2015; 191: 1–6.
18 Bruyn GA, Naredo E, Damjanov N et al.; Ultrasound Task force. An OMERACT reliability exercise of inflammatory and structural abnormalities in patients with knee osteoarthritis using ultrasound assessment. Ann Rheum Dis 2016; 75: 842-6.
19 Pelletier JP, Raynauld JP, Abram F, Haraoui B, Choquette D, Martel-Pelletier J. A new non-invasive method to assess synovitis severity in relation to symptoms and cartilage volume loss in knee osteoarthritis patients using MRI. Osteoarthritis Cartilage 2008; 16 (suppl 3): S8-13.
20 Hunter DJ, Guermazi A, Lo GH, et al. Evolution of semi-quantitative whole joint assessment of knee OA: MOAKS (MRI Osteoarthritis Knee Score). Osteoarthritis Cartilage 2011; 19: 990-1002.
21 Pham T, van der Heijde D, Altman RD, et al. OMERACT-OARSI initiative: Osteoarthritis Research Society International set of responder criteria for osteoarthritis clinical trials revisited. Osteoarthritis Cartilage 2004; 12: 389–99.

22 Goldie I, Rosengren B, Moberg E, et al. Evaluation of radiation treatment of painful conditions of the locomotor system: a double blind study. Acta Oncol (Madr) 1970; 9: 311–22.
23 Valtonen EJ, Lilius HG, Malmio K. The value of roentgen irradiation in the treatment of painful degenerative and inflammatory musculoskeletal conditions. A double-blind study. Scand J Rheumatol 1975; 4: 247–9.
24 Bannuru RR, McAlindon TE, Sullivan MC, Wong JB, Kent DM, Schmid CH. Effectiveness and implications of alternative placebo treatments: a systematic review and network meta-analysis of osteoarthritis trials. Ann Intern Med 2015; 163: 365-72.
25 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: 312–20.
26 Veenhof C, Dekker J, Bijlsma JW, van den Ende CH. Influence of various recruitment strategies on the study population and outcome of a randomized controlled trial involving patients with osteoarthritis of the hip or knee. Arthritis Rheum 2005; 53: 375–82.
27 Bevers K, Bijlsma JW, Vriezekolk JE, van den Ende CH, den Broeder AA. The course of ultrasonographic abnormalities in knee osteoarthritis: 1 year follow up. Osteoarthritis Cartilage 2014; 22: 1651–6.
28 Roemer FW, Kassim Javaid M, Guermazi A, et al. Anatomical distribution of synovitis in knee osteoarthritis and its association with joint effusion assessed on non-enhanced and contrast-enhanced MRI. Osteoarthritis Cartilage 2010; 18: 1269–74.
29 Crema MD, Roemer FW, Li L, et al. Comparison between semiquantitative and quantitative methods for the assessment of knee synovitis in osteoarthritis using non-enhanced and gadolinium-enhanced MRI. Osteoarthritis Cartilage 2017; 25: 267–71.
30 McKeown SR, Hatfield P, Prestwich RJ, Shaffer RE, Taylor RE. Radiotherapy for benign disease; assessing the risk of radiation-induced cancer following exposure to intermediate dose radiation. Br J Radiol 2015; 88: 20150405.

Supplementary file 1. Treatment protocol.

For the LDRT group, the radiation therapy regimen consisted of six fractions of 1 Gray, with a total dose of 6 Gray, from a linear accelerator (Agility, Elektra AB, Stockholm, Sweden) according to the consensus guidelines for radiation therapy of benign diseases of the German Society for Radio-oncology (DEGRO)[1]. Fractions were delivered every other weekday over a two-week period. The clinical target volume included the knee from 4.5 cm proximal of the patella to 1 cm distal of the tuberositas (figure 1). A treatment plan was calculated in a virtual water phantom using our treatment planning system (Pinnacle v.8.0h, Philips Medical Systems, Andover, MA), dose specified at the center of the phantom using two opposing 6-MV photon beams. Doses were individualised using the patients' individual measurements in three dimensions (length, width, depth). As in regular radiation therapy practice, all treatment programs were double checked by a second radiotherapist. Prior to the first fraction, the exact target locations were marked using wear-resistant ink (figure 2). During treatment, video screens that normally display treatment information were switched off to mask the patient for the treatment they received.
The sham-group received six 0 Gy fractions over a period of two weeks. During sham treatment, a sound mimicking linear accelerator sound was played, as the device was not activated. Furthermore, all instructions and proceedings were identical for both groups.

1. Ott OJ, Niewald M, Weitmann H-D, et al. DEGRO guidelines for the radiotherapy of non-malignant disorders. Strahlenther und Onkol 2015;191:1–6.

Figure 1. Target field for the knee.

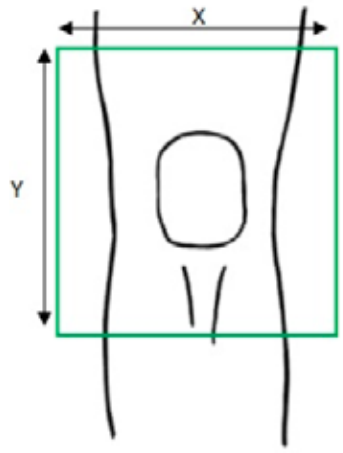
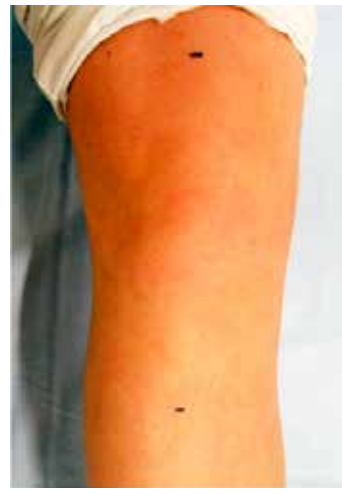


Figure 2. Treatment setup with target field markings.



Chapter 9

Summary and general discussion

Summary

Osteoarthritis (OA) is a multifactorial joint disease affecting all the tissues of the joint, and is characterized by cartilage breakdown, the formation of bony outgrowths at the joint margin (osteophytes), subchondral bone sclerosis, alterations to ligaments and muscles, and inflammation of the synovial membrane. OA is a serious health problem. It entails a high clinical burden of pain and disability, and a reduced quality of life. No cure is available for OA and treatment focuses primarily on the reduction of symptoms as pain and loss of function. Treatments encompass nonsurgical treatment modalities (such as such as education, exercise, step up analgesics, life style advices including weight loss in patients that are overweight) and surgical treatment modalities (such as total joint replacement (TJR)). Several international recommendations emphasize the importance and efficacy of starting with nonsurgical treatment modalities, before considering surgical treatment modalities. However, in general, limited effect sizes for the nonsurgical treatments have been shown. Nevertheless, nonsurgical treatment modalities are insufficiently utilized, and the number of surgeries is rising substantially. Moreover, a large variability in the surgeons' indication for TJR exists. Furthermore, TJR is considered an effective treatment, but total knee replacement is less effective compared to hip replacement and up to 25-30% of patients are dissatisfied after knee replacement. Finally, the prevalence of OA is high and rising, mainly concurrent with an ageing population and the growing obesity epidemic. Especially in developed countries, this leads to a growing impact on health care and future socioeconomic costs.

Therefore, there is a clear need to improve the management of OA. In addition, The aim of this thesis is to contribute to the body of knowledge regarding the non-surgical management of established knee and hip OA by 1) identifying appropriate outcome measures, 2) identifying risk factors, and 3) establishing evidence of new effective non-surgical treatment options. In this final chapter, the main findings of the studies comprised in this thesis are summarized and methodological considerations and insights gained are discussed.

Identifying appropriate outcome measures

In **chapter 2** we evaluated the responsiveness (i.e. the ability to detect changes over time) of four widely used patient reported outcome measures (PROMs) for the assessment of physical function in patients with symptomatic knee OA receiving optimized non-surgical treatment i.e. Knee Injury and Osteoarthritis Outcome Function Short Form (KOOS-PS), Lequesne Algofunctional index (LAI), Lower Extremity Functional Scale (LEFS) and Western Ontario and McMaster Universities Osteoarthritis Index Physical Function subscale (WOMAC-PF). Responsiveness was investigated by testing predefined hypotheses formulated by an expert group about expected (or absence of) correlations between changes in in physical function with changes in other (un)related measurements. Responsiveness was considered positive if >75% of the hypotheses could be confirmed. We could confirm 12 out of 15 predefined hypotheses (80%) about expected correlations using the physical function subscale of the WOMAC. For the KOOS-PS, LAI and LEFS, respectively 73%, 67%, and 73% of hypotheses could be confirmed. Our findings suggest that the WOMAC-PF is able to detect changes over time in physical function in patients with knee OA receiving standardized non-surgical treatment. We therefore recommend that WOMAC-PF should be the measure of first choice in clinical trials evaluating the effectiveness of an intervention on physical function in clinical knee OA patients. These results enabled us to make a well-considered choice for a PROM assessing

physical function in our randomized controlled trial evaluating the effectiveness of LDRT of symptoms in knee OA (chapter 8).

Main finding I: The WOMAC-PF is able to detect changes over time in physical function in patients with knee OA receiving standardized non-surgical treatment.

To incorporate the patients' perception in the interpretation of PROMs assessing physical function, we conducted a prospective study to estimate the patient acceptable symptom state (PASS) values (**chapter 3**). In addition, we assessed the influence of sex, age, duration of symptoms, and having depressive feelings on the estimated PASS value. Patients completed the short version of the Knee Injury and Osteoarthritis Outcome Score (KOOS-PS), Lequesne Algofunctional Index (LAI), the Lower Extremity Functional Scale (LEFS), and the physical function subscale of the WOMAC (WOMAC-PF). The PASS is considered a state and is defined as the highest level of symptoms that the majority of patient consider acceptable. PASS values were defined as the 75th percentile of the score of questionnaires for those patients who consider their state acceptable. We showed that the cut-off values with their 95% confidence intervals for the PASS for physical function were comparable across the four different standardized PROMs. We therefore concluded that PASS cut-off values in patients with clinical knee OA are robust across different PROMs assessing physical function and we suggest that different questionnaires may be used to determine PASS values for physical function of a certain population and setting. In addition, we observed that female and depressive patients have a lower chance of reaching the estimated PASS value. However, we observed also that in our knee OA cohort, patients consider a higher level of symptoms acceptable than previously reported for other rheumatic diseases. Probably, the use of a generic PASS value for different domains in patients with OA or even patients with other rheumatic diseases might not be justifiable.

Main finding II: PASS cut-off values in patients with clinical knee OA are robust across different PROMs assessing physical function.

Identification of risk factors

In **chapter 4**, we describe the development and validation of preliminary criteria for clinical worsening in knee and hip OA using a literature, expert opinion and data driven approach. The Expert Group's review of the literature yielded five sets of worsening criteria. In addition, this Expert Group reached consensus on 10 sets of newly proposed worsening criteria to be tested in observational datasets of patients with knee or hip OA. These sets included 219 patients (derivation cohort) and 296 patients (validation cohort). Newly proposed worsening criteria were first tested in the derivation cohort. Then, using patient judgement as external anchor, we examined sensitivity and specificity of the five selected literature-based as well as the ten newly proposed worsening criteria in the validation cohort. Literature-based worsening criteria were found to be specific but lacked sensitivity. Two out of 10 newly proposed worsening criteria constructed by the Expert Group showed an acceptable combination of sensitivity and specificity in the derivation cohort which was confirmed in the validation cohort (ranging from 54% to 65% and 67% to 74% respectively). The set that performed best addressed:

Worsening in:

- **pain ≥ 20% and absolute change ≥ 20* *or***
- **function ≥ 10% and absolute change ≥ 10 *or***
- **PGA ≥ 10% and absolute change ≥ 1**

*in case of using NRS pain (scale 0-10): absolute change ≥2. This set yielded a sensitivity of 0.59 (95% CI 0.47-0.71) and specificity of 0.74 (95% CI 0.66 to 0.82) with an AUC of 0.67 (95% CI 0.60 to 0.74).

We recommend to evaluate these criteria further in other knee and hip OA populations.

Main finding III: Newly developed criteria for clinical worsening in knee and hip OA showed an acceptable combination of sensitivity and specificity.

Main finding IV: Newly developed criteria for clinical worsening in knee and hip OA incorporate relatively small absolute and relative changes of patient reported outcome measures (PROMs) compared with improvement criteria.

In **chapter 5**, we investigated the added predictive value of failure of optimized multimodal non-surgical treatment for worsening at 2 years in knee and hip OA patients using the validated clinical worsening criteria presented in chapter 4. Worsening at 2 years was defined as fulfilling these clinical worsening criteria, or total joint replacement (TJR). We showed that, at 2 years, more than half of patients with established knee and hip OA (knee 59%; hip 71%) in secondary care have worsened, and 39% of these patients had undergone a TJR. A higher proportion of knee OA patients showed clinical worsening at 2 years compared to hip OA patients (40% versus 20%, $p < 0.01$). However, the proportion of patients who underwent a TJR was lower for knee than hip OA (19% versus 51% respectively, $p < 0.0001$). Having clinically worsened at 3 months appeared to be a clear independent predictor for having worsened at 2 years (odds ratio (OR) 2.8 95% confidence interval (CI) 1.5 to 5.2) with a fair discriminative ability of the area under the receiver operating characteristic curve (AUC) 0.68 (95% CI 0.57 to 0.70). Similar results were obtained when only TJR at 2 years was used as outcome measure (OR 4.1 95% CI 2.0 to 8.4) with good AUC (0.82 95% CI 0.76 to 0.87). Our findings suggest that re-assessment of symptoms after optimized non-surgical treatment could be meaningful to guide both patients and (orthopaedic) surgeons in clinical decision making about the appropriate timing of TJR. Furthermore, this information could be used to identify subgroups of patients potentially eligible for novel and advanced treatment options.

Main finding V: Short-term clinical worsening is a clear independent risk factor for worsening at 2 years in established knee and hip OA.

Since in OA low grade synovial inflammation is known to play a role, we explored in **chapter 6** whether the serum levels of a pro-inflammatory marker S100A8/A9 (also called calprotectin) are associated with clinical and structural characteristics in patients with established knee, hip or hand OA in a cross-sectional study. Serum S100A8/A9 levels in patients with established knee, hip and hand OA are not associated with clinical characteristics regarding pain, stiffness

and function, but are positively associated with ESR and negatively associated with the sum score of osteophytes. Furthermore, in the hand OA group, S100A8/A9 serum level showed a weak inverse relationship with the sum score of joint space narrowing. Our results do not support an important role for serum S100A8/A9 levels as biomarker for severity of clinical symptoms and structural abnormalities in established knee, hip and hand OA patients.

Main finding VI: Serum S100A8/A9 levels cannot be considered as relevant biomarker for severity of clinical symptoms and structural abnormalities in established knee, hip and hand OA patients.

Establishing evidence of additional non-surgical treatment

In **chapter 7**, we systematically reviewed the literature on the efficacy and safety of low dose radiation therapy on pain and functioning in patients with OA. In OA, low grade synovial inflammation is known to play a role, which may be a potential target for OA treatment. Low-dose radiotherapy (LDRT) has been widely used for treatment of non-malignant disorders since its introduction. In vitro and animal studies have shown anti-inflammatory effects in OA. LDRT may be an additional non-surgical treatment option for knee OA patients in whom non-surgical interventions are insufficiently effective and for whom surgical treatments are not (yet) an option. Seven studies were suitable for inclusion, and across these studies, in 25-90% and 29-71% of the patients pain and functioning improved, respectively. However, the methodological quality of all studies was judged as weak (no blinding, retrospective design, uncontrolled studies, and non-validated single-outcome measures). Therefore, we concluded that there is currently insufficient high-level evidence available to indisputably demonstrate the effectiveness of LDRT in OA patients.

In **chapter 8**, the results of a randomized, double-blinded, sham-controlled trial to evaluate the effectiveness of low-dose radiation therapy on symptoms in patients with knee OA are presented. The secondary aim was to examine the effects on inflammatory signs. We recruited knee OA patients through advertisements in local newspapers and from the rheumatology outpatient clinic of the Sint Maartenskliniek. Key eligibility criteria were: age ≥ 50 years, pain score ≥ 5/10, and non-response to analgesics and exercise therapy. We randomly assigned patients 1:1 to receive six times LDRT (1 Gray per fraction) or sham intervention in two weeks, using a web-response system. The primary outcome was the proportion of responders at three months post-intervention, according to the OMERACT-OARSI responder criteria. Secondary outcomes included inflammatory signs assessed by ultrasound, magnetic resonance imaging, and serum inflammatory markers. Intention-to-treat analyses was performed. We randomly assigned 55 patients: 27 (49%) to LDRT, and 28 (51%) to sham. At three months post-intervention, 12/27 patients (44%; 95% CI 26 to 63%) in the LDRT versus 12/28 patients (43%; 95% CI 25 to 61%) in the sham group responded, difference 2% (95% CI -25 to 28%), odds ratio adjusted for the stratifying variable was 1.1 (95% CI 0.4 to 3.2). No differences in any of the inflammatory signs were observed. In conclusion, we found that treatment with LDRT in knee OA patients does not lead to a substantial reduction of symptoms when compared to sham treatment. Considering the limits of the 95% confidence intervals, a difference exceeding 28% responders between groups seems unlikely. Therefore, based on this RCT and the absence of other high quality evidence, we advise against the use of LDRT as treatment for knee OA.

Main finding VII: Treatment with low-dose radiation therapy (LDRT) in knee OA patients does not lead to a substantial reduction of symptoms when compared to sham treatment.

General discussion

During the performance and interpretation of these studies, some important methodological issues were raised. In this paragraph, the following issues of our research and their possible impact on future research are discussed: 1) influence of the homogeneous patient selection, 2) influence of the study context on estimating appropriate outcome measures, 3) importance of validation, and 4) quality of the evidence on the effectiveness of low-dose radiation therapy.

Patient selection

In order to be able to generalize the study results outside the specific study setting, it is of great importance to include a sample representative of the target population in daily practice. Our clinical studies comprise four different data collections: two observational studies with multiple assessments, one cross sectional study, and one randomized controlled trial. In total, we included a large group of patients, summing up to 1000 participants, with established OA. The studies included knee and hip OA patients who fulfilled the clinical American College of Rheumatology criteria for OA and were not yet deemed eligible for total joint replacement (TJR) by their orthopaedic surgeon at the Sint Maartenskliniek Nijmegen, the Netherlands. Of importance here is that our main observations pertain to patients with established OA conducted in one specialized clinic. This homogenous population might have led to a selection of patients with a relatively high clinical burden and may have hampered the generalizability of our results. When comparing our cohort to other large and well-known OA cohorts, we found similar socio-demographic characteristics. However, patients in our cohort showed more radiological progression than other cohorts described in the literature. This is not surprising, since most of the well-known OA cohorts focus on early OA patients.[1,2,3] Second-line cohorts are rather uncommon and when we compare our cohort with these cohorts with more established OA patients, we found very similar pain and radiographic scores at inclusion.[4,5,6] Therefore, our study population is not representative for the general OA population, but generalizable to patients with established OA who are not yet deemed eligible for surgery. Additionally, most of the patients had knee OA, so generalizability for hip OA is not fully assured. Nevertheless, this sample seems representative for referrals to secondary care as mainly all patients were referred by their primary physician in the first place. However, future research is warranted and should aim to investigate other OA populations and settings, for example in OA patients from primary care who are referred to secondary care.

Relevance of the study context

Has the above-mentioned selection influenced the interpretation of our results regarding the identification of appropriate clinical outcome measures (chapters 2, 3 and 4)? In clinical research, it is considered of great importance to incorporate the patient's interpretation of outcomes in establishing the relevance of findings.[7] When using outcome measures, cut-off values are required for the definition of treatment success or failure. This is a methodological challenge and has been a topic of research for many years.[8,9] Two different concepts are currently used to estimate and interpret changes in scores of PROMs: 1) relevant amount of change i.e. minimal clinically important change (MCIC; either improvement or worsening),

9

and 2) reaching an absolute value, i.e. an acceptable symptom state. At the start of the studies evaluating PROMs for OA symptoms, I did not sufficiently realize that 1) MCIC values for worsening could be different from MCIC for improvement values, but also that 2) both MCIC and patient acceptable symptom state (PASS) values could be context specific. My understanding has gradually changed during the time-frame of this PhD project, as discussed in the next paragraphs.

The importance of the MCIC for worsening

The MCIC is part of a concept reflecting interpretability as an important measurement property of a PROM, according to the COSMIN taxonomy.[10] The interpretability is about the degree to which it is clear what the scores or change scores mean. The MCIC has been subject of many publications, mainly in the context of MCIC for improvement and rarely as MCIC for worsening. MCIC for improvement is often used when new treatment options are evaluated. However, for daily clinical practice in particular in established OA, MCIC for worsening seems more important than MCIC for improvement, because knowing if a patient has worsened may have consequences for clinical decision making, for example to apply more advanced treatment options to this patients. There are hardly any studies that have estimated the MCIC for worsening in OA, and the few studies that have examined this, showed inconsistent results.[11] Therefore, we focused on estimating and applying this fairly new concept of MCIC for worsening to established OA patients.

Values for MCIC improvement are different from values for MCIC worsening

When focusing on the concept of MCIC for worsening, there is still discussion about whether MCIC for improvement is the same as MCIC for worsening.[12,13] Some authors use the inverse of the MCIC for improvement as MCIC for worsening value, while others are convinced that the MCIC for improvement is different from the MCIC for worsening. As part of the development and validation of criteria for clinical worsening in knee and hip OA (chapter 4), we estimated the MCIC for worsening. We found the values of MCIC for worsening to incorporate relatively small absolute and relative changes compared with improvement values. This is an interesting and also unexpected finding in OA research that needs careful consideration since applying MCIC for worsening could have clinical consequences. Interestingly, recently more evidence regarding MCIC for worsening in rheumatology has been published.[14] Remarkably, their conclusion was consistent with ours that the thresholds for the MCIC for worsening were lower than those for improvement in axial spondyloarthritis. This seemingly paradoxical finding of MCIC for worsening thresholds being lower than for improvement in OA, has previously also been reported in other disease areas.[15,16] In conclusion, I would recommend to assess a separate MCIC value for worsening and improvement, and not to use the opposite of MCIC improvement values for worsening.

Applying our estimated MCIC for worsening and PASS values

Despite the fact that the MCIC has become a standard approach in the interpretation of changes in PROMs, there is still discussion whether the MCIC of a PROM should be a fixed value or not.[12,13,17,18] Different authors in mainly epidemiologic research have concluded, that the MCIC estimate depends on the context of the disease.[17] Its value can be variable across the disease severity, underlying health disorder, intervention, characteristics of the population, unit of interest (whether an individual or a group), domain being measured, time between baseline and follow-up, baseline values observed, and the change in values.[9,17,19,20,21,22] Although, it would be desirable to have one fixed MCIC value that could be applied to different diseases or circumstances, this seems not feasible for daily practice. There is currently too limited evidence regarding established MCIC values across the vast number of PROMs. In the end, I expect that MCIC values will be context-specific rather than fixed, and this might limit the application of our estimated MCIC for worsening values across other OA populations in future research.[17,20,23]

A possibility to obtain a valid and reliable MCIC value could be to determine this value during studies separately. However, this is quite labour-intensive and comprises the following two major drawbacks. First, the MCIC can only be determined per measurement instrument or per PROM. This entails that is impossible to determine one MCIC for worsening when several PROMs are important to measure the impact of a particular disease, for example in OA the domains pain, function, and patient global assessment. A composite index incorporating all relevant domains could counterbalance this drawback. Moreover, MCIC for worsening values can only be determined in the short term, such as within a time span of 3 months. A time frame of 3 months is deemed long enough to allow for worsening and brief enough to minimize the risk of a response shift.[24] This implies that when a longitudinal study of more than 3 months is performed with clinical worsening in OA as primary outcome, researchers first have to determine the MCIC for worsening value within a 3-months period before the primary outcome for the long term can be determined. This is a very cumbersome process that is also not doable for daily practice.

The same discussion is relevant for the PASS. The PASS is considered a state and is defined as the highest level of symptoms that the majority of patient consider acceptable.[25,26] In line with the MCIC, there is currently too limited evidence regarding established PASS values across the vast number of PROMs. Future research could include the standardized PASS question, whether the current state would be acceptable or unacceptable, in each study as an outcome measure.[27,28] Then, researchers are not dependent on previously estimated, and therefore context-dependent, values. However, this option does not take into account that a response shift could take place, in which interpersonal perception of the disease state changes during the assessments. Ultimately, in my view, PASS values seem to be both context-specific and interpersonal-dependent values rather than fixed values, and this hampers the use of PASS values across other OA populations in future research.[9,17,19,20,23]

In the end, the concepts of MCIC for worsening and PASS values seem of limited use for clinical practice considering the dependency of both contextual and interpersonal factors, the difficulty to assess and the complexity to interpret. This advocates the use of validated worsening criteria that combine important domains for OA and indicates the response at an individual level to facilitate prediction research in the future.

Importance of validation

The natural course of pain and physical function in OA is highly variable and heterogeneous. Most patients have been found to remain stable, while others will worsen and some even improve.[1,2,4,29,30] However, the individual course of OA is difficult to predict, because the underlying mechanisms and risk factors associated with clinical worsening are still largely unknown.[4,31,32,33] Previous research has mainly focused on determinants for improvement, because the effectiveness of new therapies was evaluated. However, for clinical practice,

9

identification of risk factors for clinical worsening is important to inform about the prognosis, and especially, to enable targeting more advanced treatments to specific subgroups of patients.[4,31,32,34] We developed criteria for clinical worsening and used these as primary outcome to measure symptom progression over time on an individual level. We showed that short-term clinical worsening is a clear risk factor for worsening at 2 years in established knee and hip OA (chapter 5). It is very important that our research is replicated in the future to evaluate whether this finding can be applied across other OA patients or whether this is a "false-positive finding".

There is increasing concern among leading researchers that in modern research, false-positive findings may be the majority of published research.[35,36,37] We developed a prediction model; the development and interpretation of predictive research is challenging for several reasons and the following three important methodological aspects could be responsible for causing false-positive results in predictive research 1) inadequate selection of predictor variables, 2) insufficient sample size of the model, and 3) lack of validation.[38,39]

1) A common source leading to false-positive findings is that selection of predictor variables for the model is based on statistical significance ($p < 0.05$) from the set of candidate predictors, instead of theoretical reasoning. An univariate screening from a large number of candidate predictors is known to lead to chance findings and thus to less performing models. We selected independent variables based on the literature and clinical relevance, which can be seen as a strength.

2) Inadequate sample size at model development. Often, small cohorts and limited total events aggravate various problems at model development. Various authors recommend to use a rule of thumb, a recommendation which is supported by simulation studies.[40] A common rule of thumb is to formulate sample size requirements as events per variable, with a minimum of 10 events per variable. Considering the 116 patients that did not worsen in our study, and using this rule of thumb, this is sufficient to build a reliable and concise prediction model including the 12 predictors we described.

3) Validation is often limited. Initially, a good internal validity could contribute to reducing this by paying close attention to the two points mentioned above. Nevertheless, risks from prediction models tend to be overestimated and it is therefore recommended to additionally correct for overfitting or optimism of performance measures to further improve the internal validity.[41,42] Moreover, external validation is recommended by applying the internal validated prediction model in an external dataset. The aforementioned point of overfitting seems to be a methodological limitation with respect to our results. We did not provide a mechanism to account for model overfitting and it would have been better if we had initially validated the model internally. Ideally, we should have performed internal validation using shrinkage techniques by performing bootstrapping. Therefore, further external validation, as commonly recommended, is of great importance to counter false positive findings and is required by evaluating the performance of the added predictive value of short term clinical worsening in another dataset, preferably by an independent team. Nevertheless, we estimated that patients who have clinically worsened at 3 months, despite optimized multimodal non-surgical treatment, have an almost threefold increased odds ratio for having worsened at 2 years. If we had applied internal validation of our model using shrinkage techniques, we expect it would still be likely that this had resulted in a clinically relevant predictor. A predictive value is generally considered clinically relevant if the odds ratios is at least > 2.0 or < 0.5.[43] With regard to our newly proposed worsening criteria, these were first tested in a development cohort, and then, validated in a different cohort. This could best be regarded as preliminary validated, hence the term preliminary worsening criteria, since it is about the same Dutch population regarding patients with established OA. However, to contribute to reducing false-positive research findings, we recommend external replication in multiple cohorts of both our newly developed worsening criteria and our predictor of short-term clinical worsening being a risk factor for worsening at 2 years.

To increase the quality of prognostic studies in the future, journals could require adherence to the TRIPOD statement (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (http://www.tripod-statement.org, which aims to improve the reporting of studies worldwide developing, validating, or updating a prediction model.[44]) Additionally, reproducibility could be improved by requiring higher reporting standards. For example, by not publishing the first studies, unless validation in another cohort is part of the original publication. More and more journals require a form of validation of results which will probably increase the quality of predictors in future studies, although a change in priorities of the academic field is also required.[37,45] Creating better incentives for replication studies would also help counter this. Nowadays, it is fairly difficult to externally validate a prediction model in another population, because it is difficult to get research funding for this topic in the first place, and moreover, this kind of research is not always rewarded in the research world. It would be desirable if journals were more equipped to publish negative results more often in case that predictors cannot be validated in a separate cohort. Furthermore, replication studies should be seen as an essential part of science.46 Above all, external replication in multiple cohorts will always verify the applicability of each predictor.

Quality of the evidence on the effectiveness of low-dose radiation therapy

At the start of the clinical studies performed in this thesis, we aimed to contribute to the knowledge about the effectiveness of low-dose radiation therapy (LDRT) as treatment for knee OA (chapters 7 and 8). LDRT for benign disorders such as knee OA is indeed widely used as treatment in some parts of the world, despite the absence of controlled studies.[47,48] In our randomized, double-blinded, sham-controlled trial, we found that treatment with LDRT does not lead to a substantial reduction of symptoms when compared to sham treatment. In view of the absence of other high-level quality evidence in favor of LDRT, we advise against its use as treatment for knee OA. Because this treatment is still widely used in some countries, future efforts should focus on deimplementation of LDRT by changing the beliefs of both patients and involved clinicians about the efficacy of LDRT. This is easier said than done. It is striking that previous research in another disease area has shown that trial results had little or no impact on the beliefs of the involved clinicians, i.e. they did not adjust their beliefs to the extent that was expected according to Bayes' theorem.[49] This raises an interesting direction for future research that could focus on methods to elicit beliefs in the effectiveness of an evaluated intervention and, ideally, to evaluate the distribution of both prior and posterior beliefs before and after publication of study results.[50,51]

Additionally, it is important that future research should also focus on the quality of the scientific evidence of LDRT treatment for other benign (musculoskeletal) disorders, for which high-quality studies are also lacking.[52] There is a need to question and discuss the necessity of treatments commonly used but not supported by evidence. In recent years this problem has gained more attention and the internationally expanding Choosing Wisely campaign

9

is a good example of the effort taken to decrease tests and treatments that does not have additional value for patients and may even cause harm.[53] Several international scientific societies provided a list yielding the most relevant treatments that lack scientific evidence. Also, an increasing number of Dutch scientific societies are making an inventory of the relevant absences of knowledge of existing treatments, because a good scientific basis for these treatments is lacking. Recently, the evidence for the effectiveness of LDRT for benign diseases has been reviewed.[54] The authors conclude that in the UK the use of radiotherapy for benign conditions is limited, in contrast to practice in Germany.[54] They also conclude that interpretation of the literature on radiotherapy for benign conditions is problematic because much of the evidence is based on case reports and single institution case series, although some randomized studies and systematic reviews do exist.[54] I therefore recommended to add the effectiveness of LDRT treatment for other benign (musculoskeletal) disorders, for which high-quality studies are also lacking to the Choosing Wisely list of the European society for radiotherapy and oncology or to that of Germany considering the relatively high use in this country.[46,47,51]

Conclusion

Considering OA being a serious health problem, this thesis aimed to contribute to the body of knowledge regarding the non-surgical management of established knee and hip OA. We focused on clinical worsening, because the possibility of an unfavorable prognosis/course is important for clinical decision-making. Further refinement and external validation of our preliminary validated worsening criteria could enhance identification of risk factors for clinical progression in the future. In addition, our clear predictor of short-term clinical worsening despite optimized non-surgical treatment for worsening at 2 years, could be used for patient information about prognosis and in particular for targeting more advanced treatment options to more severely affected patient subgroups. Of course, replication of our findings in other OA cohorts is necessary. Finally, we advise against the use of LDRT as treatment for knee OA.

References

1. Wesseling J, Bierma-Zeinstra SMA, Kloppenburg M, Meijer R, Bijlsma JWJ. Worsening of pain and function over 5 years in individuals with 'early' OA is related to structural damage: data from the osteoarthritis initiative and CHECK (Cohort Hip & Cohort Knee) study. Ann Rheum Dis 2015;74:347–53.
2. White DK, Keysor JJ, Lavalley MP, Lewis CE, Torner JC, Nevitt MC, et al. Clinically important improvement in function is common in people with or at high risk of knee OA: the MOST study. J Rheumatol 2010;37:1244-51.
3. Felson DT, Niu J, Yang T, Torner J, Lewis CE, Aliabadi P, et al. MOST and OAI investigators. Physical activity, alignment and knee osteoarthritis: data from MOST and the OAI. Osteoarthritis Cartilage. 2013;21:789-95.
4. de Rooij M, van der Leeden M, Cheung J, van der Esch M, Häkkinen A, Haverkamp D, et al. Efficacy of Tailored Exercise Therapy on Physical Functioning in Patients With Knee Osteoarthritis and Comorbidity: A Randomized Controlled Trial. Arthritis Care Res (Hoboken). 2017;69:807-16.
5. van der Esch M, Knoop J, van der Leeden M, Roorda LD, Lems WF, Knol DL, et al. Clinical phenotypes in patients with knee osteoarthritis: a study in the Amsterdam osteoarthritis cohort. Osteoarthritis Cartilage. 2015;23:544-9.
6. van Dijk GM, Veenhof C, Lankhorst GJ, van den Ende CH, Dekker J. Vitality and the course of limitations in activities in osteoarthritis of the hip or knee. BMC Musculoskelet Disord. 2011;12:269.
7. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? Ann Rheum Dis 2007;66(Suppl 3):iii40–1.
8. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. Control Clin Trials 1989;10:407–15.
9. Jayadevappa R, Cook R, Chhatre S. Minimal important difference to infer changes in health-related quality of life-a systematic review. J Clin Epidemiol. 2017;89:188-98.
10. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63:737-45.
11. Angst F, Aeschlimann A, Michel BA, Stucki G. Minimal clinically important rehabilitation effects in patients with osteoarthritis of the lower extremities. J Rheumatol 2002;29:131–9.
12. De Vet HCW, Terwee CB, Mokkink LD, Knol DL. Measurement in medicine. New York: Cambridge University Press, 2011.
13. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol. 2003;56:395-407.
14. Molto A, Gossec L, Meghnathi B, Landewé RBM, van der Heijde D, Atagunduz P, et al; ASAS-FLARE study group. An Assessment in SpondyloArthritis International Society (ASAS)-endorsed definition of clinically important worsening in axial spondyloarthritis based on ASDAS. Ann Rheum Dis. 2018;77:124-27.
15. Mannion AF, Porchet F, Kleinstück FS, et al. The quality of spine surgery from the patient's perspective: part 2. Minimal clinically important difference for improvement and deterioration as measured with the Core Outcome Measures Index. Eur Spine J 2009;18(S3):374–9.
16. Horváth K, Aschermann Z, Kovács M, et al. Minimal clinically important differences for the experiences of daily living parts of movement disorder society-sponsored unified Parkinson's disease rating scale. Mov Disord 2017;32:789–93.
17. Beaton DE, Boers M, Wells GA. Many faces of the minimal clinically important difference (MCID): a literature review and directions for future research. Curr Opin Rheumatol. 2002;14:109-14.
18. Tubach F, Ravaud P, Martin-Mola E, Awada H, Bellamy N, Bombardier C, et al. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: results from a prospective multinational study. Arthritis Care Res (Hoboken) 2012;64:1699–707.

19. Bellamy N, Hochberg M, Tubach F, Martin-Mola E, Awada H, Bombardier C. et al. Development of multinational definitions of minimal clinically important improvement and patient acceptable symptomatic state in osteoarthritis. Arthritis Care Res 2015: 67:972-80.
20. Terwee CB, Roorda LD, Dekker J, Bierma-Zeinstra SM, Peat G, Jordan KP, et al. Mind the MIC: large variation among populations and methods. J Clin Epidemiol 2010;63:524–34.
21. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimal important differences for patient-reported outcomes. J Clin Epidemiol 2008;68:102e9.
22. Angst F, Aeschlimann A, Angst J. The minimal clinically important difference raised the significance of outcome effects above the statistical level, with methodological implications for future studies. J Clin Epidemiol. 2017;82:128-36.
23. Mills KAG, Naylor JM, Eyles JP, Roos EM, Hunter DJ. Examining the minimal important difference of patient-reported outcome measures for individuals with knee osteoarthritis: a model using the knee injury and osteoarthritis score. J Rheumatol 2016;43:395–404.
24. Robling M, Hood K. Response shift, responsiveness or recall bias.Br J Gen Pract 2002;52:585.
25. 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;64:29–33.
26. Perrot S, Bertin P. 'Feeling better' or 'feeling well' in usual care of hip and knee osteoarthritis pain: determination of cutoff points for patient acceptable symptom state (PASS) and minimal clinically important improvement (MCII) at rest and on movement in a national multicenter cohort study of 2414 patients with painful osteoarthritis. Pain 2013;154:248–56.
27. Tubach F, Ravaud P, Beaton D, Boers M, Bombardier C, Felson DT, et al. Minimal clinically important improvement and patient acceptable symptom state for subjective outcome measures in rheumatic disorders. J Rheumatol 2007;34:1188–93.
28. Putrik P. Ramiro S. Hifinger M, Keszei AP, Hmamouchi I, Dougados M, et al. In wealthier countries, patients perceive worse impact of the disease although they have lower objectively assessed disease activity: results from the cross-sectional COMORA study. Ann Rheum Dis 2016;75:715-20.
29. Dekker J, van Dijk GM, Veenhof C. Risk factors for functional decline in osteoarthritis of the hip or knee. Curr Opin Rheumatol. 2009;21:520-4.
30. Holla JF, van der Leeden M, Heymans MW, et al. Three trajectories of activity limitations in early symptomatic knee osteoarthritis: a 5-year follow-up study. Ann Rheum Dis 2014;73:1369-75.
31. Conaghan PG, Kloppenburg M, Schett G, Bijlsma JWJ. Osteoarthritis research priorities: a report from a EULAR ad hoc expert committee. Ann Rheum Dis 2014;73:1142–5.
32. Bastick AN, Runhaar J, Belo JN, Bierma-Zeinstra SMA. Prognostic factors for progression of clinical osteoarthritis of the knee: a systematic review of observational studies. Arthritis Res Ther 2015;17:152.
33. Kim C, Nevitt MC, Niu J, Clancy MM, Lane NE, Link TM, et al. Association of hip pain with radiographic evidence of hip osteoarthritis: diagnostic test study. BMJ 2015;351:h5983.
34. Belo JN, Berger MY, Reijman M, Koes BW, Bierma-Zeinstra SM. Prognostic factors of progression of osteoarthritis of the knee: a systematic review of observational studies. Arthritis Rheum. 2007;57:13-26.35.
35. Ioannidis JP. Why most published research findings are false. PLoS Med 2005;2(8):e124.
36. Ioannidis JP. How to make more published research true. PLoS Med. 2014:11:e1001747
37. Ioannidis JP, Greenland S, Hlatky MA, Khoury MJ, Macleod MR, Moher D et al. Increasing value and reducing waste in research design, conduct, and analysis. Lancet 2014;383(9912):166-75.
38. Siontis GC, Tzoulaki I, Castaldi PJ, Ioannidis JP. External validation of new risk prediction models is infrequent and reveals worse prognostic discrimination. J Clin Epidemiol. 2015 ;68:25-34.
39. Collins GS, de Groot JA, Dutton S, Omar O, Shanyinde M, Tajar A, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. BMC Med Res Methodol. 2014;14:40.

40. Steyerberg EW. Clinical prediction models, New York: Springer-Verlag, 2009
41. Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. BMJ. 2009 Jun 4;338:b606.
42. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ. 2009;338:b375.
43. van der Windt DA, Thomas E, Pope DP, de Winter AF, Macfarlane GJ, Bouter LM et al. Occupational risk factors for shoulder pain: a systematic review. Occup Environ Med 2000;57(7):433-42.
44. Moons KG, Altman DG, Reitsma JB, Ioannidis JP, Macaskill P, Steyerberg EW, Vickers AJ, Ransohoff DF, Collins GS. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. Ann Intern Med. 2015;162:W1-73.
45. Chalmers I, Bracken MB, Djulbegovic B, Garattini S, Grant J, Gulmezoglu AM et al. How to increase value and reduce waste when research priorities are set. Lancet 2014;383: 156-65.
46. KNAW (2018). Replication studies – Improving reproducibility in the empirical sciences, Amsterdam, KNAW.
47. Leer JWH, Van Houtte P, Davelaar J. Indications and treatment schedules for irradiation of benign diseases: a survey. Radiother Oncol 1998; 48: 249–57.
48. Seegenschmiedt MH, Micke O, Willich N. Radiation therapy for nonmalignant diseases in Germany: current concepts and future perspectives. Strahlenther und Onkol 2004; 180: 718–30.
49. Rovers MM, van der Wilt GJ, van der Bij S, Straatman H, Ingels K, Zielhuis GA. Bayes' theorem: a negative example of a RCT on grommets in children with glue ear. Eur J Epidemiol. 2005;20:23-8.
50. Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Feldman BM. Methods to elicit beliefs for Bayesian priors: a systematic review. J Clin Epidemiol. 2010;63:355-69.
51. Johnson SR, Tomlinson GA, Hawker GA, Granton JT, Grosbein HA, Feldman BM. A valid and reliable belief elicitation method for Bayesian priors. J Clin Epidemiol. 2010;63:370-83.
52. McKeown SR, Hatfield P, Prestwich RJ, Shaffer RE, Taylor RE. Radiotherapy for benign disease; assessing the risk of radiation-induced cancer following exposure to intermediate dose radiation. Br J Radiol 2015; 88: 20150405
53. Levinson W, Kallewaard M, Bhatia RS, Wolfson D, Shortt S, Kerr EA. 'Choosing Wisely': a growing international campaign. BMJ Qual Saf 2015;24:167-74.
54. Taylor RE, Hatfield P, McKeown SR, Prestwich RJ, Shaffer R. Radiotherapy for Benign Disease: Current Evidence, Benefits and Risks. Clin Oncol (R Coll Radiol) 2015;27:433-5.

9

Chapter 10

Nederlandse samenvatting

Bijdragen aan de aanpak van artrose

Hoofdstuk 1: inleiding

Artrose is een gewrichtsaandoening die alle weefsels van het gewricht aantast. Het wordt gekenmerkt door afbraak van kraakbeen, de vorming van puntige, benige uitgroeisels aan de randen van het gewricht (osteofyten), verdikking van het bot onder het kraakbeen (subchondrale botsclerose), veranderingen in ligamenten en spieren en ontsteking van de slijmvlieslaag aan de binnenkant van het gewrichtskapsel (het synovium). Artrose kan in elk gewricht voorkomen, maar komt het meeste voor in de knie, hand of de heup. Artrose in de knie of de heup is om drie verschillende redenen een groot gezondheidsprobleem. Ten eerste ervaren mensen met artrose pijn en beperkingen bij het uitvoeren van activiteiten zoals lopen en traplopen die leiden tot een verminderde kwaliteit van leven. Ten tweede is het de meest voorkomende chronische gewrichtsaandoening. De incidentie en prevalentie van symptomatische artrose van de knie en de heup zijn de afgelopen decennia aanzienlijk gestegen, voornamelijk door een combinatie van de vergrijzende bevolking en de groeiende obesitas epidemie. Artrose is wereldwijd de snelst groeiende oorzaak van invaliditeit geworden. Ook in Nederland is artrose de snelst groeiende ziekte met een toename van 1,2 miljoen patiënten in 2015 tot een verwachte 2,2 miljoen in 2040. Ten derde is artrose een groot gezondheidsprobleem omdat dit vooral in de ontwikkelde landen leidt tot een groeiende impact op de gezondheidszorg(capaciteit) en op toekomstige sociaal economische kosten. Daardoor heeft artrose een grote maatschappelijk impact.

Voor artrose bestaat geen curatieve behandeling. De behandeling is primair gericht op het verminderen van de klachten zoals pijn en functiebeperking. De behandeling bestaat uit een niet-chirurgische component (zoals voorlichting, lichaamsbeweging, pijnstillers, leefstijladviezen inclusief gewichtsverlies bij patiënten met overgewicht) en een chirurgische component (zoals totale gewrichtsvervanging). Verschillende internationale richtlijnen benadrukken het belang en de doeltreffendheid van het starten met niet-chirurgische behandeling, voordat chirurgisch ingrijpen wordt overwogen. Toch is bekend dat niet-chirurgische behandelingen onvoldoende worden benut. Daarnaast is bekend dat het aantal operaties aanzienlijk toeneemt, waarbij geschat wordt dat bij 20-30% van de gewrichtsvervangende operaties de indicatie te vroeg is gesteld. Aan de andere kant is het wetenschappelijk tot nu toe niet mogelijk gebleken om geschikte criteria te formuleren voor het stellen van deze indicatie en wordt in het algemeen gevaren op een combinatie van pijn, beperking in het dagelijks leven en radiologische schade. In het algemeen wordt een gewrichtsvervangende operatie als een effectieve behandeling beschouwd, waarbij totale knievervanging minder effectief is in vergelijking met vervanging van de heup. Na het plaatsen van een totale knieprothese is tot 30% van de patiënten niet tevreden met het resultaat.

Gezien de zware ziektelast, de hoge en stijgende prevalentie en de toenemende economische impact, is er een duidelijke behoefte aan verbetering van de mogelijkheden van niet-chirurgische behandeling van knie- en heupartrose. Daarbij is het belangrijk om de hele aanpak mee te nemen inclusief het gebruik van de juiste meetinstrumenten om onderzoek mee te kunnen doen. Het doel van dit proefschrift is om een bijdrage te leveren aan
A) het vinden van de meest geschikte meetinstrumenten
B) de identificatie van risicofactoren voor verslechtering van klachten en
C) het bewijs voor nieuwe niet-chirurgische behandelmogelijkheden.

Dit proefschrift richt zich voornamelijk op patiënten met gevorderde knie- of heupartrose waarbij de overgrote meerderheid volgens hun orthopedisch chirurg nog niet in aanmerking komt voor een gewrichtsvervangende operatie.

A) Geschikte meetinstrumenten
Veranderingen in klachten kunnen vanuit het perspectief van de patiënt worden onderzocht door middel van zogenaamde patiënt-gerapporteerde uitkomstmaten.

Hoofdstuk 2: de responsiviteit van verschillende vragenlijsten
Voor het meten van het dagelijks functioneren bij mensen met knieartrose zijn verschillende valide en betrouwbare vragenlijsten beschikbaar. Voor het monitoren van patiënten of het evalueren van de effectiviteit van interventies, is het van belang dat verandering in dagelijks functioneren adequaat gemeten kan worden. Het meten van verandering wordt ook wel responsiviteit genoemd. In hoofdstuk 2 hebben we de responsiviteit onderzocht van vier veel gebruikte vragenlijsten om dagelijks functioneren te meten bij patiënten met symptomatische knieartrose die niet-chirurgische behandeling hadden ondergaan. We hebben de volgende vragenlijsten onderzocht: de korte versie van de Knee Injury and Osteoarthritis Outcome Score Function (KOOS-PS), Lequesne Algofunctional index (LAI), Lower Extremity Functional Scale (LEFS) en Western Ontario and McMaster Universities Osteoarthritis Index Physical Function subscale (WOMAC-PF). We hebben dit gedaan door het testen van vooraf geformuleerde hypothesen over verwachte (of afwezigheid van) correlaties tussen veranderingen in dagelijks functioneren en veranderingen in andere gerelateerde en niet-gerelateerde constructen. De resultaten hebben we gebruikt om een weloverwogen keuze te maken voor een vragenlijst die dagelijks functioneren meet (WOMAC-PF) in ons gerandomiseerde en gecontroleerde onderzoek naar de effectiviteit van een lage dosering bestraling op klachten van knieartrose (hoofdstuk 8).

Belangrijke bevinding I: De WOMAC-PF is het beste in staat om veranderingen in de tijd te detecteren in het dagelijks functioneren bij patiënten met knieartrose die gestandaardiseerde niet-chirurgische behandeling hebben ondergaan.

Hoofdstuk 3: acceptabele gezondheidstoestand
In dit hoofdstuk hebben we een prospectieve studie uitgevoerd om te onderzoeken welk niveau van klachten acceptabel is voor mensen met artrose; dit noemen we de PASS (Patient Acceptable Symptom State). De PASS is een gezondheidstoestand die wordt gedefinieerd als het hoogste niveau van klachten dat de meerderheid van de patiënten acceptabel vindt. Daarnaast hebben we de invloed beoordeeld van geslacht, leeftijd, duur van de klachten en depressieve gevoelens op de geschatte PASS-waarden. We definieerden de PASS-waarde als het 75e percentiel van de score van vragenlijsten voor die patiënten die hun toestand acceptabel vinden. We hebben de volgende vragenlijsten onderzocht: de korte versie van de Knee Injury and Osteoarthritis Outcome Score Function (KOOS-PS), Lequesne Algofunctional Index (LAI), de Lower Extremity Functional Scale (LEFS) en de fysiek functioneren subschaal van de WOMAC (WOMAC-PF).Onze resultaten laten zien dat PASS-waarden tussen verschillende vragenlijsten onderling vergelijkbaar zijn. We concluderen daarom dat PASS-afkapwaarden bij patiënten met symptomatische knieartrose robuust zijn en dat verschillende vragenlijsten kunnen worden gebruikt om de PASS-waarden voor het dagelijks functioneren te bepalen. Daarnaast hebben we vastgesteld dat vrouwelijke patiënten en patiënten met depressieve

symptomen een lagere kans hebben om de PASS-waarden te bereiken. We hebben echter ook waargenomen dat patiënten in ons knieartrose cohort een hoger niveau van klachten acceptabel vinden dan eerder gerapporteerd is voor andere reumatische aandoeningen. We concluderen dat het gebruik van één generieke PASS-waarde bij patiënten met artrose mogelijk niet juist is en dat dit ook zou kunnen gelden voor patiënten met andere reumatische aandoeningen.

Belangrijke bevinding II: PASS-afkapwaarden bij patiënten met symptomatische knieartrose zijn robuust voor verschillende vragenlijsten die het dagelijks functioneren meten.

B) Identificatie van risicofactoren voor verslechtering van de klachten
Het natuurlijke beloop van pijn en dagelijks functioneren in artrose is zeer variabel en heterogeen. De meeste patiënten lijken stabiel te blijven, terwijl anderen verslechteren en sommigen zelfs verbeteren. Het individuele beloop van artrose is moeilijk te voorspellen omdat de onderliggende mechanismen en risicofactoren voor verslechtering van de klachten nog grotendeels onbekend zijn. Het bepalen van risicofactoren zou kunnen bijdragen aan de identificatie van fenotypen oftewel specifieke subgroepen van patiënten met een ongunstig beloop die mogelijk in aanmerking komen voor specifieke en meer geavanceerde behandelopties.

Hoofdstuk 4: definitie van verslechtering van klachten
In hoofdstuk 4 beschrijven we de ontwikkeling en validatie van criteria voor symptomatische verslechtering van knie- en heupartrose op basis van literatuur, expert opinion en data. De beoordeling van de literatuur door de expertgroep leverde 5 eerder gebruikte verslechteringscriteria op. Daarnaast bereikte deze expertgroep consensus over 10 nieuwe verslechteringscriteria. Deze werden in twee databestanden van patiënten met knie- of heupartrose getest. Deze datasets bestonden uit 219 patiënten (ontwikkelcohort) en 296 patiënten (validatiecohort). Vervolgens onderzochten we, met behulp van het oordeel van de patiënt als extern anker, de sensitiviteit en specificiteit van de 5 uit de literatuur geselecteerde en de 10 nieuwe verslechteringscriteria in het validatiecohort. De uit de literatuur geselecteerde verslechteringscriteria bleken specifiek te zijn, maar niet sensitief. Twee van de 10 nieuwe verslechteringscriteria die door de expertgroep waren opgesteld toonden een acceptabele combinatie van sensitiviteit en specificiteit in beide cohorten. We bevelen aan om deze criteria verder te onderzoeken en valideren in andere knie- en heupartrose populaties.

Belangrijke bevinding III: De nieuw ontwikkelde criteria voor verslechtering van klachten bij knie- en heupartrose toonden een acceptabele combinatie van sensitiviteit en specificiteit.

Belangrijke bevinding IV: De nieuw ontwikkelde criteria voor verslechtering van klachten bij knie- en heupartrose bevatten relatief kleine absolute en relatieve veranderingen van door patiënt-gerapporteerde-uitkomstmaten in vergelijking met verbeteringscriteria.

Hoofdstuk 5: is korte termijn verslechtering van de klachten een voorspeller voor verslechtering na 2 jaar?
In dit hoofdstuk onderzochten we bij patiënten met knie- en heupartrose wat de toegevoegde voorspellende waarde is van falen van geoptimaliseerde niet-chirurgische behandeling voor verslechtering na 2 jaar. Hierbij werd gebruik gemaakt van de gevalideerde

10

verslechteringscriteria uit hoofdstuk 4. Als een patiënt aan deze criteria voldeed of een totale gewrichtsvervanging had ondergaan, kwam deze in de groep die was verslechterd na 2 jaar. We vonden dat na 2 jaar meer dan de helft van de patiënten met gevorderde knie- en heupartrose (knie 59%, heup 71%) in de tweedelijn was verslechterd en 39% van deze patiënten een gewrichtsvervangende operatie had ondergaan. Een groter aantal patiënten met knieartrose vertoonde verslechtering van de klachten na 2 jaar in vergelijking met patiënten met heupartrose. Het aandeel patiënten dat een gewrichtsvervangende operatie onderging was echter lager voor patiënten met knie- dan heupartrose. Verslechtering van de klachten na 3 maanden bleek een duidelijke onafhankelijke voorspeller te zijn voor verslechtering na 2 jaar. Vergelijkbare resultaten verkregen we wanneer alleen een gewrichtsvervangende operatie na 2 jaar werd gebruikt als uitkomstmaat. Onze bevindingen suggereren dat herbeoordeling van klachten na geoptimaliseerde niet-chirurgische behandeling van 3 maanden zinvol zou kunnen zijn om zowel patiënten als (orthopedisch) chirurgen te ondersteunen in de besluitvorming over de juiste timing van een gewrichtsvervangende operatie. Bovendien kan deze informatie worden gebruikt om subgroepen van patiënten te identificeren die mogelijk in aanmerking komen voor nieuwe en geavanceerde behandelopties.

Belangrijke bevinding V: Korte termijn symptomatische verslechtering is een duidelijke onafhankelijke risicofactor voor verslechtering van knie- en heupartrose klachten na 2 jaar.

Hoofdstuk 6: associatie tussen S100A8/A9 en verschillende kenmerken van artrose
Bij artrose speelt laaggradige synoviale ontsteking een rol. Daarom hebben we in hoofdstuk 6 in een dwarsdoorsnede onderzoek bekeken of de serumwaarde van een pro-inflammatoire marker S100A8/A9 (ook wel calprotectine genoemd) geassocieerd is met klachten en structurele afwijkingen bij patiënten met gevorderde knie-, heup- of handartrose. Serum S100A8/A9-waarden bij deze patiënten blijken niet geassocieerd te zijn met klachten zoals pijn, stijfheid en dagelijks functioneren. Wel zijn ze positief geassocieerd met de bezinking en negatief geassocieerd met de somscore van osteofyten. Bovendien werd in de groep met handartrose een zwakke omgekeerde relatie gevonden tussen de waarde van S100A8/A9 en de somscore van de vernauwing van gewrichtsruimte. Onze resultaten geven aan dat serum S100A8/A9 geen belangrijke marker is voor de ernst van klachten en structurele afwijkingen bij patiënten met gevorderde knie-, heup- en handartrose.

Belangrijke bevinding VI: S100A8/A9-waarden zijn geen relevante marker voor de ernst van klachten en ook niet voor structurele afwijkingen bij patiënten met knie-, heup- of handartrose.

C) Bewijs van een nieuwe niet-chirurgische behandeling
Een lage dosis bestraling zou een nieuwe niet-operatieve mogelijkheid van behandeling kunnen zijn voor patiënten met artrose bij wie niet-chirurgische behandeling onvoldoende effectief zijn en voor wie chirurgische behandeling (nog) geen optie is. Eerdere in vitro- en dierstudies hebben aangetoond dat een lage dosis bestraling ontstekingsremmend werkt. Behandeling met een lage dosering straling wordt veel gebruikt voor goedaardige aandoeningen zoals knieartrose in Duitsland en Oost-Europese landen. Het is dus wenselijk om meer inzicht te krijgen in de effectiviteit van deze behandeling voor artrose.

Hoofdstuk 7: systematische review over de effectiviteit van een lage dosering bestraling op pijn en dagelijks functioneren
Dit hoofdstuk is de gewijd aan een systematische beoordeling van de literatuur over de werkzaamheid en veiligheid van een lage dosis bestraling op pijn en het dagelijks functioneren bij patiënten met artrose. Zeven studies waren geschikt voor inclusie. Deze studies rapporteerden dat de pijn en het dagelijks functioneren verbeterden bij respectievelijk 25-90% en 29-71% van de patiënten. De methodologische kwaliteit van alle onderzoeken beoordeelden we echter als zwak doordat gebruikt was gemaakt van niet-geblindeerde, retrospectieve en ongecontroleerde onderzoeken en van niet-gevalideerde uitkomstmaten. Daarom concluderen we dat er momenteel onvoldoende bewijs op hoog niveau is voor de effectiviteit van behandeling met een lage dosering bestraling bij patiënten met artrose.

Hoofdstuk 8: is behandeling met een lage dosering bestraling effectief bij knieartrose?
Het doel van deze studie is om met goed opgezet onderzoek de effectiviteit van behandeling met een lage dosis bestraling op klachten bij patiënten met knieartrose te beoordelen. Het secundaire doel is om het effect op ontsteking te onderzoeken. In deze studie hebben we 55 patiënten met knieartrose geïncludeerd, met een leeftijd van minimaal 50 jaar, een pijnscore van groter dan of gelijk aan 5/10 en onvoldoende verbetering op pijnstillers en oefentherapie. Patiënten werden 1:1 gerandomiseerd om zes keer in twee weken echte of placebo bestraling te krijgen. De primaire uitkomstmaat was het percentage responders drie maanden na de interventie, volgens de OMERACT-OARSI responder criteria. De secundaire uitkomsten bestonden uit inflammatoire parameters die werden beoordeeld met behulp van echo, Magnetic Resonance Imaging (MRI) en inflammatoire markers in het bloed. We hebben geconcludeerd dat behandeling met lage dosis bestraling bij patiënten met knieartrose niet leidt tot een substantiële vermindering van de klachten in vergelijking met placebo behandeling. Daarom adviseren wij op basis van deze gerandomiseerde, geblindeerde en placebo gecontroleerde studie en de afwezigheid van ander hoogwaardig bewijs, lage dosering bestraling niet te gebruiken als behandeling van knieartrose.

Belangrijke bevinding VII: Behandeling met een lage dosis bestraling bij patiënten met knieartrose leidt niet tot een substantiële vermindering van de klachten in vergelijking met placebo.

Hoofstuk 9: discussie
In het laatste hoofdstuk bediscussiëren we de belangrijkste resultaten en methodologische aspecten van dit proefschrift. Daarnaast geven we een aantal implicaties voor de klinische praktijk en formuleren we suggesties voor toekomstig onderzoek.

Samenvattend
Artrose is een groot gezondheidsprobleem vanwege de zware ziektelast, de stijgende prevalentie en de toenemende economische impact. Dit proefschrift heeft tot doel een bijdrage te leveren aan de aanpak van niet-chirurgische knie- en heupartrose. Allereerst hebben we eigenschappen onderzocht van vragenlijsten om dagelijks functioneren te meten bij deze groep. Daarnaast hebben we ons gericht op verslechtering van de klachten omdat het risico op een ongunstige prognose van belang is voor de klinische besluitvorming. We hebben gevalideerde verslechteringscriteria ontwikkeld en gevalideerd. Verdere verfijning en externe validatie hiervan zouden de identificatie van risicofactoren voor progressie van klachten

10

in de toekomst kunnen verbeteren. Wij vonden dat verslechtering van de klachten op korte termijn, ondanks geoptimaliseerde niet-chirurgische behandeling, een duidelijke voorspeller is voor verslechtering na 2 jaar. Dit kan vooral worden gebruikt om subgroepen van patiënten te identificeren die mogelijk in aanmerking komen voor specifieke en meer geavanceerde behandelopties. Replicatie van onze bevindingen in andere artrose cohorten is noodzakelijk. Tot slot raden we behandeling van knieartrose met lage dosering bestraling sterk af.

10

Dankwoord

Dankwoord

Wat prijs ik mij gelukkig dat ik de gelegenheid heb gekregen om mijn grote wens, het combineren van patiëntenzorg en epidemiologisch onderzoek, in vervulling te laten gaan. Ik houd erg van reizen en heb mijn promotietraject ook altijd als een reis beschouwd. Hoewel de voorkant van dit proefschrift misschien anders doet vermoeden, heb ik dit traject niet ervaren als een landschap vol bergen en dalen, alhoewel de eindspurt een pittige klim was. Deze promotiereis is nu ten einde en dit boekje is met de hulp van heel veel mensen tot stand gekomen. Hen wil ik graag bedanken.

Allereerst ben ik veel dank verschuldigd aan alle artrosepatiënten die belangeloos aan de onderzoeken hebben deelgenomen. Keer op keer hebben jullie vragenlijsten ingevuld, bezoeken afgelegd, bloed afgestaan en een deel van jullie heeft zich zelfs onderworpen aan 6 keer echte of nepbestraling. Dank jullie wel!

Prof.dr. Van den Hoogen, beste Frank, zoals bij velen ben je ook bij mij de linking pin naar de reumatologie. Ik was verbaasd dat jij nog steeds enthousiast was over mij voor de eerste ANIOS reumatologie functie in de SMK, terwijl we elkaar in eerste instantie nauwelijks konden verstaan vanwege de nogal luide muziek uit mijn studentenhuis die door de hoorn galmde. Frank, jij denkt altijd in kansen en zorgt ervoor dat mensen in hun kracht komen te staan, terwijl je ondertussen wel de hoogste berg wil beklimmen. Zonder jou had ik nooit aan dit avontuur kunnen beginnen. Bedankt ook voor je altijd weer frisse blik en daaruit voortvloeiende waardevolle suggesties. Wat mooi dat je later als tweede promotor aan het team kon worden toegevoegd.

Daarnaast wil ik graag mijn copromotoren en eerste promotor bedanken voor de begeleiding.

Dr. Den Broeder en dr. Van den Ende, beste Alfons en Els, jullie vormen samen een mooi complementair promotieteam waarbij Alfons in het eerste deel en Els in de tweede helft meer de rol van eerste begeleider op zich heeft genomen. Alfons, jij was in het begin ook medisch hoofd en samen met Frank wist je dat ik een lang gekoesterde wens had om promotieonderzoek te doen. Zonder je berichtje 'hapje eten?' had dit boekje er niet gelegen. Ik ging namelijk niet met de intentie naar het etentje onder aan de berg in Tante Koosje om in de SMK te gaan werken en promoveren. Ik ben jullie voor altijd dankbaar dat jullie mij deze kans hebben gegeven en heb er geen seconde spijt van gehad. Alfons, wat fijn dat jij tijdens een overleg snel tot de kern van de zaak weet te komen en mij van praktische adviezen weet te voorzien. We delen de interesse voor de epidemiologie. Je beschikt over een soort van onuitputtelijke kennis, waarbij je altijd weer nieuwe zaken weet in te brengen. Vanwege mijn beperkte onderzoekstijd, ging ik voor doelgerichtheid, maar jij probeerde me juist te laten divergeren. Daarbij schudde je dan verrassende termen als 'de Abilene paradox' en 'de experiencing and remembering self' uit je mouw. Dank ook voor je enorme arsenaal aan grappen waarmee jij menig overleg hebt weten te verrijken. Els, jij weet precies hoe het moet en bent in staat om je kennis en ervaring, zowel op inhoudelijk als op begeleidingsniveau, op een fijne manier te delen. Ik denk dat het niet voor iedereen duidelijk is hoe groot daadwerkelijk jouw betrokkenheid is geweest. Heel veel dank voor je kritische blik en suggesties voor scherpere formuleringen. Regelmatig leverden je suggesties in eerste instantie meer werk op, maar uiteindelijk was ik je keer op keer enorm dankbaar omdat het altijd leidde tot verbetering. Bedankt dat je er altijd was,

voor je persoonlijke betrokkenheid en voor je begrip dat ik als clinicus met een beperkte onderzoeksaanstelling een beetje een vreemde eend in de bijt was.

Prof.dr. Bijlsma, beste Hans, wat heb ik het getroffen met jou als promotor. Je hebt werkelijk een dijk aan ervaring en ook op het artrosegebied heb jij je sporen ruimschoots verdiend. Je liet me, eigenlijk mijn copromotoren, redelijk vrij en indien nodig was je er. Dank voor je flexibiliteit, interesse in mij als persoon, goede raad, altijd snelle reacties en bereidheid om naar Nijmegen te komen. Tijdens onze overleggen heb ik er steeds weer van genoten hoe handig jij weet te schakelen tussen je diverse rollen. Feilloos wissel je een diepe duik in de inhoud af met het overzien van de situatie binnen een mum van tijd, hak je zo nodig soepel knopen door en weet je ondertussen mensen ook nog te verbinden en te enthousiasmeren.

Leden van de manuscriptcommissie, prof.dr. Lafeber, prof.dr. Weinans, dr. Jürgenliemk-Schulz, prof.dr. Verbeek, prof.dr. Kloppenburg, bedankt voor jullie bereidheid om het manuscript te beoordelen en zitting te nemen in de oppositie.

Prof.dr. Verbeek, beste André, bedankt dat je het mogelijk hebt gemaakt om het opleidingstraject tot epidemioloog B te volgen. Achteraf gezien zie ik het als een voorrecht om pas na vier keer te zijn uitgeloot voor geneeskunde, alsnog via de hardheidsclausule te zijn toegelaten waardoor ik ook een BMW studie heb kunnen afronden. Ik heb lang getwijfeld of ik nog wel met geneeskunde zou starten, maar de combinatie van patiëntenzorg met onderzoek inspireert mij nog altijd het meest.

Alle coauteurs, senior onderzoekers en Jaap wil ik heel hartelijk bedanken voor de samenwerking, hulp en commentaren. Michiel, bedankt voor de samenwerking en je ondersteuning bij de uitvoer en analyses van de radiotherapie trial. Ik wens je veel succes met de laatste loodjes. Nathan en Nadine, wat een cadeau om te mogen werken met twee jonge, enthousiaste en efficiënte BMW'ers. Nienke, bedankt voor je betrokkenheid bij het responsiviteitsstuk. Gijs, een deel van de artikelen komt voort uit wat begon als jouw project, dankjewel. Leden van de expertgroep en in het bijzonder Joke wil ik bedanken voor het sparren over deze ingewikkelde materie. En tenslotte Tilly en Simone, bedankt voor jullie belangrijke bijdragen aan het radiotherapie stuk. Zonder jullie was het nooit zover gekomen!

Promotieonderzoek combineren met het werk als reumatoloog is alleen mogelijk met een werkomgeving die dat ondersteunt. Alle collega reumatologen en PA's, bedankt voor veel steun, begrip en gezelligheid. Daarnaast A(N)IOS, verpleegkundigen, dokters- en poliassistentes, planners, (MT)secretariaat, ontzettend bedankt voor jullie hulp. Vincent, enorm bedankt voor je hulp en flexibiliteit bij het includeren van patiënten en Dicky voor de administratie van de verschillende studies. Marcel en Maurice, wat kwam ik goed terecht door tijdens mijn eerste baan als ANIOS onder jullie supervisie te mogen werken. Dirk-Jan, dank voor het scoren van heel veel foto's. Dank ook aan de orthopeden voor de samenwerking. Ik heb veel reuma- en farmacieonderzoekers mogen meemaken, dank voor jullie inspiraties, suggesties en samenwerking tijdens de schrijfdagen. Karen en Noortje, dank dat ik mocht aansluiten bij jullie reeds lopende Utrecht-overleggen. Jullie promoties waren wel een aderlating voor mij aangezien onze gezamenlijke overleggen en reisjes naar Utrecht daarmee ook stopten. Joost, als er iemand is die parallel met mij de wegen heeft bewandeld, ben jij het wel. We startten als AIOS in Rijnstate, gingen samen naar het Radboudumc en werden daarna weer collega's in zowel Nijmegen als Boxmeer. Wat fijn dat je mijn paranimf wilt zijn.

Mijn vrienden en familie wil ik bedanken voor de afwisseling van werk met ontspanning en gezelligheid en voor het begrip dat het contact de laatste tijd op een lager pitje heeft gestaan. Lieve Elzendalers, ondanks dat we flink verspreid over het land wonen, voelt het altijd weer als vanouds als we elkaar zien. Irene, een huis naast elkaar is jammer genoeg niet gelukt. Fijn dat je nu als paranimf naast me wil staan. Wat was die Utrecht tijd toch eigenlijk relaxed. Zeist en Nijmegen is met banen en gezinnen erbij toch best een afstand. Ik hoop dat we nog heel lang vriendinnen blijven. Lieve DMeiden, wat een fantastische studietijd hebben we gehad. Jullie vriendschap is me ontzettend dierbaar. We hebben al zo'n 20 jaarlijkse weekendjes en sinterkersten gehad en ik hoop op minstens het dubbele. Irene, speciaal bedankt voor de herintroductie in het Nijmeegse. Dat was wel nodig aangezien ik op een studentenfiets vertrok en terug kwam in een volle gezinsauto. Lieve ex-DHV-/Utrecht-groep, wat tof dat jullie meekwamen met Koen en wat heerlijk dat daar (bijna) geen medico's rondlopen.

Mijn ouders, Margje, Meindert, Elly & Ger, Saskia & Peter, neefjes en nichtje wil ik danken voor de afleiding, steun en interesse. Jullie wisten dat ik bezig was met een promotie, maar hebben vooral de drukte kunnen zien die dat met zich meebracht en zeker het laatste jaar mijn beperktere aanwezigheid daardoor. Lieve ouders, zonder jullie was ik niet gekomen waar ik nu ben. Het reizen is me met de paplepel ingegoten en jullie gingen er altijd maar vanuit dat ik niet in zeven sloten tegelijk zou lopen. (Schoon)ouders, het is een grote luxe om een beroep op jullie te kunnen doen als oppas back-up. Bedankt! Speciaal dank voor mijn moeder die vooral het laatste jaar veel oppas overuren heeft gemaakt waarbij ook ingewikkelde planningen altijd prima waren. Lieve Wout, Hille en Tom, jullie zijn het mooiste geschenk en ik ben trots op jullie. Jullie maken ogenschijnlijk belangrijke zaken relatief en kleine dagelijkse bezigheden waardevol. Zo lijkt dit boekje misschien belangrijk omdat ik hierdoor vaak afwezig was en mijn verloven eraan zijn opgegaan, maar niets is minder waar. Dankzij jullie brengt elke dag weer nieuwe verrassingen en enthousiaste verwonderingen met zich mee. Ik verheug me op meer tijd met zijn allen. Lieve Koen, zonder jou was het zeker nooit gelukt. Achteraf gezien was het best een uitdaging om als kersverse reumatoloog op twee locaties te gaan werken in combinatie met de start van een promotieonderzoek terwijl onze derde telg net geboren was en wij voor onze banen allebei forensden vanuit Arnhem. De laatste jaren zijn geleidelijk aan steeds meer gezinstaken naar jou toegegaan waarbij je vooral het laatste jaar eigenlijk het hele reilen en zeilen van ons gezin op je hebt genomen. Vanaf nu ga ik geen sabbatical en ouderschapsverlof meer opnemen om extra onderzoekstijd te creëren en ga ik het inzetten voor ons vijfjes. Bedankt voor al je optimisme, steun, relativeringsvermogen en flexibiliteit. Met je humor tover je altijd weer een lach op mijn gezicht. Maar pas op met die 1 april grappen waarbij zelfs deze promotie niet door leek te gaan. Onze reis startte in 2001 en een jaar later vertrokken we als backpackers naar het adembenemende Nepal. De prachtige bergen hebben ons sindsdien nooit meer losgelaten en reizen doen we nog steeds erg graag. We kochten een huis in Utrecht en mijn werk bracht ons naar Arnhem en Nijmegen. Deze promotiereis is nu afgelopen en ik hoop dat wij samen nog heel veel mooie reizen tegemoet gaan.

Publications

Publications:

1. **E.A.M. Mahler**, N. Boers, J.W.J. Bijlsma, F.H.J. van den Hoogen, A.A. den Broeder, C.H.M. van den Ende. Patient Acceptable Symptom State in knee osteoarthritis patients succeeds across different patient-reported outcome measures assessing physical function, but fails across other dimensions and rheumatic diseases. J Rheumatol. 2018;45:122–7.
2. **E.A.M. Mahler**, A.A. den Broeder, T.G. Woodworth, V.J.J.F. Busch, F.H.J. van den Hoogen, J.W.J. Bijlsma, C.H.M. van den Ende. How should worsening in osteoarthritis be defined? Development and initial validation of preliminary criteria for clinical worsening in knee and hip osteoarthritis. Scand J Rheumatol. 2017;46:396–406.
3. **E.A.M. Mahler**, N. Cuperus, J.W.J. Bijlsma, T.P.M. Vliet Vlieland, F.H.J. van den Hoogen, A.A. den Broeder, C.H.M. van den Ende. Authors' reply. Scand J Rheumatol. 2017;46:169–170.
4. **E.A.M. Mahler**, N. Cuperus, J.W.J. Bijlsma, T.P.M. Vliet Vlieland, F.H.J. van den Hoogen, A.A. den Broeder, C.H.M. van den Ende. Responsiveness of four patient-reported outcome measures to assess physical function in patients with knee osteoarthritis. Scand J Rheumatol. 2016;45:518–27.
5. M.J. Minten, **E.A.M. Mahler**, A.A. den Broeder, J.W. Leer, C.H.M. van den Ende. The efficacy and safety of low-dose radiotherapy on pain and functioning in patients with osteoarthritis: a systematic review. Rheumatol Int. 2016;36:133–42.
6. N. Cuperus, **E.A.M. Mahler**, T.P.M. Vliet Vlieland, T.J. Hoogeboom, C.H.M. van den Ende. Measurement properties of the Health Assessment Questionnaire Disability Index for generalized osteoarthritis. Rheumatology (Oxford). 2015;54:821–6.
7. N. Cuperus, T.P.M. Vliet Vlieland, **E.A.M. Mahler**, C.C. Kersten, T.J. Hoogeboom, C.H.M. van den Ende. The clinical burden of generalized osteoarthritis represented by self-reported health-related quality of life and activity limitations: a cross-sectional study. N Cuperus, TP Vliet Vlieland, EAM Mahler, CC Kersten, TJ Hoogeboom, CH van den Ende. Rheumatol. Int. 2015;35:871–7.
8. **E.A.M. Mahler**, M.C. Zweers, P.L. van Lent, A.B. Blom, F.H.J. van den Hoogen, W.B. van den Berg, J. Roth, T. Vogl, J.W.J. Bijlsma, C.H.M. van den Ende, A.A. den Broeder. Association between serum levels of the proinflammatory protein S100A8/A9 and clinical and structural characteristics of patients with established knee, hip and hand osteoarthritis. Scand J Rheumatol. 2015;44:56–60.
9. K. Bevers, M.C. Zweers, C.H.M. van den Ende, H.A. Martens, **E.A.M Mahler**, J.W.J. Bijlsma, R.J. Wakefield, F.H.J. van den Hoogen, A.A. den Broeder. Ultrasonic analysis in knee osteoarthritis: evaluation of inter-observer reliability. Clin Exp Rheumatol. 2012;30:673–8.
10. **E.A.M. Mahler**, M. Blom, N.C. Voermans, B.G. van Engelen, P.L.C.M. van Riel, M.C. Vonk. Rituximab treatment in patients with refractory inflammatory myopathies. Rheumatology (Oxford). 2011;50:2206–13.
11. J. Engel, **E. Mahler**, L. Anteunis, E. Marres and G. Zielhuis. Why are NICU infants at risk for chronic otitis media with effusion? Int J Pediatr Otorhinolaryngol. 2001;57:137–44.

Submitted

12. **E.A.M. Mahler**, A.A. den Broeder, N. den Broeder, J.W.J. Bijlsma, G.F. Snijders, F.H.J. van den Hoogen, C.H.M. van den Ende. Short term clinical worsening is a clear predictor for worsening at 2 years in established knee and hip osteoarthritis.
13. **E.A.M. Mahler***, M.J.M. Minten*, M.M. Leseman-Hoogenboom, J.W.H Leer, P.M.P. Poortmans, S.S. Boks, F.H.J. van den Hoogen, A.A. den Broeder, C.H.M. van den Ende
14. M.J.M. Minten, S.S. Boks, **E.A.M. Mahler**, M. Kloppenburg, F.H.J. van den Hoogen, A.A. den Broeder, C.H.M. van den Ende. Exploring associations between local pain and inflammatory and structural aspects in patients with knee osteoarthritis.

*both authors contributed equally. Effectiveness of low-dose radiation therapy on symptoms in patients with knee osteoarthritis: a randomised, double-blinded, sham-controlled trial.

Curricilum vitae

Curriculum vitae

Elien Mahler werd geboren op 26 december 1975 in Sint Anthonis, Noord-Brabant. Na het behalen van haar VWO-diploma aan het Elzendaalcollege te Boxmeer, werd ze uitgeloot voor Geneeskunde en begon ze in 1994 met de studie Biomedische Wetenschappen aan de Radboud Universiteit (RU) in Nijmegen. Zij voltooide haar onderzoeksstages aan de Cancer Council Epidemiology Research Unit in Sydney en de afdeling Epidemiologie van de RU. In 1998 behaalde ze haar MSc met als hoofdrichting epidemiologie. Daarna stroomde ze door in het verkorte doctoraal programma van Geneeskunde (RU) en studeerde in 2000 af. Tijdens de wachttijd voor de coschappen liep zij stage in het Cerrahpasa ziekenhuis in Istanbul en werkte zij bij het Integraal Kankercentrum Nederland in Nijmegen. Haar afsluitende coschap deed zij in Techiman (Ghana). Na het behalen van haar artsexamen startte ze begin 2003 als ANIOS reumatologie bij de Sint Maartenskliniek in Nijmegen (supervisie dr. M.J.A.M. Franssen en dr. M.E.C. Jeurissen). Later dat jaar begon ze aan de vooropleiding Interne Geneeskunde in het Jeroen Bosch ziekenhuis (JBZ) in Den Bosch (opleider dr. P.M. Netten). Ze vervolgde de opleiding tot reumatoloog in Rijnstate (opleider dr. M. Janssen) en Radboudumc (opleider prof. dr. P.L.C.M. van Riel). Zij was actief in diverse bestuursfuncties bij onder andere de studententennisvereniging, Medische Faculteits Vereniging Nijmegen, opleidingscommissie JBZ en het juniorenbestuur van de Nederlandse Vereniging voor Reumatologie waarbij ze deel uitmaakte van de Beroeps Belangen Commissie en het Concilium. In 2010 startte ze als reumatoloog in de Sint Maartenskliniek op de locaties Nijmegen en Boxmeer, waarna ze in 2011 startte met het onderzoek beschreven in dit proefschrift. De resultaten zijn op diverse internationale congressen gepresenteerd. Haar onderzoeksgebieden zijn artrose, artritis psoriatica en spondyloartritis. Zij is geregistreerd als epidemioloog A en volgt het opleidingstraject tot epidemioloog B (opleider prof. dr. A.L.M. Verbeek, Health Evidence RU) dat ze naar verwachting in 2018 afrondt. In 2015 werd zij plaatsvervangend opleider en in 2016 opleider in de Sint Maartenskliniek.

Elien woont samen met Koen en hun drie kinderen Wout (2005), Hille (2007) en Tom (2010).

Theses Sint Maartenskliniek

Theses Sint Maartenskliniek

De Rooij, D. (1988). *Clinical and serological studies in the connective tissue diseases.* University of Nijmegen, Nijmegen, The Netherlands.

Geurts, A. (1992). *Central adaptation of postural organization to peripheral sensorimotor impairments.* University of Nijmegen, Nijmegen, The Netherlands.

Van Lankveld, W. (1993). *Coping with chronic stressors of rheumatoid arthritis*. University of Nijmegen, Nijmegen, The Netherlands.

Tromp, E. (1995). *Neglect in action: a neuropsychological exploration of some behavioural aspects of neglect.* University of Nijmegen, Nijmegen, The Netherlands.

Van Balen, H. (1997). *A disability-oriented approach to long-term sequelae following traumatic brain injury. Neuropsychological assessment for post-acute rehabilitation.* University of Nijmegen, Nijmegen, The Netherlands.

De Kleuver, M. (1998). *Triple osteotomy of the pelvis.An anatomical, biomechanical and clinical study.* University of Nijmegen, Nijmegen, The Netherlands.

Hochstenbach, J. (1999). *The cognitive, emotional, and behavioural consequenses of stroke.* University of Nijmegen, The Netherlands.

Donker, S. (2002). *Flexibility of human walking: a study on interlimb coordination.* Groningen University, Groningen, The Netherlands.

Hendricks, H. (2003). *Motor evoked potentials in predicting motor and functional outcome after stroke.* University of Nijmegen, Nijmegen, The Netherlands.

Hosman, A. J. F. (2003). *Idiopathic thoracic spinal deformities and compensatory mechanisms.* University of Nijmegen, Nijmegen, The Netherlands.

Jongerius, P. (2004). B*otulinum toxin type-A to treat drooling. A study in children with cerebral palsy.* Radboud University, Nijmegen, The Netherlands.

Van de Crommert, H. (2004). *Sensory control of gait and its relation to locomotion after a spinal cord injury.* Radboud University, Nijmegen, The Netherlands.

Van der Linde, H. (2004). *Prosthetic prescription in lower limb amputation. Development of a clinical guideline in the Netherlands*. Groningen University, Groningen, The Netherlands.

De Haart, M. (2005). R*ecovery of standing balance in patients with a supratentorial stroke.* Radboud University, Nijmegen, The Netherlands.

Den Otter, R. (2005). *The control of gait after stroke: an electromyographic approach to functional recovery.* Groningen University, Groningen, The Netherlands.

Weerdesteyn, V. (2005). *From the mechanisms of obstacle avoidance towards the prevention of falls.* Radboud University, Nijmegen, The Netherlands.

Baken, B. (2007). *Reflexion on reflexes. Modulation during gait.* Radboud University, Nijmegen, The Netherlands.

Gaasbeek, R. (2007). *High tibial osteotomy. Treatment of varus osteoarthritis of the knee.* Radboud University, Nijmegen, The Netherlands.

Koëter, S. (2007). *Patellar instability. Diagnosis and treatment.* Radboud University, Nijmegen, The Netherlands.

Langeloo, D. (2007). *Monitoring the spinal cord during corrective spinal surgery: a clinical study of TES-MEP.* Radboud University, Nijmegen, The Netherlands.

Ruiter, M. (2008). *Speaking in ellipses. The effect of a compensatory style of speech on functional communication in chronic agrammatism*. Radboud University, Nijmegen, The Netherlands.

Van den Bemt, B. (2009). *Optimizing pharmacotherapy in patients with rheumatoid arthritis: an individualized approach*. Radboud University, Nijmegen, The Netherlands.

Van Nes, I. (2009). *Balance recovery after supratentorial stroke. Influence of hemineglect and the effects of somatosensory stimulation.* Radboud University, Nijmegen, The Netherlands.

Aarts, P. (2010). *Modified constraint-induced movement therapy for children with unilateral spastic cerebral palsy: the Pirate group intervention.* Radboud University, Nijmegen, The Netherlands.

Groen, B. (2010). *Martial arts techniques to reduce fall severity.* Radboud University, Nijmegen, The Netherlands.

Van Koulil, S. (2010). *Tailored cognitive behavioral therapy in fibromyalgia.* Radboud University, Nijmegen, The Netherlands.

Boelen, D. (2011). *Order out of chaos? Assessment and treatment of executive disorders in brain-injured patients*. Radboud University, Nijmegen, The Netherlands.

Heesterbeek, P. (2011). *Mind the gaps! Clinical and technical aspects of PCL-retaining total knee replacement with the balanced gap technique.* Radboud University, Nijmegen, The Netherlands.

Hegeman, J. (2011). *Fall risk and medication. New methods for the assessment of risk factors in commonly used medicines.* Radboud University, Nijmegen, The Netherlands.

Smulders, E. (2011). *Falls in rheumatic diseases. Risk factors and preventive strategies in osteoporosis and rheumatoid arthritis.* Radboud University, Nijmegen, The Netherlands.

Snijders, G. (2011). *Improving conservative treatment of knee and hip osteoarthritis.* Radboud University, Nijmegen, The Netherlands.

Vriezekolk, J. (2011). *Targeting distress in rheumatic diseases.* Utrecht University, Utrecht, The Netherlands.

Willems, P. (2011). *Decision making in surgical treatment of chronic low back pain. The performance of prognostic tests to select patients for lumbar spinal fusion.* Maastricht University, Maastricht, The Netherlands.

Beijer, L. (2012). *E-learning based speech therapy (EST). Exploring the potentials of E-health for dysarthric speakers.* Radboud University, Nijmegen, The Netherlands.

Hoogeboom, T. (2012). *Tailoring conservative care in osteoarthritis.* Maastricht University, Maastricht, The Netherlands.

Brinkman, M. (2013). F*ixation stability and new surgical concepts of osteotomies around the knee.* Utrecht University, Utrecht, The Netherlands.

Kwakkenbos, L. (2013). *Psychological well-being in systemic sclerosis: Moving forward in assessment and treatment.* Radboud University, Nijmegen, The Netherlands.

Severens, M. (2013). *Towards clinical BCI applications: assistive technology and gait rehabilitation.* Radboud University, Nijmegen, The Netherlands.

Stukstette, M. (2013). *Understanding and treating hand osteoarthritis: a challenge.* Utrecht University, Utrecht, The Netherlands.

Van der Maas, A. (2013). *Dose reduction of TNF blockers in Rheumatoid Arthritis: clinical and pharmacological aspects.* Radboud University, Nijmegen, The Netherlands.

Zedlitz, A. (2013). *Brittle brain power. Post-stroke fatigue, explorations into assessment and treatment.* Radboud University, Nijmegen, The Netherlands.

Koenraadt, K. (2014). *Shedding light on cortical control of movement*. Radboud University, Nijmegen, The Netherlands.

Smink, A. (2014). *Beating Osteoarthritis. Implementation of a stepped care strategy to manage hip or knee osteoarthritis in clinical practice.* VU University Amsterdam, Amsterdam, The Netherlands.

Stolwijk, N. (2014). *Feet 4 feet. Plantar pressure and kinematics of the healthy and painful foot.* Radboud University, Nijmegen, The Netherlands.

Van Kessel, M. (2014). *Nothing left? How to keep on the right track. Spatial and non-spatial attention processes in neglect after stroke.* Radboud University, Nijmegen, The Netherlands.

Altmann, V. (2015). *Impact of trunk impairment on activity limitation with a focus on wheelchair rugby.* Leuven University, Leuven, Belgium.

Bevers, K. (2015). *Pathophysiologic and prognostic value of ultrasonography in knee osteoartrhitis.* Utrecht University, Utrecht, The Netherlands.

Cuperus, N. (2015). *Strategies to improve non-pharmacological care in generalized osteoarthritis.* Radboud University, Nijmegen, The Netherlands.

Kilkens, A. (2015). *De ontwikkeling en evaluatie van het Communicatie Assessment & Interventie Systeem (CAIS) voor het aanleren van (proto-)imperatief gedrag aan kinderen met complexe ontwikkelingsproblemen.* Radboud University, Nijmegen, The Netherlands.

Penning, L. (2015). *The effectiveness of injections in cuffdisorders and improvement of diagnostics.* Maastricht University, Maastricht, The Netherlands.

Stegeman, M. (2015). *Fusion of the tarsal joints: outcome, diagnostics and management of patient expectations.* Utrecht University, Utrecht, The Netherlands.

Van Herwaarden, N. (2015). *Individualised biological treatment in rheumatoid arthritis.* Utrecht University, Utrecht, The Netherlands.

Wiegant, K. (2015). *Uitstel kunstknie door kniedistractie.* Utrecht University, Utrecht, The Netherlands.

Willems, L. (2015). *Non-pharmacological care for patients with systemic sclerosis.* Radboud University, Nijmegen, The Netherlands.

Witteveen, A. (2015). *The conservative treatment of ankle osteoarthritis.* University of Amsterdam, Amsterdam, The Netherlands.

Zwikker, H. (2015). *All about beliefs. Exploring and intervening on beliefs about medication to improve adherence in patients with rheumatoid arthritis.* Radboud University, Nijmegen, The Netherlands.

Lesuis, N. (2016). *Quality of care in rheumatology. Translating evidence into practice.* Radboud University, Nijmegen, The Netherlands.

Luites, J. (2016). *Innovations in femoral tunnel positioning for anatomical ACL reconstruction.* Radboud University, Nijmegen, The Netherlands.

Pakvis, D. (2016). *Survival, primary stability adn bone remodeling assessment of cementless sockets. An appraisal of Wolff's law in the acetabulum.* Radboud University, Nijmegen, The Netherlands.

Schoenmakers, K. (2016). *Prolongation of regional anesthesia. Determinants of peripheral nerve block duration.* Radboud University, Nijmegen, The Netherlands.

Van Hooff, M. (2017). *Towards a paradigm shift in chronic low back pain? Identification of patient profiles to guide treatment.* VU University Amsterdam, Amsterdam, The Netherlands.

Remijn, L. (2017). *Mastification in children with cerebral palsy.* Radboud University, Nijmegen, The Netherlands.

Selten, E. (2017). *Beliefs underlying treatment choices in osteoarthritis.* Radboud University, Nijmegen, The Netherlands.

Bouman, C. (2018). *Dose optimisation of biologic DMARDs in rheumatoid arthritis - Long-term effects and possible predictors.* Radboud University, Nijmegen, The Netherlands.

Tweehuysen, L. (2018). *Optimising biological treatment in inflammatory rheumatic diseases. Predicting, tapering and transitioning.* Radboud University, Nijmegen, The Netherlands.