Osteoarthritis and Cartilage 19 (2011) 475–477

Osteoarthritis and Cartilage

Editorial

Introduction to OARSI FDA initiative OAC special edition

Since July 1999, the Food and Drug Administration (FDA) has provided guidance for industry on "Clinical Development Programs for Drugs, Devices, and Biological Products Intended for the Treatment of Osteoarthritis (OA)"[1]. This draft guidance document was intended to assist sponsors who were developing drugs, devices or biological products for OA and included a number of issues for sponsor consideration including the utility of animal models and the measurement of improvement in OA. The draft guidance discussed the types of label claims that could be considered for OA products and provided guidance on the clinical development programs to support these claims. It was recognized that OA is a disease with a complex pathophysiology and thus, multiple clinical outcomes for product claims could be considered such as an improvement in signs and symptoms or a delay in structural progression. Additionally, the 1999 draft guidance also proposed, in principal, a claim for prevention of OA. The FDA solicited comments to the draft document and until 2007 no further update was published. It should be noted that under this draft guidance, no products have been approved in the US for the indications of either delay in structural progression or prevention of OA.

On August 14, 2007, a request for proposals (RFP) was posted by the FDA in the Federal Register seeking an updated critical appraisal on the issues related to clinical development programs for the treatment and prevention of OA that would help inform their internal discussions and subsequent finalization of the 1999 OA draft guidance[2].

In response to this solicitation, the Osteoarthritis Research Society International (OARSI) submitted a proposal outlining a specific approach to the management and coordination of a critical appraisal of the science related to the design of clinical development programs for the treatment and prevention of OA. In June 2008, OARSI received approval from the FDA to embark on an 18-month review of the current literature resulting in a series of recommendations for the FDA's consideration as they embark upon finalizing the 1999 OA draft guidance.

Under the direction of an executive committee and through a series of individual committee meetings and teleconferences as well as two open public meetings, eight working groups comprised of individuals from academia, professional societies, industry and governmental agencies [Appendix 1] addressed specific questions outlined within the FDA's original notice. The result was a comprehensive report encompassing the recommendations by each working group based on the current state of knowledge on the pre-defined topics outlined with the original notice as well as a series of on-going research recommendations to further inform the evolving areas of structural change and the role of biomarkers in the context of clinical trials.

This special edition of Osteoarthritis and Cartilage provides an insightful, evidence based exploration and discussion on important issues related to current and future OA clinical program development. While much has been learned since 1999, OA still remains a disease characterized by a prolonged pre-radiographic phase followed by evident structural joint changes, associated with frequent pain and loss of function. The research discussed herein recognizes how far we have come and charts the course for future research into the development of new therapies and devices for OA, as well as the potential for disease modifying drugs.

Declaration of funding and role of funding source

The OARSI FDA OA Initiative received financial support from the following professional organization:

American College of Rheumatology

Additionally the OARSI FDA OA Initiative received financial support from the following companies:

Amgen
ArthroLab
AstraZeneca
Bayer Healthcare
Chondrometrics
CombinatoRx
Cypress BioScience
DePuy Mitek
Expanscience
4QImaging
Genevrier/IBSA
Genzyme
King (Alpharma)
Merck
Merck Serono
NicOx
Pfizer
Rottapharm
Smith & Nephew
Wyeth

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the FDA.

1063-4584/$ – see front matter
doi:10.1016/j.joca.2010.12.013

Acknowledgment

We gratefully acknowledge financial support received from following:

Professional organization: American College of Rheumatology.

Industry: Amgen Genevrier/IBSA, ArthroLab Genzyme, AstraZeneca King (Alpharma), Bayer Healthcare Merck, Chondrometrics Merck Serono, CombinatoRx NicOx, Cypress BioScience Pfizer, DePuy Mitek Rottapharm, Expanscience Smith & Nephew, 4QImaging Wyeth.

Appendix 1. Membership of OARSI FDA initiative committees

Executive Committee/Steering Committee
Francis Berenbaum, MD, PhD
Steven Abramson, MD
Marc Hochberg, MD, MPH
Roland Moskowitz, MD
David Hunter, MD, PhD
Gillian Hawker, MD, MSc
C. Kent Kwoh, MD
Maxime Dougados, MD

Industry Representatives to Steering Committee
Richard Leff, MD
Sean Curtis, MD, MPH

Consultant to Executive Committee
Lee S. Simon, MD

Definition of Disease State
Nancy Lane, MD (Chair)
Kenneth Brandt, MD (Chair)
Paul Dieppe, MD
Michael Doherty, MD
John Esdaile, MD
Gillian Hawker, MD
Marc Hochberg, MD, MPH
Alan Silman, MD
Shepard Hurwitz, MD
Gayle Lester, PhD
James Witter, MD, PhD
John Hardin, MD
Wayne Tsuji, MD
Edward Schreyer, BS
Elena Peeva, MD, MSc
Alexandre Valentin, MD
Wim Scheele, MD
Richard Polisson, MD, MHSc
Bernard Guillou, MD
Massimo D'Amato, MD
Drew Burdon, BSc, PhD
Srinivas Rao, MD, PhD
Martin Michaelis, MD
Brooks Story, PhD
Matt Fisher, PharmD

Safety Considerations
Lee S. Simon, MD (Chair)
Vibeke Strand, MD
Steven Abramson, MD
Francis Berenbaum, MD, PhD
Roy Altman, MD
Clifton Bingham, III, MD
Daniel Furst, MD
Kenneth Saag, MD, MSc
James Herndon, MD, MBA
Phillip Davidson, MD
Paul Peloso, MD, MSc
David Hovland, PhD
Richard Leff, MD
Judith Johnson, MD
Gianfranco Caselli, PhD
Stephen Wax, MD, PhD
Graeme Howling, PhD
Macil Krieser, PhD
Eileen Barry, PharmD

Claim of Symptomatic Relief
Allan Gibofsky, MD, JD (Chair)
Robert Dworkin, PhD (Chair)
Lee S. Simon, MD
Maarten Boers, MSc, MD, PhD
James McWilliam, MD
James Witter, MD
Gary Williams, PhD
Joe Stauffer, DO
Mark Sostek, MD
Paul Peloso, MD, MSc
Marie Pierre Hellio Le Graverand-Gastineau, MD, PhD, DSc
Sonya Glasson, B.V.Sc
Christopher Murray, PhD
John Randle, PhD
Lucio Rovati, MD
Michael Gendreau, MD, PhD
Bruno Boezennec, MD
Diann White, BS
Martin Michaelis, MD
Sean Lilienfeld, MD
Shirley Chen, PharmD

Prevention or Risk Reduction
Joanne Jordan, MD, MPH (Chair)
Cyrus Cooper, MD
C. Kent Kwoh, MD
Steven Messier, PhD
Mary Fran Sowers, PhD
Tim Spector, MD, MSc
Weiya Zhang, BSc, MSc, PhD
George Arangio, MD
Robert Pedowitz, MD, PhD
Chad Helmick, MD
Lynn Baird, PhD
Mona Whaba, MD
John Bradley, MD
Jennifer Lee Gardiner, PhD
Nathan Bachtell, MD
Lucio Rovati, MD
Yuan Wang, PharmD
Sean Lilienfeld, MD

Assessment of Structural Change
Philip Conaghan, MB, BS, PhD (Chair)
David Hunter, MD, PhD
Jeffrey Duryea, PhD
Garry Gold, MD
Steven Mazzuca, PhD
Jean Francis Maillefert, MD
Timothy Mosher, MD
Hollis Potter, MD
David Felson, MD, MPH
Ali Guermazi, MD
Helen Keen, MD
Gayle Lester, PhD
Wayne Tsuji, MD
John Randle, PhD
Felix Eckstein, MD
Erika Schneider, PhD
Elena Losina, PhD
Sarah Kingsbury, PhD
William Reichman, PhD
Jean Pierre Pelletier, MD
Saara Totterman, MD, PhD
Rose Maciewicz, PhD
Bernard Dardzinski, PhD
Mona Wahba, MD
Marie Pierre Hellio Le Graverand-Gastineau, MD, PhD, DSc
Elisabeth Morris, DVM
Jeffrey Kraines, MD
Lucio Rovati, MD
Don Dreher, MD, PhD
James Huckle, PhD
Mary-Ann Preston, PhD
Brooks Story, PhD

Biomarkers
Virginia Kraus, MD, PhD (Chair)
David Eyre, PhD
Dick Reingard, MD, PhD
Yves Henrotin, PhD
Stefan Lohmander, MD, PhD
Roland Moskowitz, MD
Robin Poole, PhD, DSc
John Hardin, MD
Bruce Burnett, PhD
Patrick Garnero, PhD, DSc
Sue Cottrell, PhD
Javier Coindreau, MD
Jennifer Lee Gardiner, PhD
Gloria Matthews, DVM, PhD
Stefano Persiani, PhD
Michael Gendreau, MD, PhD
Martin Todman, PhD
Joseph Menetski, PhD
Bill Parrish, PhD
Matt Fisher, PharmD

Devices
Victor Goldberg, MD (Chair)
Joseph Buckwalter, MD
Michael Pinzur, MD
William Jiranek, MD
William Macaulay, MD
Russell Windsor, MD
Thomas Vail, MD
Timothy Wright, PhD
Peter Walker, PhD
Nizar Mahomed, MD
William Mihalko, MD, PhD
John Bradley, MD
Michael Halpin, MEngr
Antonino Santoro, PhD
Barbara Rohan
Izi Bruker, PhD
Anish Patel, PhD

Statistical Considerations
Daniel Bloch, PhD (Chair)
Elena Losina, PhD
Amy Ko, PhD
Eric Chi, PhD
Melissa Nichols, PhD
Clare Elkins, PhD
Giampaolo Giacovelli, PhD

References

1. www.fda.gov/.../GuidanceComplianceRegulatoryInformation/Guidances/ucm071577.pdf; [accessed 10.05.10].
2. http://www.access.gpo.gov/su_docs/fedreg/a070814c.html; Health and Human Services Department, Food and Drug Administration [Human drugs, biological products, and medical devices] [accessed 10.05.10].

S.B. Abramson†* F. Berenbaum‡, M.C. Hochberg§, R.W. Moskowitz||
†*Department of Rheumatology and Medicine, New York University School of Medicine, New York, NY, USA*

‡*Department of Rheumatology, Paris VI University, Saint-Antoine Hospital, APHP, Paris, France*

§*Division of Rheumatology and Clinical Immunology, University of Maryland School of Medicine, Baltimore, MD, USA*

||*Division of Rheumatology, University Hospitals, Case Western Reserve School of Medicine, Cleveland, OH, USA*

* Address correspondence and reprint requests to: Steven B. Abramson, Department of Rheumatology and Medicine, New York University School of Medicine, New York, NY, USA.
E-mail address: madeline.rios@nyumc.org (S.B. Abramson)

Osteoarthritis and Cartilage 19 (2011) 478–482

Osteoarthritis and Cartilage

OARSI-FDA initiative: defining the disease state of osteoarthritis

N.E. Lane †*, K. Brandt ‡, G. Hawker §, E. Peeva ||, E. Schreyer ¶, W. Tsuji #, M.C. Hochberg ††

† University of California at Davis Medical Center, Sacramento, CA, USA
‡ University of Kansas Medical Center, Kansas City, KS, USA
§ Women's College Hospital, University of Toronto, Toronto, Canada
|| Merck Research Laboratories, Whitehouse Station, NJ, USA
¶ 4QImaging, Rochester, NY, USA
# Amgen Inc, University of Washington, Seattle, WA, USA
†† University of Maryland School of Medicine, Baltimore, MD, USA

ARTICLE INFO

Article history:
Received 16 June 2010
Accepted 20 September 2010

Keywords:
Biomarker
Clinical trial design
MRI
OA phenotype
Patient-reported outcome

SUMMARY

Objective: To respond to a pre-specified set of questions posed by the United States Food and Drug Administration (FDA) on defining the disease state to inform the clinical development of drugs, biological products, and medical devices for the prevention and treatment of osteoarthritis (OA).
Methods: An Osteoarthritis Research Society International (OARSI) Disease State working group was established, comprised of representatives from academia and industry. The Working Group met in person and by teleconference on several occasions from the Spring of 2008 through the Autumn of 2009 to develop consensus-based, evidence-informed responses to these questions. A report was presented at a public forum in December 2009 and accepted by the OARSI Board of Directors in the Summer of 2010.
Results: An operational definition of OA was developed incorporating current understanding of the condition. The structural changes that characterize OA at the joint level were distinguished from the patients' experience of OA as the 'disease' and 'illness', respectively. Recommendations were made regarding the evaluation of both in future OA clinical trials. The current poor understanding of the phenotypes that characterize OA was identified as an important area for future research.
Conclusions: The design and conduct of clinical trials for new OA treatments should address the heterogeneity of the disease, treatment-associated structural changes in target joints and patient-reported outcomes.

Introduction

In 2007, the United States Food and Drug Administration (FDA) published a request for proposals to conduct a critical appraisal providing information on various issues related to clinical development programs for drugs, biological products, and medical devices for the prevention and treatment of osteoarthritis (OA). The Osteoarthritis Research Society International (OARSI) was selected to prepare the response to these issues and, in 2008, constituted an Executive Committee, Steering Committee and eight Working Groups to address questions posed by the FDA[1]. One of these Working Groups, "Defining Disease State" was asked to provide recommendations based on current evidence on a pre-specified set of issues:

- What is OA?
- How do we define OA for the purposes of treatment or prevention?
- Are oligoarticular, monoarticular and polyarticular OA the same disease?
- Is hand OA different than hip OA and knee OA?
- Where does degenerative disc disease (DDD) fit in?
- How many sites need to be studied for approval of: a systemic (oral) therapy? How many for a local therapy?
- Should there be uniform inclusion and exclusion criteria for OA clinical trials?
- What is the research agenda required to inform each of the above questions?

This document summarizes the group's recommendations to the FDA.

* Address correspondence and reprint requests to: N.E. Lane, Endowed Professor of Medicine and Rheumatology, Department of Medicine, University of California at Davis School of Medicine, Sacramento, CA 95817, USA. Tel: 1-916-734-0758; Fax: 1-916-734-4773.
E-mail address: nelane@ucdavis.edu (N.E. Lane).

1063-4584/$ – see front matter
doi:10.1016/j.joca.2010.09.013

Methods

The Disease State Working Group was assembled by the project Steering and Executive Committees to ensure adequate representations from academia and industry. The Working Group met in person and by teleconference on several occasions from the Spring of 2008 through the Autumn of 2009 to develop consensus-based, evidence-informed responses to these questions. A report was presented at a public forum in December 2009 and accepted by the OARSI Board of Directors in the Summer of 2010.

Results

What is OA?

Defining OA has important implications for prevention, diagnosis, and treatment of this condition. Based on evidence to date, there was consensus that OA is usually a progressive disease of synovial joints that represents failed repair of joint damage that results from stresses that may be initiated by an abnormality in any of the synovial joint tissues, including articular cartilage, subchondral bone[2–4], ligaments, menisci (when present)[5,6], periarticular muscles[7], peripheral nerves, or synovium[8,9]. This ultimately results in the breakdown of cartilage and bone[10], leading to symptoms of pain, stiffness and functional disability[11]. Abnormal intra-articular stress and failure of repair may arise as a result of biomechanical[12], biochemical[13] and/or genetic factors[14]. This process may be localized to a single joint, a few joints, or generalized, and the factors that initiate OA likely vary depending on the joint site. The complexity and variability of OA etiology suggests the need for patient-specific, etiology-based treatment.

How do we define OA for the purposes of treatment or prevention?

While late-stage OA is often characterized by both demonstrable structural damage and patient reports of joint pain, stiffness and disability[17]. There is only a weak correlation between symptoms and pathology, particularly in early stages of the disease[18]. Further, FDA-approved treatments directed at reducing the symptoms of OA have not been shown, to date, to prevent ongoing joint structural damage. For this reason, the Working Group felt that future development of treatments for OA should consider the effects of the treatment on the structural changes at the joint level (the *disease* OA) separately from the effects on patient-reported symptoms (the *illness* OA). Future pharmacotherapy for OA may therefore be considered to be 'structure modifying' (i.e., designed to prevent the development of joint failure), symptom modifying, or both.

Classification of patients based on the presence/absence of the *disease* (structural changes demonstrated on imaging studies) and the *illness* (patient-reported symptoms of OA) may be useful in trial design and recruitment. For example, in the absence of longitudinal observations of asymptomatic non-arthritic individuals who have recognized risk factors for OA, it may be difficult to differentiate pathology that increases intra-articular stress and thus the risk for OA (e.g., abnormalities in bone shape, such as subtle acetabular dysplasia) from pathologic changes of OA that are a consequence of the damage caused by an increase in intra-articular stress (e.g., bony remodeling that may occur in response to abnormal stress or meniscal degeneration)[19]. Studying subjects with neither structural changes nor symptoms would be most appropriate for primary prevention studies, those with only the 'disease' for interventions designed to prevent symptomatic OA, and studies of those with both the illness and disease for treatments designed to prevent joint failure, defined, for example, as the need for total joint replacement.

Considerations regarding the disease

The proposed definition of OA, above, incorporates the current understanding of the role of intra-articular stress, which may result from abnormal biomechanical forces, in the etiology of OA. In knee OA, biomechanical changes, including varus/valgus angulation[20] and rotational abnormalities after acute anterior cruciate ligament injury[21], contribute to OA progression and may serve as useful biomarkers of structural damage[22]. Due to the strong influence of these biomechanical changes on OA progression, the type and the degree of abnormality (e.g., varus/valgus deformity) should be included in the definition of OA since they may influence treatment outcome.

However, while substantial progress has been made in this area, knowledge gaps remain regarding the influence of corrections to joint loading on structural changes in OA, including whether or not relief of joint pain and improvement in function are outcomes of the structural healing process. Promising work evaluating biomechanics of the knee and hip *via* functional gait analysis, stereo fluoroscopy, and adduction moment analysis may ultimately permit identification of risk factors or markers of joint abnormality that contribute to the *development* of OA[20,23–25]. Adequately powered, randomized, controlled trials in patients with symptomatic knee OA may help to determine whether amelioration of impulsive loads and/or peak dynamic loading reduces joint pain and improves function, and thus could serve as alternatives to pharmacologic therapies.

As OA is characterized by synovial joint abnormalities that may include structural and compositional changes to bone, cartilage, meniscus, synovium, and other soft tissues of the joint[3–5,26,27], defining joint abnormality in OA by plain X-ray evaluation alone does not provide a complete description of the disease; changes to soft tissues in the joint cannot be visualized on plain film X-rays. Indeed, the destruction and loss of cartilage characteristic of knee OA is often inferred from narrowing of the tibio-femoral joint, which reflects loss of tissue from the joint space[28–30]. However, destruction and loss of cartilage and meniscus in the joint space is characteristic of late-stage disease[11]. The Working Group agreed that additional objective criteria, beyond plain radiographs, such as use of other imaging techniques, including magnetic resonance imaging (MRI), and molecular biomarkers – the focus of other FDA OA working groups – are needed to characterize and define the onset and early progression of OA[19], when intervention may be more likely to achieve joint preservation. It is hoped that the definition of OA will continue to evolve to encompass these new biomarkers, once identified and validated, and contribute to our understanding of the pathobiology of OA in different patients.

Considerations regarding the illness

OA is characterized by joint pain, stiffness and functional limitations resulting in reduced participation in valued activities, and downstream effects on fatigue, mood, sleep and overall quality of life[31,32]. Symptom onset may occur years after that for OA *disease*, when structural deterioration is more difficult to treat. The symptoms of OA are largely evaluated using patient self-report measures, scores on which may be influenced by a number of factors, e.g., measures of coping[33], and by performance-based measures of physical functioning. Consideration of the influence of these other factors on treatment response is important in the evaluation of OA treatments.

Traditionally, OA treatment studies have focused on pain intensity and/or physical functioning as their primary outcomes of interest. However, this approach fails to consider the full OA illness

experience. Future trials should incorporate evaluation of the presence and severity of additional common features of OA to determine which aspect(s) of the illness (e.g., pain, sleep quality, depressed mood) that a given intervention may improve, and those for which it is unlikely to provide benefit. This approach may ultimately permit treatment to be matched to the symptoms that are most important to the patient.

Among those who seek help from a physician, joint pain is the most common complaint[34]. Although OA-related pain has traditionally been assumed to be nociceptive in origin, resulting from joint tissue destruction[35], there is increasing evidence from both animal and human studies that neuropathic type pain exists in some individuals with OA, likely due to peripheral and/or central pain sensitization[35,36]. Thus, OA patients may have pain that is predominantly nociceptive, predominantly neuropathic, or of mixed nociceptive/neuropathic etiology. Research is under way to further elucidate pain mechanisms in OA, and to facilitate better characterization of the nature of the pain in OA patients. Ultimately, the goal will be to stratify OA patients in intervention trials on the basis of the type of pain present, with the goal of better targeting pain therapies to individual patients.

OA most commonly affects older adults, who often have co-morbid medical conditions. Both age and co-morbidity may independently influence, or be associated with, pain, fatigue, depressed or anxious mood, muscle strength, fitness, and level of physical activity. While it is often difficult to distinguish OA-related effects from those of other conditions, clinical trials evaluating the efficacy and safety of treatments for OA should consider these issues. The heterogeneity of the pathophysiology of OA may further impact variability in response to treatment outcomes in clinical trials of OA; such heterogeneity may be reduced in the future by improved phenotyping of OA that enables focused recruitment of subjects with the same 'phenotype'.

Are oligoarticular/monoarticular and polyarticular OA the same disease? Is hand OA different from hip OA and from knee OA?

Existing data are insufficient to answer these questions. Which OA joints and/or patterns of OA joint involvement are associated with greater illness than others is not well understood. Furthermore, it remains unclear whether erosive hand OA is part of a spectrum of nodal hand OA or a distinct entity. As noted above, OA has no common etiology; it is likely that the risk factors for both structural changes and symptoms of OA differ not only for different joints, for example, because of differences in the local protective mechanisms[16], but even within a single joint (e.g., varus malalignment leading to medial tibiofemoral changes and lateral meniscal tear inciting lateral tibiofemoral disease). As a result, therapeutic responses are likely to vary.

In this respect, there is a growing consensus that OA is not a single disease of the joints, but rather a collection of diseases with many causes and potential treatments[15,37]. The concept of "OA phenotypes" has been advanced to address the variability of OA with respect to pattern and site of joint involvement (knee, hand, hip, spine) and characteristics (inflammation, mal-alignment, cartilage erosion, osteophyte formation)[38–41]. There is an urgent need for improved characterization of these OA phenotypes; until such work is completed, trials in OA should evaluate a primary joint site, but should ideally also collect data on the presence/absence of structural changes and symptoms at other typical OA sites, e.g., using a joint homunculus. This would enable evaluation of the effect of new interventions on incident OA as well as the number of OA joints that undergo significant progression over the treatment period.

Where does DDD fit in?

Spinal OA refers to the degeneration of the cartilage and surrounding tissues in the synovial facet joints of the spine[42]. DDD describes the deterioration of the intervertebral disc. While advanced DDD can contribute to the onset of spinal OA, and while the two conditions are often seen together[43], DDD and spinal OA are anatomically distinct. DDD treatments currently in development are aimed at restoring disc height. However, it is unclear whether the restoration of disc anatomy will result in significant pain relief[44].

Identifying the source of back pain in patients with facet joint OA and DDD remains a challenge. As for peripheral joint OA, there is a lack of concordance between symptoms and structural changes on radiographs in spinal OA; many patients with advanced facet joint OA and/or DDD have no symptoms, while others with only mild structural changes present with extreme pain and disability[42]. Further complicating matters, many factors other than OA or DDD can contribute to back pain, including vertebral fracture, congenital spinal deformity, and muscle strains and imbalances.

Back pain resulting from DDD, facet joint OA, or other conditions is often present in patients with OA involving other joints, particularly the knee[43]. The relationship between knee pain and back pain is complex and not well understood. Both back and knee pain can result in gait changes that can negatively impact each other. OA is often polyarticular and may be present in facet joints and knees in the same patient. Due to the complex and inconsistent relationships among DDD, facet joint OA, back pain, and extremity pain, DDD should be considered separately. Relief of pain specifically from DDD, as well as therapies directed towards disc restoration, should be addressed in separate clinical trials. However, due to the high frequency of concomitant back pain in individuals with lower extremity OA, and the potential for back pain to influence OA treatment effects, clinical trials should ascertain the presence of back pain as a potential confounder, and possibly the effect of OA therapies on back pain, when present, but only as secondary or tertiary endpoints.

How many joints need to be studied for regulatory approval of an oral therapy? For a local therapy?

An advantage of a systemic oral therapy is the potential to treat multiple arthritic joints in the same patient. A potential disadvantage to this approach is the difficulty in achieving sufficient concentrations for efficacy in the target joints. Local therapies, including intra-articular and topical therapies, can achieve high concentrations in specific joints but may require multiple treatments. In either case, because of the variability in causes and characteristics of OA between joints, efficacy in one joint does not assure efficacy in another. Thus, approval of local therapies should be based on joint-specific efficacy. With systemic treatment, however, more than one joint could be evaluated.

Should there be a uniform definition of inclusion and exclusion criteria in OA clinical trials?

Inclusion and exclusion criteria should be joint-specific, reflecting factors that are known to affect the incidence and progression of OA in that joint. Inclusion and exclusion criteria for clinical trials may also vary, based on the mechanism of action of the drug or device being studied. It is recommended that inclusion/exclusion criteria be similar for therapeutic interventions with the same mechanism of action. To permit comparison between agents, assessment for systemic toxicity should be as similar as possible for all interventions.

What is the research agenda required to inform each of the above questions?

To standardize the evaluation of OA, research is needed to:

1. Define the phenotypes of OA. Such phenotypes should ideally take into consideration the patterns and sites of joint involvement, pathophysiology, clinical presentation (e.g., severity and quality of pain), the presence of specific distinguishing biomarkers, and potentially rate of progression/prognosis. Information on genotypes linked to differences in OA onset and progression may help to further refine the OA phenotypes and improve study subject selection.
2. Gain consensus on a core set of measures to evaluate the spectrum of the *illness* OA beyond pain intensity and physical disability. Specifically, research is needed to develop and test measures to evaluate changes in response to treatment in OA-related fatigue, poor sleep, depressed and anxious mood, and participation in valued activities.

Conclusions

Over the past decade, the paradigm has shifted away from a 'chondrocentric' view of OA, with recognition that OA is a complex disease with no common pathological pathway. However, biomechanical joint stress has a substantial etiologic role. Because risk factors for OA differ not only for different joints, but even within a single joint, therapeutic responses may vary. More recently, the concept of OA as a *disease*, manifest in the structural changes occurring at the joint level, and an *illness*, reflecting the patient's experience living with OA, has evolved, with the recognition that OA interventions may have discordant effects on OA illness vs disease.

Therefore, both should be considered in development and evaluation of new therapeutic interventions. Finally, the heterogeneity of OA as an entity, and the resulting inherent difficulty in classifying patients into distinct subgroups or phenotypes, has prevented the targeting of clinical trials of OA therapies to those OA phenotypes most likely to benefit and may explain, in part, the relatively modest effect sizes of most OA therapies to date[45]. Improved understanding of the phenotypes of OA has the potential to enhance the specificity of treatment selection and is sorely needed.

Author contributions

All of the authors contributed to the research, writing and editing of the manuscript.

Declaration of funding and role of funding source

The OARSI FDA OA Initiative received financial support from the following professional organization:

American College of Rheumatology

Additionally the OARSI FDA OA Initiative received financial support from the following companies:

Amgen
ArthroLab
AstraZeneca
Bayer Healthcare
Chondrometrics
CombinatoRx
Cypress BioScience
DePuy Mitek
Expanscience
4QImaging
Genevrier/IBSA
Genzyme
King (Alpharma)
Merck
Merck Serono
NicOx
Pfizer
Rottapharm
Smith & Nephew
Wyeth

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the FDA.

Conflict of interest

KB: Consultant, Pfizer, Inc.
EP: Employee of Merck.
ES: Employee of 4QImaging.
WT: Employee of Amgen.

References

1. Abramson S, Berenbaum F, Hochberg MC, Moskowitz RW. Introduction to the OARSI FDA OA Initiative. Osteoarthritis Cartilage 2011 Mar 8 [Epub ahead of print].
2. Radin EL, Abernethy PJ, Townsend PM, Rose RM. The role of bone changes in the degeneration of articular cartilage in osteoarthrosis. Acta Orthop Belg 1978;44:55–63.
3. Felson DT, McLaughlin S, Goggins J, LaValley MP, Gale ME, Totterman S, *et al*. Bone marrow edema and its relation to progression of knee osteoarthritis. Ann Intern Med 2003;139:330–6.
4. Oegema Jr TR, Carpenter RJ, Hofmeister F, Thompson Jr RC. The interaction of the zone of calcified cartilage and subchondral bone in osteoarthritis. Microsc Res Tech 1997;37:324–32.
5. Hunter DJ, Zhang YQ, Niu JB, Tu X, Amin S, Clancy M, *et al*. The association of meniscal pathologic changes with cartilage loss in symptomatic knee osteoarthritis. Arthritis Rheum 2006;54:795–801.
6. Crema M, Roemer F, Marra M, Guermazi A, Eckstein F, Hellio Le Graverand MP, *et al*. The association of prevalent medial meniscal mucoid degeneration and tears with cartilage loss in the medial tibiofemoral compartment over a 2-year period assessed with 3.0T MRI. Eur Radiol 2009;19:S247.
7. Bennell KL, Hunt MA, Wrigley TV, Lim BW, Hinman RS. Role of muscle in the genesis and management of knee osteoarthritis. Rheum Dis Clin North Am 2008;34:731–54.
8. Hill CL, Gale DG, Chaisson CE, Skinner K, Kazis L, Gale ME, *et al*. Knee effusions, popliteal cysts, and synovial thickening: association with knee pain in osteoarthritis. J Rheumatol 2001;28:1330–7.
9. Hill CL, Hunter DJ, Niu J, Clancy M, Guermazi A, Genant H, *et al*. Synovitis detected on magnetic resonance imaging and its relation to pain and cartilage loss in knee osteoarthritis. Ann Rheum Dis 2007;66:1599–603.
10. Eckstein F, Burstein D, Link TM. Quantitative MRI of cartilage and bone: degenerative changes in osteoarthritis. NMR Biomed 2006;19:822–54.
11. Link TM, Steinbach LS, Ghosh S, Ries M, Lu Y, Lane N, *et al*. Osteoarthritis: MR imaging findings in different stages of disease and correlation with clinical findings. Radiology 2003;226:373–82.

12. Wilson DR, McWalter EJ, Johnston JD. The measurement of joint mechanics and their role in osteoarthritis genesis and progression. Med Clin North Am 2009;93:67–82.
13. Garnero P, Aronstein WS, Cohen SB, Conaghan PG, Cline GA, Christiansen C, *et al*. Relationships between biochemical markers of bone and cartilage degradation with radiological progression in patients with knee osteoarthritis receiving risedronate: the knee osteoarthritis structural arthritis randomized clinical trial. Osteoarthritis Cartilage 2008;16: 660–6.
14. Dai J, Ikegawa S. Recent advances in association studies of osteoarthritis susceptibility genes. J Hum Genet 2010;55: 77–80.
15. Brandt KD, Dieppe P, Radin EL. Etiopathogenesis of osteoarthritis. Rheum Dis Clin North Am 2008;34:531–59.
16. Brandt KD, Radin EL, Dieppe PA, Van De PL. Yet more evidence that osteoarthritis is not a cartilage disease. Ann Rheum Dis 2006;65:1261–4.
17. McGonagle D, Tan AL, Carey J, Benjamin M. The anatomical basis for a novel classification of osteoarthritis and allied disorders. J Anat 2010;216:279–91.
18. 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.
19. Ding C, Jones G, Wluka AE, Cicuttini F. What can we learn about osteoarthritis by studying a healthy person against a person with early onset of disease? Curr Opin Rheumatol 2010;22:520–7.
20. Sharma L, Song J, Felson DT, Cahue S, Shamiyeh E, Dunlop DD. The role of knee alignment in disease progression and functional decline in knee osteoarthritis. JAMA 2001;286:188–95 [erratum appears in JAMA 2001 Aug 15;286(7):792.].
21. Lohmander LS, Englund PM, Dahl LL, Roos EM. The long-term consequence of anterior cruciate ligament and meniscus injuries: osteoarthritis. Am J Sports Med 2007;35:1756–69.
22. Thorp LE, Sumner DR, Block JA, Moisio KC, Shott S, Wimmer MA. Knee joint loading differs in individuals with mild compared with moderate medial knee osteoarthritis. Arthritis Rheum 2006;54:3842–9.
23. Wilson DR, McWalter EJ, Johnston JD. The measurement of joint mechanics and their role in osteoarthritis genesis and progression. Rheum Dis Clin North Am 2008;34:605–22.
24. Astephen JL, Deluzio KJ, Caldwell GE, Dunbar MJ, Hubley-Kozey CL. Gait and neuromuscular pattern changes are associated with differences in knee osteoarthritis severity levels. J Biomech 2008;41:868–76.
25. Astephen JL, Deluzio KJ, Caldwell GE, Dunbar MJ. Biomechanical changes at the hip, knee, and ankle joints during gait are associated with knee osteoarthritis severity. J Orthop Res 2008;26:332–41.
26. Hunter DJ, Conaghan P, Peterfy C, Bloch D, Guermazi A, Woodworth T, *et al*. Responsiveness, effect size, and smallest detectable difference of magnetic resonance imaging in knee osteoarthritis. Osteoarthritis Cartilage 2006;14(Suppl 1): 112–5.
27. Felson DT, Chaisson CE, Hill CL, Totterman SM, Gale ME, Skinner KM, *et al*. The association of bone marrow lesions with pain in knee osteoarthritis. Ann Intern Med 2001;134:541–9.
28. Mazzuca SA, Brandt KD, Dieppe PA, Doherty M, Katz BP, Lane KA. Effect of alignment of the medial tibial plateau and x-ray beam on apparent progression of osteoarthritis in the standing anteroposterior knee radiograph. Arthritis Rheum 2001;44:1786–94.
29. Mazzuca SA, Brandt KD, Buckwalter KA. Detection of radiographic joint space narrowing in subjects with knee osteoarthritis: longitudinal comparison of the metatarsophalangeal and semiflexed anteroposterior views. Arthritis Rheum 2003;48:385–90.
30. Auleley GR, Duche A, Drape JL, Dougados M, Ravaud P. Measurement of joint-space width in hip osteoarthritis: influence of joint positioning and radiographic procedure. Rheumatology (Oxford) 2001;40:414–9.
31. Hawker GA, French MR, Waugh EJ, Gignac MA, Cheung C, Murray BJ. The multidimensionality of sleep quality and its relationship to fatigue in older adults with painful osteoarthritis. Osteoarthritis Cartilage 2010;18:1365–71.
32. Hawker GA, Gignac MA, Badley E, Davis AM, French MR, Li Y, *et al*. A longitudinal study to explain the pain-depression link in older adults with osteoarthritis. Arthritis Care Res 2010 Jul 26 [Epub ahead of print].
33. Sale JE, Gignac M, Hawker G. The relationship between disease symptoms, life events, coping and treatment, and depression among older adults with osteoarthritis. J Rheumatol 2008;35: 335–42.
34. Gwilym SE, Pollard TC, Carr AJ. Understanding pain in osteoarthritis. J Bone Joint Surg Br 2008;90:280–7.
35. Dieppe PA, Lohmander LS. Pathogenesis and management of pain in osteoarthritis. Lancet 2005;365:965–73.
36. Hochman JR, French MR, Bermingham SL, Hawker GA. The nerve of osteoarthritis pain. Arthritis Care Res 2010;62: 1019–23.
37. Arden N, Nevitt MC. Osteoarthritis: epidemiology. Best Pract Res Clin Rheumatol 2006;20:3–25.
38. Carroll GJ. Polyarticular osteoarthritis–two major phenotypes hypothesized. Med Hypotheses 2006;66:315–8.
39. Valdes AM, McWilliams D, Arden NK, Doherty SA, Wheeler M, Muir KR, *et al*. Involvement of different risk factors in clinically severe large joint osteoarthritis according to the presence of hand interphalangeal nodes. Arthritis Rheum 2010;62: 2688–95.
40. Kerkhof HJ, Meulenbelt I, Akune T, Arden NK, Aromaa A, Bierma-Zienstra SM, *et al*. Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. Osteoarthritis Cartilage 2010 Nov 6 [Epub ahead of print].
41. Akune T, Kawaguchi H. Human genetic studies on osteoarthritis from the clinicians' viewpoint. Osteoarthritis Cartilage 2010 Dec 21 [Epub ahead of print].
42. Modic MT, Ross JS. Lumbar degenerative disk disease. Radiology 2007;245:43–61.
43. Stupar M, Côté P, French MR, Hawker GA. The association between low back pain and osteoarthritis of the hip and knee: a population-based cohort study. J Manipulative Physiol Ther 2010;33:349–54.
44. Madigan L, Vaccaro AR, Spector LR, Milam RA. Management of symptomatic lumbar degenerative disk disease. J Am Acad Orthop Surg 2009;17:102–11.
45. 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 the research published through January 2009. Osteoarthritis Cartilage 2010;18:476–99.

Osteoarthritis and Cartilage 19 (2011) 483–492

Osteoarthritis and Cartilage

Outcome measures in placebo-controlled trials of osteoarthritis: responsiveness to treatment effects in the REPORT database

R.H. Dworkin †*, S. Peirce-Sandner †, D.C. Turk ‡, M.P. McDermott §, A. Gibofsky ||, L.S. Simon ¶, J.T. Farrar #, N.P. Katz ††‡‡

† Department of Anesthesiology, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA
‡ Department of Anesthesiology, University of Washington, Seattle, WA, USA
§ Department of Biostatistics and Computational Biology, University of Rochester School of Medicine and Dentistry, Rochester, NY, USA
|| Division of Rheumatology, Hospital for Special Surgery, Weill Medical College of Cornell University, New York, NY, USA
¶ SDG, LLC, Cambridge, MA, USA
# Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, PA, USA
†† Analgesic Solutions, Natick, MA, USA
‡‡ Tufts University, Boston, MA, USA

ARTICLE INFO

Article history:
Received 24 December 2010
Accepted 14 February 2011

Keywords:
Outcome measures
Randomized clinical trials
Osteoarthritis
Responsiveness
Treatment effect
Standardized effect size

SUMMARY

Introduction: Treatment response in randomized clinical trials (RCT) of osteoarthritis (OA) has been assessed by multiple primary and secondary outcomes, including pain, function, patient and clinician global measures of status and response to treatment, and various composite and responder measures. Identifying outcome measures with greater responsiveness to treatment is important to increase the assay sensitivity of RCTs.
Objective: To assess and compare the responsiveness of different outcome measures used in placebo-controlled RCTs of OA.
Search strategy: The Resource for Evaluating Procedures and Outcomes of Randomized Trials database includes placebo-controlled clinical trials of pharmacologic treatments (oral, topical, or transdermal) for OA identified from a systematic literature search of RCTs published or publicly available before August 5, 2009, which was conducted using PubMed, the Cochrane collaboration, publicly-available websites, and reference lists of retrieved publications.
Data collection and analysis: Data collected included: (1) pain assessed with single-item ratings and the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain subscale; (2) patient and clinician global measures of status, improvement, and treatment response; (3) function assessed by the WOMAC function subscale; (4) stiffness assessed by the WOMAC stiffness subscale; and (5) the WOMAC and Lequesne Algofunctional Index composite outcomes. Measures were grouped according to the total number of response categories (i.e., $<$10 categories or $\geq$10 categories). The treatment effect (difference in mean change from baseline between the placebo and active therapy arms) and standardized effect size (SES) were estimated for each measure in a meta-analysis using a random effects model.
Results: There were 125 RCTs with data to compute the treatment effect for at least one measure; the majority evaluated non-steroidal anti-inflammatory drugs (NSAIDs), followed by opioids, glucosamine and/or chondroitin, and acetaminophen. In general, the patient-reported pain outcome measures had comparable responsiveness to treatment as shown by the estimates of treatment effects and SES. Treatment effects and SESs were generally higher for patient-reported global measures compared with clinician-rated global measures but generally similar for the WOMAC and Lequesne composite measures.
Conclusions: Comparing different outcome measures using meta-analysis and selecting those that have the greatest ability to identify efficacious treatments may increase the efficiency of clinical trials of treatments for OA. Improvements in the quality of the reporting of clinical trial results are needed to facilitate meta-analyses to evaluate the responsiveness of outcome measures and to also address other issues related to assay sensitivity.

* Address correspondence and reprint requests to: Robert H. Dworkin, Department of Anesthesiology, University of Rochester Medical Center, 601 Elmwood Ave., Box 604, Rochester, NY 14642, USA. Tel: 1-585-275-8214; Fax: 1-585-276-0122.
E-mail address: robert_dworkin@urmc.rochester.edu (R.H. Dworkin).

1063-4584/$ – see front matter
doi:10.1016/j.joca.2011.02.020

Introduction

An appreciable percentage of patients with osteoarthritis (OA) are refractory to existing analgesic treatments, and the patients who do respond to these treatments often obtain only partial relief of their pain[1,2]. Considerable effort is therefore being devoted to the development of new treatments for OA and to conducting randomized clinical trials (RCTs) to evaluate their efficacy and safety. In designing these trials, it is critically important that methodological factors are identified that might improve their assay sensitivity, which has been defined as "the ability to distinguish an effective treatment from a less effective or ineffective treatment"[3]. Assuming that the treatment studied is efficacious, RCTs with greater assay sensitivity are less likely to have falsely negative study results and can detect treatment effects with smaller sample sizes. This not only hastens the time to study completion and reduces costs, but also exposes fewer subjects to the unknown risks of novel treatments.

An essential aspect of the assay sensitivity of a clinical trial involves the responsiveness to treatment of its outcome measures. The importance of assay sensitivity in identifying efficacious treatments as efficiently as possible provides a compelling rationale for identifying and then selecting measures that have the greatest responsiveness (assuming other characteristics of the measures do not offset this, for example, lack of clinical importance or substantially increased patient burden).

Response to treatment in RCTs of patients with OA has been measured by patient-reported assessments of pain, function, and stiffness as well as by patient and clinician global evaluations of disease status and response to treatment[4–6]. Various visual analog scales (VAS) and numeric rating scales (NRS) as well as disease-specific outcome measures, such as the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC)[7] and the Lequesne Algofunctional Index[8], have been widely used as primary and secondary outcome measures in OA trials. To encourage standardization among the diverse measures that are available, the Osteoarthritis Research Society International (OARSI) and Outcome Measures in Rheumatology (OMERACT) have recommended core outcome domains and measures for OA clinical trials[9–12]. For RCTs of chronic pain conditions in general, the Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials (IMMPACT) has recommended consideration of specific outcome domains[13], measures[14], approaches to developing new measures[15], and strategies for evaluating the clinical importance of treatment outcomes[16,17].

IMMPACT has also created the Resource for Evaluating Procedures and Outcomes of Randomized Trials (REPORT) to examine relationships between clinical trial research methods and study outcomes and thereby contribute to the development of an evidence-based approach to analgesic clinical trial design[18,19]. REPORT consists of comprehensive databases of RCTs of acute and chronic pain conditions (e.g., neuropathic pain, fibromyalgia, chronic low back pain, and acute post-operative pain), which are at present limited to trials of pharmacologic treatments. In this article, we evaluate and compare the responsiveness to treatment of commonly used outcome measures in placebo-controlled RCTs of pharmacologic treatments in the REPORT database of OA trials.

Materials and methods

Literature search

The REPORT database of OA clinical trials includes placebo-controlled trials of pharmacologic treatments identified from a systematic literature search of RCTs published or publicly available before August 5, 2009, which was conducted using PubMed, Cochrane collaboration systematic reviews, publicly-available websites, and references from published reports of trials that met inclusion criteria, and other retrieved publications[20]. The search terms included "osteoarthritis", "degenerative joint disease", "coxarthrosis", and "gonarthrosis," with limits of "Randomized Controlled Trial", "Human", and "English" applied. Only trials that met the following criteria were included: (1) results reported in publicly-available sources, including publications and websites (e.g., www.clinicaltrialresults.org); (2) evaluated oral, topical, or transdermal pharmacologic treatments; (3) had treatment durations of at least 7 days; (4) used a parallel group design; (5) included patients with OA of the knee or hip; and (6) were placebo-controlled and double blinded (except for one single blind trial). Clinical trials reported only in abstract form were not included.

Information was extracted on standard forms and entered into a spreadsheet. Variables collected from each trial that were used in the present analyses included the following, when available: (1) eligibility criteria, including joint(s) studied; (2) active treatments; (3) baseline and endpoint mean values and either the respective standard deviation (SD) or information from which this SD could be derived (e.g., standard error, confidence interval); (4) specific outcome measures and scales used for all primary and secondary endpoints related to pain, function, stiffness, and patient and clinician global assessments of status, improvement, and treatment response; and (5) responder outcomes based on pain reduction, patient and clinician global or treatment response assessments, and OMERACT–OARSI responder criteria[5,11]. The statistical significance for the comparison of each active treatment group with the placebo group was recorded for all primary and secondary endpoints.

Study outcomes

Data from primary and secondary endpoints using 0–10 NRS, 0–10 cm VAS, and other measures were transformed to a 0–100 scale[21]. Although NRSs and VASs may have somewhat different psychometric properties, responses to these two types of measures of pain intensity are highly correlated and there is no evidence that their responsiveness to change and to treatment effects differs[14].

The placebo group and treatment group mean changes from baseline and standardized effect sizes (SESs) were determined as follows:

1. Placebo group mean change from baseline, as reported or computed as the difference in mean responses at the baseline and final visits.
2. Active treatment group mean change from baseline, as reported or computed as the difference in mean responses at the baseline and final visits.
3. Treatment effect: active treatment group mean change from baseline − placebo group mean change from baseline.
4. Pooled SD:

$$\frac{(N_{Treatment}-1)\times(SD_{Treatment})^2+(N_{Placebo}-1)\times(SD_{Placebo})^2}{N_{Treatment}+N_{Placebo}-2}$$

5. SES: treatment effect/pooled SD

When only final visit means were available (i.e., no mean baseline or change from baseline values were present), expressed as either actual mean values or mean values adjusted for baseline, the treatment effect was calculated by subtracting the placebo

group final mean value from the active treatment group final mean value. The SES was calculated only for trials that reported the corresponding measure of variability (e.g., SD) of the change from baseline[22] or of the final values adjusted for baseline.

Analyses focused on RCTs of efficacious treatments, which were considered those medications that are "recommended treatments" for OA in prominent national and international guidelines[23,24]. Given the difficulty of establishing which other treatments truly lack efficacy, supplementary analyses are presented in Appendix I for RCTs examining all treatments irrespective of whether they are recommended or not. For trials with multiple arms examining the same treatment (e.g., different dosages or titration schedules), a mean value for each treatment was calculated by computing the average response, weighted according to sample size, across all arms of the same treatment within a trial so that each active treatment contributed only one value per trial for each outcome measure.

The outcome measures were included in the analyses of treatment effects and SESs if there were at least five RCTs with sufficient data[25]. Pain-related outcome measures included: (1) "pain", representing spontaneous pain, pain at rest, or otherwise undesignated pain; (2) "active pain", representing pain during a weight-bearing activity, for example, when walking or standing, but not in response to passive movement by an examiner or clinician; (3) the WOMAC pain subscale[7]; and (4) item one of the WOMAC pain subscale (pain walking on a flat surface)[7]. Patient and clinician measures were based on global assessments (e.g., of OA status or overall improvement) and on response to treatment of OA. Function and stiffness endpoints were measured by the WOMAC function and stiffness subscales[7], and composite outcomes were represented by the WOMAC total score[7] and the Lequesne Algofunctional Index[8,26], a 10-item composite measure of pain, stiffness, walking distance, and other aspects of function. All of these measures were also evaluated according to the number of response categories for ratings made by the patient or clinician during the trial, specifically (1) scales with 10 response categories or more (e.g., NRS with a 0–10 scale or VAS with a 0–100 mm scale); and (2) scales with fewer than 10 response categories (e.g., Likert scale with a 0–3 verbal rating scale). For measures that consist of multiple individual rating scales each having fewer than 10 response categories, the number of response categories was considered the total number of possible categories; for example, a measure composed of four 0–3 Likert scales with a maximum total score of 12 was considered a scale with $\geq$10 response categories. Responder outcomes were classified according to the instrument used to categorize responders and non-responders, specifically, pain (i.e., $\geq$30% reduction), patient and clinician global ratings and treatment response assessments, and OMERACT–OARSI responder criteria[5,11], which are based on absolute and percentage improvements in pain, function and patient global assessments.

Statistical analyses

Due to the expected heterogeneity among the different studies (including eligibility criteria, intervention studied, evaluation protocol, concomitant treatments, and other factors), overall estimates of treatment effects and SESs for each outcome measure were obtained from random effects models that treated study as a random effect[27,28]. These overall estimates were computed as weighted averages of the individual study estimates, with the weights being inversely proportional to the estimated variances of the individual study estimates[27,28]. Responder measures with binary outcomes (i.e., responder vs non-responder, however defined) were evaluated in the same manner, with the overall percentage of responders estimated by the weighted average of the study-specific estimated percentages, with the weights being inversely proportional to the estimated variances of the individual study estimates[27,28].

The number of response categories used by outcome measures could be associated with differences in either treatment effects or SESs, for example, fewer response categories could be associated with less responsiveness to change. To evaluate this possibility, mixed effects models with number of scale categories as a fixed effect (0 = scales with <10 response categories, 1 = scales with $\geq$10 response categories) and study as a random effect was used. Because the effect of number of scale response categories was significant for clinician global status for treatment effect and SES and showed trends for the other measures that used both of these two groups of number of response categories, treatment effects and SESs were summarized overall as well as according to number of scale categories when an outcome measure had adequate data for both groups of number of response categories.

The estimates of treatment effect and SES for selected measures were compared using a mixed effects model with study as a random effect and the measure of interest as a fixed effect. The measures selected for comparison were the most commonly used outcome measures that were similar with respect to the underlying construct being assessed (i.e., pain, global status, overall composite outcome): (1) spontaneous pain vs WOMAC pain; (2) patient vs clinician global assessment; and (3) WOMAC total score vs Lequesne index. In the comparisons of these selected measures, when there were studies that had values for both of the measures, the measure with the greatest number of values in each of the analyses was discarded from the study to preserve independence. For the random effects models, when a trial included two or more different active treatments, only one active treatment arm per placebo group was retained in order to preserve independence. A single active treatment arm was randomly selected from trials with two or more different active treatments and any other active treatment arms were discarded.

For each outcome measure, the following data are presented: (1) estimates from the random effects models overall and for each of the two groups of measures differing in number of response categories, when data from five or more trials were available[25]; (2) overall unweighted means across all arms to show treatment effects and SESs for the greatest number of treatment arms, including those for which no measure of variability was available. The numbers of treatment arms available for the random effects models are lower than the number of arms used in calculating the unweighted means because only one active arm per trial was used and the random effects model requires an estimate of variance that was not always available. The sample sizes for estimates of the SESs are often lower than for the treatment effects because SES was computed only for studies that reported variance adjusted for baseline values[22]. In comparing selected measures (e.g., spontaneous pain and WOMAC pain) for treatment effects and for SESs, the sample size is further reduced because one measure was dropped from studies that reported both measures in order to preserve independence. In the following presentation and discussion of the results, we focus on estimates from the random effects models.

Results

A total of 1774 articles were retrieved from the PubMed search, Cochrane collaboration reviews related to pharmacotherapy for OA, and the clinicaltrialresults.org website, from which 167 blinded, randomized, and placebo-controlled trials of oral, topical, and transdermal therapies for OA of the knee and/or hip were

Table I
Active treatment groups according to outcome measure and treatment type for trials with a non-missing value for treatment vs placebo difference or responder outcome

Measure	Total number of active treatment groups	Non-steroidal anti-inflammatory drug, N (%)*	Acetaminophen (paracetamol), N (%)	Glucosamine/chondroitin†, N (%)	Opioid analgesic‡, N (%)	Other, N (%)
Pain	83	45 (54.2)	0 (0)	11 (13.3)	11 (13.3)	16 (19.3)
Pain (with activity)	30	24 (80.0)	1 (3.3)	2 (6.7)	1 (3.3)	2 (6.7)
WOMAC pain subscale	99	63 (63.6)	3 (3.0)	4 (4.0)	11 (11.1)	18 (18.2)
WOMAC pain walking§	21	19 (90.5)	0 (0)	0 (0)	2 (9.5)	0 (0)
Patient global rating	66	53 (80.3)	1 (1.5)	3 (4.5)	4 (6.1)	5 (7.6)
Patient response to therapy	20	17 (85.0)	0 (0)	2 (10.0)	1 (5.0)	0 (0)
Clinician global rating	50	44 (88.0)	0 (0)	2 (4.0)	2 (4.0)	2 (4.0)
Clinician response to therapy	11	10 (90.9)	0 (0)	0 (0)	1 (9.1)	0 (0)
Lequesne algofunctional index	32	16 (50.0)	1 (3.1)	7 (21.9)	0 (0)	8 (25.0)
WOMAC total score	65	36 (55.4)	3 (4.6)	7 (10.8)	5 (7.7)	14 (21.5)
WOMAC function subscale	93	59 (63.4)	4 (4.3)	8 (8.6)	8 (8.6)	14 (15.1)
WOMAC stiffness subscale	82	51 (62.2)	2 (2.4)	7 (8.5)	8 (9.8)	14 (17.1)
≥30 % pain reduction responder	14	4 (28.6)	1 (7.1)	2 (14.3)	4 (28.6)	3 (21.4)
Patient responder	61	44 (72.1)	1 (1.6)	7 (11.5)	4 (6.6)	5 (8.2)
Clinician responder	35	26 (74.3)	0 (0)	5 (14.3)	3 (8.6)	1 (2.9)
OMERACT–OARSI responder	18	12 (66.7)	2 (11.1)	4 (22.2)	0 (0)	0 (0)

* Values for numbers of active treatment groups and percentages, which may not sum to 100% due to rounding.
† Individually or in combination.
‡ Includes tramadol or tramadol in combination with acetaminophen.
§ Item one of the WOMAC pain subscale.

identified. Of these 167 reports, 125 studies representing 184 different active treatment arms had sufficient data to compute the treatment effect for at least one measure. The majority of these RCTs examined only knee OA (66%), followed by the combination of knee OA and hip OA (30%), and hip OA only (5%). The categories of recommended treatments examined in these trials were predominantly non-steroidal anti-inflammatory drugs (NSAIDs) followed by opioid analgesics, glucosamine and/or chondroitin, and acetaminophen/paracetamol (Table I).

Pain

For the pain-related outcome measures presented in Table II, estimated treatment effects ranged from 7.6 to 9.5 (all outcomes converted to a 0–100 scale) and estimated SESs ranged from 0.21 to 0.45. Treatment effect estimates were generally comparable between typically single-item NRS or VAS pain intensity measures and the WOMAC five-item pain subscale. The SES for the WOMAC pain subscale (0.45) was appreciably higher than the SES for the measures of spontaneous pain (0.27), but this difference was not statistically significant ($p = 0.09$; Table V) and was based on data from 32 studies (13 for spontaneous pain, 19 for the WOMAC pain scale).

Responder analyses based on a reduction in pain of ≥30% from baseline to endpoint showed a treatment effect of 16.1 (i.e., the difference in percentages of responders between the active treatment and placebo arms) for scales with ≥10 categories (there were inadequate data for scales with <10 categories). These responder analyses were based on a variety of different pain outcome measures, including spontaneous pain, pain with activity, WOMAC pain subscale scores, and pain walking (item one of the WOMAC pain subscale).

Patient and clinician global measures and response to treatment measures

Most trials in the database used patient and clinician global measures of disease activity to assess response, typically as secondary endpoints. The range in estimates for patient and clinician global measures was 5.6–13.2 for treatment effects and 0.20–0.68 for SESs (Table III). The estimates of treatment effect and SES were significantly higher for patient global measures compared with clinician global measures using scales with ≥10 categories, but the differences were not significant for scales with <10 categories (Table V). For patient measures of response to treatment, five trials provided the basis for estimates of treatment

Table II
Treatment vs placebo group differences and SESs for pain-related outcome measures for recommended treatments

Measure	Pain			Pain (with activity)			WOMAC pain subscale			WOMAC pain walking			≥30% pain reduction (% patients)		
	N*	Mean	95% CI†	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI
Treatment vs placebo group difference															
All scales, random effects model‡	29	9.4	7.2, 11.6	11	7.7	3.6, 11.8	36	8.4	6.3, 10.5	6	9.5	6.7, 12.3	6	16.1	4.4, 27.8
Scales ≥10 categories, random effects model‡	27	8.7	6.9, 10.5	10	7.6	3.1, 12.1	36	8.4	6.3, 10.5				6	16.1	4.4, 27.8
All scales, unweighted, all arms§	67	11.4		28	10.3		81	8.9		21	9.9		11	15.7	
SES															
All scales, random effects model‡	13	0.27	0.11, 0.43	7	0.22	0.08, 0.36	26	0.40	0.27, 0.53	6	0.39	0.25, 0.52			
Scales ≥10 categories, random effects model‡	13	0.27	0.11, 0.43	6	0.21	0.05, 0.37	26	0.40	0.27, 0.53						
All scales, unweighted, all arms§	19	0.29		8	0.25		44	0.43		9	0.45				

* Number of active treatment arms (one value per treatment).
† CI = confidence interval.
‡ Estimate from random effects model with one randomly selected treatment included per trial.
§ Confidence intervals are not provided because their calculation would assume the statistical independence of the results from all of the trials, which does not hold in this case due to the inclusion of multiple treatment comparisons with the same placebo group in some of the trials.

Table III
Treatment vs placebo group differences and SESs for patient- and clinician-rated outcome measures for recommended treatments

Measure	Patient global			Patient treatment response			Patient responder (% patients)			Clinician global			Clinician treatment response			Clinician responder (% patients)		
	N*	Mean	95% CI†	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI
Treatment vs placebo group difference																		
All scales, random effects model‡	26	10.7	8.5, 12.9	5	17.1	13.2, 21.0	41	23.1	19.8, 26.4	18	8.6	6.6, 10.6				28	24.5	19.6, 29.4
Scales ≥10 categories, random effects model‡	20	10.1	7.5, 12.7							7	5.6	3.4, 7.8						
Scales < 10 categories, random effects model‡	6	13.2	10.0, 16.4	5	17.1	13.2, 21.0	41	23.1	19.8, 26.4	11	10.9	8.8, 13.0				28	24.5	19.6, 29.4
All scales, unweighted, all arms§	61	12.0		20	15.5		56	23.1		48	10.9		11	18.4		34	24.5	
SES																		
All scales, random effects model‡	19	0.38	0.27, 0.49	5	0.68	0.48, 0.88				15	0.34	0.24, 0.45						
Scales ≥10 categories, random effects model‡	15	0.38	0.24, 0.52							6	0.20	0.14, 0.26						
Scales < 10 categories, random effects model‡				5	0.68	0.48, 0.88				9	0.46	0.33, 0.59						
All scales, unweighted, all arms§	44	0.44		9	0.66					33	0.40		7	0.74				

* Number of active treatment arms (one value per treatment).
† CI = confidence interval.
‡ Estimate from random effects model with one randomly selected treatment included per trial.
§ Confidence intervals are not provided because their calculation would assume the statistical independence of the results from all of the trials, which does not hold in this case due to the inclusion of multiple treatment comparisons with the same placebo group in some of the trials.

effect (17.1) and SES (0.68) for scales with <10 response categories. There were inadequate data to compute these estimates for scales with ≥10 categories and the clinician measures of response to treatment.

All of the patient and clinician responder outcomes were based on scales with <10 categories. Treatment effects (i.e., the difference in percentages of responders between the active treatment and placebo arms) for patient (23.1) and clinician (24.5) responder outcomes were very similar.

Function and composite measures

For the function, stiffness, and composite outcome measures, the estimates of treatment effects (5.3–8.3) and SESs (0.25–0.37) were relatively modest (Table IV), and the SESs for the Lequesne index and WOMAC total score were similar (Table V). There were 11 studies with a total of 18 different active treatment arms that provided data for the OMERACT–OARSI responder criteria. The estimated treatment effect (i.e., the difference in percentages of responders between the active treatment and placebo arms) was 12.7, which was somewhat lower than the treatment effects for the responder analyses of ≥30% pain reduction presented in Table II and the responder outcomes based on patient and clinician measures presented in Table III.

Discussion

We conducted a meta-analysis of the responsiveness of the outcome measures that are used most frequently in RCTs of OA. These measures include patient-reported assessments of pain, physical function, stiffness, global status, treatment response, and

Table IV
Treatment vs placebo group differences and SESs for function, stiffness, and composite outcome measures for recommended treatments

Measure	WOMAC function subscale			Lequesne algofunctional index			WOMAC stiffness subscale			WOMAC total score			OMERACT–OARSI responder (% patients)		
	N*	Mean	95% CI†	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI	N	Mean	95% CI
Treatment vs placebo group difference															
All scales, random effects model‡	36	6.8	5.5, 8.2	14	5.3	3.5, 7.1	33	7.4	5.7, 9.2	24	5.5	4.4, 6.6	11	12.7	7.8, 17.6
Scales ≥10 categories, random effects model‡	36	6.8	5.5, 8.2	14	5.3	3.5, 7.1	22	8.3	5.9, 10.7	24	5.5	4.4, 6.6			
Scales < 10 categories, random effects model‡							11	5.9	4.1, 7.8						
All scales, unweighted, all arms§	79	7.6		24	5.7		68	8.2		55	6.9		18	13.2	
SES															
All scales, random effects model‡	26	0.33	0.25, 0.41	6	0.34	0.20, 0.48	25	0.32	0.23, 0.41	15	0.30	0.24, 0.36			
Scales ≥10 categories, random effects model‡	26	0.33	0.25, 0.41	6	0.34	0.20, 0.48	16	0.37	0.25, 0.49	15	0.30	0.24, 0.36			
Scales < 10 categories, random effects model‡							9	0.25	0.17, 0.33						
All scales, unweighted, all arms§	44	0.35		8	0.37		43	0.34		27	0.34				

* Number of active treatment arms (one value per treatment).
† CI = confidence interval.
‡ Estimate from random effects model with one randomly selected treatment included per trial.
§ Confidence intervals are not provided because their calculation would assume the statistical independence of the results from all of the trials, which does not hold in this case due to the inclusion of multiple treatment comparisons with the same placebo group in some of the trials.

Table V
Comparison of pain, global, and composite outcome measures for recommended treatments

Measure	Pain		WOMAC pain subscale				Patient global		Clinician global				WOMAC total score		Lequesne			
	N*	Mean	N	Mean	Difference (95% CI†)	P‡	N	Mean	N	Mean	Difference (95% CI)	P	N	Mean	N	Mean	Difference (95% CI)	P
Treatment vs placebo group difference																		
All scales, random effects model§	29	9.6	26	9.4	0.2 (−3.2, 3.6)	.90	11	11.8	18	8.7	3.1 (−0.5, 6.7)	.09	22	5.7	14	5.2	0.6 (−1.5, 2.6)	.40
Scales ≥10 categories, random effects model§	27	9.0	27	9.4	0.3 (−2.9, 3.6)	.83	13	12.3	7	5.8	6.5 (2.0, 11.1)	.01	22	5.7	14	5.2	0.6 (−1.5, 2.6)	.40
Scales <10 categories, random effects model§							6	13.2	8	10.5	2.7 (−1.3, 6.7)	.19						
SES																		
All scales, random effects model§	13	0.27	19	0.45	0.19 (−0.03, 0.40)	.09	7	0.42	15	0.34	0.07 (−0.12, 0.26)	.55	14	0.31	6	0.34	0.03 (−0.11, 0.17)	.65
Scales ≥10 categories, random effects model§	13	0.27	19	0.45	0.19 (−0.03, 0.40)	.09	9	0.50	6	0.20	0.30 (0.07, 0.43)	.01	14	0.31	6	0.34	0.03 (−0.11, 0.17)	.65
Scales <10 categories, random effects model§							4	0.41	8	0.47	.06 (−0.15, 0.27)	.56						

* Number of active treatment arms (one value per treatment).
† CI = confidence interval.
‡ *P* value for difference between measures.
§ All values estimated from random effects model with one randomly selected treatment included per trial. When both measures were included in the same trial, the measure with the highest N in each of the three analyses (all scales, scales with ≥10 categories, scales with <10 categories) was deleted from that trial.

composite outcome, clinician-rated global assessments, and various responder outcomes. For patient-reported pain outcomes, there were generally comparable treatment effect estimates for the different measures, including single VAS or NRS ratings of overall pain or pain with activity (e.g., pain walking) and the multi-item WOMAC pain subscale. However, the mean SES for pain rated on single scales with ≥10 response categories was considerably lower than the mean SES for the WOMAC pain subscale (0.27 vs 0.45; $p = .09$, 95% confidence interval (CI) for difference = −0.03, 0.40). Although this difference was not statistically significant and was based on a total of only 32 trials, SES differences of this magnitude would have important implications with respect to sample size, with considerably fewer patients being required for adequate statistical power if the WOMAC pain subscale were to be used rather than a single pain rating. However, these results need to be interpreted with caution because it is possible that confounding factors — that is, systematic differences between trials using a single pain rating and those using the WOMAC pain subscale — may have influenced these results.

The conclusion that different patient-reported outcome measures of pain severity may have generally comparable responsiveness to treatment is consistent with the results of data analyses from single clinical trials[29,30]; for example, the difference between improvement in pain on the WOMAC pain subscale and on a VAS following knee lavage was not significant[29]. In research comparing VAS and Likert versions of the WOMAC[31,32], and in other studies of OA pain assessed using VAS and Likert scales[33,34], generally comparable responsiveness to change of these different rating scales has been found. Different pain measures and scales may not always be interchangeable[35], however, and there are circumstances in which one type of assessment might be preferred[14]. Although we found non-significant differences in favor of the WOMAC pain subscale, which will need to be examined in future research, considered together with the results of previous research, our analyses suggest that different measures of pain in patients with OA may have generally comparable ability to identify efficacious treatments.

The analyses of global outcome measures showed larger treatment effects and a trend toward larger SESs for patient-reported vs clinician-rated measures (but only for measures with ≥10 response categories). This result may not be surprising. It is likely that global assessments made by clinicians are based, in large part, on what patients report to them, and improvement in pain, which is a subjective experience, appears to account for a major portion of the variation in patient global assessments of outcome and treatment satisfaction[36]. These considerations[5,13,14] and our data suggest that patient global measures are likely to provide more valid and responsive outcomes in analgesic RCTs than clinician global measures in most circumstances.

Treatment effects and SESs were generally lower or comparable for the function and composite measures compared with the pain, global, and responder outcomes, which is consistent with the results of other studies in patients with OA[29,34,37]. Treatment effects for the OMERACT–OARSI responder criteria were somewhat lower than for the other responder outcomes, which were typically based on single-item pain or global ratings. Multidimensional measures of outcome — such as the WOMAC, Lequesne index, and OMERACT–OARSI responder criteria — may provide a more comprehensive assessment of the patient's overall response to treatment, but it is not clear why this might lead to lower responsiveness to treatment effects than found with unidimensional measures. One possibility is that existing medications for pain in OA have analgesic effects that are generally modest and do not reduce pain to low enough levels for improvements in function to become apparent. In addition, some of these treatments may not have meaningful benefits on the additional outcome dimensions included in the composite outcome measures. Of course, these observed differences also may be due to chance or confounding.

Our results are based on a meta-analysis of clinical trials that were conducted using different research designs, treatments, and outcomes, an approach that has also been used recently to evaluate the "discriminating power" of outcome measures in clinical trials of fibromyalgia[38]. A different approach to examining the assay sensitivity of outcome measures involves evaluating treatment effects and SESs in a single clinical trial in which each patient completes all of the measures and patient-level data, rather than the group means used in our analyses, provide the basis for comparing measures[29,39–41]. However, the generalizability of such results is potentially limited by specific features of the clinical trial,

including patient demographic and clinical characteristics and study methodology (e.g., trial duration[42]) as well as the specific treatment examined. The present meta-analysis provides a comparative evaluation of the responsiveness of outcome measures across a broad range of patients, clinical trial characteristics, and treatments.

It is important to emphasize that the heterogeneity of the sample of trials we examined is also a limitation of our analyses. The treatment effects and SESs for the different outcome measures could reflect not only potential differences among the measures in responsiveness to treatment but also differences among trials in methods, treatment efficacy and safety, imputation of missing data, study duration, and random variation[19,43,44]. For example, we did not adjust for whether trials used a flare design, which has recently been shown to accentuate the treatment effects of NSAIDs[21], perhaps by enriching for those patients who are most likely to respond to treatment; differences in the use of this design across the outcome measures we examined could have influenced our results. In addition, the frequency with which the outcome measures we examined were administered varied greatly among trials; for example, some RCTs conducted pain ratings on a daily basis and examined the means of such multiple ratings, whereas others only captured pain weekly or monthly and examined single ratings. Measures that use multiple assessments — on different occasions or within a single measure, for example, the five items of the WOMAC pain subscale — generally have greater reliability, which might be associated with increased responsiveness to treatment effects.

There are other important limitations of our analyses. We considered glucosamine and chondroitin to be efficacious treatments, although recent evidence suggests that they might not be[45] (this conclusion, however, has been disputed[46]). In addition, it is widely recognized that negative trials are less likely to be published, and our analyses were limited to published and publicly-available RCTs; estimates of treatment effects based on the published literature are therefore likely to be higher than if all RCTs could be examined[47,48] and this could have influenced our results. Our analyses were also limited to trials published or reported in English, and it is therefore possible that the inclusion of RCT results that are only available in other languages might have altered our conclusions. Finally, our results must also be viewed with sample size limitations in mind. Although the number of studies available for computing estimates of treatment effects and SESs was limited for some measures, all of the mean estimates we have presented are based on the results of at least five trials[25] and the treatment effect estimates represent total numbers of patients ranging from 1,296 (clinician treatment response) to 13,486 (WOMAC function subscale). Nevertheless, it is important to recognize that there were a substantial number of trials in the database that did not report information from which an appropriate measure of variability could be determined for calculating treatment effect and SES estimates. This made the sample sizes for our meta-analyses considerably smaller than if there had been more complete reporting of the results of the clinical trials in the database. Improvements in the quality of the reporting of clinical trial results are needed to facilitate meta-analyses such as those performed here.

The assay sensitivity of an outcome measure is a function of the separation between measured improvement in the active treatment group and in the placebo group. It is widely appreciated that substantial improvements in pain occur in the placebo groups of OA trials[20,49], and "excessive" placebo group improvement for an outcome measure could compromise its responsiveness. Benefit in placebo groups can be due to multiple factors alone and in combination, including placebo effects, natural history, regression to the mean, and various subject, study site, and research design factors. In future research, it would be worthwhile to examine whether different types of outcome measures vary in the extent to which improvement is demonstrated with placebo treatment. Identifying specific outcome measures that are less responsive to placebo treatment than are other measures (while showing comparable responsiveness as the other measures to active treatments) has the potential to show greater treatment effects and thereby improve the assay sensitivity of analgesic trials[19].

There is little question that our analyses will need to be updated in several years. One important reason for this is that new outcome measures and new approaches to evaluating outcome in RCTs of pain in OA are being developed. For example, recent research has examined electronic pain diaries of various types[50], including mobile phones[51], and such methods may increase the convenience of collecting more frequent, and therefore, more reliable pain ratings (although there is little evidence to date that these measures show greater responsiveness to treatment). There has also been increasing attention to evaluating the clinical importance of outcome measures[14,52], including the identification of low or acceptable levels of symptoms[53,54]. In addition, the importance of considering the patient's perspective has been emphasized[15,55], and initial attempts have been made to develop patient-centered outcome measures that assess the specific treatment goals of individual patients[56–58]. These efforts may lead to the identification of outcome measures with greater reliability, validity, and responsiveness, which could increase the assay sensitivity of clinical trials of treatments for OA.

Declaration of funding and role of funding source

The OARSI FDA OA Initiative received financial support from the following professional organization:

American College of Rheumatology.

Additionally the OARSI FDA OA Initiative received financial support from the following companies: Amgen, ArthroLab, AstraZeneca, Bayer Healthcare, Chondrometrics, CombinatoRx, Cypress, BioScience, DePuy Mitek, Expanscience, 4QImaging, Genevrier/IBSA, Genzyme, King (Alpharma), Merck, Merck Serono, NicOx, Pfizer, Rottapharm, Smith & Nephew, Wyeth.

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the Food and Drug Administration.

Author contributions

Conception and design of the study: RHD, SP, DCT, MPM, AG, LSS, JTF, NPK.

Acquisition and analysis of data: RHD, SP, MPM.

Interpretation of data: RHD, SP, DCT, MPM, AG, LSS, JTF, NPK.

Drafting the manuscript: RHD, SP, DCT.

Revision of manuscript for important content: RHD, SP, DCT, MPM, AG, LSS, JTF, NPK.

Final approval of the submitted manuscript: RHD, SP, DCT, MPM, AG, LSS, JTF, NPK.

Conflicts of interest

The authors do not have financial conflicts of interest related to the material presented in this article.

Appendix I

Treatment vs. placebo group differences, SESs, and study outcomes for osteoarthritis outcome measures for all treatments and arms

	Pain		Pain (with activity)		WOMAC pain subscale		WOMAC pain walking		≥30% pain reduction (% patients)	
	N*	Mean†	N	Mean	N	Mean	N	Mean	N	Mean
Treatment vs placebo group difference, unweighted	83	10.9	30	10.7	99	8.6	21	9.8	14	14.8
SES, unweighted	22	0.32	8	0.25	48	0.42	9	0.44		
% positive‡	101	75.3	38	86.8	118	76.3	21	90.5	15	66.7

	Patient global		Patient treatment response		Patient responder (% patients)		Clinician global		Clinician treatment response		Clinician responder (% patients)	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Treatment vs placebo group difference, unweighted	66	11.5	20	15.5	61	23.7	50	10.6	11	18.4	35	24.3
SES, unweighted	34	0.39	9	0.66			29	0.37	5	0.74		
% positive‡	93	74.1	26	80.1	63	84.1	68	80.9	14	71.4	34	88.2

	WOMAC function subscale		Lequesne algofunctional index		WOMAC stiffness subscale		WOMAC total score		OMERACT–OARSI responder criteria (% patients)	
	N	Mean	N	Mean	N	Mean	N	Mean	N	Mean
Treatment vs placebo group difference, unweighted	93	7.6	32	5.8	82	7.8	65	7.8	18	13.2
SES, unweighted	47	0.35	10	0.39	46	0.33	29	0.34		
% positive‡	120	70.0	40	62.5	104	65.4	78	71.8	23	82.6

* Number of active treatment arms (one value per treatment).

† Confidence intervals are not provided because their calculation would assume the statistical independence of the results from all of the trials, which does not hold in this case due to the inclusion of multiple treatment comparisons with the same placebo group in some of the trials.

‡ For "% positive," the outcome for each treatment arm was categorized as positive if the comparison with placebo for a given measure yielded a statistically significant ($p \leq 0.05$) treatment effect, and the percentage of positive outcomes was calculated by dividing the number of treatment arms (one for each active treatment per trial) with a positive outcome by the total number of treatment arms. The outcome for multiple arms of the same treatment in a trial was categorized as positive if at least one of the arms was positive. The number of treatment arms for this variable can be higher than for the unweighted treatment effect because some studies provided values for outcome but insufficient information to compute treatment effects.

References

1. Bjordal JM, Klovning A, Ljunggren AE, Slørdal L. Short-term efficacy of pharmacotherapeutic interventions in osteoarthritic knee pain: a meta-analysis of randomised placebo-controlled trials. Eur J Pain 2007;11:125–38.
2. 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.
3. U.S. Department of Health and Human Services. Guidance for industry: E10 choice of control group and related issues in clinical trials, 2001; Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm073139.pdf.
4. U.S. Department of Health and Human Services. Guidance for industry: clinical development programs for drugs, devices, and biological products intended for the treatment of osteoarthritis (OA), 1999; Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071579.pdf.
5. 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.
6. European Medicines Agency. Guideline on clinical investigation of medicinal products used in the treatment of osteoarthritis, 2009; Available at: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003443.pdf.
7. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833–40.
8. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee: validation: value in comparison with other assessment tests. Scand J Rheumatol 1987;65(Suppl):85–9.
9. 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. Osteoarthritis Cartilage 1996;4:217–43.
10. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al.* Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
11. 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 Response Criteria Initiative. Osteoarthritis Cartilage 2000;8:395–403.
12. Maheu E, Altman RD, Bloch DA, Doherty M, Hochberg M, Mannoni A, *et al.* Design and conduct of clinical trials in patients with osteoarthritis of the hand: recommendations from a task force of the Osteoarthritis Research Society International. Osteoarthritis Cartilage 2006;14:303–22.
13. Turk DC, Dworkin RH, Allen RR, Bellamy N, Brandenburg N, Carr DB, *et al.* Core outcome domains for chronic pain

clinical trials: IMMPACT recommendations. Pain 2003;106: 337–45.
14. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, *et al*. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. Pain 2005;113:9–19.
15. Turk DC, Dworkin RH, Burke LB, Gershon R, Rothman M, Scott J, *et al*. Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. Pain 2006;125:208–15.
16. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, *et al*. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain 2008;9:105–21.
17. Dworkin RH, Turk DC, McDermott MP, Peirce-Sandner S, Burke LB, Cowan P, *et al*. Interpreting the clinical importance of group differences in chronic pain clinical trials: IMMPACT recommendations. Pain 2009;146:238–44.
18. Dworkin RH, Turk DC, Peirce-Sandner S, McDermott MP, Farrar JT, Hertz S, *et al*. Placebo and treatment group responses in postherpetic neuralgia vs. painful diabetic peripheral neuropathy clinical trials in the REPORT database. Pain 2010;150:12–6.
19. Dworkin RH, Turk DC, Katz NP, Rowbotham MC, Peirce-Sandner S, Cerny I. Evidence-based clinical trial design for chronic pain pharmacotherapy: a blueprint for ACTION. Pain 2011;152(Suppl):S107–15.
20. Zhang W, Robertson J, Jones AC, Dieppe PA, Doherty M. The placebo effect and its determinants in osteoarthritis: meta-analysis of randomised controlled trials. Ann Rheum Dis 2008;67:1716–23.
21. Trijau S, Avouac J, Escalas C, Gossec L, Dougados M. Influence of flare design on symptomatic efficacy of non-steroidal anti-inflammatory drugs in osteoarthritis: a meta-analysis of randomized placebo-controlled trials. Osteoarthritis Cartilage 2010;18:1012–8.
22. Cochrane Handbook for Systematic Reviews of Interventions, Version 5.0.2, 2009; Available at: http://www.mrc-bsu.cam.ac.uk/cochrane/handbook.
23. National Collaborating Centre for Chronic Conditions. Osteoarthritis: National Clinical Guideline for Care and Management in Adults. London: Royal College of Physicians; 2008.
24. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *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.
25. Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, *et al*. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;336:601–5.
26. Faucher M, Poiraudeau S, Lefevre-Colau MM, Rannou F, Fermanian J, Revel M. Algo-functional assessment of knee osteoarthritis: comparison of the test-retest reliability and construct validity of the WOMAC and Lequesne indexes. Osteoarthritis Cartilage 2002;10:602–10.
27. Normand SL. Meta-analysis: formulating, evaluating, combining, and reporting. Stat Med 1999;18:321–59.
28. Whitehead A. Meta-analysis of Controlled Clinical Trials. Chichester, UK: John Wiley; 2002.
29. Gentelle-Bonnassies S, Le Claire P, Mezieres M, Ayral X, Dougados M. Comparison of the responsiveness of symptomatic outcome measures in knee osteoarthritis. Arthritis Care Res 2000;13:280–5.
30. Krebs EE, Bair MJ, Damush TM, Tu W, Wu J, Kroenke K. Comparative responsiveness of pain outcome measures among primary care patients with musculoskeletal pain. Med Care 2010;48:1007–14.
31. Bellamy N. WOMAC: a 20-year experiential review of a patient-centered self-reported health status questionnaire. J Rheumatol 2002;29:2473–6.
32. Villanueva I, del Mar Guzman M, Javier Toyos F, Ariza-Ariza R, Navarro F. Relative efficiency and validity properties of a visual analogue vs a categorical scaled version of the Western Ontario and McMaster Universities Osteoarthritis (WOMAC) Index: Spanish versions. Osteoarthritis Cartilage 2004;12:225–31.
33. Bellamy N, Campbell J, Syrotuik J. Comparative study of self-rating pain scales in osteoarthritis patients. Curr Med Res Opin 1999;15:113–9.
34. Bolognese JA, Schnitzer TJ, Ehrich EW. Response relationship of VAS and Likert scales in osteoarthritis efficacy measurement. Osteoarthritis Cartilage 2003;11:499–507.
35. Lund I, Lundeberg T, Sandberg L, Budh CN, Kowalski J, Svensson E. Lack of interchangeability between visual analogue and verbal rating pain scales: a cross sectional description of pain etiology groups. BMC Med Res Methodol 2005;5:31.
36. Dworkin RH, Jensen MP, Gould E, Jones BA, Xiang Q, Galer BS, *et al*. Treatment satisfaction in osteoarthritis and low back pain: the role of pain, physical and emotional functioning, sleep, and adverse events. J Pain 2011;12:416–24.
37. Bingham III CO, Bird SR, Smugar SS, Xu X, Tershakovec AM. Responder analysis and correlation of outcome measures: pooled results from two identical studies comparing etoricoxib, celecoxib, and placebo in osteoarthritis. Osteoarthritis Cartilage 2008;16:1289–93.
38. Carville SF, Choy EHS. Systematic review of discriminating power of outcome measures used in clinical trials of fibromyalgia. J Rheumatol 2008;35:2094–105.
39. Dunkl PR, Taylor AG, McConnell GG, Alfano AP, Conaway MR. Responsiveness of fibromyalgia clinical trial outcome measures. J Rheumatol 2000;27:2683–91.
40. Angst F, Aeschlimann A, Steiner W, Stucki G. Responsiveness of the WOMAC osteoarthritis index as compared with the SF-36 in patients with osteoarthritis of the legs undergoing a comprehensive rehabilitation intervention. Ann Rheum Dis 2001;60:834–40.
41. Soohoo NF, Vyas RM, Samimi DB, Molina R, Lieberman JR. Comparison of the responsiveness of the SF-36 and WOMAC in patients undergoing total hip arthroplasty. J Arthroplasty 2007;22:1168–73.
42. Quessy SN, Rowbotham MC. Placebo response in neuropathic pain trials. Pain 2008;138:479–83.
43. Katz N. Methodological issues in clinical trials of opioids for chronic pain. Neurology 2005;65:S32–49.
44. Katz J, Finnerup NB, Dworkin RH. Clinical trial outcome in neuropathic pain: relationship to study characteristics. Neurology 2008;70:263–72.
45. Wandel S, Jüni P, Tendal B, Nüesch E, Villiger PM, Welton NJ, *et al*. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. BMJ 2010;341. c4675.
46. Pelletier J-P, Hochberg MC, du Souich P, Kahan A, Michel BA. Effect size is encouraging. BMJ 2010;341. c6328.
47. Rowbotham MC. The impact of selective publication on clinical research in pain. Pain 2008;140:401–4.
48. Rowbotham MC. The case for publishing 'negative' trials. Pain 2009;146:225–6.

49. Scott-Lennox JA, McLaughlin-Miley C, Lennox RD, Bohlig AM, Cutler BL, Yan C, *et al*. Stratification of flare intensity identifies placebo responders in a treatment efficacy trial of patients with osteoarthritis. Arthritis Rheum 2001;44:1599–607.
50. Allen KD, Coffman CJ, Golightly YM, Stechuchak KM, Voils CI, Keefe FJ. Comparison of pain measures among patients with osteoarthritis. J Pain 2010;11:522–7.
51. Bellamy N, Wilson C, Hendrikz J, Whitehouse SL, Patel B, Dennison S. Osteoarthritis Index delivered by mobile phone (m-WOMAC) is valid, reliable, and responsive. J Clin Epidemiol 2011;64:182–90.
52. 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.
53. Dougados M, Moore A, Yu S, Gitton X. Evaluation of the Patient Acceptable Symptom State in a pooled analysis of two multicentre, randomized, double-blind, placebo-controlled studies evaluating lumiracoxib and celecoxib in patients with osteoarthritis. Arthritis Res Ther 2007;9. R11.
54. Bellamy N, Bell MJ, Goldsmith CH, Lee S, Maschio M, Raynauld J-P, *et al*. BLISS index using WOMAC index detects between-group differences at low-intensity symptom states in osteoarthritis. J Clin Epidemiol 2010;63:566–74.
55. U.S. Department of Health and Human Services. Guidance for industry: patient-reported outcome measures: use in medical product development to support labeling claims, 2009; Available at: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf.
56. Ruta DA, Garratt AM, Russell IT. Patient centred assessment of quality of life for patients with four common conditions. Qual Health Care 1999;8:22–9.
57. Clinch J, Tugwell P, Wells G, Shea B. Individualized functional priority approach to the assessment of health related quality of life in rheumatology. J Rheumatol 2001;28:445–51.
58. Jolles BM, Buchbinder R, Beaton DE. A study compared nine patient-specific indices for musculoskeletal disorders. J Clin Epidemiol 2005;58:791–801.

Osteoarthritis and Cartilage 19 (2011) 493–499

Osteoarthritis and Cartilage

Safety issues in the development of treatments for osteoarthritis: recommendations of the Safety Considerations Working Group

V. Strand †*, D.A. Bloch ‡, R. Leff §, P.M. Peloso ||, L.S. Simon ¶

† Division of Immunology and Rheumatology, Stanford University School of Medicine, Palo Alto, CA, USA
‡ Department of Biostatistics, Stanford University School of Medicine, Palo Alto, CA, USA
§ AstraZeneca, Wilmington, DE, USA
|| Clinical Research – Analgesia and Inflammation, Merck Research Laboratories, Rahway, NJ, USA
¶ SDG, LLC, Boston, MA, USA

ARTICLE INFO

Article history:
Received 26 November 2010
Accepted 8 February 2011

Keywords:
Safety assessments
Serious adverse events
Cardiovascular risks
NSAIDs
COX-2

SUMMARY

Objective: The symptomatic treatment of osteoarthritis (OA) remains to be improved, as many patients do not respond well to current palliative therapies and/or suffer unacceptable adverse events. Given the unmet need for innovative, effective and well-tolerated therapies, it is important to develop the means to estimate the ongoing safety profile of novel therapeutic agents over short- and longer term use.
Design: Methods are presented to estimate the number of serious adverse events (SAEs) of interest considered as "acceptable" per 1000 patient-years exposure and to estimate the numbers of patient-years needed in a randomized controlled trial (RCT) to meet objectives. As exposure is increased, more evidence is accrued that the overall risk is within study limits. It is equally important that requirements for delineating the safety of promising new therapies not create barriers that would preclude their development. Therefore, ongoing surveillance of occurrence of SAEs of interest during clinical development is proposed, for example after every incremental 500 patient-years exposure are accrued.
Results: This paper and others in this special issue focus on identification of safety signals for symptomatic treatments of OA. Much less information is available for agents aimed at slowing/preventing structural progression but it is expected that a higher risk profile might be considered acceptable in the context of more promising benefit.
Conclusion: This paper provides a proposal and supporting data for a comprehensive approach for assessing ongoing safety during clinical development of both palliative and disease-modifying therapies for OA.

Introduction

Osteoarthritis (OA) is a common and heterogeneous disease that occurs worldwide, predominantly in older individuals[60,37,52]. The pain, impairment in physical function, and disability associated with OA vary greatly from mild and intermittent to severe and continuous[51,50,53], prompting patients to seek a wide variety of treatments, ranging from intermittent use of analgesics to total joint arthroplasties, with greatly varying associated risks[61,34,32].

As cyclooxygenase-2 selective (COX-2) agents were developed that decreased the risk of gastrointestinal (GI) bleeding[1,45], it became apparent that both non-selective nonsteroidal anti-inflammatory drugs (nsNSAIDs) and selective COX-2s were associated with increased risks for cardiovascular (CV) events[43,47,49,25]. Results to date have led to the conclusion that treatment-associated increases in CV risk vary according to patient characteristics, underlying risk factors, specific NSAID/COX-2 administered, and dose and duration of treatment[13,2,56,55,31,25,26]. Recognizing that absolute rates of risk are small and the large number of factors influencing NSAID/COX-2-associated increases in CV risk[12,31] means that the incidence of treatment-associated CV events require evaluation, not only in multinational randomized control trials (RCTs), but also in large post-approval, randomized pragmatic trials and longitudinal observational studies (LOS)[4,28,56]. RCTs, cohort studies and case control series contribute information to the evolving safety profile of a novel therapeutic, once approved, and

* Address correspondence and reprint requests to: Vibeke Strand, Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, CA, USA
E-mail addresses: vstrand@stanford.edu (V. Strand), dbloch@stanford.edu (D.A. Bloch), Richard.leff@astrazeneca.com (R. Leff), Paul_peloso@merck.com (P.M. Peloso).
URL: http://www.sdgltd.com

1063-4584/$ – see front matter © 2011 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.joca.2011.02.019

each offer different types of information[29,7,4,28]. To the extent that assessment of uncommon serious adverse events (SAEs) relies upon voluntarily reported events in LOS, this may underestimate adverse event (AE) frequency and/or be confounded by channeling bias, other unidentified comorbidities and risk factors[17,40]. In the case of liver toxicity associated with NSAIDs and in particular with a recent COX-2 selective inhibitor, lumiracoxib, large numbers of patients needed be studied to characterize this rare risk.

Uncommon and/or less easily predicted complications of OA treatment (e.g., idiosyncratic skin rashes, including Stevens - Johnson syndrome and/or Toxic Epidermal Necrolysis)[48], those reflective of comorbidities (e.g., hypertension (HTN), diabetes)[6], and/or polypharmacy frequently present in subjects with OA[5,44] should also be considered. Recent data and RCTs indicate as many as 40–50% of OA subjects have HTN; they are twice as likely to develop a myocardial infarction (MI) and 70% more likely to suffer a cerebrovascular accident (CVA). In addition there is an associated increased risk of type II diabetes, with its own attendant CV risks. Other common comorbidities in the OA population include chronic obstructive pulmonary disease (COPD), peptic ulcer and other GI diseases, increased risk of obesity and metabolic syndrome and increased incidence of CV disease with increasing age, impairment in renal function and osteoporosis[11,58,59]. Thus, it is important that novel therapies under development include drug–drug interaction studies in this patient population as well as information regarding instability in blood pressure, blood glucose, and/or renal function during RCTs – and that CV events be carefully surveilled.

A question posed by the U.S. Food and Drug Administration (FDA) in 2007 regarding the Draft 1999 Guidance Document for development of novel agents for treatment of OA was "What should the size and duration of exposure of the safety database be for agents offering symptomatic relief?"[18]. This paper outlines recommendations for ongoing evaluation of the safety of novel agents for symptomatic treatment of OA. Other therapies, including topical or intra-articular agents that do not result in significant systemic drug exposure are not considered. In addition, pure analgesics without anti-inflammatory effects are not addressed. Simple analgesics, over-the-counter (OTC) acetaminophen and oral opioid drugs, have significant well-recognized safety risks, and are not included in this discussion[46,42,41,10].

Recommendations provided address studies of novel agents and it is acknowledged that the known safety profile for any new therapy will almost certainly evolve after approval and subsequent administration to thousands and millions of patients rather than the limited numbers typically involved in a clinical development program, also including those with comorbidities that would otherwise preclude their participation in pre-approval RCTs. The following discussion is also undertaken with the assumption that any new molecular entity in development for symptomatic relief of OA should have no evidence of risks beyond those identified with currently approved therapies since these agents are palliative in nature and do not alter the natural history of the disease. An acceptable safety profile for a disease-modifying agent may also be very different than that for a symptomatic therapy and a certain degree of greater risk may be acceptable for achievement of this benefit[57]. The magnitude of benefit with a novel palliative therapy for OA should be an important determinant of the number of patients required to demonstrate an acceptable understanding of its associated risk.

Current guidances

Safety databases vary according to size and populations studied, whether pre- or post-approval, by recognized risks, and class of therapeutic agent (Table I). Depending upon an *a priori* concern regarding SAEs based upon nonclinical information or results from early trials, larger studies may be required to better characterize the safety profile of a new therapy. As it is difficult to predict the safety profile of a novel agent and accurately determine the 95% confidence intervals (CIs) around the incidence of uncommon to rare SAEs, it is recommended that ongoing estimates of risks during clinical development be performed to inform decisions regarding the size of the database required for approval.

The previous 1999 FDA OA Draft Guidance Document did not specifically address safety recommendations[19] and International Conference on Harmonization (ICH) recommendations published in 1994[33] were generally applied for development of novel agents that would be used both intermittently and regularly on a chronic basis. The ICH guidelines recommend 1500 patients as the minimum number of subjects to have received a new therapeutic at any dose for any time period; 300–600 patients be treated for 6 months and a minimum of 100 patients for at least one year at the proposed dose.

With identification of relatively rare SAEs of variable incidence, the evaluation of risk based upon exposure (e.g., number of events per 100 patient-years) has become important. For example, clinical development programs with 3-month RCTs in OA aimed at assessing symptomatic relief typically have resulted in databases with approximately 1000 patient-years of exposure. As noted above, these limited databases may not permit identification of rare but likely important SAEs. For example, early biologic inhibitors of tumor necrosis factor (TNFα) for treatment of rheumatoid arthritis (RA) were approved with limited databases, and post-marketing surveillance was required to identify uncommon SAEs such as opportunistic infections, lymphomas and malignances[38,27]. Post-approval recognition of these SAEs motivated the requirement for 2500 patient-years of exposure for approval of subsequent new disease-modifying anti-rheumatic drugs (DMARDs) for treatment of RA[22–24].

Similarly, FDA has recently issued two guidances for evaluation of new agents for treatment of diabetes mellitus (DM), recommending a minimum of 3000 patient-years of exposure[20] and based on recognition of increased CV risk in subjects with Type 2 DM, 5000 patient-years of exposure[21].

Requirements for safety assessments are based upon point estimates of relative risk and the 95% CIs estimated around that risk. In the past relatively rare risks were better defined and 95% CIs narrowed by performance of large post-marketing safety trials conducted with the goal of increasing exposure by ~5000 patient-years and/or by studies which included subjects with more

Table I
Size of safety databases

	Patient-years exposure (approximate)
Osteoarthritis efficacy studies and ICH guidelines (estimated summation)	1000
DMARD approvals in RA (disease-modifying anti-rheumatic drugs: synthetics: 1998–; biologic agents: 2002–)	2500
DM CV risk guidance for approval (based upon RR 95% upper CI <1.8)	3000
DM CV risk guidance safety study (based upon RR 95% upper CI <1.3)	5000
OA CV outcome studies (TARGET, MEDAL, PRECISION studies)	10,000–30,000

underlying comorbidities than those enrolled in pre-approval RCTs. Based on "signals" identified during clinical development, such post-approval studies are requested on a more frequent basis. The recent guidance from the FDA regarding CV risk assessment for new therapies in Type 2 DM provides insight into why such recommendations have been issued.

The current guidance to assess pre-approval CV risk of therapies for Type II DM proposes that the upper limit of the 95% two-sided CI of the risk ratio relative to control must be <1.8 and the absolute risk ratio <1.5, representing a nominally significant increase. A risk ratio <1.3 may not require a post-marketing safety trial and those intermediate between 1.3 and <1.8 will require a clearly demonstrated positive benefit/risk ratio to support approval. Some scenarios of numbers of patient-years of exposure accrued that are unlikely to meet risk criteria are illustrated in Fig. 1, based on identification of major adverse cardiac events (MACEs) of CV death, fatal or non-fatal MI or CVA. Increasing numbers of patient-years of follow-up are required to be included in RCTs to identify SAEs with sufficient power (say 80%), especially if the predetermined incidence of SAEs of interest is rare and if the aim is to exclude risk ratios <1.8 with the incidence of interest equal to 2%, Fig. 1 shows the power is approximately 78% with 3000 treated and 2000 control patient-years of follow-up; if the incidence is 1%, then the power is approximately 83% with 7000 treated and 4000 control patient-years of follow-up. Clearly this poses a challenge as CV event rates in RCTs are generally ≤1% and do not approach 2% except in particularly high-risk populations. Similarly, event rates are lower in subjects with newly diagnosed or earlier disease, with fewer comorbidities; recognizing also that high-risk patients are generally excluded from RCTs early in clinical development.

In Appendix 1 a formula is presented to statistically estimate the number of patient-years of follow-up needed in RCTs to meet these objectives. For example, if approximately 14,800 treated patients-years and 7400 control patient-years have been accrued with an SAE rate of 0.5%, then the power to exclude a rate ratio of 1.8 is 80%. Effective implementation of a guidance requiring absolute risk be determined as <1.3–1.5 will be challenging, likely increasing clinical development times by 1–3 years and a minimum of $150–300 million in costs. Clearly these cost implications will limit the incentives for identifying and proving new therapies.

Symptomatic treatment of OA

Symptomatic agents for the treatment of OA are administered to large numbers of patients in a primary care setting[15]. Typically they are systemically active and may interact pharmacokinetically and/or pharmacodynamically with other therapies/drug classes and non-pharmaceutical agents (e.g., herbal remedies)[8,14]. Thus, the safety of oral treatments for OA must be carefully characterized in multiple settings with chronic intermittent and daily use.

Risk assessments for newly approved agents in a given therapeutic class continue to change over time following introduction of the first products into the clinic to identification of SAEs with new members of the class still in clinical development. The probability of identifying rare SAEs, typically identified *via* the Adverse Event Reporting System, increases with long-term exposure of larger numbers of patients with more diverse demographic and clinical characteristics than those enrolled in RCTs[4].

Based upon the above discussion, the following is proposed for assessment of novel therapies for OA. Minimum requirements by ICH guidelines require a database of 2500–3000 patients followed for one year if the SAE incidence rate is 0.1% (Table II).

A database of 2500–3000 patient-years will detect at least one SAE with high certainty if the incidence rate is of 0.1%. A standard phase 3 program, which meets current FDA requirements, has approximately 50%–75% power to exclude a hazard ratio of ≥1.8 in terms of CV events, assuming a 1%–2% rate in the study population (Fig. 1). Assuming that the observed event rate in such a phase 3 program is approximately 1%, additional clinical work will be necessary to better define the overall risk. To increase power, options could include either increasing exposure and number of patients in the pivotal trials or conducting a separate safety study (initiated prior to submission but likely not completed before approval). At present, a CV outcomes trial would likely require >20,000–30,000 patients treated for at least 3 years, akin to that planned in OA by the PRECISION (Prospective Randomized Evaluation of Celecoxib Integrated Safety versus Ibuprofen or Naproxen) study[3].

Predicated on the above guidance for the diabetic population, it is possible to extrapolate this requirement to patients with OA, who are usually older but with similar comorbidities and overall 10-year risk for development of coronary heart disease[36,16,39]. It is possible to develop a statistical estimate of the number of SAEs of interest considered as "acceptable" per a chosen number of patient-years exposure. As an example, for an SAE rate of 1/1000 patient-years, one can calculate that after 3065 patient-years of exposure, there should be no more than 6 SAEs, or the estimated lower limit of a the 95% CI of the true SAE rate exceeds 1/1000 patient-years. (see Appendix 1)

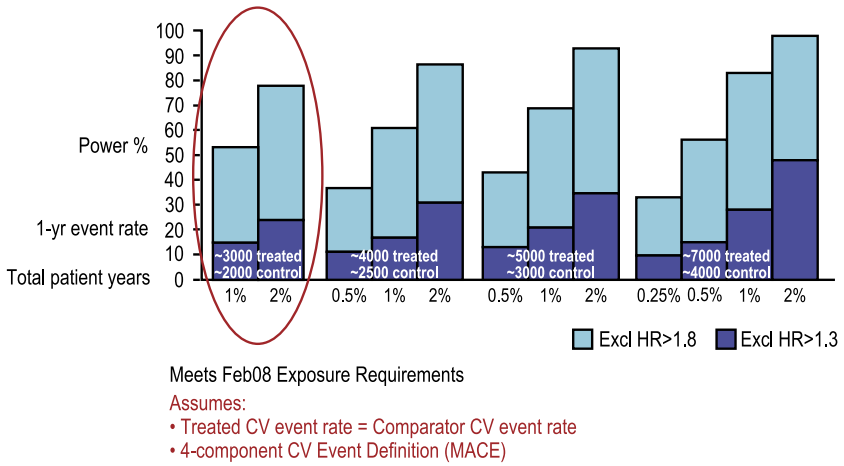

Fig. 1. Patient Exposure Recommendations from DM. Guidance Unlikely to Meet CV Risk Criteria. Power to exclude a Hazard Ratio or Risk Ratio of ≥1.8 or ≥1.3 based on SAEs of MACEs based on 3000 – 7000 patient-years exposure of the new therapy vs 2000 – 4000 of an active control with incidence of adverse events between 0.25 to 2%.

Table II
ICH estimates for study duration, exposure, and characterization of AE incidence rates (ICH, 1994)

Duration	Time	Exposure (patients)	Incidence Rate characterized
Short-term	≤3 months	1500	~1%
Mid term	6 months	300–600	0.5–5%
Long-term	1 year	100	3%
Not ICH characterized	≥1 year	2500–3000	0.1%

ICH guidelines (E1, 1994) are considered "minimums" to characterize the safety of a new agent, but:
- Don't reveal rare (<1/1000) or long-ter AEs nor,
- AEs in at risk or special populations (for example those with HTN, on low-dose aspirin or other concomitant medications)

This acceptable SAE rate can be calculated repeatedly as patient-years are accumulated throughout the clinical development program. Hence, one could propose that once an SAE of interest was identified, the number accrued as each RCT is completed could be calculated. As exposure is increased the 95% CIs around the risk would narrow, and either lend more or less confidence to the estimate that the overall risk remains within the predefined acceptable rate, for example <1/1000. These assumptions permit establishment of CI's for actual SAE rates, to decide if they violate the predefined "acceptable rate". For the example, after 1000 patient-years of follow-up one expects 10 such SAEs, but as many as 15 SAEs could occur before establishing with 95% certainty that the true rate exceeds the predefined rate. Once 3000 patient-years have been accrued, as many as 39 SAEs could occur before establishing with 95% certainty that the true rate exceeds the predefined rate. The example in Appendix 1 illustrates how a proposal could be pre-specified that after every incremental 500 patient-years exposure are accrued, the number of observed SAEs of interest would be compared to the allowable limit to determine if the SAE rate is in danger of violating the "acceptable" rate. This requires an ongoing surveillance of the occurrence of SAEs of interest in the clinical development program, but may preclude the need for large expensive post-marketing surveillance studies.

Statistically, this same metric can be applied to patient populations in post-marketing surveillance. Practically, this requires an agreement regarding an estimate of patient-years and confidence that all SAEs are reported and adjudicated. More realistically, this metric for RCTs can be applied to LOS to monitor SAEs of interest after more patient-years of follow-up are accrued, as in current registries for RA and other health provider databases. Critical considerations include the definition of SAEs of interest, such as MACEs as well as the definition of an "acceptable" SAE rate per 100 patient-years exposure and the power available to rule out excessive rate ratios. As an example, Table III presents CV risk estimates based on RCTs, cohort studies and case control series with selected NSAIDs.

Table III
CV risk estimates from various databases for selected NSAIDs

Drug	Database	CV risk or hazard rate (95% CI)
Celecoxib	RCTs	1.10 (0.70–1.60)
		1.30 (0.60–2.60)
		2.30 (0.90–5.50)
	Cohort studies	1.32 (0.69–2.16)
	Case control series	1.01 (0.90–1.13)
Naproxen	RCTs	1.57 (0.87–2.61)
	Cohort studies	0.94 (0.85–1.04)
	Case control series	0.96 (0.84–1.10)
Ibuprofen	RCTs	1.18 (0.93–1.19)
	Cohort studies	1.12 (0.90–1.38)
	Case control series	1.06 (0.95–1.18)
Diclofenac	RCTs	1.05 (0.93–1.19)
	Cohort studies	1.36 (0.51–3.65)
	Case control series	1.36 (1.21–1.54)

Strand Lancet 2007 [McGettigan *et al.* summarized with permission and Solomon, *et al.* updated for Lancet 2007 publication]

Post-marketing commitments for RCTs and/or observational studies should be focused upon patient populations likely to be at higher risk for uncommon AEs and, therefore, not frequently studied in sufficient numbers prior to approval to estimate such risk. There should be a commitment to collect safety information after approval to narrow the "window" of CI estimates around actual risks (known or unknown) to <1:10,000–1:30,000. To achieve this, RCTs and other means (e.g., claims databases) should be used. The present recommendations also take the position that an outcomes study prior to registration should not be required if the SAE rate of interest remains low and within the acceptable predefined limits.

Remaining questions

There are a number of questions that remain to be addressed regarding the safety of OA treatment. The most important of these include:

- Do the comorbidities in OA require considerations similar to those in diabetic populations?
- What approach to safety analysis do we employ for non-systemically absorbed, topical and/or intra-articular-administered products?
- How should we assess safety for potentially structure-modifying and/or "preventive" agents?
- How do we strike a "reasonable balance" between potential risk and promising benefit?
- What would be the impact on sample size calculations of identified genetic polymorphisms potentially affecting both safety and efficacy?
- In this regards, what role might population pharmacokinetics play?
- If we can enrich enrolled patient populations for extreme examples of the above (e.g., extensive and poor metabolizers and/or patients expressing higher vs lower levels of target receptor), how might we reduce sample sizes?

Expensive large scale RCTs may not answer the appropriate questions, due to types of patients accrued and well-recognized confounders such as dropouts and appropriate comparator agents, as well as the duration needed to generate sufficient data. It is hoped that this proposal offers a flexible approach to assess the safety of a promising novel therapy in OA that is also pragmatic.

Declaration of Funding and Role of Funding Source

The OARSI FDA OA Initiative received financial support from the following professional organization:

American College of Rheumatology

Additionally the OARSI FDA OA Initiative received financial support from the following companies:

Amgen
ArthroLab
AstraZeneca
Bayer Healthcare
Chondrometrics
CombinatoRx
Cypress BioScience

DePuy Mitek
Expanscience
4QImaging
Genevrier/IBSA
Genzyme
King (Alpharma)
Merck
Merck Serono
NicOx
Pfizer
Rottapharm
Smith & Nephew
Wyeth

While individuals from pharmaceutical, biotechnology and device companies actively participated in ongoing working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the Food and Drug Administration.

Author contributions

VS role: conception, drafting, critical revision and final approval of the manuscript
DB role: conception, critical revision
RL role: critical revision
PP role: critical revision
LS role: drafting, critical revision and final approval of the manuscript

Conflicts of interest

VS, DB, LS: no conflicts of interest to declare
RL: employee of AstraZeneca
PP: employee of Merck & Co.

Appendix 1. Statistical methodology

Let t_i denote the person-years of observation of the i-th patient in a cohort of N patients followed within the interval of time [0, T], and $M = \sum_1^n t_i$ denote the number of patient-years of observation of the entire cohort. During the follow-up period each patient either does or does not have the serious adverse event (SAE) of interest.

Let $\delta_i = 1$ if the i-th patient has a serious adverse event and $\delta_i = 0$ if the patient does not have a serious adverse event.

Let K represent the random variable for the number of SAEs and let $k = \sum_1^N \delta_i$; that is, $k =$ number of events observed in the interval $[0, T]$.

Assume that K is distributed as a Poisson random variable with mean and variance equal to the theoretical rate (λ) that we wish to estimate, multiplied by M. Thus

$$\text{Probability}\,(K = k) = e^{-\lambda M}(\lambda M)^k \frac{1}{k!} \tag{1}$$

The mean ("expected value") and variance of K are both equal λM. The rate estimator is given by K/M, which is estimated by k/M.

The variable K/M is approximately normally distributed and a confidence interval for the rate is more accurately estimated by natural logarithmic transformation. The lower and upper one-tailed bounds of the 95% CI of ln (K/M) are estimated by

$$\begin{matrix} \ln(k/M) - 1.645/\sqrt{k} \\ \ln(k/M) + 1.645/\sqrt{k} \end{matrix} \tag{2}$$

Should two-sided intervals be required, then replace 1.645 by 1.96 in the equation (2) expressions.

Number of SAEs allowed for a given number of patient-years of follow-up

By taking the exponential of the above interval one obtains the lower bound of the 95% CI for the rate itself,

$$\frac{k}{M} e^{-1.645/\sqrt{k}} \tag{3}$$

Let $L =$ minimum total number of patient-years of follow-up that the study subjects could have accrued by time T, at which time k SAEs have occurred, to establish with 95% confidence that the true rate does not exceed **"Y/1000"** patient follow-up years. The choice of "Y" is pre-specified; for example, if $Y = 1$, then the rate is 1/1000 years of follow-up, or 0.1% per year of follow-up. Formula (3) above is the one used to derive numerical results. This is done in two steps as follows:

(i) Set $\frac{k}{L} e^{-1.645/\sqrt{k}} = Y/1000$
(ii) For a fixed number of SAEs (k), solve equation (i) for L

Example 1: If $Y = 1$ and $k = 6$,

$$L = \frac{k e^{-1.645/\sqrt{k}}}{Y/1000} = \frac{6}{1/1000} e^{-1.645/\sqrt{6}} = 6(0.5109)(1000)/1$$
$$= 3065.4 \text{ years.}$$

This would be rounded up to $L = 3066$ years. In words, if there are least 3066 years of patient follow-up at which time at most 6 patients have been identified as having had the SAE, then the Investigator can be assured (with 95% confidence) that the evidence at that time does not support a conclusion that the true SAE rate exceeds 1/1000 patient-years of follow-up. However, if either 6 SAEs have occurred before 3066 patient-years have been accrued or if 7 SAEs have occurred in the first 3066 years of follow-up, then there is 95% confidence that the true rate exceeds 1/1000 patient-years of follow-up.

Example 2: How a proposal could be pre-specified during drug development that after every incremental 500 patient-years of exposure are accrued, the number of observed SAEs of interest would be compared to the allowable limit, to determine if, with 95% confidence, the SAE rate is in danger of violating a pre-registration "acceptable" rate. The SAE rate of interest chosen for this example equals 1% (10/1000 patient-years of follow-up).

Number of allowed SAEs	8	15	21	27	33	39	44	50
Number of patient-years	500	1000	1500	2000	2500	3000	3500	4000

Calculating the number of patient-years of observation for a controlled clinical trial comparing SAEs with a rate ratio

Assume the two groups to be compared are a treatment group followed for M_T patient-years and a control group followed for M_C patient-years in the ratio $r = M_C/M_T$ and the number of SAEs after M_T (or M_C) patient-years of follow-up is Poisson distributed; that is, with mean and variance equal to the theoretical rate λ_T (or λ_C) multiplied by M_T (or M_C), as represented by equation (1) above. Assume k_T and k_C are the observed numbers of SAEs observed in the treatment group and control group, respectively. The aim is to demonstrate that the true ratio, $R = \lambda_T/\lambda_C$ is not greater than a pre-specified amount, R_U say.

Formally, the null and alternative hypotheses are $H_O: \lambda_T/\lambda_C \geq R_U$ and $H_1: \lambda_T/\lambda_C < R_U$. The estimates of λ_T and λ_C are denoted by $\hat{\lambda}_T = k_T/M_T$ and $\hat{\lambda}_C = k_C/M_C$. The null hypothesis is rejected in favor of the alternative hypothesis if $\hat{\lambda}_T - \hat{\lambda}_C R_U$ is "small enough."

If the alternative hypothesis is true, assume $\lambda_T = \lambda_C = \lambda$, so that $R = 1$. In what follows we present formulas for M_T and M_C assuming the hypotheses are tested at the α-level of statistical significance with power = (1-β). Using the methods presented by Miettinen and Nurminen (Reference: Miettinen O and Nurminen M. "Comparative analysis of two rates." *Statistics in Medicine* 1985;4:213–226) and by Laster and Johnson (Reference: Laster LL and Johnson MF. "Non-inferiority trials: the 'at least as good as' criterion." *Statistics in Medicine* 2003;22:187–200), denote the null and alternative variances of $\hat{\lambda}_T - \hat{\lambda}_C R_U$ by V_0 and V_1, respectively. $V_1 = \lambda(\frac{1}{M_T} + \frac{R_U^2}{M_C})$ and $V_0 = \frac{\bar{\lambda}_T}{M_T} + \frac{\bar{\lambda}_C}{M_C}R_U^2$ where $\bar{\lambda}_T$ and $\bar{\lambda}_C$ are maximum likelihood estimators of λ_T and λ_C under the null hypothesis restriction $\bar{\lambda}_T/\bar{\lambda}_C = R_U$. The restricted maximum likelihood estimators are $\bar{\lambda}_C = \lambda(1+r)/(R_U + r)$ and $\bar{\lambda}_T = \bar{\lambda}_C R_U$. Then

$$M_T = \frac{\left\{Z_\alpha\sqrt{\bar{\lambda}_T + \bar{\lambda}_C\left(R_U^2/r\right)} + Z_\beta\sqrt{\lambda\left[1 + \left(R_U^2/r\right)\right]}\right\}^2}{\lambda^2(1-R_U)^2} \quad (4)$$

and

$$M_C = rM_T \quad (5)$$

where Z_u is the upper u-th percentile of the Standard Normal distribution.

Example 1. A trial has accrued 5000 patient-years of follow-up in the treatment group ($M_T = 5000$) and 3000 patients-years of follow-up in the control group ($M_C = 3000$). The predetermined SAE rate of interest is 1%, expected to be equal in both groups ($\lambda = 0.01$). Choose 1.8 as the upper limit of the rate ratio ($R_U = 1.8$) that is of interest to exclude at the 5% 2-tailed significance level ($Z_\alpha = 1.96$). Then $r = M_C/M_T = 0.6$, $\bar{\lambda}_C = 0.0067$ and $\bar{\lambda}_T = 0.012$. Solving equation (4) for Z_β gives $Z_\beta = 0.54$. Referring to the cumulative Standard Normal distribution, power = 70.5%. This is in close agreement with the power presented for this approximate scenario in Fig. 1.

Example 2. The aim is to design a post-registration study with 80% power ($Z_\beta = 0.84$) at the 5% 2-tailed significance level ($Z_\alpha = 1.96$) to rule out a rate ratio of 1.8 or larger ($R_U = 1.8$) where the predetermined SAE rate of interest is 0.5%, expected to be equal in both groups ($\lambda = 0.005$). Assume twice as many patients will be treated than controls, so $r = 0.5$. Then $\bar{\lambda}_C = 0.0033$ and $\bar{\lambda}_T = 0.0059$. From equations (4) and (5), $M_T = 14{,}775$ patient-years and $M_C = 7388$ patient-years. As presented in Fig. 1, if approximately 7000 treated patients-years and 4000 control patient-years have been accrued with the predetermined SAE rate of interest equal to 0.5%, then the power to exclude the rate ratio of 1.8 is approximately 56%.

References

1. Altman RD. Practical considerations for the pharmacologic management of osteoarthritis. Am J Manag Care 2009;15(Suppl 8):S236–43.
2. Beaulieu M, Choquette D, Rahme E, Bessette L, Carrier R. CURATA: A patient health management program for the treatment of osteoarthritis in Québec: an integrated approach to improving the appropriate utilization of anti-inflammatory/analgesic medications. Am J Manag Care 2004;10:569–75.
3. Becker MC, Wang TH, Wisniewski L, Wolski K, Libby P, Lüscher TF, *et al.* Rationale, design, and governance of prospective randomized evaluation of celecoxib integrated safety versus ibuprofen or naproxen (PRECISION), a cardiovascular end point trial of nonsteroidal antiinflammatory agents in patients with arthritis. Am Heart J 2009;157:606–12.
4. Berlin JA, Glasser SC, Ellenberg SS. Adverse event detection in drug development: recommendations and obligations beyond phase 3. Am J Public Health 2008;98:1366–71.
5. Beyth RJ, Shorr RI. Epidemiology of adverse drug reactions in the elderly by drug class. Drugs Aging 1999;14:231–9.
6. Blackshear JL, Davidman M, Stillman MT. Identification of risk for renal insufficiency from nonsteroidal anti-inflammatory drugs. Arch Intern Med 1983;143:1130–4.
7. Brewster W, Gibbs T, Lacroix K, Murray A, Tydeman M, Almenoff J. Evolving paradigms in pharmacovigilance. Curr Drug Saf 2006;1:127–34.
8. Brouwers JR, de Smet PA. Pharmacokinetic-pharmacodynamic drug interactions with nonsteroidal anti-inflammatory drugs. Clin Pharm 1994;27:462–85.
10. Chan AT, Manson JE, Albert CM, Chae CU, Rexrode KM, Curhan GC, *et al.* Nonsteroidal antiinflammatory drugs, acetaminophen, and the risk of cardiovascular events. Circulation 2006;113:1578–87.
11. Cimmino MA, Sarzi-Puttini P, Scarpa R, Caporali R, Parazzini F, Zaninelli A, *et al.* Clinical presentation of osteoarthritis in general practice: determinants of pain in Italian patients in the AMICA study. Semin Arthritis Rheum 2005;35(Suppl 1):17–23.
12. Cunnington M, Webb D, Qizilbash N, Blum D, Mander A, Funk MJ, *et al.* Risk of ischaemic cardiovascular events from selective cyclooxygenase-2 inhibitors in osteoarthritis. Pharmacoepidemiol Drug Saf 2008;17:601–8.
13. Curhan GC, Willett WC, Rosner B, Stampfer MJ. Frequency of analgesic use and risk of hypertension in younger women. Arch Intern Med 2002;162:2204–8.
14. Di YM, Li CG, Xue CC, Zhou SF. Clinical drugs that interact with St. John's wort and implication in drug development. Curr Pharm Des 2008;14:1723–42.
15. Dziedzic KS, Hill JC, Porcheret M, Croft PR. New models for primary care are needed for osteoarthritis. Phys Ther 2009;89:1371–8.
16. Erb N, Pace AV, Douglas KM, Banks MJ, Kitas GD. Risk assessment for coronary heart disease in rheumatoid arthritis and osteoarthritis. Scand J Rheumatol 2004;33:293–9.
17. Figueiras A, Tato F, Fontaiñas J, Gestal-Otero JJ. Influence of physicians' attitudes on reporting adverse drug events: a case-control study. Med Care 1999;37:809–14.
18. Food and Drug Administration Federal Register Notice. Docket No.1998D-0077 [formerly 98D-0077]; August 14, 2007.
19. FDA Guidance for Industry. clinical development programs for drugs, devices, and biological products intended for the treatment of osteoarthritis (OA), http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM071577.pdf; July 1999.
20. FDA. Guidance for industry diabetes mellitus: developing drugs and therapeutic biologics for treatment and prevention HHS/FDA/CDER DRAFT, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071624.pdf; February 2008.
21. FDA. Guidance for industry diabetes mellitus – evaluating cardiovascular risk in antidiabetic therapies to treat type 2 diabetes, http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071627.pdf; December 2008.
22. FDA Product Labeling: adalimumab. http://www.accessdata.fda.gov/drugsatfda_docs/label/2008/125057s0110lbl.pdf
23. FDA Product Labeling: abatacept. http://www.accessdata.fda.gov/drugsatfda_docs/label/2009/125118s0086lbl.pdf
24. FDA Product Labeling: tocilizumab. http://www.accessdata.fda.gov/drugsatfda_docs/label/2010/125276lbl.pdf

25. Fosbøl EL, Folke F, Jacobsen S, Rasmussen JN, Sørensen R, Schramm TK, *et al*. Cause-specific cardiovascular risk associated with nonsteroidal antiinflammatory drugs among healthy individuals. Circulation: Circ Cardiovasc Qual Outcomes 2010;3:395–405.
26. Fosbøl EL, Gislason GH, Jacobsen S, Folke F, Hansen ML, Schramm TK, *et al*. Risk of myocardial infarction and death associated with the use of nonsteroidal anti-inflammatory drugs (NSAIDs) among healthy individuals: a nationwide cohort study. Clin Pharmacol Ther 2009;85:190–7.
27. Furst DE. The risk of infections with biologic therapies for rheumatoid arthritis. Semin Arthritis Rheum 2010;39:327–46.
28. Glasser SP, Salas M, Delzell E. Importance and challenges of studying marketed drugs: what is a phase IV study? Common clinical research designs, registries, and self-reporting systems. J Clin Pharmacol 2007;47:1074–86.
29. Hammond IW, Gibbs TG, Seifert HA, Rich DS. Database size and power to detect safety signals in pharmacovigilance. Expert Opin Drug Saf 2007;6:713–21.
31. Hernández-Díaz S, Varas-Lorenzo C, García Rodríguez LA. Nonsteroidal antiinflammatory drugs and the risk of acute myocardial infarction. Basic Clin Pharmacol Toxicol 2006;98:266–74.
32. Hochberg MC, Perlmutter DL, Hudson JI, Altman RD. Preferences in the management of osteoarthritis of the hip and knee: results of a survey of community-based rheumatologists in the United States. Arthritis Care Res 1996;9:170–6.
33. ICH Guidelines: the Extent of Population Exposure to Assess Clinical Safety for Drugs Intended for Long-Term Treatment of Non-Life Threatening Conditions. E1; http://www.ich.org/LOB/media/MEDIA435.pdf
34. Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P, *et al*. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2003;62:1145–55.
36. Lanas A, Tornero J, Zamorano JL. Assessment of gastrointestinal and cardiovascular risk in patients with osteoarthritis who require NSAIDs: the LOGICA study. Ann Rheum Dis 2010;69:1453–8.
37. Lawrence RC, Felson DT, Helmick CG, Arnold LM, Choi H, Deyo RA, *et al*. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Arthritis Rheum 2008;58:26–35. Part II.
38. Lee JH, Slifman NR, Gershon SK, Edwards ET, Schwieterman WD, Siegel JN, *et al*. Life-threatening histoplasmosis complicating immunotherapy with tumor necrosis factor alpha antagonists infliximab and etanercept. Arthritis Rheum 2002;46:2565–70.
39. Lipscomb ER, Finch EA, Brizendine E, Saha CK, Hays LM, Ackermann RT. Reduced 10-year risk of coronary heart disease in patients who participated in a community-based diabetes prevention program: the DEPLOY pilot study. Diabetes Care 2009;32:394–6.
40. Lopez-Gonzalez E, Herdeiro MT, Figueiras A. Determinants of under-reporting of adverse drug reactions: a systematic review. Drug Saf 2009;32:19–31.
41. Mazer M, Perrone J. Acetaminophen-induced nephrotoxicity: pathophysiology, clinical manifestations, and management. J Med Toxicol 2008;4:2–6.
42. McClain CJ, Price S, Barve S, Devalarja R, Shedlofsky S. Acetaminophen hepatotoxicity: an update. Curr Gastroenterol Rep 1999;1:42–9.
43. McGettigan P, Henry D. Cardiovascular risk and inhibition of cyclooxygenase: a systematic review of the observational studies of selective and nonselective inhibitors of cyclooxygenase 2. J Am Med Assoc 2006;296:1633–44.
44. The National Collaborating Centre for Chronic Conditions. Osteoarthritis National Clinical Guideline for Care and Management in Adults, http://www.nice.org.uk/nicemedia/pdf/CG059FullGuideline.pdf; 2008
45. Niculescu L, Li C, Huang J, Mallen S. Pooled analysis of GI tolerability of 21 randomized controlled trials of celecoxib and nonselective NSAIDs. Curr Med Res Opin 2009;25:729–40.
46. Pergolizzi J, Böger RH, Budd K, Dahan A, Erdine S, Hans G, *et al*. Opioids and the management of chronic severe pain in the elderly: consensus statement of an International Expert Panel with focus on the six clinically most often used World Health Organization Step III opioids (buprenorphine, fentanyl, hydromorphone, methadone, morphine, oxycodone). Pain Pract 2008;8:287–313.
47. Ray WA, Varas-Lorenzo C, Chung CP, Castellsague J, Murray KT, Stein CM, *et al*. Cardiovascular risks of nonsteroidal antiinflammatory drugs in patients after hospitalization for serious coronary heart disease. Circulation: Cardiovascular Quality and Outcomes 2009;2:155–63.
48. Roujeau JC. Clinical aspects of skin reactions to NSAIDs. Scand J Rheumatol 1987;65:131–4.
49. Roumie CL, Choma NN, Kaltenbach L, Mitchel Jr EF, Arbogast PG, Griffin MR. Non-aspirin NSAIDs, cyclooxygenase-2 inhibitors and risk for cardiovascular events-stroke, acute myocardial infarction, and death from coronary heart disease. Pharmacoepidemiol Drug Saf 2009;18:1053–63.
50. Rosemann T, Laux G, Kuehlein T. Osteoarthritis and functional disability: results of a cross sectional study among primary care patients in Germany. BMC Musculoskelet Dis 2007;8:79.
51. Rosemann T, Laux G, Szecsenyi J, Wensing M, Grol R. Pain and osteoarthritis in primary care: factors associated with pain perception in a sample of 1,021 p.tients. Pain Med 2008;9:903–10.
52. Sacks JJ, Luo YH, Helmick CG. Prevalence of specific types of arthritis and other rheumatic conditions in the ambulatory health care system in the United States, 2001-2005. Arthritis Care Res 2010;62(4):460–4.
53. Sharma L, Cahue S, Song J, Hayes K, Pai YC, Dunlop D. Physical functioning over three years in knee osteoarthritis: role of psychosocial, local mechanical, and neuromuscular factors. Arthritis Rheum 2003;48:3359–70.
55. Solomon DH, Glynn RJ, Rothman KJ, Schneeweiss S, Setoguchi S, Mogun H, *et al*. Subgroup analyses to determine cardiovascular risk associated with nonsteroidal antiinflammatory drugs and coxibs in specific patient groups. Arthritis Rheum 2008;59:1097–104.
56. Strand V. Patients with cardiovascular risk receiving low dose aspirin, is a COX-2 selective agent preferable to nsNSAIDs for antiinflammatory therapy? Lancet 2007;370:2138–57.
57. Thompson MS. Willingness to pay and accept risks to cure chronic disease. Am J Public Health 1986;76:392–6.
58. Buominen U, Blom M, Hirvonen J, Seitsalo S, Lehto M, Paavolainen P, *et al*. The effect of comorbidities on health related quality of life in patients placed on the waiting list for total joint replacement. Health Qual Life Outcomes 2007 Mar 15;5:16. doi: 10.1186/1477-7525-5-16.
59. Van Dijk GM, Veenhof C, Schellevis F. Comorbidity, limitations in activities and pain in patients with osteoarthritis of the hip or knee. BMC Musculoskelet Disord 2008 Jun 26;9:95. doi: 10.1186/1471-2474-9-95.
60. Woolf AD, Pfleger B. Burden of major musculoskeletal conditions: policy and practice. Bone and joint decade 2000–2010. Bull World Health Organ 2003;81:646–56.
61. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *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.

Osteoarthritis and Cartilage 19 (2011) 500–508

Osteoarthritis and Cartilage

Methodologic issues in clinical trials for prevention or risk reduction in osteoarthritis

J.M. Jordan †*, M.F. Sowers ‡, S.P. Messier §, J. Bradley ||, G. Arangio ¶, J.N. Katz #††‡‡§§, E. Losina ‡‡§§|||, L. Rovati ¶¶, N. Bachtell ##, C. Cooper †††, T. Spector ‡‡‡, W. Zhang §§§, J. Gardiner |||||, M. Wahba ¶¶¶

† *Thurston Arthritis Research Center, University of North Carolina, Chapel Hill, NC, USA*
‡ *University of Michigan, School of Public Health, Department of Epidemiology, Ann Arbor, MI, USA*
§ *Department of Health and Exercise Science, Wake Forest University, Winston-Salem, NC, USA*
|| *Pfizer Global Research & Development, Groton, CT, USA*
¶ *Pennsylvania State Hershey Medical College, Hershey, PA, USA*
# *Department of Orthopedic Surgery, Harvard Medical School, Boston, MA, USA*
†† *Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA*
‡‡ *Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA*
§§ *Orthopedic and Arthritis Center for Outcomes Research, Department of Orthopedic Surgery, Brigham and Women's Hospital, Boston, MA, USA*
||| *Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA*
¶¶ *Division of Research & Development, Rottapharm SpA, Monza, Italy*
Genzyme Corporation, Cambridge, MA, USA
††† *MRC Environmental Unit, Southampton General Hospital, Oxford, UK*
‡‡‡ *Department of Twin Research, St Thomas' Hospital Campus, Kings College London School of Medicine, London, UK*
§§§ *Academic Rheumatology, University of Nottingham, Clinical Sciences Building, City Hospital, Nottingham NG5 1PB, UK*
||||| *Discovery Translational Medicine, Wyeth Research, Collegeville, PA, USA*
¶¶¶ *US Arthritis, Novartis Pharmaceutical Corporation, East Hanover, NJ, USA*

ARTICLE INFO

Article history:
Received 20 June 2010
Accepted 26 October 2010

Keywords:
Prevention
Clinical trial
Obesity
Injury
Osteoarthritis

SUMMARY

The design and execution of prevention trials for OA have methodological issues that are distinct from trials designed to impact prevalent disease. Disease definitions and their precise and sensitive measurement, identification of high-risk populations, the nature of the intervention (pharmaceutical, nutraceutical, behavioral) and its potential pleiotropic impacts on other organ systems are critical to consider. Because prevention trials may be prolonged, close attention to concomitant life changes and co-morbidities, adherence and participant retention in the trial is of primary importance, as is recognition of the potential for "preventive misconception" and "behavioral disinhibition" to affect the ability of the trial to show an effect of the intervention under study. None of these potential pitfalls precludes a successful and scientifically rigorous process and outcome. As technology improves the means to measure and predict the OA process and its clinical consequences, it will be increasingly possible to screen individuals for high-risk phenotypes, combining clinical factors with information from imaging, genetic, metabolic and other biomarkers and to impact this high-risk condition to avoid or delay OA both structurally and symptomatically.

Osteoarthritis (OA) is the most common specific arthritis condition, affecting 27 million people in the United States in 2005[1]. Knee and hip OA are generally considered to have the greatest impact due to effects on ambulation[2]. OA of these joints accounted for 97% of the total knee replacements and 8% of the total hip replacements for arthritis in 2004[3]. OA, however, is frequently a generalized condition, involving multiple joint sites, including the hand, knee, hip, great toe, and spine, all of which can be associated with significant symptoms and disability[4–6].

In 2007, the Osteoarthritis Research Society International (OARSI) was awarded a contract from the Federal Drug Administration (FDA) to review issues related to the design and conduct of clinical trials for OA, particularly pertaining to agents purporting to effect disease modification (See Introduction to Issue). Several categories of

* Address Correspondence and reprint requests to: J.M. Jordan, Thurston Arthritis Research Center, University of North Carolina, Chapel Hill, NC 27599-7280, USA.
E-mail address: joanne_jordan@med.unc.edu (J.M. Jordan).

1063-4584/$ – see front matter © 2011 Published by Elsevier Ltd on behalf of Osteoarthritis Research Society International.
doi:10.1016/j.joca.2010.10.031

inquiries, and Working Groups to examine them, were established, including Imaging, Biomarkers, Definition of Disease State, Safety, and Prevention and Risk Reduction. This paper will discuss the outcome of deliberations by the Working Group for Prevention and Risk Reduction. This Working Group was composed of individuals from academia and the pharmaceutical industry. The remit of this group was to examine potential outcome measures, the desirable duration of, and population for, an OA prevention trial, and the safety database and acceptable risk that would be required for prevention. Lastly, a research agenda to inform these issues was requested. Through a series of face-to-face meetings, telephone conferences, and electronic mail exchanges over almost 2 years, the members of the Working Group discussed these relevant questions, reviewed literature as required to inform answers, and presented the final product to a public forum attended by representatives from the FDA, the OARSI, and the National Institutes of Health (NIH), and academic and industry/private foundation communities.

Generally, clinical trials in OA have addressed three major types of outcomes: (1) symptoms of pain, function, and stiffness; (2) structural disease progression; and (3) replacement of affected joints. The clinical trials in OA to relieve symptoms of pain or stiffness and to improve function may involve pharmaceutical or nutraceutical agents[7], devices (i.e., braces, shoe orthotics), or behavioral interventions, such as weight-reduction, exercise, or increased physical activity[8–12]. The less common disease-modifying trials aim to demonstrate slowing of the rate of structural progression, (frequently measured by change in joint space narrowing on radiographs of the knee or hip[13–15]) and have employed pharmaceuticals or neutraceutical with, for example, putative anti-oxidant properties, the ability to inhibit cartilage degradative enzymes, impact bone turnover, modulate inflammation, or enhance or induce cartilage repair and/or lubrication[16,17]. The goal of these trials is to prevent *structural progression of established disease*, or to prevent disability or the need for total joint replacement, an indicator of total joint failure, in those with established disease, ie tertiary prevention. A third major type of OA trial involves evaluation of actions intended to assure the safety, efficiency, and efficacy of joint replacement.

This report will address the *primary and secondary prevention* and *risk reduction* of structural and symptomatic indicators of OA. These types of trials face specific hurdles because the onset of OA can be insidious and progression slow, with consequently, the need for trials of long duration, or the use of proxy measures with imperfect sensitivity and specificity for development of OA clinical outcomes, to allow the trial to be feasible. This report will discuss definitions, eligible populations and high-risk groups to whom initial prevention efforts might be directed for proof of concept, and possible outcome/surrogate outcome measures for primary and secondary prevention and risk reduction (Fig. 1). Then, an example of a prevention and a risk reduction trial for knee OA, directed at the high-risk group of those who are overweight or obese, and young athletes at risk of knee injury, respectively, will be proposed. These example trial designs are directed at knee OA with the understanding that OA in other joint sites (i.e., hands and hip) may have different prevalences, different risk factor profiles, different natural history of development and unique measures to define the disease state. Therefore, the approaches in these examples may not be generalizable to OA affecting joints other than the knee. In these examples, the recommended duration of a trial and appropriate database for safety will be outlined. Finally, ethical issues surrounding the conduct of clinical trials for OA prevention will be introduced.

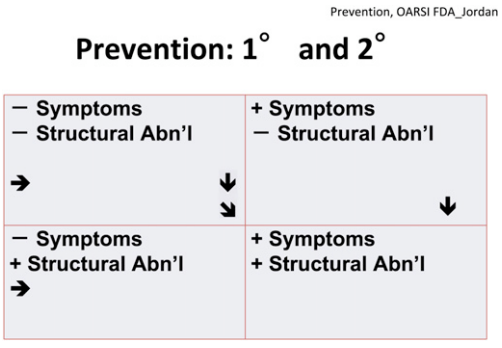

Fig. 1. Structural abn'l = structural abnormality.

Definitions of prevention and risk reduction

For the purposes of this report, *prevention* refers to those agents or actions that curtail or delay the onset or new occurrence of clinically diagnosed OA at the joint site of interest, in someone initially without evidence satisfying the clinical definition of the condition. Components of this definition may include structural evidence, e.g., on radiographs, and characteristic signs and symptoms, e.g., bony enlargement, crepitus, and/or pain. This report will **not** address tertiary prevention, or treatment, to modify the progression of established disease or achieve the maximum accommodation of living with established disease. *Risk reduction* refers to decreasing specific and modifiable risk factors associated with the development of OA, in an attempt to decrease the likelihood of developing OA or to delay its onset. For example, since obesity and overweight are strong risk factors for knee OA, a weight loss intervention could be evaluated to determine its ability to reduce the risk of developing knee OA in the obese. Similarly, since joint trauma, with its frequently resultant altered biomechanics, is a strong risk factor for the development of OA, an intervention to alter abnormal biomechanics in those with joint injury could also be considered in a preventive context for OA. Further, an intervention to prevent joint injury in the first place would be an example of risk reduction. It must also be acknowledged that an intervention may be both a preventive measure and a risk reduction measure, i.e., a weight loss intervention would fit both categories though the outcomes would differ (incident OA vs loss of weight).

Because the presentation of OA is frequently generalized, i.e., occurs in more than one joint in more than one joint group, an intervention could be applied in someone with OA in one joint site, in order to prevent the development of OA in another joint site unaffected at the start of the trial. For example, those with hand OA could be the subject of a prevention trial to prevent the development of OA in the knees or hips[6]. This situation blurs the distinction between incidence of new disease and progression of established disease, and may need to be considered on a case-by-case basis, with statistical methodology applied to allow for the non-independence of multiple joints within the same person. This also suggests that collection of information about joints beyond the target joint should be considered at the beginning and throughout the trial, both for the purpose of recognizing important secondary effects of the intervention and for identifying potential safety signals of the intervention.

Study populations

In a prevention trial, the optimal study population to demonstrate efficacy most efficiently would be at high risk for future OA, but free of full evidence satisfying an accepted and operational disease definition. However, the initial testing of an intervention on

a high-risk population is not without drawbacks, as this may limit generalizability, necessitating further testing on others with varying degrees of risk. Or, the efficacy of agents which might be effective in those of lower risk, but prove ineffective in the "high-risk" population, would remain undiscovered. Ultimately, the study population selection cannot be dictated and is dependent upon the definition of disease that is employed and overall goals of the trial.

A prevention trial study population can be selected to represent the three major domains of disease definition related to OA: (1) structural compromise, (2) pain and other symptoms, and (3) impaired function. Additionally, physiological/immunological/genetic locally or systemically measured biomarkers, such as synovial fluid aggrecan, serum C-reactive protein (CRP) or cartilage oligomeric matrix protein (COMP), urinary type II collagen telopeptides (uCTX-II), or combinations of biomarkers, might be incorporated to either define an at-risk population or to exclude individuals from selection into a prevention trial[18]. Further, population selection can be predicated on addressing each of these domains singularly or in combination[19]. For discussion of the current state of qualification of biomarkers for OA, the reader is referred to the article in this issue on Biomarkers.

Eligible study populations for trials to prevent structurally-defined OA

If the eligible population for a prevention trial is to be free of structurally-defined OA, one option for defining a "disease-free" population includes enrollment of persons with Kellgren–Lawrence (K–L) radiographic grades 0 or 1. Decision-making based on the selection of a population with a K–L score of K–L = 0 vs K–L = 1, which is designated as "doubtful OA" must acknowledge that there is an embedded probability that individuals with a K–L = 1 have early OA[20], or the underlying conditions leading to OA, but have not yet been identified definitively radiographically. This probability should be factored into estimating the sample size and developing data analytic strategies. Similar concepts apply if the study population lacks knee OA, defined as the absence of a definite osteophyte. Currently, there are very limited data organized to inform these design issues; this is the rationale for this report including a call to identify and organize data to support making evidence-based design choices.

An example of the type of data needed comes from the 15-year study of the natural history of knee OA development (the Michigan Bone Health and Metabolism Study) encompassing 660 women who were aged 24–44 at the 1992 study inception. The women were recruited from a population-based sample to increase the likelihood of generalizability of the findings. Radiographs, taken every 3 years, were scored by two radiologists using K–L definitions for OA knee severity.

The probability of moving from one K–L score to the same or a different K–L score in a 3-year period was estimated using Markov transition modeling. Estimating the probabilities of transitioning from a K–L score of 1 (proposed here as an example for a prevention trial) to other K–L scores reveals the impact of age and body mass index (BMI) and provides evidence to define inclusion and exclusion criteria in the prevention trial.

At age 50, the probability that a K–L = 1 score would remain at a K–L = 1 score 3 years later was 54–59% when BMIs ranged from 25 kg/m^2 to 35 kg/m^2 [Fig. 2(A)]. The probability of transitioning from K–L = 1 score to K–L = 2 score in a 3-year period ranged from 8% in non-obese women to 15% in women with a BMI $\geq$ 35 kg/m2 [Fig. 2(B)]. The probability of transitioning from a K–L = 1 score to a K–L = 3 score in a 3-year period is less than 2% (data not shown graphically). This evidence-based approach increases the likelihood

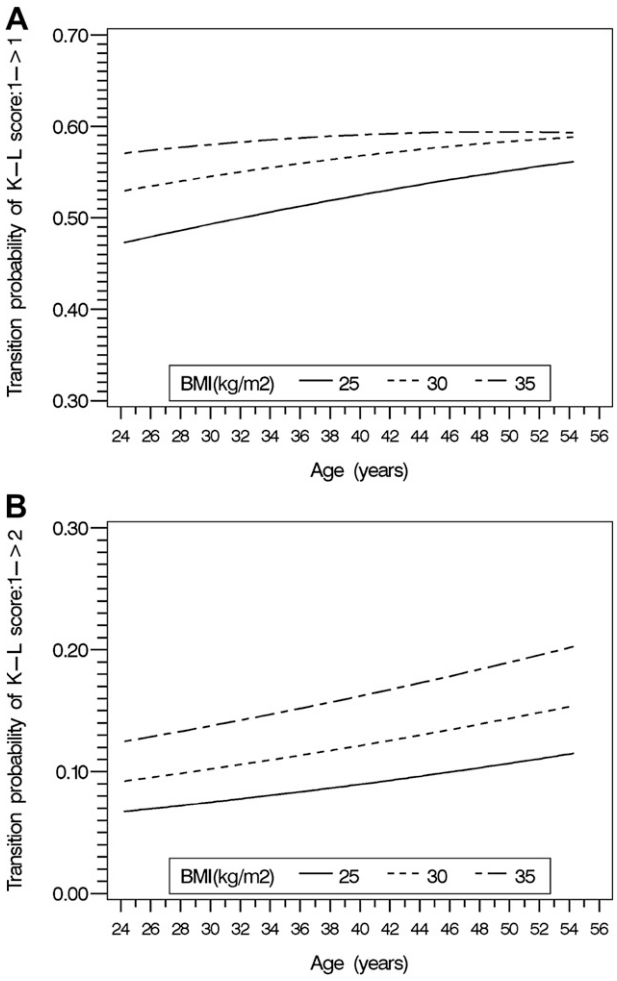

Fig. 2. Three-year transition probabilities of K–L score of 1 (doubtful OA) staying at a K–L = 1 or progressing to a K–L = 2 to OA as a function of age (years) and BMI (kg/m^2): designing a prevention trial of knee OA. (A) Transition probability of K–L = 1 score staying at K–L = 1; (B) Transition probability of K–L = 1 score to a K–L = 2 score, indicative of OA.

of having efficiently designed trials of prevention practices to forestall the development of knee OA.

Efforts are underway to define structural changes of knee OA by techniques other than the standing knee radiograph. For instance, static magnetic resonance imaging (MRI) to define OA based on morphologic changes in cartilage, bone or other soft tissues[21] or functional magnetic resonance imaging (fMRI) or other types of MRI measures (such as delayed gadolinium-enhanced MRI of cartilage [dGEMRIC], T2-mapping, T1rho, sodium imaging, etc) to define OA based on compositional changes in cartilage, bone or other soft tissues[22,23] may become modalities of choice. Currently, there is no agreed upon definition of OA based on these technologies. However, the field is rapidly evolving, i.e., the OARSI FDA Initiative Imaging Working Group is currently developing criteria for the early diagnosis of knee OA using MRI, and these developments must be anticipated in developing future trials (See article on Imaging in this issue).

Eligible study populations for trials to prevent symptoms of OA

If the eligible population lacks characteristic defining symptoms, especially pain or stiffness, the limits of allowable symptoms must be carefully defined, including how pain is to be assessed, its severity and duration, and the allowable frequency for transient pain, and potentially whether or not pain in joints apart from the target joints are

considered informative. The use of usual and rescue medications, such as analgesics or non-steroidal anti-inflammatory drugs (NSAIDs), also needs to be factored into the methodologic strategy to assess symptoms of OA[24]. Depending upon the mode of action of the agent under study, it may be necessary to disallow some usual or rescue medications to decipher the effect of the intervention unambiguously. For example, if the mechanism of action of the agent to be used effects pain relief through disruption of bone turnover, it might be necessary to exclude the use of drugs that affect bone turnover, such as bisphosphonates. If a drug is related to narcotics, it might be necessary to exclude the use of narcotics as rescue medication. Again, these issues, critical to the design of prevention trials, cannot be dictated and must be decided in the context of the trial under consideration.

Eligible study populations for trials to prevent functional decline from OA

If the eligible study population is to be free of functional performance impediments, investigators will need to determine whether inclusion criteria are based on self-report instruments or performance-based assessments. There are numerous questionnaire-based instruments to characterize functional status (See article on Functional status measures in this issue). For the selection of a study population, it is particularly important to choose an instrument or combination of instruments that have a known specificity (the known probability of truly being free of functional compromise), and that specificity should be relevant to the population from which the prevention trial population will be recruited. The use of performance-based assessment in prevention trial recruitment is limited by the relative absence of normative data in persons younger than age 65, thereby precluding the ability to estimate the probability of any specific assessment value's actually representing the disease-free state for a prevention trial. Further, there are many determinants of function which may or may not be directly relevant to OA. Alternatively, these measures may be considered to be estimates of an "at-risk" state and therefore eligible for study in a prevention trial; it is important that the predictive capacity of these performance measures over a period of time for increased compromise be known.

Use of biomarkers to define eligible study populations for prevention trials in OA

If the eligible study population will be selected based on physiological or immunological biomarker measures, there are at least two expectations. First, there must be adequate information to discern when a specific value of the biomarker(s) truly represents a "disease-free" state and, second, information about the rapidity of the biomarker change (if treated as a continuous variable) or conversion (if treated as a discrete variable) in relation to the development of disease, must be known and available. Additionally, the biomarker must have been previously validated against a clinically relevant endpoint for its use as a surrogate measure[25]. Even if the biomarker is used only as a criterion for inclusion or exclusion for participation in a prevention trial, it must have sufficient evidence of predictive relevance to warrant its application. Further discussion of this topic is found in the article on Biomarkers.

High-risk groups to target for prevention and risk reduction

For primary prevention and risk reduction, careful characterization of the relevant risk factor of interest is as critical as being able to define the absence of OA. It is important to be able to (1) define and measure the risk factor unambiguously, and (2) know the relative contribution of the risk factor to OA disease development, the average duration to disease manifestation among those with and without the risk factor, and the prevalence of the risk factor in the population. Clinical risk factors for OA may be joint-site specific, i.e., rupture of the anterior cruciate ligament (ACL) as a risk factor for the development of knee OA. Other risk factors may exert systemic effects on risk of OA in multiple joints. The latter situation includes factors such as age, female gender, overweight and obesity, endocrine disorders, and family history or genetically-defined population subgroups. Although not all of these are modifiable, they may influence participant selection criteria in certain trials.

As our measurement tools become increasingly sensitive and precise, it may be possible to classify the risk status of individuals and groups based on characteristics such as cartilage lesions on MRI, levels of biomarkers associated with OA development, or possession of a specific genotype.

Sample trial design for prevention of knee OA in overweight and obese

- The following is presented for illustrative purposes only, and should NOT be considered a prescriptive mandate for the design of a prevention trial. Further, as definitions of at risk populations change and measurements of the disease process and outcomes advance, it is expected that design features of such a trial would necessarily evolve as well. It is critical to enrich the probability of including individuals who may develop knee OA in a shorter and feasible time frame that acknowledges that clinical trials of long duration are not only costly, but are difficult to implement (i.e., to conduct an intervention without drift or maintain a study group compliant with the protocol, etc.). Including persons with an increased likelihood of developing disease will improve the ability to determine the intervention's effectiveness in preventing disease, but may limit generalizability.

Proposed study population

The study population for a primary or secondary prevention trial should be structured to the proposed intervention. A reasonable "high-risk" study population for a prevention trial could consist of ambulatory, community-dwelling men and women aged 50–65 years with: (1) no more than a "questionable" osteophyte (K–L = 1) in the medial or lateral tibiofemoral compartment (2) knee varus or valgus malalignment (angle $\geq 2^\circ$ and $\leq 10^\circ$); (3) BMI ≥ 30 kg/m^2 and ≤ 45 kg/m^2; (4) sedentary lifestyle, i.e., no participation within the past 6 months in an exercise program that incorporated more than 30 min/week of formal exercise; and, (5) the absence of interview-determined knee pain or limited function for a month-long time period. Scores on either questionnaire-based or performance-based functional assessment will reflect values considered in the "normal" range for men and women in the 50–65 year age range. A detailed record of medication use should be collected at baseline and for each specific follow-up testing interval.

Rationale for criteria

- Use of K–L = 1 rather than K–L = 0, 1 should increase the likelihood that individuals will develop OA[20].
- A 30 < BMI < 45 kg/m^2 is likely to include a population that is obese but able to participate in a designated intervention, and for which normative measures are interpretable. The range of BMI should be evaluated for population groups of shorter stature, such as first generation Asian enrollees to BMI of 25–35 kg/m^2.

- Potential study participants with a BMI > 45 kg/m^2 should be considered for exclusion because of the difficulty in using computerized tomography and MRI equipment to characterize hard and soft tissue structures. Additionally, in this group, there is a lower exercise compliance rate associated with high BMIs[26].
- Including only those with moderate malalignment (varus or valgus knee angle ≥ 2 and $\leq 10^{\circ}$) will potentially allow more rapid development of disease, because medial and lateral knee OA progression is strongly associated with moderate malalignment, and this may or may not be independent of body size[27]. However, this is not absolute, since data supporting the role of malalignment in the development of new knee OA is controversial[28,29].

Possible interventions

Interventions could be pharmacologic or non-pharmacologic. Importantly, an intervention in a primary or secondary prevention trial has unique elements in that (1) implementation is likely, but not definitely, to require a protracted administration period; (2) the administration cannot generate risk of accentuating other potential on-going disease processes; and (3) a careful weighing of the costs and benefits must occur. For example, bariatric surgery might be considered a candidate intervention for a primary prevention trial for OA, but its use imposes unique consideration for other heath costs and risks of morbidity and mortality. One way to address this issue could be to append an ancillary study for prevention of knee OA to an on-going trial of bariatric surgery for other outcomes. An active drug (unknown at this point) could be directed toward decreasing inflammation and/or pain or improving weight management. A functional intervention might include measures to modify alignment and/or build strength. A behavioral intervention could be directed toward increasing physical activity, changing the type of physical activity, or modifying dietary practices.

It is also possible that a preventive intervention might not have to be administered over prolonged periods of time. Such a situation might obtain in the setting of acute joint injury, in which hypothetical Agent X might be injected into the injured joint weekly for 4 weeks. Assessment of such a regimen could improve the feasibility and tolerability of delivering the intervention itself, but would not eliminate the need for prolonged assessments to ascertain whether the agent inhibited the onset of OA and whether it is safe.

Primary outcomes

If the trial hypothesis is that an intervention in a prevention trial among obese adults with no or doubtful evidence of radiographic knee OA (K–L $= 0, 1$) will be associated with a delayed onset of knee OA compared to the *placebo* group, this delay could be reflected in two co-primary outcomes: less symptom report and minimal structural change in relation to the untreated group. Candidate measures to detect these areas include changes in: (1) K–L score or minimal joint space and (2) questionnaire-based pain assessment. Other potentially relevant outcome measures could include newer technologies once they have been validated, such as MRI with or without T2-mapping to assess morphological changes in joint structures or articular cartilage degradation and/or bone marrow lesions. As imaging and molecular techniques advance to the stage where they could be surrogates of downstream clinical outcomes, it may be that an intervention might be able to show a primary effect on structure of the OA process, regardless of its immediate effect on symptoms. While examples of prevention in other medical conditions abound, e.g., interventions directed toward lowering serum cholesterol or altering lipid profiles to prevent future cardiovascular events[30,31], or altering bone mineral density to prevent osteoporotic fractures[32], it is unlikely that requirements for a proposed intervention to affect relevant clinical outcomes would be waived entirely.

Secondary outcomes

Secondary outcomes could include some or all of the following largely predicated on the nature of the proposed intervention: (1) clinical measures of function, pain and mobility; (2) mechanistic measures of the OA disease pathways such as knee alignment, knee external adductor moment, knee joint compressive and shear forces, and; (3) biomarker measures of pro-inflammatory molecules (e.g., interleukin-6 (IL-6), Tumor Necrosis Factor-α, CRP) and joint metabolism (e.g., uCTX-II, COMP); (4) lower extremity strength and power; (5) limb proprioception; and (6) abdominal and thigh fat depots measured by CT; (7) adverse effects associated with the intervention; and (8) quality of life.

In addition to OA outcome measures, investigators need to select or develop appropriate measures of intervention-related processes and adherence to the intervention.

Study time line

Depending upon the factors discussed above, a primary prevention trial is likely to require a 10-year follow-up with data collected from participants at 1 or 2-year intervals. The interval distance should be based on time required to detect meaningful differences in the measures of interest and motivate subjects to maintain optimal participation in the trial. For example, MRI, knee X-ray, gait, and strength might be measured biannually (years 2, 4, 6, 8, and 10), while biomarker levels might be assessed every 3–6 months. Proposed trials of shorter duration, with proper justification of clinically relevant outcomes and safety monitoring, would likely improve feasibility.

Sample trial design for prevention of knee OA by preventing knee injury

Many of the issues above also apply to trials of injury prevention, but the latter have a number of unique, key design features worthy of separate discussion and illustrated by a trial of an educational/exercise intervention vs an attention control to prevent knee injury in high school female basketball players. This is an example of a risk reduction trial to prevent injury that might later lead to knee OA.

Selecting a sample/population

Injury prevention trials must identify populations at considerable risk of the relevant injury. Low risk populations are inefficient to study because event rates are minimal, requiring very large samples or longer trial duration, which may lead to contamination across study arms and considerable attrition of study participants. Sports teams, military trainees and other such groups exposed to high levels of demanding physical activity are appropriate at-risk populations.

Unit of randomization

Such prevention studies often require cluster-randomized designs, in which the unit of randomization is the group, not the individual subject. These may include sports teams, schools, sports leagues, or even towns. For the trial of female basketball players, the cluster group is the high school team. The rationale for randomizing at the group level is to reduce contamination, or diffusion of the intervention to the control group. For example, if two basketball

players on the same team are randomized to separate arms, the player randomized to receive exercises may show the exercises to the player in the control group. Cluster randomization reduces this risk. Furthermore, cluster randomization permits the group to be incorporated into the intervention. When an entire school is randomized to an educational intervention arm, the investigators can display injury prevention educational posters in the school and not worry about contamination. Cluster-randomized designs are typically costly in terms of sample size because each observation is not independent. The more similar the outcomes are among members of the group, and the larger the cluster group, the greater the sample size needed to overcome the non-independence.

Intervention protocol

Interventions in injury prevention trials must be delivered in a standardized fashion at all intervention sites. This requires training, reliability assessment, site visits, and logistical work to ensure that the intervention is administered similarly across diverse settings. In this example, a basketball injury prevention program allocated at the school level needs to be delivered identically despite differences in the gyms, practice schedules and coaches' styles in different schools.

Outcome assessment

Assessments must be done in a standardized fashion at regular intervals using well-defined, reproducible outcome definitions. In many trials the outcome is injury, but investigators must clarify what constitutes an injury (a sore knee for an hour? a day? with swelling, defined by whom? had to leave practice or game? had to miss next game? radiographic or imaging findings, e.g., ligamentous injury, meniscal injury, fracture?). This assessment should ideally be made by an observer blinded to random intervention assignment.

Statistical analysis

The analysis of cluster-randomized intervention trials must use techniques (such as generalized estimating equations) that account for clustered observations[33] or risk artificially lowering variance estimates and over-stating the statistical significance.

Methodological considerations for prevention trials

Study design

As these examples illustrate, the double blinded, randomized, *placebo* or active comparator study design is the gold standard, but its appropriateness is dependent upon the agent and the availability of known effective interventions for primary preventions. Many likely interventions may be difficult or impossible to blind completely. Further, many potential primary prevention interventions, such as the injury prevention trial, may be more effectively delivered using cluster randomization, where the community is the unit of analysis, rather than the individual[33]. In this case, contamination related to community behavioral or other change can influence results and must be rigorously addressed[34]. Therefore, selection of study design for the trial will be dependent upon intervention, the degree to which individual implementation is feasible, and the capacity to include an effective *placebo*.

Adherence

It is a particular problem for long-term interventions, particularly if participants do not readily perceive benefit from continued participation or experience other barriers. Both the active intervention and *placebo* groups will require supplemental behavioral components to maintain adherence, and the inclusion of an adherence specialist on the study team may be wise.

There are also organic factors that may influence adherence. It may be appropriate to assess for depression symptoms and design interventions and intervention monitoring to address their impact in terms of both individual behaviors as well as interactions of depression therapy with the intervention for OA. Female enrollees are likely to be in the midst of the menopause transition and the degree of symptoms and stage of the transition are likely to influence both behaviors and potentially structural tissue responses. This suggests that adherence management needs to be prepared to deal with concomitant symptomatic conditions and potential interventions associated with those symptoms. The proposed age range is likely to reflect other competing illness processes that may affect adherence, as well as directly impact intervention effectiveness and potential for side-effects.

Recruitment

It is the life-blood of any clinical trial; however, recruiting for a primary prevention trial imposes requirements that are not always evident in treatment trials. Recruitment could be enhanced by using complementary strategies coupled with a system that provides feedback on each strategy's effectiveness and cost[35]. Mass mailings and media (newspaper, television, internet) may be effective in some settings. Depending upon the age of the primary prevention target population, having strong ties with local aging service networks and access to senior centers, churches, drug stores, shopping malls, and other sites where older adults gather could be important but may be ineffective for the population 50–65 years. Most health science centers maintain a large database of adults who have signed consent to be contacted about participating in future clinical trials; however, it is important to identify why these adults are associated with such registries and if their registration is associated with diseases that may impinge on the intervention or decrease the likelihood that they are going to be free of OA.

Experience has proven that *on-going monitoring of the recruitment process is necessary* to achieve study goals and to review recruitment activities, plan new activities, and monitor the number of contacts[24]. Close attention should be given to the gender and minority frequencies of those who qualify for, and enroll in, the study.

Safety database for trials of prevention of OA

Because a prevention trial for OA could involve an intervention with active agents administered to otherwise healthy individuals, or to individuals with co-morbid conditions, for extended periods of time, the safety database must be extensive and involve information from multiple organ systems. The extent of this safety database may depend upon the intervention. For example, systemically-administered interventions may have pleiotropic effects, e.g., statins or bisphosphonates[36–39], reinforcing the need to monitor multiple organ systems for toxicity. A more localized intervention, such as an unloading brace, might not require the same degree of vigilance for safety in remote organ systems. Observations must also be long in duration, particularly for agents that might impact the immune system and be associated with infections or subsequent development of cancer. Finally, when trials are of considerable duration, such as in these cases of OA prevention, careful monitoring of evolving technology that might impact the long-term assessment of outcome must also occur. See article on Safety as part of this issue.

Ethical issues for prevention trials

As recently reviewed[40], rheumatology clinical trials may involve some issues that pose specific ethical concerns. This may particularly be the case in prevention trials for OA. First, since currently no clearly effective agents exist, novel agents to be used in primary prevention must first include substantial testing on healthy volunteers or people with early disease to establish viability. Prevention trials necessarily involve people who may not have the disease in question, who may not ever get the disease, or who might experience a relatively benign course even with no intervention. Further, for a condition such as OA, which develops over years, any effective agent for prevention would likely need to be administered for a prolonged time, possibly beginning at an early age. The potential for multi-system toxicity must be monitored, especially in younger individuals who may be of reproductive age when the agent is started.

Some preventive interventions may be directed at the population level, rather than provide benefit to a specific individual. An example of this would be a vaccine study. In this instance, studying a treatment in a person with disease can be profoundly different than studying an intervention in healthy people. Studies in other diseases have shown that study participants may have misconceptions about the potential effectiveness of a preventive intervention and/or may have inflated estimates of the likelihood that they will be randomized to get the active agent, and may have exaggerated impressions of the likelihood that the intervention will be personally effective for them. Simon and colleagues have called this the "preventive misconception," defined as "the overestimate in probability or level of personal protection that is afforded by being enrolled in a trial of a preventive intervention"[41]. This can be particularly problematic when accompanied by "behavioral disinhibition" or the adoption of behaviors that may pose a risk to the participant or others. This has been observed in persons participating in a HIV vaccine trial, in which individuals had an increase in risky behaviors[41]. In the case of OA, various scenarios could be imagined, in which behavioral disinhibition could occur. One could imagine that someone with a strong family history of OA, or even someone who possessed a very high-risk genotype, might be less vigilant about maintaining a normal weight because of a false expectation that the preventive agent he/she received in a trial will be effective and protect him/her from his/her increased risk of OA. These issues emphasize the critical importance of the informed consent process in OA trials, particularly those for prevention.

Recommendations for future research

First and foremost is the requirement that research continue to work to refine definitions of OA, utilizing genetic, biochemical, and imaging biomarkers and psychometrically valid questionnaires and performance measures, with the goal of diminishing ambiguity in the currently used metrics and increasing their clinical relevance. Collection of extensive biological specimens, e.g., serum, plasma, DNA, RNA, urine, should be a part of all of these on-going and future studies.

Observational studies with both short and long-term follow-up can be particularly helpful in this regard, to define molecular, structural and symptomatic correlates of disease and to identify risk factors predictive of the development of disease and its clinical impact. Attention to gender and minority inclusion, with the requisite consideration of distinct issues regarding their propensity to participate in prevention trials, should be a part of this research agenda. Observational studies can be particularly helpful in the following activities:

- Evaluation of existing datasets with particularly long follow-up times (10, 20 or more years) in order to identify risk factors that may be exposed long in advance of disease onset.
- Extended follow-up as adults of cohorts established during childhood, adolescence, and early adulthood, for the development of OA.
- Extended, detailed follow-up of inception cohorts of those with acute joint injury, with detailed information regarding the events and treatment modalities applied in the acute setting, as well as other potential risk factors.
- Evaluation of existing datasets with detailed genetic, biomarker, and imaging data to expand our information about various OA phenotypes along the continuum from molecular to pre-radiographic OA, to radiographic to symptomatic OA.
- Addition of short follow-up times (i.e., months), to studies of existing cohorts to obtain sensitive, dynamic imaging and other biomarker data to aid prediction of the development of structural and clinical disease.
- Evaluation of distinct ethnic/racial sub-populations to ascertain accurate assessment of the burden of disease in these groups, differences in risk factor profiles, and genetic, imaging, and biomarker sub-types in order to tailor trials to relevant groups, (i.e., differences in BMI that might be used to screen Asians or African Americans into prevention trials for the overweight/obese).
- Methodological studies of distinct threats to validity of prevention trials and their execution, related to cultural differences in attitudes toward trial participation and risk factor reduction; techniques to maximize adherence and retention; and ways to measure and overcome biases such as preventive misconception and behavioral disinhibition. As one example, the use of technology, such as hand-held devices and the Internet, for participant recruitment, retention and data collection, is becoming more widespread and will continue to evolve. The study of the impact of such methods upon prevention trials in general will likely inform future prevention trials for OA.

An additional future direction may be a multi-center clinical trial of a non-pharmacologic intervention, alone and in combination with a pharmacologic co-therapy, that can alter mechanisms in the pathological pathway (e.g., decrease knee joint loading and reduce inflammation) to OA and thus lower its incidence. The 2009 NIAMS Roundtable presented a roadmap for how prevention trials should be organized. For large multi-center trials, NIAMS will identify the most qualified investigators who will be required to first establish the need for a larger trial with results from a planning grant or similar study. This will allow applicants to demonstrate their abilities to design and manage clinical trials before launching a full-scale project. The large-scale project should be comprehensive, incorporating clinical (e.g., pain, function), mechanistic (e.g., inflammation and knee joint loading), and structural (e.g., quantitative cartilage morphology with qMRI, semi-quantitative whole joint scoring) outcomes. Demonstrating the ability to identify and target people who are at high-risk of OA will be crucial as this will lay the foundation for primary prevention efforts.

Secondary prevention is equally important. Knee trauma, such as ACL or meniscus injury, is a strong predictor of subsequent knee OA. Considering the young age at which many of these injuries occur, knee joint replacement at a relatively young age is a distinct possibility, possibly followed by a second replacement after 10–15 years. A secondary prevention trial with outcomes related to the risk of knee replacement would have important public health implications. Knee and hip strengthening in young adults with knee trauma to reduce the risk of knee replacement would be an example of a secondary

prevention trial. A synopsis of the NIAMS roundtable can be found at the location listed below.

- (http://www.niams.nih.gov/news_and_events/Meetings_and_Events/Roundtables/2009/ortho_OA.asp).

Declaration of funding and role of funding source

The OARSI FDA OA Initiative received financial support from the following professional organization:

American College of Rheumatology.

Additionally the OARSI FDA OA Initiative received financial support from the following companies: Amgen; ArthroLab; AstraZeneca; Bayer Healthcare; Chondrometrics; CombinatoRx; Cypress BioScience; DePuy Mitek; Expanscience; 4QImaging; Genevrier/IBSA; Genzyme; King (Alpharma); Merck; Merck Serono; NicOx; Pfizer; Rottapharm; Smith & Nephew; Wyeth.

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the Food and Drug Administration.

Author contributions

JM Jordan: Chair, Prevention and Risk Reduction Working Group; conception and design; drafting of the article; critical revision of the article for important intellectual content; final approval of the article.

MF Sowers: conception and design; drafting of the article; critical revision of the article for important intellectual content; final approval of the article.

SP Messier: conception and design; drafting of the article; critical revision of the article for important intellectual content; final approval of the article.

J Bradley: conception and design; critical revision of the article for important intellectual content; final approval of the article.

G Arangio: conception and design; drafting of the article; critical revision of the article for important intellectual content; final approval of the article.

JN Katz: drafting of the article; critical revision of the article for important intellectual content; final approval of the article.

E Losina: drafting of the article; critical revision of the article for important intellectual content; final approval of the article.

L Rovati: conception and design; critical revision of the article for important intellectual content; final approval of the article.

N Bachtell: conception and design; critical revision of the article for important intellectual content; final approval of the article.

C Cooper: critical revision of the article for important intellectual content; final approval of the article.

T Spector: critical revision of the article for important intellectual content; final approval of the article.

W Zhang: critical revision of the article for important intellectual content; final approval of the article.

J Gardiner: critical revision of the article for important intellectual content; final approval of the article.

M Wahba: critical revision of the article for important intellectual content; final approval of the article.

Conflict of interest

MFS, TS, GA, JNK, EL, and WZ: Authors have nothing to disclose.

JJ: received research support or compensation from Cerimon Pharmaceuticals, Endo Pharmaceuticals, GlaxoSmithKline, Novartis, Eli Lilly, Interleukin Genetics, Wyeth.

SPM: received research support or compensation from General Nutrition Company.

CC: received research support or compensation from Alliance for Better Bone Health, MSD, Eli Lilly, Servier Laboratories, Roche Pharmaceuticals, GSK, Amgen, Novartis.

JG: previous employee of Wyeth.

LR: employee of Rottapharm SpA.

NB: employee of Genzyme Corp.

Acknowledgment

Valorie J. Thompson of Innovations Consulting Group, LLC for administrative and logistic support.

References

1. Lawrence RC, Helmick CG, Arnett FC, Deyo RA, Felson DT, Giannini EH, *et al*. Estimates of the prevalence of arthritis and selected musculoskeletal disorders in the United States. Arthritis Rheum 1998;41(5):778–99.
2. Dillon CF, Rasch EK, Gu Q, Hirsch R. Prevalence of knee osteoarthritis in the United States: arthritis data from the Third National Health and Nutrition Examination Survey 1991–94. J Rheumatol 2006;33(11):2271–9.
3. The Burden of Musculoskeletal Diseases in the United States: Prevalence, Societal and Economic Cost. Rosemont, IL: Bone and Joint Decade; 2008. p. 79.
4. Kraus VB, Jordan JM, Doherty M, Wilson AG, Moskowitz RW, Hochberg MC, *et al*. The Genetics of Generalized Osteoarthritis (GOGO) study: study design and evaluation of osteoarthritis phenotypes. Osteoarthritis & Cartilage 2007;15(2):120–7.
5. Cimmino MA, Sarzi-Puttini P, Scarpa R, Caporali R, Parazzini F, Zaninelli A, *et al*. Clinical presentation of osteoarthritis in general practice: determinants of pain in Italian patients in the AMICA study. Semin Arthritis Rheum 2005;35(Suppl 1):17–23.
6. Englund M, Paradowski PT, Lohmander LS. Association of radiographic hand osteoarthritis with radiographic knee osteoarthritis after meniscectomy. Arthritis Rheum 2004;50(2):469–75.
7. Sawitzke AD, Shi H, Finco MF, Dunlop DD, Bingham III CO, Harris CL, *et al*. The effect of glucosamine and/or chondroitin sulfate on the progression of knee osteoarthritis: a report from the glucosamine/chondroitin arthritis intervention trial. Arthritis Rheum 2008;58(10):3183–91.
8. Messier SP, Loeser RF, Miller GD, Morgan TM, Rejeski WJ, Sevick MA, *et al*. Exercise and dietary weight loss in overweight and obese older adults with knee osteoarthritis: the arthritis, diet, and activity promotion trial. Arthritis Rheum 2004;50(5):1501–10.
9. Messier SP, Legault C, Mihalko S, Miller GD, Loeser RF, DeVita P, *et al*. The Intensive Diet and Exercise for Arthritis (IDEA) trial: design and rationale. BMC Musculoskelet Disord 2009;10:93. PMCID: PMC2729726.
10. Vitiello MV, Rybarczyk B, Von KM, Stepanski EJ. Cognitive behavioral therapy for insomnia improves sleep and decreases pain in older adults with co-morbid insomnia and osteoarthritis. J Clin Sleep Med 2009;5(4):355–62. PMCID: PMC2725255.
11. Rannou F, Poiraudeau S, Beaudreuil J. Role of bracing in the management of knee osteoarthritis. Curr Opin Rheumatol 2010;22(2):218–22.

12. Wang C, Schmid CH, Hibberd PL, Kalish R, Roubenoff R, Rones R, *et al*. Tai Chi is effective in treating knee osteoarthritis: a randomized controlled trial. Arthritis Rheum 2009;61(11):1545–53.
13. Reginster JY, Deroisy R, Rovati LC, Lee RL, Lejeune E, Bruyere O, *et al*. Long-term effects of glucosamine sulphate on osteoarthritis progression: a randomised, placebo-controlled clinical trial. Lancet 2001;357(9252):251–6.
14. Pavelka K, Gatterova J, Olejarova M, Machacek S, Giacovelli G, Rovati LC. Glucosamine sulfate use and delay of progression of knee osteoarthritis: a 3-year, randomized, placebo-controlled, double-blind study. Arch Intern Med 2002;162(18):2113–23.
15. Brandt KD, Mazzuca SA, Conrozier T, Dacre JE, Peterfy CG, Provvedini D, *et al*. Which is the best radiographic protocol for a clinical trial of a structure modifying drug in patients with knee osteoarthritis? J Rheumatol 2002;29(6):1308–20.
16. Brandt KD, Mazzuca SA, Katz BP, Lane KA, Buckwalter KA, Yocum DE, *et al*. Effects of doxycycline on progression of osteoarthritis: results of a randomized, placebo-controlled, double-blind trial. Arthritis Rheum 2005;52(7):2015–25.
17. Bingham III CO, Buckland-Wright JC, Garnero P, Cohen SB, Dougados M, Adami S, *et al*. Risedronate decreases biochemical markers of cartilage degradation but does not decrease symptoms or slow radiographic progression in patients with medial compartment osteoarthritis of the knee: results of the two-year multinational knee osteoarthritis structural arthritis study. Arthritis Rheum 2006;54(11):3494–507.
18. Garnero P, Aronstein WS, Cohen SB, Conaghan PG, Cline GA, Christiansen C, *et al*. Relationships between biochemical markers of bone and cartilage degradation with radiological progression in patients with knee osteoarthritis receiving risedronate: the Knee Osteoarthritis Structural Arthritis randomized clinical trial. Osteoarthritis Cartilage 2008;16(6):660–6.
19. Dam EB, Loog M, Christiansen C, Byrjalsen I, Folkesson J, Nielsen M, *et al*. Identification of progressors in osteoarthritis by combining biochemical and MRI-based markers. Arthritis Res Ther 2009;11(4): R115.
20. Hart DJ, Spector TD. Kellgren & Lawrence grade 1 osteophytes in the knee – doubtful or definite? Osteoarthritis Cartilage 2003;11(2):149–50.
21. Davies-Tuck M, Wluka AE, Forbes A, Wang Y, English DR, Giles GG, *et al*. Development of bone marrow lesions is associated with adverse effects on knee cartilage while resolution is associated with improvement – a potential target for prevention of knee osteoarthritis: a longitudinal study. Arthritis Res Ther 2010;12(1): R10.
22. Yao W, Qu N, Lu Z, Yang S. The application of T1 and T2 relaxation time and magnetization transfer ratios to the early diagnosis of patellar cartilage osteoarthritis. Skeletal Radiol 2009;38(11):1055–62.
23. Taylor C, Carballido-Gamio J, Majumdar S, Li X. Comparison of quantitative imaging of cartilage for osteoarthritis: T2, T1rho, dGEMRIC and contrast-enhanced computed tomography. Magn Reson Imaging 2009;27(6):779–84. PMCID: PMC2722506.
24. Brandt KD, Mazzuca SA. Lessons learned from nine clinical trials of disease-modifying osteoarthritis drugs. Arthritis Rheum 2005;52(11):3349–59.
25. Bauer DC, Hunter DJ, Abramson SB, Attur M, Corr M, Felson D, *et al*. Classification of osteoarthritis biomarkers: a proposed approach. Osteoarthritis Cartilage 2006;14(8):723–7.
26. van Gool CH, Penninx BW, Kempen GI, Rejeski WJ, Miller GD, Van Eijk JT, *et al*. Effects of exercise adherence on physical function among overweight older adults with knee osteoarthritis. Arthritis Rheum 2005;53(1):24–32.
27. Felson DT, Nevitt MC, Yang M, Clancy M, Niu J, Torner JC, *et al*. A new approach yields high rates of radiographic progression in knee osteoarthritis. J Rheumatol 2008;35(10):2047–54. PMCID: PMC2758234.
28. Hunter DJ, Niu J, Felson DT, Harvey WF, Gross KD, McCree P, *et al*. Knee alignment does not predict incident osteoarthritis: the Framingham osteoarthritis study. Arthritis Rheum 2007;56(4):1212–8.
29. Tanamas S, Hanna FS, Cicuttini FM, Wluka AE, Berry P, Urquhart DM. Does knee malalignment increase the risk of development and progression of knee osteoarthritis? A systematic review. Arthritis Rheum 2009;61(4):459–67.
30. Amarenco P, Labreuche J. Lipid management in the prevention of stroke: review and updated meta-analysis of statins for stroke prevention. Lancet Neurol 2009;8(5):453–63.
31. Narla V, Blaha MJ, Blumenthal RS, Michos ED. The JUPITER and AURORA clinical trials for rosuvastatin in special primary prevention populations: perspectives, outcomes, and consequences. Vasc Health Risk Manag 2009;5:1033–42. PMCID: PMC2801627.
32. Ensrud KE, Lui LY, Taylor BC, Schousboe JT, Donaldson MG, Fink HA, *et al*. A comparison of prediction models for fractures in older women: is more better? Arch Intern Med 2009;169(22):2087–94.
33. Green SB. The advantages of community-randomized trials for evaluating lifestyle modification. Control Clin Trials 1997;18(6):506–13.
34. Kouyate B, Some F, Jahn A, Coulibaly B, Eriksen J, Sauerborn R, *et al*. Process and effects of a community intervention on malaria in rural Burkina Faso: randomized controlled trial. Malar J 2008;7:50. PMCID: PMC2287184.
35. Brandt KD, Mazzuca SA. Experience with a placebo-controlled randomized clinical trial of a disease-modifying drug for osteoarthritis: the doxycycline trial. Rheum Dis Clin North Am 2006;32(1):217. xii.
36. Athyros VG, Kakafika AI, Tziomalos K, Karagiannis A, Mikhailidis DP. Pleiotropic effects of statins–clinical evidence. Curr Pharm Des 2009;15(5):479–89.
37. McGuinness B, Craig D, Bullock R, Passmore P. Statins for the prevention of dementia. Cochrane Database Syst Rev 2009 Apr 15;(2). CD003160.
38. Garrett IR, Gutierrez G, Mundy GR. Statins and bone formation. Curr Pharm Des 2001;7(8):715–36.
39. Saag KG. Bisphosphonates for osteoarthritis prevention: “Holy Grail” or not? Ann Rheum Dis 2008;67(10):1358–9.
40. Sugarman J, Bingham III CO. Ethical issues in rheumatology clinical trials. Nat Clin Pract Rheumatol 2008;4(7):356–63.
41. Simon AE, Wu AW, Lavori PW, Sugarman J. Preventive misconception: its nature, presence, and ethical implications for research. Am J Prev Med 2007;32(5):370–4.

Osteoarthritis and Cartilage 19 (2011) 509–514

Osteoarthritis and Cartilage

Recommendations of the OARSI FDA Osteoarthritis Devices Working Group

V.M. Goldberg †*, J. Buckwalter ‡, M. Halpin §, W. Jiranek ||, W. Mihalko ¶, M. Pinzur #, B. Rohan ††, T. Vail ‡‡, P. Walker §§, R. Windsor ||||, T. Wright ¶¶

† Department of Orthopaedics, Case Western Reserve University School of Medicine, Cleveland, OH, USA
‡ Department of Orthopaedics and Rehabilitation, University of Iowa, Iowa City, IA, USA
§ Genzyme Corporation, Cambridge, Massachusetts, USA
|| Department of Orthopaedic Surgery, Virginia Commonwealth University Medical Center, Richmond, VA, USA
¶ Department of Orthopaedic Surgery, Campbell Clinic, University of Tennessee Health Science Center, Memphis, TN, USA
# Loyola University Medical Center, Maywood, IL, USA
†† Smith & Nephew Inc., Memphis, TN, USA
‡‡ Department of Orthopaedic Surgery, University of California, San Francisco, CA, USA
§§ New York University Hospital, Hospital for Joint Diseases, New York, NY, USA
|||| Weill Medical College of Cornell University, Hospital for Special Surgery, New York, NY, USA
¶¶ Hospital for Special Surgery, Department of Orthopaedics, Weill Medical College of Cornell University, New York, NY, USA

ARTICLE INFO

Article history:
Received 22 December 2010
Accepted 11 February 2011

Keywords:
Osteoarthritis
Devices
Arthroplasty
Hyaluronans
Biological devices

SUMMARY

Osteoarthritis (OA) is the most common type of arthritis and a major cause of chronic musculoskeletal pain and functional disability. While both pharmacologic and non-pharmacologic modalities are recommended in the management of OA, when patients with hip or knee OA do not obtain adequate pain relief and/or functional improvement, joint replacement surgery or other surgical interventions should be considered. Total joint arthroplasties are reliable and cost-effective treatments for patients with significant OA of the hip and knee. Evidence from cohort and observational studies has confirmed substantial improvements in pain relief with cumulative revision rates at 10 years following total hip (THA) and total knee arthroplasties (TKA) at 7% and 10%, respectively. Joint replacements have been used in most every synovial joint, although results for joints other than hip and knee replacement have not been as successful. The evolution of new device designs and surgical techniques highlights the need to better understand the risk to benefit ratio for different joint replacements and to identify the appropriate methodology for evaluating the efficacy and optimal outcomes of these new devices, designed to treat OA joints.

Introduction

1n 1999, the Food and Drug Administration (FDA) drafted a guidance document for industry on clinical development programs for drugs, devices, and biological products intended for the treatment of Osteoarthritis (OA)[1]. Since then much progress has been made in the development of devices and drugs for the treatment of OA leading the FDA to request additional information to assist in their ongoing work to finalize the draft guidance.

Beginning in 2007, the Osteoarthritis Research Society International (OARSI) convened a number of working groups represented by leading researchers, academicians and clinicians to lead a critical appraisal of the scientific advances made over the past decade related to OA. The Working Group on Devices considered a number of key issues, including the appropriate study design and outcome measurements that should be considered in clinical development programs for new devices designed to treat OA joints.

The Medical Device Amendments of 1976 defined devices as an entity intended for diagnosis, cure mitigation or prevention of a disease or condition or an entity intended to affect the function or structure of the body that does not achieve its primary intended use through chemical action or metabolism.

Overview

Total hip (THA) and total knee arthroplasty (TKA) have both shown effectiveness in relieving arthritis associated pain, improving physical function, and enhancing health-related quality of life[2,3]. Favorable outcomes following THA and TKA are well

* Address correspondence and reprint requests to: V.M. Goldberg, University Hospitals Case Medical Center, Department of Orthopaedics, 11100 Euclid Avenue, Cleveland, OH 44106, USA.
E-mail address: victor.goldberg@UHhospitals.org (V.M. Goldberg).

1063-4584/$ – see front matter © 2011 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.joca.2011.02.017

established within the literature based not only on the effectiveness of the surgical and technical aspects of arthroplasty, but also on patient centered outcomes, including patient satisfaction[4].

Evidence from cohort and observational studies has confirmed substantial improvements and durability of the devices with cumulative revision rates at 10 years following THA and TKA at 7% and 10%, respectively[5]. Joint replacements have been used in almost every synovial joint; however, results for other joints have not been as successful as those observed after hip and knee replacement.

Newer device designs and surgical techniques are rapidly evolving. For example, spinal OA presents a significant healthcare problem in the USA resulting in severe disability and enormous societal costs. New products for spinal pathologies are being developed, including cervical and lumbar disc replacement, lumbar dynamic internal fixation stabilization, facet replacement, and interspinous distraction devices. To date, these new products have met with variable outcomes in clinical use[6,7].

Hyaluronans (HA) represent another area of research. While HA have been used intra-articularly as an approach to the treatment of the pain associated with knee OA, recent preclinical studies suggested that HA may also have a disease modifying effect on articular cartilage[8]. However, no clinical studies have confirmed this potential mechanism of action[9–11].

Implantable biological devices, such as cell-based treatments for repair of articular cartilage, have also recently been introduced with variable results[12].

FDA: regulatory pathway for device approval

Many of the devices recently introduced to the marketplace followed the FDA 510(k) guidance. A 510(k) is a premarket submission made to the FDA demonstrating that the device to be marketed is as safe and effective (i.e., substantially equivalent) and is similar to a currently marketed device (a predicate device). Much of the necessary information required for approval of the proposed device is based on the previous assessment of the predicate device. Companies must compare their proposed device to one or more similar marketed devices and provide support of their "substantial equivalency" claim. Once a new device is determined by the FDA to be substantially equivalent, it can be marketed in the USA. The parameters by which a device is considered to be equivalent include:

- Having the same intended use as the predicate device; and
  Having the same technological characteristics as the predicate; or
- Having the same intended use as the predicate; and
  Having different technological characteristics and the information submitted to the FDA:
  - Does not raise new questions of safety and effectiveness; and
  - Demonstrates that the device is at least as safe and effect as the currently marketed device.

If a device is not considered substantially equivalent, a company may be required to submit a premarket approval application (PMA). The PMA requires far more scientific and regulatory documentation to the FDA demonstrating the safety and effectiveness of the device than is necessary for a 510(k). Occasionally, the FDA will require a short-term clinical trial as part of a 510(k) clearance; however, randomized controlled trials, registries, and retrospective reviews performed to further assess the efficacy of the device are typically required post approval[13–15].

In 1999, Dr Henrik Malchau presented a conceptual approach to device approval that spans both the 510(k) and PMA process, without the intention of fitting the framework into the current regulatory process. This approach included a phased innovation process of preclinical study followed by rigorous quantitative metrics to assess the true effectiveness of the device[16,17]. Preclinical metrics would be established through existing standards produced by the American Society for Testing and Materials (ASTM) or the International Organization for Standardization (ISO), as well as other guidelines developed specifically for each device. Examples of rigorous clinical metrics that he suggested would include validated physician and patient directed clinical assessment tools and quantitative measures to assess the functional abilities of the patient. These might include kinematic studies, unique quantitative measures such as Radiostereometric Analysis (RSA) methodologies, or simple tools such as muscle testing and the sit-to-stand and 6-min walk tests.

Regardless of the care with which safety and effectiveness of a new device are determined through preclinical studies or pre-approval clinical trials, the ultimate assessment of its performance emerges through post-market surveillance. The Swedish Hip and Knee Registries, for example, have provided data for the successful assessment of the survival of devices used by a broad spectrum of surgeons in sufficient numbers of patients leading to more cost-effective approaches for the use of devices[18–20]. A limitation of many registries, however, is that they use revision surgery as the primary and often only outcome measure, with little or no information on patient-reported outcomes. Nonetheless, the use of registries provides a valuable tool in establishing performance and could be adapted to the study of unique implantable biological devices and products that combine biologics, drugs, and devices.

In anticipation of the increased demand for total joint arthroplasties and the continued advancements in the development of other devices, biologics, and new surgical techniques for the treatment of OA, it is important to consider how the efficacy, benefit to risk ratio, and clinical outcomes of these new products will be assessed.

Measuring the efficacy of devices

Devices, especially orthopaedic devices, do not fit into the same definition of efficacy defined by the FDA for pharmacological treatments. The time course for showing efficacy in a device such as a total joint replacement is usually much longer than pharmacologic treatments, since devices are intended to demonstrate pain relief and return of function over a period of years. Efficacy early in the time course may reflect the variables associated with the surgical procedure and not the performance of the device itself. The compromise for assessing the efficacy of joint replacement has been to select a suitable time frame combined with acceptable outcomes that reflect efficacy. For joint arthroplasty, for example, orthopaedic surgeons and the editorial boards of respected peer-reviewed orthopaedic journals have accepted 2 years as the minimum acceptable time period for assessing efficacy. Acceptable measures of efficacy are less agreed upon, though most joint arthroplasty surgeons rely on hip or knee scores that encompass both pain and function. However, specific instruments to measure efficacy of devices that have been validated and reflect outcomes include the Western Ontario McMaster University Osteoarthritis Index (WOMAC) , the Knee Society Clinical Rating System and the Oxford-1 Item Questionnaire for the knee. The hip outcome measures that are validated include the Harris Hip Score and the Hospital for Special Surgery Hip Rating. These latter scoring systems, however, do not take into consideration changes in the patient's medical condition and age so that a Short-Form Health Survey (SF-36) is useful to assess the health-related quality of life aspects of outcome[21].

This general approach has been accepted by the FDA and device manufacturers and should remain. However, the effectiveness of the implant should be differentiated from that of the surgical procedure. Additional research should focus on what specific endpoints can be assessed to determine failure modes. Currently a device with a relatively low failure rate, as identified by data collected from a large number of surgeons/surgical procedures, is considered efficacious. It is unlikely, however, that a common time point such as 2 years is appropriate for all devices used in the treatment of OA. Certain failure modes for joint replacements, such as osteolysis, are know to emerge only after many years of service, while new devices that incorporate biologics or drugs may demonstrate effectiveness over time courses much shorter than 2 years.

HA-based viscosupplements are currently indicated for the treatment of pain related to knee OA that is unresponsive to simple analgesics. These products require clinical safety and efficacy data as part of their PMA application to obtain FDA approval. Because they are intended to treat pain, viscosupplements have effectiveness endpoints that are similar to that expected for pharmacological OA pain treatments. However, viscosupplements and biologic devices differ from some of the other pharmacological therapies in that the treatments are confined to the joint in question and the expected effect is sustained pain relief for 3 months or longer in the treated joint.

For purposes of drug approval, the current FDA draft guidance for HA viscosupplements should allow for measurements specific to the treated joint. For example, in patients with OA in more than one knee joint or one hip joint, the patient-reported outcome of pain, function, and stiffness should be for the treated joint alone. Several approaches could be implemented to address this issue during clinical development programs for the treatment of OA pain with a medical device, including enrolling patients with a single joint disease (which may be daunting since bilateral disease is far more common), treating all affected joints, analyzing patients separately based on whether all symptomatic OA was treated or not, and using subscales that may be more specific for the treated joint, such as the WOMAC A-1 pain on walking[22].

Determining the relative risk to benefit of new devices

It is important for device manufacturers to work closely with the FDA and the clinical community to adopt special controls aimed at minimizing risk while providing an avenue for maximizing benefit. Establishing with the FDA what additional data (preclinical, clinical, and post-market) are needed to show substantial equivalence or claimed improvements with both the surgery and the device and gaining consensus on the appropriate control groups are important steps. Minimal standards should be obtained from the existing ASTM and ISO standards. While a claimed improvement by a new device over an existing device (e.g., improved wear performance, kinematics, or fixation) could be tested with existing standards, situations will arise for which an appropriate, validated test is unavailable. Therefore, tests that have been developed within the scientific community and published in the peer-reviewed literature could be considered. Changing over to a relative risk vs benefit would enhance the present 510K pathway that in many circumstances is used by device manufacturers to circumvent the more difficult Class III pathway.

Clearly, risk is associated with every device and every procedure. For a new Class II device, the risks should be minimized by comparing with a predicate device plus utilizing additional tests to demonstrate the safety and efficacy of any new claims. However, risk is not constant for a given device for every patient. For example, the risk of loosening and wear is higher with heavier and more active patients[21]. Yet it is unrealistic or even impossible to design every device to function indefinitely without problems in the "worst case scenario". In this circumstance, the patient–physician relationship is central in importance.

HA-based supplements, since they are indicated for pain relief, require a patient-reported outcome. To adequately assess the safety and effectiveness of HA-based supplements, the patient and other reporters need to be blinded to treatment selection. An appropriate control would need to incorporate an intra-articular or sham injection and be perceived by the patient as being the same as an intra-articular injection of the study device. Viscosupplement trials have included the following treatments and controls in an effort to blind the patients to treatment: intra-articular injections of phosphate buffered saline, an already approved viscosupplement, or glucocorticoid, or a sham injection. Depending on the trial design (e.g., non-inferiority to an active available pain relief product or superiority to a non-active treatment or an active treatment) any of these control options are acceptable. However, improvement compared to some other therapy requires not only statistical superiority (or non-inferiority as the case may be) but also clinically meaningful improvements in the outcome. Since clinical relevance is open to debate, but must be established prior to initiating the regulatory pathway for new products, early consultation and regular communication with the appropriate review division at the FDA is desirable. If non-inferiority to an active comparator is to be pursued, then establishing an acceptable margin of non-inferiority is critical: the smaller the margin, the larger the trial. The approach is applicable to other injectables and biological device, e.g., BMP-2.

Optimal outcome parameters for evaluation

Pain relief, restoration of function (range of motion (ROM), 6 min walk), other functional observed and measured performance outcomes as previously discussed, assessment of radiographs and/or other images, complications and complication rates, revision and revision rates are all validated and optimal approaches to defining outcomes[22,23]. Each of these measures is important and valuable to provide a comprehensive quantitative assessment of the device outcome from the physicians' perspective. However, patient-reported self-assessment outcomes are critical in that they give a specific measure of performance in relation to patient expectations, independent of evaluation by medical staff. Independent living, work status, and return to recreational sports and related activities may also be indicators for restoration of function. A number of instruments have been validated to define outcome measures from either physician-derived or patient-reported measures. These include the Harris Hip Score, the new Knee Society Scoring System, and the WOMAC. The quality of life measures that are important in determining the effectiveness of medical treatments, including biological devices, include the SF-36 (medical outcome study short form) and the quality of life evaluation and health assessment questionnaire (HAQ).

Optimal global assessment tools for evaluating the outcome of HA-based viscosupplements on pain and function in a single target joint include but are not limited to the visual analog scale (VAS) for pain, the WOMAC scale, and patient directed study instruments such as the SF-36 outcome form, the Lequesne Functional Index, and the OARSI-Outcome measures in rheumatology clinical trials (OMERACT) responder rate[24,25]. Until now the most prevalent primary endpoint has been a VAS pain measurement in the treated joint and patient-reported outcome tools in the WOMAC and PTTA instruments. Outcome measures of biological devices may require specific structural outcome measures such as magnetic resonance imaging (MRI) biomarkers for cartilage quality as defined in the

biomarker section to assess their efficacy. All of the above measures incorporate validated and verifiable outcome tools and can be used to measure a benefit of biological devices for the treatment of OA.

Are outcome parameters substantially different with respect to different joints?

The intended goals for surgical treatment of OA across all joints include pain relief, restoration of function, independent living and return to productive employment, and a low re-operation rate within 10 years of the incident procedure. However, specific parameters and criteria exist for success directly related to each joint. These parameters are best reflected in the numerous validated assessment tools both from the physician and the patient point of view that have been developed by the specialty societies addressing each anatomical area. For example, the glenohumeral joint is addressed by a number of specific shoulder assessment tools including the Western Ontario OA Scale for the shoulder. The American Academy of Orthopaedic Surgeons (AAOS) subscale for the ankle is another example of a joint specific assessment tool[26].

The general goals for treatment with HA-based viscosupplements are also the same across all joints: pain relief and functional improvement. However, as with devices, specific measurement tools have been designed for load bearing vs non-load bearing joints. These tools as described above should be considered on a case-by-case basis for specific joints.

Assessment of short-term vs long-term benefits

Orthopaedic devices used for joint replacement have a goal of long-term success, that is the permanent replacement of the joint bearing surfaces. Short-term benefits are important in regard to complications that accompany the surgical implantation; from a regulatory standpoint, these include complications directly related to surgical instrumentation or to the device itself. However, the ultimate goal of long-term benefit must be considered. In this respect, orthopaedic devices differ from most other medical devices. A 2-year time period with suitable evaluation methods will define problems such as premature loosening, instability, or inadequate motion. Additional evaluation might be required for Class II devices with special claims.

Biological devices are similar to HA-based viscosupplements in that short-term benefits are critical. These products should be studied over a 3–6 month period, and repeat injections should be performed to assess the safety of repeated injections. Post-marketing surveillance is critical, as these products have not shown pain relief in all patients. A specific responder analysis should be carefully monitored to determine the outcomes of success or failure. Their measured benefit should not only be in terms of statistical improvement to a comparator but should also be clinically relevant.

Assessment of complications and other adverse events and their prevention

The complications of orthopaedic devices are well documented and numerous studies have provided data on their incidence, causes and preventive measures[27,28]. The prevention of device-related complications begins with rigorous preclinical testing. The importance of well-designed clinical programs cannot be over emphasized. There are, however, complications that are associated not with the device, but introduced by inadequate instrumentation, by surgical factors such as incorrect ligament balancing, poor cement technique, mal-rotations of components, or even by the patient themselves. Complications must be clearly attributed to the device as opposed to related to surgical or patient issues. If possible the preclinical testing should include studies which expose the sensitivity of devices to such occurrences. Specific device adverse events would include premature wear, breakage and early loosening.

HA-based supplements also have a history of adverse events. The adverse events are typically divided into those related to the injection procedure itself and those related to the HA material injected into the intra-articular joint space. The same is true with biological devices. Although significant clinical data exist on the knee, less clinical data are available for other joints and new safety issues or signals may exist for these joints. Typical complications occurring in the knee include injection site pain, erythema, effusion, stiffness, or potential allergic reaction to the material. However, the severity of these adverse events is usually mild to moderate, and the reported problems resolve spontaneously. No long-term complications have been reported. Continued post-market surveillance with standardized criteria should continue.

Clinical indications

The clinical indications for joint replacement are well documented and include limitation of function of any given joint either due to pain or malfunction to justify the risk of surgical intervention and introduction of a foreign body with the intent of relief of pain and restoration of function of the joint[29,30].

The clinical indications for the use of HA-based viscosupplements include treatment of patients with pain from OA of the knee who have failed to respond to conservative non-pharmacological therapy and simple analgesics[31]. Presently no viscosupplements are approved for non-knee joint involvement in the United States, though a number of clinical trials are being conducted with the goal of extending these treatment modalities to other joints. No well-documented indications exist for biological treatments (devices), used for cartilage repair and considerable work must be completed by the FDA in developing guidance documents for the regulation of these products. Collaboration with the scientific community is required.

Balancing the cost associated with devices against conservative therapy

This is a timely, though difficult area to address. Essential to understanding this relationship should be a quality of life estimate so that both economic costs and patient derived satisfaction are considered. There is a strong subjective element, which relates to the patient–physician relationship. A central question is whether a patient is prepared to continue with conservative treatment for an extended time period, functioning sub optimally and experiencing pain, because they believe that a total joint or a biological device has a limited lifetime with a measurable risk for an early revision; or does the patient prefer early treatment with restoration of function and relief of pain, but with the risk of a failure of the device requiring revision in the future. No well-validated studies address the issue of defining the true risks and benefits of these procedures in the long term and how they affect cost. Careful consideration of the issues of quality of life and its concomitant economic costs must be considered for any evidence-based decision making process.

HA-based viscosupplements and biological intra-articular treatment may be indicated for pain relief in patients where simple analgesics have failed. Data indicate that HA-based viscosupplements may provide long-term pain relief with only one treatment and could be cost effective. However, further studies should be done specifically to address the issue of cost-effectiveness of intra-

articular biological treatments compared to traditional conservative therapy.

Moving forward

To further our understanding of orthopaedic devices, their safety and efficacy, relative benefits and risk, and long-term outcomes, the following recommendations have emerged from literature reviews, clinical experience, and group consensus:

- Meta-analyses of clinical results with current technologies should be undertaken with emphasis on demonstrating safety and efficacy by identifying types of complications, their prevalence, their timing, and their relationship (if any) to the device. Meta-analysis should also be undertaken for revisions. Current national and Medicare registries do not contain enough information to make these recommended efficacy determinations.
- A meta-analysis of current outcome measures (patient- and surgeon-derived questionnaires, objective measures like the 6 min walk, *etc.*) should be performed and a consensus reached on the most appropriate outcome measures to be utilized. Multi-center prospective studies using non-developer participants are needed. This effort might be of importance in revising existing national registries outside the USA and in establishing a US registry as a means of including outcome measures beyond revision surgery as the endpoint of a joint replacement.
- Research is necessary to establish the efficacy of existing and proposed standards within the context of regulatory science. Existing standards should be challenged on the basis of objective scientific and cost data to establish whether the standards have demonstrated device performance in a clinically meaningful way. Efforts by FDA to accept voluntary standards as part of the special controls used in the regulatory process should be encouraged through collaborative efforts among FDA, device manufacturers, and the scientific and clinical communities.
- A national device registry should be established with well-defined goals. This registry would as example characterize practice patterns, identify failures, establish benchmarks, develop guidelines and assess utilization issues[20].
- Consensus building needs to take place in the surgical and scientific community to clearly define the primary modes of device failure. This would allow the FDA to establish more meaningful guidance to device manufacturers than might currently exist and create a framework for continual evaluation of the current consensus as new information becomes available.

For HA-based viscosupplements, a significant number of clinical trials have been conducted on multiple injection viscosupplements used in the knee. These trial results have varied dramatically, as have the trial designs. It is important to collect improved randomized, controlled, double-blind patient-reported outcomes on these products to establish a class effect. A constant comparator such as a saline injection should be incorporated into the clinical trials.

Viscosupplement development should be focused on reducing the number of injections required for treatment, increasing intra-articular residence time through cross-linking, and providing effective treatments of other synovial joints beyond the knee such as the hip, shoulder, ankle, temporomandibular joint, carpometacarpal joint, and the facet joints of the spine.

Potential research topics might include:

- Definition of a responder on a patient-reported outcome for a viscosupplement treatment from 3 to 6 months, stating limitation of current OARSI-OMERACT responder rate criteria.
- Consideration on whether or not repeated measures or a landmark analysis is more appropriate for 3 and 6 months viscosupplement trials.
- Definition of an appropriate placebo comparator (e.g., saline control, sham injection, or phosphate balance solution (PBS) control with lidocaine).
- Increased understanding of appropriate injection volumes for different joints and appropriate endpoints measures for different joints.
- Increased understanding of the importance of residence time for viscosupplements and mechanism of action of synovial fluid replacement with HA-based viscosupplement material.

Conclusions

It is hoped that the work undertaken by the members of the OARSI devices working group will be helpful in enhancing the FDA process of assessing new devices. An ordered sequential approach to the introduction of any “device” is critical. Additionally, a National Registry is important but should have well-defined research objectives, a valid protocol design, clear inclusion and exclusion criteria, a comprehensive collection of variables necessary to answer the registry objectives, mechanisms implemented to track patients and to insure a high level of data integrity, and finally a blinding of data collection personnel and a method to rectify methodological problems. Finally, appropriate dissemination and data sharing procedures must be put in place to benefit the consumers, which includes patients, surgeons, and device manufacturers. The feedback process should result in an enhanced quality of care and cost- and comparative-effectiveness of any new treatment.

Author contributions

VG role: conception, drafting, critical revision and final approval of the manuscript.

MH, WJ, WM, MP, BR, TV, PW, RW, TW: critical revision.

Declaration of funding and role of funding source

The OARSI FDA OA Initiative received financial support from the following professional organization:

American College of Rheumatology

Additionally the OARSI FDA OA Initiative received financial support from the following companies:

Amgen
ArthroLab
AstraZeneca
Bayer Healthcare
Chondrometrics
CombinatoRx
Cypress BioScience
DePuy Mitek
Expanscience 4QImaging Genevrier/IBSA
Genzyme
King (Alpharma)
Merck
Merck Serono
NicOx
Pfizer
Rottapharm
Smith & Nephew
Wyeth

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group

discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the FDA.

Conflicts of interest

JB, WJ, MM, MP, TV, RW, TW: no conflicts of interest to declare.
VG: Royalties, Zimmer.
MH: employee of Genzyme.
BR: employee of Smith & Nephew.
PW: Consultant, royalties Zimmer.
TW: Funding Synthese and Stryker; Royalties Mathys.

References

1. Food and Drug Administration. Guidance for Industry, Clinical Development Programs for Drugs, Devices, and Biological Products Intended for the Treatment of Osteoarthritis (OA) July 1999.
2. Ethgen O, Bruyere O, Richy F, Dardennes C, Reginster J. Health-related quality of life in total hip and total knee arthroplasty. J Bone Joint Surg May 2004;86A(5).
3. Jones CA, Beaupre LA, Johnston DW, Suarez-Almazor ME. Total joint arthroplasties: current concepts of patient outcomes after surgery. Clin Geriatr Med 2005;21:527–41.
4. Lee K, Goodman SB. Current state and future of joint replacements in the hip and knee. Expert Rev Med Devices 2008;5(3):383–93.
5. Rand JA, Trousdale RT, Ilstrup D, Harmsen WS. Factors affecting the durability of primary total knee prostheses. J Bone Joint Surg 2003;85A:259–65.
6. Bono CM, Kadaba M, Vaccaro AR. Posterior pedicle fixation-based dynamic stabilization devices for the treatment of degenerative diseases of the lumbar spine. J Spinal Disord Tech 2009 Jul;22(5):376–83.
7. American Academy of Orthopaedic Surgeons. Update on Pedicle Screws, http://www2.aaos.org/aaos.archives/bulletin/apr96/pedicle.htm; April 1996 [accessed 28.02.08].
8. Goldberg VM, Buckwalter JA. Hyaluronans in the treatment of osteoarthritis: evidence for disease-modifying activity. Osteoarthritis Cartilage 2005 Mar;13(3):216–24.
9. Ghosh P, Guidolin D. Potential mechanism of action of intra-articular hyaluronan therapy in osteoarthritis: are the effects molecular weight dependent? Semin Arthritis Rheum 2002;32:10–37.
10. Listrat V, Ayral X, Patarnello F, Bonvarlet J, Simonnet J, Amor B, *et al.* Arthroscopic evaluation of potential structure modifying activity of yaluronon (Hyalgan®) in osteoarthritis of the knee. Osteoarthritis Cartilage 1997;5:153–60.
11. Jubb RW, Piva S, Beinat L, Dacre J, Gishen P. A randomized placebo (saline)-controlled clinical trial of the structure modifying effect of 500–730 KDa sodium hyaluronate (Hyalgan) in osteoarthritis of the knee. Int J Clin Pract 2003;57:467–74.
12. Vasiliadis HS, Wasiak J. Autologous chondrocyte implantation for full thickness articular cartilage defects of the knee. Cochrane Database Syst Rev 2010 Oct 6;10. CD003323.
13. Food and Drug Administration. Premarket Notification 510(k). http://www.fda.gov/CDRH/DEVADVICE/314.html.
14. Food and Drug Administration. Premarket Approval. http://www.fda.gov/CDRH/DEVADVICE/pma/.
15. Kirkpatrick John S, Stevens Theodore. The FDA process for the evaluation and approval of orthopaedic devices. J AAOS 2008;16:260–6.
16. Malchau H. Introducing new technology: a stepwise algorithm. Spine 2000 Feb 1;25(3):285.
17. Malchau H. On the importance of stepwise introduction of new hi implant technology: assessment of total hip replacement using clinical evaluation, radiostereometry, cigitised radiography and a national hi registry. *Thesis* Univ of Gothenburg, 1995.
18. Robertsson O, Dunbar MJ, Knutson K, Lewold S, Lidgren L. The Swedish knee arthroplasty register: 25 years experience. Bull Hosp Jt Dis 1999;58:133–8.
19. Rolfson O. Patient-reported outcome measures and health-economic aspects of total hip arthroplasty. A study of the Swedish hip arthroplasty registry. Institute of Clinicl Sciences at Sahlgrenska Academy University of Bothen burg. Thesis, 2010.
20. Malchae H, Garllick G, Eisler T, Karrholm J, Herberts P. Presidential guest address: the Swedish hip registry: increasing the sensitivity by patient outcome data. Clin Orthop Relat Res 2005;441:19–29.
21. Mullhall KJ, Saleh KJ. Outcomes assessment in hip and knee arthroplasty. In: Orthopaedic Knowledge Update: Hip and Knee Reconstruction 2006;3:271–80.
22. VibStrand S, Hochberg M. Study design and outcome measures in osteoarthritis clinical trials. In: Osteoarthritis. 4th edn. Lippincott, Williams & Wilkins; 2007:313–25.
23. Foran JRH, Mont MA, Etienne G, Jones L, Hungerford D. The outcome of total knee arthroplasty in obese patients. J Bone Joint Surg 2004;86:1609–15.
24. Pham T, van der Heljde D, Altman R, Bonvarlet J, Simonnet J, Amor B, *et al.* OMERACT-OARSI initiative: osteoarthritis research society international set of responder criteria for osteoarthritis\clinical trials revisited. Osteoarthritis Cartilage 2004;12:389–99.
25. Lequeane MG. The algofunctional indices for hip and knee osteoarthritis. J Rheumatol 1997;24:779–81.
26. Foot and Ankel Outcomes Questionnaire Developed by AAOS, AAHKA, Hip Society, Knee Society et al 2005
27. Scuderi G, Trousdale RT. Complications after total knee arthroplasty. In: Orthopaedic Knowledge Update: Hip and Knee Reconstruction, AAOS 2006:147–56.
28. Masri B, Davidson D, Duncan C, Lewallen D, Noiseux N, Ranawat C, *et al.* Total hip arthroplasty complications. In: Orthopaedic Knowledge Update: Hip and Knee Reconstruction, AAOS 2006:457–74.
29. Zhang W, Moskowitz RW, Nuki G, Abransom S, Altman R, Arden N, *et al.* OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis Cartilage 2007;15(9):981–1000.
30. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman R, Arden N, *et al.* OARSI recommendations for the management of hip and knee osteoarthritis, part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008;16(2):137–62.
31. Goldberg VM, Goldberg L. Intra-articular hyaluronans the treatment of knee pain in osteoarthritis. J Pain Res 2010;3:51–6.

Osteoarthritis and Cartilage 19 (2011) 515–542

Osteoarthritis and Cartilage

Application of biomarkers in the development of drugs intended for the treatment of osteoarthritis

V.B. Kraus †*, B. Burnett ‡, J. Coindreau §, S. Cottrell ||, D. Eyre ¶, M. Gendreau #, J. Gardiner ††, P. Garnero ‡‡§§|||¶¶, J. Hardin ##†††‡‡‡, Y. Henrotin §§§, D. Heinegård |||||, A. Ko ¶¶¶, L.S. Lohmander ###, G. Matthews ††††, J. Menetski ¶¶¶, R. Moskowitz ‡‡‡‡, S. Persiani ‡‡‡‡§§§§, A.R. Poole |||||||, J.-C. Rousseau §§|||, M. Todman ¶¶¶¶ OARSI FDA Osteoarthritis Biomarkers Working Group

† *Department of Medicine, Duke University Medical Center, Durham, NC, USA*
‡ *Duke Translational Medicine Institute, Duke University School of Medicine, Durham, NC, USA*
§ *Novartis Pharmaceuticals Corporation, East Hanover, NJ, USA*
|| *Amgen, Seattle, WA, USA*
¶ *Department of Orthopaedics and Sports Medicine, University of Washington, Seattle, WA, USA*
# *Cypress Biosciences, Inc, San Diego, CA, USA*
†† *Discovery Translational Medicine, Wyeth Research, Collegeville, PA, USA (previously)*
‡‡ *INSERM , UMR1033, F-69008, Lyon, France*
§§ *Université de Lyon, F-69008, Lyon, France*
||| *Hospices Civils de Lyon, France*
¶¶ *Cis Bioassays, Codolet, France*
Albert Einstein College of Medicine, Bronx, NY, USA
††† *Montefiore Medical Center, Bronx, NY, USA*
‡‡‡ *Arthritis Foundation National Office, Atlanta, GA, USA*
§§§ *Bone and Cartilage Research Unit, Institute of Pathology, Level 5, CHU Sart-Tilman, 4000 Liège, Belgium*
||||| *Department of Clinical Science, Lund, Section Rheumatology, Lund University, Lund, Sweden*
¶¶¶ *Merck & Company, Inc, Rahway, NJ, USA*
Department of Orthopaedics, Clinical Sciences Lund, Lund University, Lund, Sweden
†††† *Genzyme Corporation, Cambridge, MA, USA*
‡‡‡‡ *University Hospitals, Case Western Reserve University, Cleveland, OH, USA*
§§§§ *Department of Translational Sciences and Pharmacokinetics, ROTTAPHARM-MADAUS Monza, Italy*
||||||| *Department of Surgery, McGill University, Montreal, Quebec, Canada*
¶¶¶¶ *Smith & Nephew Research Centre, Heslington, York, UK*

ARTICLE INFO

Article history:
Received 22 June 2010
Accepted 28 August 2010

Keywords:
Osteoarthritis
Biomarkers
Clinical trial

SUMMARY

Objective: Osteoarthritis (OA) is a chronic and slowly progressive disease for which biomarkers may be able to provide a more rapid indication of therapeutic responses to therapy than is currently available; this could accelerate and facilitate OA drug discovery and development programs. The goal of this document is to provide a summary and guide to the application of *in vitro* (biochemical and other soluble) biomarkers in the development of drugs for OA and to outline and stimulate a research agenda that will further this goal.
Methods: The Biomarkers Working Group representing experts in the field of OA biomarker research from both academia and industry developed this consensus document between 2007 and 2009 at the behest of the Osteoarthritis Research Society International Federal Drug Administration initiative (OARSI FDA initiative).
Results: This document summarizes definitions and classification systems for biomarkers, the current outcome measures used in OA clinical trials, applications and potential utility of biomarkers for development of OA therapeutics, the current state of qualification of OA-related biomarkers, pathways for biomarker qualification, critical needs to advance the use of biomarkers for drug development, recommendations regarding practices and clinical trials, and a research agenda to advance the science of OA-related biomarkers.

* Address correspondence and reprint requests to: Virginia Byers Kraus, Box 3416, Duke University Medical Center, Durham, NC 27710, USA. Tel: 1-919-681-6652; Fax: 1-919-684-8907.
E-mail address: vbk@duke.edu (V.B. Kraus).

1063-4584/$ – see front matter
doi:10.1016/j.joca.2010.08.019

Conclusions: Although many OA-related biomarkers are currently available they exist in various states of qualification and validation. The biomarkers that are likely to have the earliest beneficial impact on clinical trials fall into two general categories, those that will allow targeting of subjects most likely to either respond and/or progress (prognostic value) within a reasonable and manageable time frame for a clinical study (for instance within 1–2 years for an OA trial), and those that provide early feedback for preclinical decision-making and for trial organizers that a drug is having the desired biochemical effect. As *in vitro* biomarkers are increasingly investigated in the context of specific drug treatments, advances in the field can be expected that will lead to rapid expansion of the list of available biomarkers with increasing understanding of the molecular processes that they represent.

Introduction

It is said that a disease starts when detected by the best marker available to define it. To date, this usually requires the presence of a clinical symptom, which often occurs well into the progression of an illness or disease. However, there is significant evidence that there are often early, pre-symptomatic biomarkers of illness and disease, which if detected, may allow for earlier treatment. Therein lies the power and importance of applying biomarkers to osteoarthritis (OA), a disease often characterized by a prolonged asymptomatic molecular phase, a pre-radiographic phase, and a recalcitrant later radiographic phase with evident structural joint changes, frequent pain, and loss of function (Fig. 1). Biomarkers have the potential to provide an early warning of the initiation of matrix breakdown that could prompt earlier treatment to prevent the cartilage and bone destruction that leads to disability. Thus, there currently exists a great need and opportunity for biomarkers to provide a method for earlier diagnosis of OA, and to inform the prognosis, monitoring and therapeutic strategies for OA. Wagner has predicted that the next few years will see a rapid increase in the number of drugs approved with biomarker data in their labels, and older drugs that will have biomarker data added to their labels[1]. OA may be chief among them due to the current lack of a gold standard that comprehensively captures the disease in all of its manifestations. In addition, OA is a chronic and slowly progressive disease for which biomarkers may be able to provide a more rapid indication of therapeutic response to disease structure modifiers than is available through currently established means; this could streamline and optimize the discovery and development programs of new therapeutic agents. The mandate of the Osteoarthritis Research Society International Federal Drug Administration (OARSI FDA) Biomarkers Working Group was 2-fold. First to create a critical appraisal of fundamentals of the science related to biomarkers of OA, particularly as they relate to the development of drugs intended for the treatment of OA. Second, to address specific queries posed by the FDA related to OA biomarkers, namely: What biomarkers now exist? What is their utility? What evidence is available to support surrogacy for clinical outcomes? What is the face validity? What is the practicality? What is the research agenda required to inform each of the above questions? Thus this document is intended to address this 2-fold purpose in the hopes of helping to advance the development of drugs for OA.

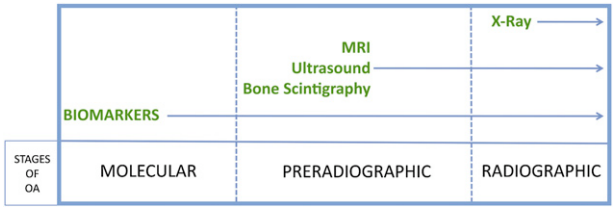

Fig. 1. Continuum of OA stages as paradigms for structure modifying OA trials.

Scope of the document

A previous broad ranging biomarker white paper was commissioned and prepared for the launch of the National Institutes of Health Osteoarthritis Public/Private Research Initiative and was published on line in 2000 (and now found at the OARSI website http://www.oarsi.org/index2.cfm?section=OARSI_Initiatives&content=Biomarkers) The present document has a much more specific focus. It also covers the great increase in biomarker research activity in the present decade and utilizes definitions and nomenclature that are harmonized with and expand upon those proposed to date in FDA draft guidance documents. This current paper covers biochemical/molecular and genomic (RNA-gene expression, DNA-genetic polymorphisms) biomarkers of OA but excludes imaging biomarkers and clinical risk factors such as obesity, malalignment, and gender because other working groups are covering these topics in companion documents. We include a brief summary of issues related to the current methods of OA diagnosis, treatment and response criteria for therapeutic trials, and the challenges posed by the current 'gold standard' radiographic trial criteria, in order to provide a framework in which to conceptualize the role to be played by biomarkers in the development of drugs for OA. The concept of OA as a continuum includes early stages that may be amenable to treatment if appropriate biomarkers are defined, which in turn could complement current treatment paradigms for established radiographic OA; prevention versus treatment of established disease has traditionally been referred to as primary and secondary prevention, respectively.

Potential uses and challenges for each type of biomarker based on the BIPEDS classification scheme (described below) in the drug development process are discussed. Summary tables illustrating study power for treatment effects based on varying effect sizes are provided utilizing a theoretical biomarker as well as known soluble biochemical OA biomarkers, and their current level of qualification based on published clinical trials.

A summary of the pathways required for biomarker qualification is included that lists the regulatory agencies involved with biomarker development, as well as recommendations for biomarker endpoints in trials. Clinical and scientific issues are also raised that would benefit from more research. Appendices are provided containing recommendations for sample collection, processing and storage, as well as a glossary of biomarker terms.

Definition of biomarkers

A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention[2]. This is in contrast to a clinical endpoint that is a marker or variable that measures how a patient feels, functions or survives. A biomarker becomes a surrogate endpoint when it is appropriately qualified to substitute for a clinical endpoint. The technical

revolution in molecular biology has led to the expansion of the notion of what constitutes a potential biomarker to include, not only proteins and protein fragments, but also metabolites, carbohydrate biomarkers, genomic biomarkers (RNA and DNA)[3], cellular biomarkers (could be captured for instance as the cell pellet from body fluids), and imaging biomarkers. Based on their characteristics, we can divide biomarkers into two major groups: the so-called soluble or "wet biomarkers", usually measured in a selected body fluid such as blood, serum, plasma, urine, or synovial fluid (SF) and usually representing modulation of an endogenous substance in these fluids; and the so-called "dry biomarkers" usually consisting of visual analog scales (VASs), questionnaires, performed tasks, or imaging. These two types of biomarkers can also be referred to as *in vitro* biomarkers (derived from *in vitro* diagnostics) vs *in vivo* biomarkers respectively. Although many of the concepts presented here are applicable to all of these types of biomarkers, imaging biomarkers are dealt with more specifically in a companion document so we focus herein on the non-imaging, *in vitro*, soluble biomarkers.

Processes of biomarker qualification and validation

Qualification is a process applied to a particular biomarker to support its use as a surrogate endpoint in drug discovery, development or post-approval and, where appropriate, in regulatory decision-making[2]. In contrast, validation of a biomarker is much broader and can relate to verification of analytical performance characteristics (such as precision, accuracy, stability, etc) as well as clinical correlation of a biomarker with a biological process or clinical outcome. Current practice however is to supplant the term validation with qualification when the focus is on the portent (meaning) as opposed to the performance (analytical aspects) of the biomarker. A major difference between validation and qualification resides in the fact that the latter only has meaning in a context. For example, qualification of a biomarker may take into consideration the particular level of progression of the disease and its severity, thereby leading to the qualification for some states of the disease, but not for others. A systematic process has been in development for accurate and comprehensive qualification of biomarkers for use in drug development[4]. To date, draft guidelines exist on qualification of genomic biomarkers[2], produced by the International Conference on Harmonisation (ICH), whose goal has been to create a harmonized structure for qualifying the biomarkers that will lead to consistent applications and discussions among regulatory authorities and sponsors. Qualification endpoints in OA could include structural outcomes [identified with magnetic resonance imaging (MRI), or X-ray etc], and/or clinical outcomes (pain, function etc); biochemical and/or genomic biomarkers are linked to modifications in these outcomes through the process of biomarker qualification.

Classification systems for biomarkers

BIPEDS

In this document we refer to and use two main classification systems for biomarkers with modifications as described here. The first, a system called BIPED, classifies the major types of biomarkers[5] into five categories corresponding to Burden of Disease, Investigational, Prognostic, Efficacy of Intervention, and Diagnostic biomarkers. We have added a Safety category to the BIPED system, and hereafter, throughout this document, refer to the BIPEDS classification system. This change facilitates the goal of this document to provide a guide to the comprehensive application of biomarkers to the study and treatment of OA. Biomarkers of safety can be considered biomarkers able to reflect tissue and/or organ toxicity of an agent or intervention and are analogous to biomarkers of toxicity in the process of evaluation and validation by the Critical Path Initiative for diverse organ systems (see home page http://www.fda.gov/ScienceResearch/SpecialTopics/CriticalPathInitiative/default.htm).

Qualification levels for biomarkers

The second useful classification system referred to here divides biomarkers into four categories according to their current level of qualification described further in Pathways for Biomarker Qualification-Levels of qualification of biomarkers for drug development use [1]:

Exploration level biomarkers are research and development tools accompanied by *in vitro* and/or preclinical evidence for which there is no consistent information linking the biomarker to clinical outcomes in humans (these are used for hypothesis generation);

Demonstration level biomarkers are associated with clinical outcomes but have not been reproducibly demonstrated in clinical studies (this category corresponds to "probable valid biomarkers" in nomenclature suggested in draft guidance from the FDA[6] and are useful for decision-making by providing evidence to support the primary clinical evidence);

Characterization level biomarkers are reproducibly linked to clinical outcomes in more than one prospective clinical study in humans (this category corresponds to "known valid biomarkers" in nomenclature suggested in guidance by the FDA[6] and are useful for decision-making, dose finding, and secondary and tertiary claims); and

Surrogacy level biomarkers can substitute for a clinical endpoint (this category corresponds to "surrogate end point" and requires agreement with regulatory authorities as an FDA registrable endpoint).

Summary

As noted in a recent FDA guidance document[7], the use of biomarkers in drug discovery, development and post-approval has the potential to facilitate development of safer and more effective medicines; in fact, one of the main objectives of a biomarker for drug development is to allow the construction of the dose-exposure–response curve in patients for both the therapeutic and toxicity effects. This will facilitate dose selection in order to reach the best benefit-risk ratio of an approved medicine. In the OA field, the potential also exists for biomarkers to enhance the probability of obtaining early indications of success during clinical drug development for OA. The selection of a new biomarker test depends critically upon the ability of the test to link the mechanism of action of a new agent with a therapeutic response. The therapeutic response usually addresses an unmet medical need, and in the case of OA, there are currently no qualified biomarkers that can be considered as surrogate clinical endpoints. Thus it is a two-edged sword: the ultimate degree of biomarker uptake and use is intimately tied to the ability to act on the biomarker information provided, which in turn is dependent on the ability of biomarkers to enhance the success of clinical trials to achieve the actionable result needed for biomarkers to be adopted for clinical use.

It is worth noting here that the field of drug development for OA is currently analogous to osteoporosis 30 years ago[8], namely a disease in search of a robust gold standard outcome measure to inform clinical trials. The 1979 FDA Osteoporosis Guidelines acknowledged that evaluating the clinical effectiveness of osteoporosis drugs posed special challenges because of the "difficulties in assessing the state of skeletal bone quantitatively *in vivo*, the relatively small changes that are usually encountered and the duration of studies necessary to show significant effects"[8,9]. By 1984, the FDA Osteoporosis Guidelines upgraded dual-energy photon absorptiometry from investigational

to a valid and reliable method for measuring trabecular bone mass of the spine and this was critical to the subsequent approach to the development and regulation of osteoporosis drugs[8,9]. OA is at a similar crossroads to which biomarkers may contribute substantively at this time. Given the urgent need for OA therapies, it is hoped that the concepts advanced in this document will facilitate and stimulate the inclusion of biomarkers as secondary endpoints in all future OA trials, and lay the groundwork for the evolution to the use of biomarkers, in some cases, as primary endpoints.

OA diagnosis, treatment and trials

Diagnosis

The American College of Rheumatology (ACR) has developed a set of clinical and radiological criteria for the diagnosis of hip, knee and hand OA[10–12]. The ACR diagnostic criteria are based on the association of many clinical, or clinical and radiological criteria, and are commonly used for patient inclusion in clinical trials. These ACR criteria are very specific and thus are useful for differentiating patients with OA from those with inflammatory joint diseases. Their sensitivity is less impressive, illustrating their limited ability to discriminate patients with early OA from healthy controls. The most commonly used radiographic grading system is that of Kellgren and Lawrence (KL)[13], based on the presence of osteophytes, joint space narrowing (JSN), subchondral bone sclerosis and cyst formation. This scoring system divides OA into five grades (0–4) mainly based on the presence and number of osteophytes. A score of 2 or more has traditionally been considered to be a definitive radiographic diagnosis of OA and has been widely used in clinical trials as an inclusion criterion. However, evidence suggests that KL grade 1 is *bona fide* OA and distinct from KL grade 0 based on subsequent risk of progression[14]. Based on the concept of the disease continuum that includes a molecular stage and a pre-radiographic stage of OA as presented in Fig. 1 and supported by the literature[15,16], even with inclusion of KL grade 1 as *bona fide* OA, radiographic criteria will identify only late-stage OA. Because the KL scoring system relies predominantly on osteophytes to determine OA severity, the atrophic form of OA, which consists mainly of JSN, is underestimated. The KL grading system is also known for its poor correspondence of radiographic severity with hip or knee pain. MRI, ultrasound or biochemical markers are not yet included in any set of diagnostic criteria for OA.

Treatment

A cure for OA remains elusive and the management of OA is largely palliative, focusing on the alleviation of symptoms. Current recommendations by the European League Against Rheumatism (EULAR), the ACR, and the OARSI for the management of OA include a combination of non-pharmacological interventions and pharmacological treatments. One of the main obstructions to the efficient development of new structure modifying therapies for OA is the low sensitivity of change in plain radiographic endpoints that necessitates long-term trials involving a large number of patients to show a significant difference between *placebo* and active-drug treated groups. Biomarkers are promising sensitive tools, but they have to demonstrate specificity for OA pathology and ideally, provide earlier information than JSN measurement by X-ray. The current paucity of (1) biomarker data from human OA clinical trials (summarized in Table II), and (2) data on the role of biochemical markers for monitoring the treatment of OA, can chiefly be ascribed to the absence of therapies with structure modifying activity. Without a structure modifying agent and a practically useful gold standard for monitoring structural change, it is challenging to qualify a biomarker to be "fit for purpose" for monitoring structural modification. Nevertheless, preclinical studies of disease-modifying OA drugs (DMOADs) using biomarkers offer significant promise in terms of early indications of responses to treatment that may translate into the clinic. Experiences with biomarkers in the context of biologic therapies in rheumatoid arthritis offer promise for OA in that short-term changes in serum levels of biomarkers following initiation of therapy may predict long-term clinical and radiographic outcomes[17]. These kinds of data need to be generated in OA trials[18,19].

Another issue regarding treatment monitoring using biochemical markers is the heterogeneity of OA subsets. Results may differ considerably between subsets with differences in pathobiology. OA may be localized in one joint or generalized, hypertrophic with osteophytes and subchondral bone sclerosis or atrophic, slowly or rapidly progressing or showing no progression. Finally, a therapy may act on OA through a variety of mechanisms and pathways. This suggests that a biomarker may need to be specific for the particular molecular target of the therapy in question. For instance, neo-epitopes generated by collagenase activity could be sensitive to collagenase inhibitors but not to drugs acting on proteoglycan turnover. Even if a biomarker reflects the effects of a particular therapy, it may not reflect all the mechanisms of action of the drug, thus underestimating the therapeutic efficacy or missing the toxicity of the particular therapy. This means that the sensitivity of change in a biomarker in a clinical trial may be dependent on the characteristics of the population and the mechanisms of action of the therapy. For these reasons, it would be advantageous to develop a panel of biomarkers and use a wide variety of biomarkers during the preclinical and clinical drug development processes.

Therapeutic trials

OA clinical trials are commonly focused on the investigation of symptoms or structure modification. In general, trial participants fulfill the validated OA criteria of the ACR. In addition, trials of symptom-modifying agents include patients whose disease is likely to respond to treatment, for example those with at least moderate intensity of symptoms (VAS $\geq$50 mm), and those with a flare of symptoms upon withdrawal of their standard therapy (flare trials). These trials are generally limited to 3 or 6-month follow-up.

Trials of structure-modifying agents include patients without end-stage disease and often those with a perceived high risk for structural progression, for example, middle age, overweight women, although these traditional selection criteria are generally poor for identifying risk of knee OA progression[20,21]. Structure-modifying trials generally span 1–3 years. A series of disappointing late-stage terminations of clinical trials investigating new potential DMOADs has led to the call for a new development paradigm for DMOADs, with a stronger focus on the biology of the joint and the redesign of clinical trials to include new and more sensitive biomarkers[22].

One very important issue that is usually ignored in recruiting patients for clinical trials is the phasic nature of OA in some patients resulting in much variability in rates of disease progression. Some patients with knee OA observed over prolonged periods (5 years) may experience periods of progressive structural damage and then relative inactivity[23]. Often non-progressors have been found to predominate in OA clinical trials for disease modification making the detection of therapeutic efficacy very difficult if not impossible. Importantly, there are a few studies that demonstrate the potential ability to identify progressors using biomarkers thereby enabling enrichment of trial populations with disease progressors as opposed to non-progressors, and providing a significant advantage over existing practice[18,19,24,25]. In future, recruitment for clinical

trials should take advantage of such biomarker-directed opportunities to enrich for progressors.

The Outcome Measures in Rheumatoid Arthritis Clinical Trials (OMERACT)–OARSI consensus has recommended a core set of clinical outcome measures that should be included in clinical trials in OA. No OMERACT–OARSI guidelines have yet been developed for the use of non-imaging *in vitro* biomarkers in clinical trials. The core set of clinical items includes pain, physical function, patient global assessment, and for studies of at least 1-year duration, joint imaging[26]. It was subsequently found that successful trial designs must include both absolute and relative change, as well as measures of pain and function as primary domains[27]. Each of these types of clinical outcome measures (pain, physical function, patient or physician global responses), as well as imaging outcomes, can serve as clinical trial endpoints and endpoints for biomarker qualification.

The success of biomarker qualification on a structural modifying endpoint depends critically on the performance and specificity of the endpoint. Although the methodological limitations are well recognized[28], to date, assessment of the inter-bone distance and loss of joint space on a plain radiograph of the hip or knee is the only validated measure of OA progression recommended for use in randomized clinical trials in OA. Unfortunately, the limitations of the traditional clinical trial outcome, JSN, are considerable and have hampered the qualification of biomarkers as well as the registration of DMOADs. To date, no therapeutic agent has met this definition, and it remains unclear how best to identify structural outcomes, whether by radiographs, MRI, biomarkers, or direct visualization using arthroscopy or a combination of these approaches.

General limitations of JSN that hamper the qualification of biomarkers include the following:

- It is an indirect measure of the alterations in articular cartilage;
- It fails to measure a dynamic process;
- Assessment of knee OA is confounded by the presence of meniscal lesions and meniscal extrusion[29];
- Changes in the knee over time are small, and typically occur in only a subset (progressor) of patients (mean estimated annual JSN rate 0.13 ± 0.15 mm/year for knee OA)[30];
- It is poorly reproducible when measured from conventional weight-bearing radiographs of the hip or especially, of the knee in full extension;
- Apparent JSN can occur in the absence of structural changes due to varying degrees of knee flexion;
- Bone marrow and synovial abnormalities may go undetected;
- X-ray features appear only after deterioration of surrounding hard and soft tissues;
- It is poorly correlated with joint function and pain.

A variety of methodological approaches have been proposed to improve the reproducibility of the assessment of the joint space width in randomized controlled trials (RCTs) including semi-flexed views[31], and fluoroscopically assisted protocols; it remains unclear which approaches are preferable, or whether other imaging techniques are preferable and more promising. Among the new imaging techniques, MRI is the most promising and a more sensitive imaging modality for use in the immediate future. MRI allows assessment of cartilage biochemical and biomechanical integrity. It permits quantification of cartilage volume and changes in cartilage contour and can be tailored to assess pathological changes in associated joint structures, and tissues including bone, synovium (inflammation), ligament, menisci and muscle as well as effusions. Correlations between serum biomarkers and MRI data have already been reported for knee OA[32]. Moreover, a combination of MRI and soluble biomarkers have recently been used to improve the ability to identify patients at highest risk of knee OA progression over either modality used independently[33]. MRI has not yet been recommended as a primary endpoint in structural modifying RCTs in OA. A review of its potential, and recommendations regarding the use of MRI for OA clinical trials, is the subject of a companion OARSI FDA white paper.

Although the consensus reached at OMERACT 3 advocated continued study of biological markers of bone and cartilage degradation and repair, none was recommended for inclusion in clinical trials. Nonetheless, in view of the duration required for phase III structure-modifying trials, identification of a surrogate biomarker for use in earlier phase II trials could considerably improve the safety, cost, and efficiency of clinical development programs. Osteoporosis trials provide a good example in which molecular biomarkers are increasingly used as adjunct measures of effect before initiation of multi-year long phase III trials[34].

Biomarker applications in development of therapeutics for OA

Qualified biomarkers of OA have the potential to greatly expand the knowledge gained from preclinical and clinical trials of disease modifying agents. The BIPEDS system classifies potential OA biomarkers into six categories and encompasses the array of biomarkers that could be used for enhancing clinical trials. The most immediate hurdle facing researchers wishing to test a potential DMOAD in humans is the lack of early information in a clinical trial. In order to test a DMOAD, a trial must presently have a lengthy follow-up, enroll many subjects and rely upon an insensitive method of assessment of disease progression. The level of financial investment is daunting, resulting in a negative impact on research and development. In this section, we consider how each category of the BIPEDS classification scheme could be used to improve clinical trial design and outcome. We also address the challenges in developing and qualifying such biomarkers for clinical use.

Burden of disease

Burden of disease biomarkers indicate the extent or severity of disease and could be considered tools for the staging of the disease. They reflect the state of the disease at the time of assessment, but do not necessarily predict a likelihood of progression or change in disease burden. A burden of disease biomarker is typically qualified by comparison to a clinically defined gold standard assessment method. A burden of disease biomarker assessed locally, such as from analysis of SF, would be expected to reflect the disease status in a single joint, while assessment in blood or urine would more likely indicate the extent of the disease in all joints as well as normal physiology. Some molecular biomarkers, such as biomarkers of cartilage turnover, can provide information on the nature and extent of the current active process, but will not indicate the level of tissue damage already accrued or its precise location.

Uses

- To provide a global measure of disease burden from all joints and skeletal and soft tissue components thereof;
- Potentially to discriminate between mono- and poly-articular OA;
- To identify patients with high burden of active disease for inclusion into clinical trials of DMOADs expected to improve later stage disease;
- To help identify patients with low burden of active disease but with no or limited tissue alterations or structural alterations for

inclusion in clinical trials of DMOADs expected to prevent progression of early OA;
- To balance treatment arms in a DMOAD trial for metabolic activity or stage of disease that would not otherwise be obvious from usual randomization criteria;
- To identify where in the body the burden of disease lies and aid in patient stratification, made possible when joint-specific biomarkers or patterns of biomarker expression are discovered.

Challenges

- Requires comparison to gold standard for qualification, but there is no clear gold standard;
- A biomarker may be more sensitive than imaging, picking up a signal of early OA in asymptomatic joints with no obvious imaging changes;
- Uncertainty about what level of burden of disease is the optimal target for a DMOAD as early pathology may differ from more advanced pathology;
- There may be molecular subsets of disease – a biomarker might accurately reflect the burden of disease in one patient but not another;
- The level of a biomarker may change with the disease progression, such that some will be particularly elevated in early phases and others in late phase;
- Due to the complex nature of the joint organ comprised of different tissue types, a true burden of disease measurement might require multiple biomarkers.

Investigative

Investigative biomarkers are those that may not yet have enough evidence accumulated to be assigned to a particular BIPEDS category but nevertheless show sufficient promise to be incorporated in drug research at early stages to determine utility for subsequent use. In general, investigative biomarkers should be included, along with better-qualified biomarkers, in preclinical studies and clinical trials to advance our understanding of the disease and drug and to provide opportunities for biomarker development and qualification.

Uses

- To explore novel biomarkers that could be informative in future preclinical and clinical trials;
- To contribute to biomarker data packages that support qualification of a biomarker or biomarker set for a particular outcome;
- To further understand the pathobiology of OA;
- To further understand the mechanism of action of a DMOAD.

Challenges

- Assays for investigative biomarkers might not be well validated and the data produced might not be robust;
- Conversely, investigative assays could produce highly reproducible, robust data that turn out to lack specificity for the molecular or tissue target;
- Clinical trials are not currently designed for testing of investigative biomarkers, making it difficult to achieve statistical power for biomarker evaluation;
- Biomarkers studied in preclinical disease models might not translate to human OA.

Prognostic

A prognostic biomarker indicates whether a patient's disease is likely to progress and may also indicate how quickly the progression will occur. A prognostic biomarker may also provide an early response to treatment that is prognostic of subsequent, much later, clinical responses. Similarly, a prognostic biomarker could indicate who is at risk for developing symptomatic OA. There is a need for such markers since current clinical trials designed without the aid of biomarkers, often contain a minority of progressors (mean annual risk 6%, range 1–20% based on KL grade)[30]. Predictive biomarkers, used to identify a subset of patients likely to respond to a particular drug, constitute a particularly useful subset of prognostic biomarkers. For instance, a threshold PGE2 level in SF might correlate with the ability of a COX-2 antagonist to be effective in that joint. Prognostic biomarkers include the largest variety of biomarker types, including variant biochemical biomarkers and invariant genetic biomarkers, although the latter may at some point in the future be considered risk factors as opposed to biomarkers.

Uses

- To select subjects likely to progress rapidly ('high-risk' patients by biomarker measurement) to reduce the length of time required to see an effect of a DMOAD in a clinical trial thereby shortening the trial and to improve the chances of observing efficacy;
- To select subjects likely to progress rapidly ('high-risk' patients by biomarker measurement) for purposes of stratification;
- To increase the power of a trial to detect a significant drug effect with a limited number of subjects;
- To select subjects likely to progress rapidly ('high-risk' patients by biomarker measurement) who would benefit most from therapy with structure modifying agents;
- To select subjects for primary prevention trials (screen for at risk for developing OA to demonstrate reduction of incidence);
- To select patients likely to respond to a given drug for inclusion in a clinical trial. For instance, patients with high levels of a matrix metalloproteinase (MMP)-13 specific collagen cleavage product could be selected for inclusion in a trial of an MMP-13 inhibitor;
- As a companion diagnostic, to select likely responders for treatment with a marketed product;
- To provide predictive evidence that disease processes have been beneficially impacted by serving as an early indicator of a later trial outcome or response to therapy; this category of markers would therefore form a specific subset of efficacy of intervention markers described below.

Challenges

- The prognostic effect of a biochemical biomarker must be distinguished from prognostic clinical (weight, injury) or genetic variables that may influence biomarker levels;
- Qualification of a prognostic biomarker would require a large, long and financially daunting prospective trial although this challenge may be overcome with the use of legacy samples from the many excellent existing OA epidemiology studies.

Efficacy of intervention

Biomarkers of efficacy of intervention can range from target engagement and pharmacodynamic assays (which assess whether the compound is hitting the desired target and is having the desired downstream biochemical effects) to strict surrogate endpoints that

indicate the drug is having an impact on the clinical manifestations of the disease. Slowly progressive diseases, such as OA, pose a range of drug development challenges, particularly in phase II dose-finding studies[35]. Target engagement and pharmacodynamic biomarkers are likely to have the earliest impact on drug development of all the BIPEDS biomarkers by influencing decisions on dose selection and advancement of drugs to later phase trials. While a surrogate biomarker would be highly desirable, the path to generation and qualification for a 'characterization level' biomarker is likely to be shorter and provide benefit to programs in the near term at decision points in early preclinical studies and clinical trials. In contrast, qualification of a biomarker as a surrogate biomarker will be a painstaking but highly valuable effort (see the section on Pathways for Biomarker Qualification).

Uses

- To demonstrate that a drug is having the desired immediate downstream biochemical effect;
- To understand the pharmacodynamics of a drug intervention and the relationship between pharmacodynamics and pharmacokinetics;
- To provide a basis for the selection of lead candidates for clinical trials;
- To contribute to the understanding of the pharmacology of candidates;
- To characterize subtypes of disease for which a therapeutic intervention is most appropriate;
- To choose a dose and dose schedule *via ex vivo* and *in vivo* studies;
- To support an efficacy endpoint;
- To support go/no go decisions in advance of preclinical and clinical studies and trials;
- To serve as a surrogate biomarker for delay of structural worsening, reduction of pain, or improvement in function.

Challenges

- For drugs administered intra-articularly to treat a single joint, it may be difficult to monitor efficacy of intervention using systemic biomarker assessments (blood or urine), particularly if other joints are involved in OA;
- Qualification as a surrogate biomarker is difficult in the absence of a gold standard;
- In order for a pharmacodynamic or target engagement biomarker to be informative, it must be specific for the mechanism of action of drug being assessed;
- A biomarker might provide an accurate assessment of target engagement, but might not be related to clinical response.

Diagnostic

A diagnostic biomarker usually indicates whether an individual has the disease or a specific subtype of the disease, but may not reflect disease severity. It also has the potential to identify people at risk for OA based on genetic or other considerations. A biochemical biomarker could be more sensitive than an imaging marker, by detecting the process leading to OA before it is detectable by radiography or other imaging modalities.

Uses

- To select subjects with molecular pre-radiographic OA for primary prevention trials;
- To identify patients with different disease subtypes;
- To identify individuals unlikely to have OA as controls in case–control studies.

Challenges

- The processes in OA vary with time and may vary in nature, although common pathobiology is identifiable. A single diagnostic biomarker may therefore not be informative in all patients;
- Qualification of a diagnostic biomarker requires a gold standard. A biochemical assay could potentially be more sensitive than an imaging gold standard. The qualification would then depend on long term cohorts where the diagnosis can be verified in follow-up;
- Given the insidious onset and slow progression of OA structural changes, it may take many years, patients, trials, and dollars to achieve correlation between a biochemical biomarker and disease. The National Institutes of Health/National Institute of Arthritis and Musculoskeletal and Skin Diseases/National Institute on Aging (NIH/NIAMS/NIA) Public/Private Osteoarthritis Initiative is an example of an effort that could contribute to this end of assessing the correlation between biomarkers and OA.

Safety

There exist important opportunities to use biomarkers to detect pathological changes and cytotoxicity. Safety biomarkers could be used in preclinical and clinical applications to monitor the health of the joint tissues, the whole joint organ, or the skeleton in general. For instance, biomarkers reflecting the synthesis of the main proteins of the joint might provide an index of the "joint-protective" effect of a potential treatment. There are currently no studies exploring specifically this aspect of joint tissue related biomarkers. Potential complications obviously exist with regard to discriminating toxic or pathological effects from beneficial effects in the case of skeletal biomarkers. In the absence of contrary evidence, increased cartilage degradation or decreased synthesis of cartilage based on biomarker data would be considered as potential "red flags" in any treatment regimen. A special circumstance is represented by repair, exemplified by collagen fibrillogenesis, where molecules catalyzing and enhancing this process, may instead prevent fibril formation and hamper repair when produced in relative excess[36,37].

In contrast, there are emerging examples of toxicity monitoring in OA trials with biomarkers of other organ systems. A notable recent example is provided by the pilot trial of Brune 2009[38] wherein N-terminal pro-B-type natriuretic peptide concentrations were shown to predict the risk of cardiovascular adverse events from NSAIDs and glucocorticoid rescue medications in a trial of an MMP inhibitor for OA. We anticipate that this will be a growing area that will enhance the goal of personalized medicine and patient safety. Clearly, a broad spectrum of biomarkers will be necessary for a full safety assessment. The safety biomarkers should also be chosen to demonstrate any effects on other similar structural anatomical elements, e.g., tracheal cartilage, intervertebral disc, and rib, to name a few.

Uses

- To support other more generalized organ system safety indicators in preclinical and clinical trials;
- To monitor for local and systemic adverse effects both early and advanced;
- To set therapeutic dosages that do not impact on physiology.

Challenges

- Understanding what 'safe' ranges are for joint tissue biomarkers;
- Safety biomarkers will need to be qualified against accepted clinical standards, including pain assessments, functional testing, and imaging;
- The safety threshold for each biomarker might be different across individuals.

Summary

With the BIPEDS scheme, the biomarkers that are likely to have the earliest beneficial impact on clinical trials fall into two general categories. The first are those that will allow us to target trials to subjects that are likely to either respond and/or progress within a short time frame. For instance, a patient population with high levels of an MMP-13 cleavage product, but without end-stage cartilage loss, would be ideal for a trial with an MMP-13 inhibitor. The second category of biomarkers includes those that provide early feedback for preclinical decision-making and for trial organizers that a drug is having the desired biochemical effect. This category of biomarkers is particularly desirable in chronic diseases, such as OA, where clinical outcomes may take years to present[39]. In some cases, the biomarker might be sufficiently qualified that the researchers have confidence in using it to justify advancement to phase 2 trials and to determine a dosing schedule. These two categories reduce the burden and risk of early stage trials by delivering essential early information, making OA a more manageable and therefore a more attractive target for drug developers.

Qualification of known OA biomarkers

Biomarker validation vs qualification

The validation and qualification of a biomarker are two essential processes involved with assessing the level of confidence in a specific biomarker. For scientists who develop new biomarkers, validation means assessing all technical aspects of a specific assay to address the following question: "Under what conditions can we trust this assay and what it tells us?" Conversely, qualification consists of assessing the clinical value of a specific assay and answers the question: "Is this marker useful for learning more about the disease pathobiology or the efficacy of the treatment tested?" Currently there are no biomarkers that have been formally qualified and cleared by the FDA for OA-related outcomes.

Validation

Standard laboratory-based biomarker assays are typically quantitative in nature. Analytical validation of a specific quantitative assay is usually established by five tests: intra- and inter-assay variation, dilution recovery, determination of the detection and quantification limits and spiking recovery, although this latter test is often not performed, especially when standards are synthetic peptides. In addition, the stability of the biomarker (with storage and freeze–thaws) and key reagents should be established to determine the parameters and stringency of storage necessary to assure reliability of measurements. The exact assay validation process will depend on the intended use of the assay, with assays for surrogate markers undergoing more rigorous validation than assays for exploratory endpoints. Not all biomarker assays are "definitive" quantitative measurements. Some biomarker assays generate "relative" results, due to the nature of the reference materials or sample matrix[39]. One example would be genomic data generated from microarray analysis of RNA. For these sorts of relative quantitative assays it is appropriate to place greater emphasis on relative and temporal changes in biomarker concentrations rather than the absolute concentrations. Another example would be an Enzyme Linked Immunosorbent Assay (ELISA) that uses a crude extract as standard and for which biomarker results are reported in arbitrary units. For these sorts of assays, the availability and sharing of a common international standard for normalization is highly desirable.

In contrast to quantitative biomarkers, qualitative biomarkers are discrete (discontinuous) and reported in either ordinal or nominal formats. An example of a qualitative assay would be a method to detect the presence of a single nucleotide polymorphism or gene mutation in a sample of DNA[39]. Assay validation for a qualitative assay is more limited than for a quantitative assay since concepts such as precision and dilutional recovery are not relevant[39]. Just as important as pre-study method validation is in-study validation (run acceptance), appropriate control samples and run/sample acceptance criteria should be incorporated into the analytical method for each assay to ensure quality data.

The specificity of the antibody(ies) used in the immunoassay is a very important factor, although this has not been carefully investigated for most biomarkers. Indeed, recognition and cross-reactivity experiments are usually performed using synthetic peptides or *in vitro* generated degradation fragments, which are probably of a different structure than the native immunoreactive forms detected in biological fluids. To date, published results of the structure of the immunoreactive form have only been partly determined for one OA-related biomarker, TIINE, which involves type II collagen cleavage by collagenase[40]. This information can be difficult to generate because the concentrations of the analytes found in serum and/or urine are usually very low and their determination requires complex analysis. This aspect of the biomarker validation process is however of critical importance for correct interpretation of biomarker results[41].

Other critical information is that which concerns the tissue and site(s) of origin of the biomarker. Incorrect assumptions regarding tissues of origin have led to misinterpretation of biomarker data. Mistakes of this kind may in part account for lack of correlation between clinical and biomarker outcomes.

The Standards for Reporting of Diagnostic Accuracy (STARD) statement[42] has provided a checklist of specific information about biomarker measurement, and the subjects tested, that should be provided in any study validating a biomarker regardless of its intended use. These include the following specific requirements (summarized by Felson *et al.*[43]): to blind those measuring the biomarker as to disease status (in a study of prognosis, this would mean blinding to progression status); to define the rationale for and selection of cutoffs differentiating 'normal' from 'abnormal' biomarker levels; and importantly, to note the source of subjects in a study, reporting whether they were selected because of their biomarker status or unique clinical findings.

Qualification

Previously, the process of linking a surrogate endpoint to a clinical endpoint has been referred to as *validation* or *evaluation*[2]. However the use of the term validation has now been confined to the assessment of the performance characteristics of a biomarker assay, while linking a biomarker to a clinical endpoint is referred to as qualification[44]. The use of biomarkers as surrogate endpoints in a clinical trial requires the qualification of the biomarker for specific clinical endpoints (such as pain, loss of mobility, or need for a total joint replacement) in a specific population with a particular disease state and/or in the context of a specific class of therapeutic intervention (adapted from Ref.[2]). Loss of mobility and total joint replacement occur only after a very long time in most patients (Fig. 2), and vary by nation and region due to differences in patient

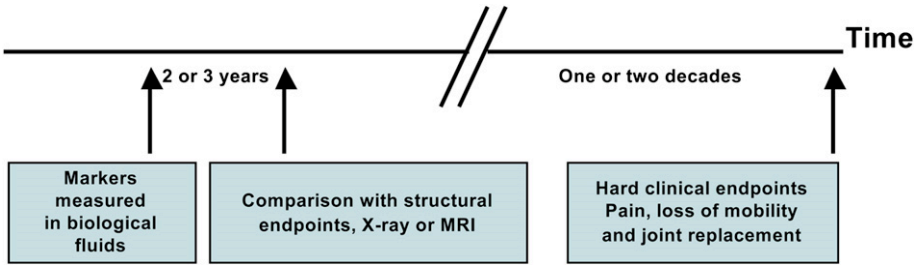

Fig. 2. Conceptual framework for biomarker qualification.

expectations and health-care policies. Consequently, to reduce the time needed to qualify a biomarker, studies use structural endpoints derived from X-ray and more recently from MRI.

For drug development, 'efficacy of intervention' ('E' of BIPEDS) biomarkers are sought. In theory, the optimal efficacy of intervention biomarker would be a perfect clinical outcome surrogate. In the case of the perfect surrogate:

- The effect of the intervention on the surrogate predicts the effect on the clinical outcome;
- The surrogate is in the only causal pathway of the disease process;
- The intervention's entire effect on the true clinical outcome is mediated through its effect on the surrogate;
- The surrogate fully captures the treatment effect.

In reality, it is likely that few if any biomarkers will ultimately achieve surrogate status let alone perfect surrogate status. Several different methods have been proposed for quantifying the strength of the surrogate[45]. This method provides a quantitative score for a biomarker. Wagner *et al.* categorize the strength of a surrogate based on four levels[1]: Exploratory, Demonstration, Characterization and Surrogate biomarkers (summarized in Pathways for Biomarker Qualification-Levels of qualification of biomarkers for drug development use). This mark of the strengths of surrogacy is used in this document.

As the Wagner classification implies, robust linkage of a biomarker with a clinical endpoint is not essential in early clinical development when the goal is confirmation of pharmacologic activity or optimization of dose regimens[2]. As stated by the Biomarkers Definitions Working Group in 2001: "Reliance on a biomarker early in the drug development process, for instance for candidate selection, entails the hazard that failure of a biomarker may lead to the elimination of potentially effective agents. On the other hand, substantial evidence that a biomarker will predict clinical benefit or risk is needed when use of the biomarker as a surrogate endpoint is proposed as the basis for regulatory approval. In this case, erroneous decisions based on invalid surrogate endpoints may have broad public health consequences"[2].

Qualification endpoints for OA biomarkers

As described above, there are many possible qualifying endpoints for an OA-related biomarker including signs (inflammation) and symptoms (pain), structure or functional outcomes in OA. A biomarker could be qualified for different stages of OA such molecular, pre-radiographic, or radiographic stages of OA. In theory, a biomarker could be qualified for an outcome in a specific joint if the biological findings supported such specificity. We are only beginning to appreciate cartilage matrix biochemistry in this level of detail as exemplified by the differences in matrix biochemistry and response to injury of ankle vs knee cartilage[46]. In practice, the qualification process is an empiric and gradual one, correlating changes in a biomarker with change in state of a joint(s). To date the process of biomarker qualification has tended to relate a biomarker to a specific tissue component of the whole joint organ such as bone, cartilage or synovial tissue.

Sources of biomarker variability

Biochemical markers in blood and urine provide information on systemic skeletal tissue turnover[47] and are not necessarily specific for the alterations occurring in the signal joint[48]. For example, it has been shown that degenerative disease of the knees, hips, hands and lumbar discs contributed independently and additively to urinary CTX-II levels illustrating the total body contribution to systemic levels[48,49]. The potential contribution of intervertebral discs is of particular relevance because disc degeneration is common in aging. Systemic biomarker levels cannot be assumed to reflect total body OA burden based on radiographic damage or cartilage volume estimated by quantitative MRI because these factors alone do not fully account for the differential contribution of soluble biomarkers from different joints[50]. Serum and urinary levels of most markers also vary with gender, age, menopausal status, ethnicity, and OA risk factors such as body mass index. Specific examples include the effects of gender, ethnicity and age on COMP[51,52] and the effect of BMI on CPII[53].

Biomarker levels can also be influenced by other skeletal alterations, such as osteoporosis or by concomitant medications. It is likely that differential processing by the liver or kidneys occurs before systemic biomarkers reach a steady state in body fluids, and this metabolism may not occur reproducibly in all patients, particularly in the presence of systemic disease[54,55]. Measurements in urine require correction by creatinine to adjust for variability related to hydration and renal status. One of the main factors affecting pre-analytical variability is diurnal change. The magnitude of diurnal-related changes in the concentration of seven markers (serum HA, COMP, KS-5D4, TGFß1, CPII, and urinary CTX-II and C2C) has been shown to be greater than the analytical inter- and intra-assay related variability, indicating that the diurnal-related variation was predominantly a result of biological variability rather than assay variability[56,57]. For the biomarkers found to be significantly associated with radiographic severity (serum COMP, KS-5D4, C2C, C1, 2C, and urinary CTX-II), the biomarker concentrations at the T2 or T3 time points showed the most consistent correlation with radiographic knee OA when the sampling was performed during the afternoon (T2) and the early evening (T3). A study on serum PIIANP and serum HELIX-II concluded that concentrations of these two markers increased significantly from T0 (before arising from bed) to T1 (1 h after arising)[58]. It was also shown that serum CTX-I and serum HA markers levels are markedly influenced by food intake which also does increase intra-subject variability[59]. These and other data (prior biomarkers white paper http://www.oarsi.org/index2.cfm?section=OARSI_Initiatives&content=Biomarkers) provide a rationale for standardization of sample collection procedures for OA clinical trials.

Table Ia
Sample sizes to achieve 80% and 90% power to detect assumed differences between two parallel groups

Number of patients/group required for 90% power*	Number of patients/group required for 80% power*	Underlying treatment difference to detect	SD (between-subject)	Effect size (difference/SD)
15	12	1.84	1.5	1.23
30	23	1.28	1.5	0.85
60	45	0.90	1.5	0.60
100	76	0.69	1.5	0.46
15	12	2.45	2.0	1.23
30	23	1.70	2.0	0.85
60	46	1.19	2.0	0.60
100	76	0.92	2.0	0.46

* Based on 2-sample *t*-test (2-sided, alpha = 0.05) for difference between groups with null hypothesis that treatment difference = 0.

Table Ib
Sample sizes to achieve 80% and 90% power to detect assumed underlying ratio of treatment effect between two parallel groups

Number of patients/group required for 90% power*	Number of patients/group required for 80% power*	Underlying mean ratio between groups to detect	CV (SD/Mean) in original scale
114	85	0.65	1.3
165	124	0.70	1.3
253	189	0.75	1.3
418	313	0.80	1.3
80	60	0.65	1.0
116	87	0.70	1.0
177	133	0.75	1.0
294	220	0.80	1.0
47	35	0.65	0.7
67	51	0.70	0.7
103	77	0.75	0.7
170	127	0.80	0.7

* Based on 2-sample *t*-test (2-sided, alpha = 0.05) for ratio of treatment effect between groups with null hypothesis that ratio = 1, and common CV.

Limited research has been done to analyze the effects of diet and dietary supplements on biomarker levels. As described above, serum hyaluronan showed significant variation related to food consumption in healthy volunteers[60] and circadian variation of CTX-I was found to be reduced by fasting[61], suggesting that fasting can have a significant effect on the circadian variation of markers of bone resorption. Gordon *et al.* 2008[57] showed that urinary CTX-II was not affected by food consumption or physical activity and may offer an advantage in the context of clinical trials incorporating morning body fluid sampling. Clearly, pre-analytical factors contribute to intra- and inter-assay variability of biochemical markers levels and consequently need to be investigated and controlled as tightly as possible. Taken together, these studies point to the need for standardization of sample collection within a trial to minimize non-treatment related variation. Recommended methods of sample acquisition, handling and storage are provided in Appendix A.

Summary of OA biomarkers

Biochemical markers of bone and cartilage turnover are presently the most advanced with respect to matrix remodeling[35]. Several excellent recent reviews provide a summary of biomarkers in general and several summarize the data to support classification into one or more of the particular BIPEDS categories[35,62–68]. In this section, we focus on "soluble biomarkers" studied to date in human OA clinical trials, and not genetic/genomic or imaging biomarkers or biomarkers studied in the absence of an intervention. Although a few soluble biomarkers are quantified by mass spectroscopy approaches, most are currently assessed by immunoassay. Tables Ia and b provide a look at the sample sizes required for biomarker studies. Table II presents data for all known peer-reviewed publications to date of pharmacologic OA trials with either structural or clinical trial outcomes that included published biomarker analyses, and an indication of the success or failure of the trial for the primary and biomarker outcomes. The reported assay coefficients of variation (CVs) are provided when they were reported, which may be helpful for assessing needed sample sizes for future studies. In addition, the reported concentrations [and standard deviations (SDs) when available] before and after treatment are listed to begin to provide a benchmark for comparison across studies, albeit limited at the present time. Table III provides a summary of the known tissue sources and current BIPED classification for many of the most common and best-qualified OA-related biomarkers.

Statistical issues and sample size estimates for biomarker studies

Table Ia provides a look at the sample sizes required if the between-subject variability (SD) increases from 1.5 to 2.0 or the power desired changes from 90% to 80% given the same treatment differences. Biomarkers are often not normally distributed due to the potential for a high incidence of values below the limit of quantification. To normalize the distribution the values are usually log-transformed and Table Ib provides some sample size estimates when the biomarker is expressed as ratio or percent differences and analyzed on the log scale. In the papers summarized in Table II (below) and others (reviewed by van Spil *et al.*[62]), many biomarkers, such as those measured by radiography, e.g., JSN, were explored for their ability to predict the progression of OA or to change concurrently with OA. However, results were generally not consistent across the studies for multiple reasons: large variability of the assays, unpredictable variability of the biomarkers, under-powering of the study, or slow progression of OA were the most often cited reasons for non-significant or inconsistent findings.

The under-powering of the studies was generally due to the fact that the biomarkers were regarded as exploratory endpoints or the basis for subgroup analyses, hence, were not powered sufficiently at the planning stage. Some studies were designed as pilot studies, which relied on detecting statistical significance instead of meaningful difference as a measure of the importance of the biomarker. These types of studies serve the purpose of hypothesis generation; however, as experiences with the biomarkers accumulate, an organized effort is necessary to define the following elements so that standards can be established for future studies against which to benchmark:

1. Identify clinically meaningful differences between two active treatments or between an active treatment and *placebo* with respect to validated clinical endpoints.
2. Define meaningful correlation between the biomarkers and the clinical endpoints, i.e., how large the magnitude of the correlation has to be.
3. Define the meaningful difference between two active treatments or between an active treatment and *placebo* with respect to the biomarker once it is demonstrated to correlate with the clinical endpoints.

Consideration of these three elements is important to ensure sufficient numbers of subjects in the study, and hence, sufficient power to detect the underlying meaningful difference based on biomarkers. They also prevent statistical significance being reached only because of the large sample size while meaningful difference is not observed. A critical component for the success of these aims will be the establishment of clinical meaningful endpoints related

Table II
Summary of biomarker data generated in OA clinical trials to date

Trial − intervention (duration)	Study ref	Patient numbers	Sample type		CV% (biomarker units)	Treatment		*Placebo*		Comments	Assay/cut-points
						Pre	Post	Pre	Post		
Ibuprofen 2400 mg qd for knee pain × 4−6 w (E)	Gineyts 2004[70]	Human 156/45	NF morning urine	uCTX-II (E)	<10% (ng/mmol Cr)	225 ± 2.16	229 ± 2.06	226 ± 1.88	265 ± 2.06	Patients with high levels were responsive to therapy	C-ELISA Cartilaps − Christgau 2001[71]
				uGlc-Gal-PYD (⌀E)	<11% (nmol/mmol Cr)	6.0 ± 1.5	6.2 ± 1.5	5.7 ± 1.4	6.3 ± 1.4		HPLC Gineyts 2001[72]
Glucosamine sulphate 1500 mg/d × 3 y (E)	Christgau 2004[18]	Human 106/106 [$n = 61$ above 1 SD cut-off]	NF second morning void urine	uCTX-II/Cr (E)	8.4% (ng/mmol)	All: 216.5 ± 9 at baseline	All: Loss of joint space (0.06 mm) over 3 years	All: 219.5 ± 9 at baseline	Loss of joint space (0.31 mm) over 3 years		C-ELISA Cartilaps with mAb F46 *per* Christgau 2001[71]; High turnover group defined as baseline ≥261.3 (i.e., ≥1 SD above mean of 169.1 ± 92.3 in reference population)
						High turnover group mean 413 ± 28	High turnover group mean 336 ± 26 [Gain of joint space (0.083 mm; $P = 0.07$) over 3 years; Global WOMAC decreased 24.5%]	High turnover group mean 375 ± 33	High turnover group mean 411 ± 252 [Loss of joint space (0.44 mm) over 3 years; Global WOMAC decreased 4.5%]	Promising approach; larger sample size (>61) of high turnover patients likely needed for statistical significance	
						For high turnover group: change in uCTX-II from baseline to 12 m correlated with average joint space width loss over 3 years ($r = 0.43$; $P < 0.05$)		For high turnover group: change in uCTX-II from baseline to 12 m correlated with average joint space width loss over 3 years ($r = 0.27$; $P < 0.03$)			
Salmon calcitonin (oral) 0.5−1.0 mg/d × 48d for knee OA patients with positive knee bone scans (E)	Manicourt 2006[73]	Human 27/14	F serum & second morning void between 9 and 11 AM [all median values reported show baseline and day 84 values]	uCTX-II/Cr (E)	<6% (ng/mM)	395	290	368	370		ELISA Cartilaps − Nordic Bioscience (Herlev, Denmark)
				sHA (E)	<6% (μg/ml)	61	48	60	69		ELISA method of Manicourt 1999
				sC2C (E)	<6% (ng/ml	30	23	27	30		ELISA-IBEX (Montreal, CA)
				uNTX-I/Cr (⌀E)	<4% (BCE mM/mM Cr)	48	43	57	56		ELISA-OSTEX Intl (Seattle, WA)
				sOC (⌀E)	<9% (ng/ml)	12	16	18	16		ELISA BioSource, (Nivelles, Belgium)
				sMMP-1 (⌀E)	<8% (ng/ml)	8	9	8	9		ELISA-GE Healthcare (Little Chalfont, UK)
				sMMP-3 (E)	<5% (ng/ml)	20	19	19	24		ELISA-GE Healthcare (Little Chalfont, UK)
				sMMP-8 (⌀E)	<5%	5	5	4	4		ELISA-GE Healthcare (Little Chalfont, UK)
				sMMP-13 (E)	<5% (pg/ml)	100	64	52	76		ELISA-GE Healthcare (Little Chalfont, UK)
				sTIMP-1 (⌀E)	<5%	173	184	151	149		ELISA-GE Healthcare (Little Chalfont, UK)
				TIMP-2 (⌀E)	<5%	11	11	18	14		ELISA-GE Healthcare (Little Chalfont, UK)
BRISK study: Risedronate 5 mg/d or 15 mg/d × 12 m (⌀E for JSN, E for WOMAC)	Spector 2005[74]	Human	F early morning urine and serum	uCTX-II	(ng/mmol Cr) (E)	340.1 (24.0)	−22.8 ± 5.35% (15 mg)	312.5 (19.9)	±14.5 ± 5.4% (15 mg)		ELISA Cartilaps − Nordic Bioscience (Herlev, Denmark)
				uNTX-1	(nmol/mmol Cr) (E)	38.6 (2.2)	−32.9 ± 4.92% (15 mg)	40.3 (2.8)	±17.2 ± 4.9% (15 mg)		ELISA Osteomark-OrthoClinical Diagnostics (High Wycombe, Bucks, UK)
				sAlk Phos (bone specific)	NR	NR	−29.1 ± 2.6% (15 mg)	NR	−2.7 ± 2.5% (15 mg)		ELISA Ostase-Beckman-Coulter (San Diego, USA)

(*continued on next page*)

Table II (*continued*)

Trial − intervention (duration)	Study ref	Patient numbers	Sample type		CV% (biomarker units)	Treatment		*Placebo*		Comments	Assay/cut-points
						Pre	Post	Pre	Post		
KOSTAR study: Risedronate 5 mg/d, 15 mg/d, or 35−50 mg/w × 24 m (⌀E)	Bingham 2006[75]	Human 1861/622 (from two cohorts)	F second morning void	uCTX-II/Cr (E)	<10% (ng/nmole Cr)	297.16−360.70 ± 14.87−12.06	(−) 17.9−19.6% (decrease at 24 m)	296.47−376.72 ± 17.09−13.72	(±) 10.1−26.3% (increase at 24 m)	Treatment effect on biomarkers but not X-ray progression	ELISA Cartilaps − Nordic Bioscience (Herlev, Denmark)
				uNTX-I/Cr (E)	<10% (nmol BCE/nmole Cr)	38.80−49.91 ± 1.07−2.10	(−) 39.2−41.7% (decrease at 24 m)	37.48−49.43 ± 1.96−1.36	(±) 3.0−7.3% (increase at 24 m)		Osteomark-Ortho Clinical Diagnostics (Rochester, NY)
	Garnero 2008[19]	Human 1885 (subset of two cohorts)	F early morning urine	uCTX-II/Cr	<10% (ng/mmol Cr)	(−) 39.9 ± 3.0 (treatment effect of biomarker shown for all doses by a mean decrease from baseline to 6 m, $P < 0.05$ compared to baseline and *placebo* [baseline CTX-II and change from baseline to 6 m associated with radiographic progression at 24 m as absolute change or for progression defined as JSN ≥0.6 mm)				Early biomarker endpoint to predict long-term progressor/non-progressor status	ELISA Cartilaps − Nordic Bioscience (Herlev, Denmark); high turnover defined as >150 ng/mmol Cr
				uNTX-I/Cr	<10% (nmol/mmol Cr)	(values reported graphically)					ELISA Osteomark-OrthoClinical Diagnostics (Rochester, NY)
Chondroitin sulfate 500 mg bid × 24 w (E)	Mazieres 2007[76]	Human 139/140	F serum between 7:30 and 10 AM & second morning void urine	uCTX-II/Cr (⌀E)	<15% (ng/mmol Cr)	389 ± 247	406 ± 302	375 ± 238	376 ± 214		ELISA Cartilaps − Nordic Bioscience (Herlev, Denmark)
				sHA (⌀E)	<9% (ng/ml)	86 ± 71	100 ± 86	79 ± 61	89 ± 78		ELISA HA-Corgenix, (CA, USA)
				sCTX-I (⌀E)	<10% (ng/ml)	0.44 ± 0.27	0.44 ± 0.23	0.39 ± 0.22	0.40 ± 0.22		Automated analyzer Elecsys 2010-Roche (Mannheim, Germany)
Acute activity (⌀E)	Andersson 2006[77]	Human 29/29	NF serum twice with 1 h apart (after activity and after rest)	sCOMP (E)	NR (U/L)	11.03	(±) 1.3 [median change score after 1-h activity]	11.29	(−) 0.6 [median change score after 1-h rest]		S-ELISA COMP-Anamar (Uppsala, Sweden)
ADAPT: Exercise and/or diet × 18 m (E)	Chua 2008[78]	Human 138/53 (193 studied)	F between 7 and 9 AM	sCOMP (⌀E) (baseline values reported graphically)	NR (U/L)	10.80 ± 0.49 (at 6 m diet & exercise)	11.81 ± 0.46 (at 18 m diet & exercise)	11.75 ± 0.45 (at 6 m)	11.72 ± 0.42 (at 18 m)		S-ELISA COMP-Anamar (Uppsala, Sweden)
				sHA (⌀E) (baseline values reported graphically)	NR (ng/ml)	42.28 ± 3.79 (at 6 m diet & exercise)	45.33 ± 3.63 (at 18 m diet & exercise)	40.46 ± 3.58 (at 6 m)	47.67 ± 3.35 (at 18 m)		Immunosorbent assay Li 1989[79]
				sKS (⌀E) (baseline values reported graphically)	NR (ng/ml)	310.22 ± 7.62 (at 6 m diet & exercise)	310.93 ± 7.32 (at 18 m diet & exercise)	308.67 ± 7.17 (at 6 m)	286.66 ± 6.71 (at 18 m)		ELISA with mAb 5-D-4 *per* Method of Thonar 1985[80]
				sTGF-ß1 (⌀E) (baseline values reported graphically)	NR (ng/ml)	38.89 ± 1.14 (at 6 m diet & exercise)	39.06 ± 1.07 (at 18 m diet & exercise)	40.93 ± 1.04 (at 6 m)	39.41 ± 0.98 (at 18 m)		ELISA-Quantikine R&D (Minn, USA)
Glucosamine sulfate discontinuation × 6 m (OE)	Cibere 2005[81]	Human 63-65/63-65	NF urine or serum	sC2C (⌀E)	5.5% (pmol/ml)	Mean change: −3.5 ± 28.5		Mean change: 3.7 ± 23.6			ELISA-IBEX *per* Method of Poole 2004[82]
				uC2C/Cr (⌀E)	NR (pmol/μmol Cr)	Mean change: −6.9 ± 54.1		Mean change: −0.6 ± 11.8			
				sC1,2C (⌀E)	NR (pmol/ml)	Mean change: 8.5 ± 64.2		Mean change: 9.5 ± 80.0			ELISA-IBEX with pAb *per* Method of Billinghurst 1997 using pAb[83]
				uC1,2C/Cr (⌀E)	NR (pmol/μmol Cr	Mean change: −20.2 ± 144.9		Mean change: 0.4 ± 17.1			

MMP inhibitor, variable dosage × 3 w prior to knee replacement (⌀E)	Leff *et al.* 2003[84]	Human 22/11	F articular cartilage at arthroplasty	cCPII (⌀E)	NR (ng/μg DNA)		2.35 (1.07–8.34) [median and range for max dose]		1.42 (0.33–3.86) [median and range]		Method of Nelson 1998[85]
				cC1, 2C (⌀E)	NR (pmol/μg DNA)		29.8 (9.3–134) [median and range for max dose]		25.7 (5.1–45.7) [median and range]		ELISA Billinghurst 1997[83]
				cCol2-3/4 m (⌀E)	NR-5% *per* RP (nmol/μg DNA)		0.21 (0.09–0.64) [median and range for max dose]		0.17 (0.06–1.25) [median and range]		ELISA Hollander 1994[86]
				cCS-846 (E)	NR (μg/μg DNA)		0.78 (0.18–5.62) [median and range for max dose]		0.35 (0.24–2.86) [median and range]		ELISA Rizkalla 1992[87]
				cKS (⌀E)	NR (μg/μg DNA)		66.0 (28.7–258) [median and range for max dose]		90.7 (24.1–177) [median and range]		ELISA Rizkalla 1992[87]
Doxycycline × 30 m (E increased)	Lohmander 2005[88]	Human 60/60 subset of main study [21/39 progressors; 30/30 non-progressors]	NF plasma and second morning void urine	pMMP-3 (E)	19.4% ng/ml	Contrary to *placebo* group – every SD increase in mean MMP-3 was associated with lower rate of JSN (−0.11 mm)		Baseline upper tertile (11.86–41.00) more likely to progress than lower tertile (<6.43); for every SD (4.6 ng/ml) increase in mean MMP-3 – JSN increased 0.18 mm ($P = 0.001$); increase over time in MMP-3 associated with concurrent JSN			ELISA Method of Walakovits 1992[89]
	Mazzuca 2006[90]			uCTX-II (⌀E)	27.7% (ng/mg Cr)	No association between uCTX-II and JSN progression; Mean values 63.5–66.8; change from baseline to 30 m (mean ± SEM): doxycycline group = 1.14 ± 1.93; *placebo* group 0.53 ± 1.75; progressors −0.03 ± 1.88; non-progressors 1.69 ± 1.78				Study designed for 80% power to detect 35% difference between highest and lowest tertiles of baseline uCTX-II in frequency of JSN progression	ELISA with mAb 2B4 and plates coated with matrilysin digested type II collagen (different antibody from Cartilaps assay)
	Mazzuca 2006[91]			sC2C, sCPII, sCS846, sC1, 2C (⌀E)	9.7%, 6.4%, 11.5%, 10% respectively (all ng/ml)	1SD change in CS846 associated with concurrent JSN; no biomarker was significant predictor of JSN progression					ELISAs-IBEX (Montreal, CA)
	Le Graverand 2006[69]			uTIINE/Cr (E)	Up to 12.3% (ng/mM Cr)	1SD (64–68 ng/mM Cr) increase in baseline uTIINE associated with lower rate of JSN (not significant in either group)					Two dimensional LC-MS/MS
	Otterness 2007[92]	Human 51/69 (subset)		uTIINE/uCr (⌀E – increased with treatment)	8% inter-assay, 30 ± 17% within patient (ng/mmole)	109 ± 68	144 ± 81 ng/mmole	125 ± 62 (overall baseline mean)	115 ± 49		Two dimensional LC-MS/MS; increase due to treatment due possibly to decreased fragment metabolism or change in clearance

(*continued on next page*)

Table II (*continued*)

Trial − intervention (duration)	Study ref	Patient numbers	Sample type		CV% (biomarker units)	Treatment		*Placebo*		Comments	Assay/cut-points
						Pre	Post	Pre	Post		
Chondroitin sulfate × 1 y (E)	Uebelhart 1998[93]	Human 21/19 (21/20 for sOC but 23/23 overall)	NF serum & second morning void urine	sKS (E)	NR (ng/ml)	449 ± 119	420 ± 100	386 ± 133	403 ± 142		C-ELISA with mAb 1/20/5-D-4 Method of Thonar 1984[80]
				uPYD/Cr (E)	NR (nmol/L/mmol	56 ± 25	53 ± 19	59 ± 40	70 ± 30		RP-HPLC Uebelhart 1990[94]
				uDPD/Cr (E)	NR (nmol/L/mmol)	7.7 ± 3.0	7.7 ± 2.3	8.5 ± 5.4	11.7 ± 8.1		RP-HPLC Uebelhart 1990[94]
				sOC (E)	NR (ng/ml)	16 ± 7	16 ± 6	21 ±13	26 ±29		RIA, ELSA-OSTEO CisBiointernational, Gif/Yvette, France
Intra-articular hyaluronan × 5 weekly injections	Hasegawa 2008[95]	Human 28 (all treated)	SF time not specified	sfKS (E)	NR (nmol/ml)	61.2 ± 35.8	52.8 ± 25.3			ND − no vehicle control	HPLC Method of Yamada 2000[96]
				sfC6S (E)	NR (nmol/ml)	19.1 ± 6.7	17.8 ± 6.1			ND − no vehicle control	HPLC Method of
				sfC4S (E)	NR (μg/ml)	6.1 ± 3.7	5.2 ± 2.9			ND − no vehicle control	Yoshida 1989[97] & Shinmei 1992[98]
				sfTenascin-C (⌀E)	NR (ng/ml)	37.4 ± 59.1	39.0 ± 58.1			ND − no vehicle control	ELISA − IBA (Gunma, Japan)
Supplemental soy protein 40 g/d × 3 m (E)	Arjmandi 2004[99]	Human 44/44	F serum	YKL-40 (E in men)	6.8% (ng/ml)	All: 89.9 ± 7.6; men: 91.0 ± 10.3; women: 93.4 ± 11.4	Change (decrease) in YKL-40 from baseline to 3 months only significant in men (compared to *placebo*)	All: 67.8 ± 6.3 men: 71.3 ± 10.2; women: 64.6 ± 7.8	Increased in all groups	Clinical and biomarker effects in men, not women	S-ELISA-Metra Biosystems (Mountain View, CA)
				IGF-1 (E in men)	7.6% (ng/ml)	All: 113.3 ± 8.2; men: 125.0 ± 10.7; women: 97.6 ± 12.9	Change (increase) in IGF-1 from baseline to 3 months only significant overall and in men, not women (compared to *placebo*)	All: 135.6 ± 10.6; men: 158.7 ± 14.8; women: 107.9 ± 9.3	Increased in all groups	Clinical and biomarker effects in men, not women	Radioimmunoassay-Diagnostic Systems Labs Inc (Webster, TX)

F = fasting; NF = non-fasting; h = hour; d = day; bid = twice daily; w = week; m = month; y = year; pAb = polyclonal antibody; LC-MS/MS = liquid chromatography followed by low then high energy mass spectroscopy; RP = reversed-phase; HPLC = high pressure liquid chromatography; s = serum; p = plasma; u = urine; c = cartilage; NR = not reported; S-ELISA = sandwich-ELISA; C-ELISA = competitive (inhibition)-ELISA.
(E) means showed evidence for change with intervention and (⌀E) means no evidence of statistical difference of biomarker with intervention (did not meet efficacy of intervention criteria); when the trial produced disease modification an (E) is listed in the first column.
CTX-II = carboxy-telopeptide of type II collagen; COMP = cartilage oligomeric matrix protein; HA = hyaluronan; C2C = collagenase-generated neoepitope of type II collagen; C1, 2C = collagenase-generated neoepitope of types I and II collagens; TIINE (mAbs 9A4/5109) = type II collagen neoepitope; CPII/PIICP = type II collagen carboxy-propeptide; Col2-3/4m = type II collagen denaturation epitope; KS = keratan sulfate; CS-846 = aggrecan chondroitin sulfate 846 epitope; NTX-I = N-telopeptide of type I collagen; CTX-I = carboxy-telopeptide of type I collagen; PYD = pyridinoline; DPD = deoxy-pyridinoline; OC = osteocalcin; Glc-Gal-PYD = glucosyl-galactosyl-pyridinoline; C4S and C6S = chondroitin-4 and -6
sulfate; Tenascin-C; YKL-40 = human cartilage glycoprotein 39; IGF-1 = insulin growth factor-1; MMP = metalloproteinases: -1 (collagenase-1), -3 (stromelysin), -8 (neutrophil collagenase), -13 (collagenase-3); TIMP = tissue inhibitor of metalloproteinase: -1 or -2; TGF-ß1 = transforming growth factor-ß1; Alk Phos = alkaline phosphatase.

Table III
Recommended panel of informative commercially available OA-related biomarkers qualified for various OA outcomes*

Biomarker	Process† (preliminary)	Tissues of origin (see discussion below Table)	BIPEDS classifications†	Surrogacy based on human clinical trials (preliminary)	ELISA assay type
Urinary CTX-II	Type II collagen degradation, osteophyte burden of disease	Mineralized and non-mineralized cartilage, growth plate cartilage, bone	Knee: BPED Hip: BPD	Characterization: changed significantly in three pharmacologic trials that met primary clinical endpoints[18,70,73]	Competitive-inhibition for human urinary samples and sandwich for animal serum samples
Serum COMP	Cartilage degeneration	Cartilage > tendon, meniscus, synovium, osteoblasts, arterial wall	Knee: BPD Hip: BPD	Exploration: not used to date in published pharmacologic trial	Competitive-inhibition & sandwich
Serum HA	Osteophyte burden of disease and synovitis	Cartilage, meniscus, synovium and ubiquitous in body	Knee: BPED Hip: P	Demonstration: changed significantly in one pharmacologic trial that met primary clinical endpoints[73]	Sandwich protein binding assay
Serum and urine C1, 2C	Types I and II collagen degradation	Cartilage, bone, synovium, etc.	Knee: D(u) Hip: none	Exploration: non-significant change in one pharmacologic trial that met primary clinical endpoint[73,91]	Competitive-inhibition
Serum and urine C2C	Type II collagen degradation	Cartilage	Knee: E(s), D(u) Hip: B(s)	Demonstration: changed significantly in one pharmacologic trial meeting primary clinical endpoints[73]	Competitive-inhibition
Serum and urine Coll2-1 and Coll2-1NO2	Type II collagen degradation	Cartilage	Knee: D(s), B(u), P(u) Hip: D(s)	Exploration: not used to date in published pharmacologic trial	Competitive-inhibition
Serum CPII or PIICP	Type II collagen synthesis	Cartilage	Knee: D(s) Hip: B(s)	Exploration: non-significant change in one pharmacologic trial that met primary clinical endpoint[91]	Competitive-inhibition
PIIANP	Type II collagen synthesis	Cartilage	Knee: BPD Hip: none	Exploration: not used to date in published pharmacologic trial	Competitive-inhibition
Urine/serum NTX-1	Bone resorption	Bone turnover	Knee: P(u), E(u) Hip: P(s)	Demonstration: changed significantly in one pharmacologic trial that met primary clinical (WOMAC) endpoint[74]	Competitive-inhibition
Urine/serum CTX-1	Bone resorption	Bone turnover	Knee: B(u), D(s/u), P(u) Hip: none	Exploration: not used to date in published pharmacologic trial	Competitive-inhibition
Serum CS846	Cartilage aggrecan synthesis/turnover	Cartilage	Knee: P Hip: none	Exploration: non-significant change in one pharmacologic trial that met primary clinical endpoint[91] but changed associated with concurrent JSN	Competitive-inhibition
Serum MMP-3	Protease stromelysin involved with joint tissue degradation and inflammation	Synovium, cartilage	Knee: E Hip: none	Characterization: changed significantly in two pharmacologic trials that met primary clinical endpoints[73,88]	Sandwich for total MMP-3 assay

PIIANP = type IIA procollagen amino propeptide; see footer Table II for additional abbreviations.

* This list does not include many emerging biomarkers that may prove useful in the future nor cytokines and chemokines that are also worthy of consideration.

† These are general recognized processes for which these biomarkers are known. This is very preliminary information at this time and should not be considered definitive but rather in evolution. This information is derived from van Spil[62]; Cibere 2009[15]; Conrozier 2008[101]; Kraus 2010[50]. References in Table as follows:[18,70,73,74,88,90].

to imaging and symptom-related outcomes which serve as the qualifying endpoints for biomarker studies.

Le Graverand, *et al.* 2006[69] had also suggested the possibility that no single biomarker is sensitive enough to serve as a surrogate for radiographic outcomes in OA, but the combination of multiple biomarkers, representing different aspects of articular cartilage biochemistry, may significantly improve the detection and prediction of radiographic changes of knee OA. A natural extension of the three elements stated above, therefore, is to identify groups of biomarkers that are correlated with each other and that, in combination, have good predictive value for the progression of OA or change concurrently with radiographic outcomes.

Summary of biomarker data generated in OA clinical trials

A few details are worth noting regarding the use of biomarkers in published clinical trails. In 137 individuals with knee OA, no significant difference was seen between patients in the *placebo* or glucosamine sulfate treated groups with respect to the ratio of markers of collagen type II breakdown (lnC1, 2C/C2C) in serum or in urine[81]. This study used flare/no flare status as the clinical endpoint. In a study of 201 patients with inflammatory knee OA, a decrease in the levels of urinary Glc-Gal-PYD was observed following treatment with the nonsteroidal anti-inflammatory drug ibuprofen but not with *placebo*[70]. Finally, a group of 35 patients with OA were randomly selected to receive a potent inhibitor of MMP-3 (BAY 12-9566) or *placebo*. Levels of the aggrecan 846 epitope were higher in the treated group compared to the control group implying that aggrecan synthesis improved[84]. This study used an original protocol, measuring markers directly in the cartilage samples obtained at the time of surgery 3 weeks after the start of treatment. The advantage of such an approach is direct analysis of cartilage, short duration of treatment, and small numbers of patients.

Most past studies have used structural and/or clinical endpoints to investigate the usefulness of a biomarker. A 30-month study of a subset (60 progressors and 60 non-progressors) of the patients in a clinical trial assessed by radiography (progression limit: JSN $\geq$0.33 mm) showed a reduction of JSN in the doxycycline treatment group but a paradoxical increase in uTIINE with treatment[92]. In a study testing the effects of risedronate on 1885 patients suffering from knee OA, CTX-II levels decreased with risedronate in patients with knee OA although there were no differences in the traditional JSN radiologic outcome or in symptoms in response to the treatment. There was however a dose-related preservation and improvement in tibial subchondral bone architecture[100] with treatment. The utilization of more sensitive imaging methods such as MRI, in future clinical trials, may clarify and resolve such

apparent inconsistencies, providing a way forward for biomarkers qualification. In another study, CTX-II levels reached at 6 months were associated with radiological progression at 24 months[19], defined *a priori* as a JSN of $\geq$0.6 mm from baseline which corresponds to three times the SD of the X-ray measurement method for joint space[19]. Three clinical trials in OA[18,19,88] have used baseline levels, or early change of a biomarker (CTX-II, MMP-3), to predict subsequent progression of radiological damage. These studies demonstrate the advantages of selecting a high matrix turnover/progressor patient population for trial inclusion.

Other studies evaluated biomarkers with patient centered (self-reported) clinical endpoints. A study with 53 patients receiving oral calcitonin daily for 84 days showed that CTX-II, C2C, and MMP-13 levels were decreased in the group of patients receiving 1 mg/day of calcitonin. The efficacy of the treatment was evaluated by Lequesne's index[73]. A small Japanese study with 28 patients with knee OA evaluated the effects of repeated injections of hyaluronan and showed a significant reduction of C6S, C4S and KS relative to baseline. However, a vehicle treated control group was not evaluated and it would be important to rule out changes in biomarkers due to SF aspiration alone. The effect of this treatment was evaluated by change in knee pain assessed by a VAS[95].

Level of qualification of OA-related biomarkers

The following, Tables III and IV, summarize OA-related biomarkers used to date in human clinical trials described in Table II and/or commercially available. Specifically, Table III lists commercially available biomarkers currently recommended as a panel for study in past and future clinical trials (discussed in Conclusions and Recommendations-Recommendations to advance the science of biomarkers), and Table IV lists other OA-related biomarkers qualified for various OA outcomes. The BIPEDS classification assignments are based on studies in which the biomarker showed a statistically significant difference for a clinical or structural outcome as summarized primarily by van Spil[62] but also by Cibere 2009[15]; Conrozier 2008[101]; and Kraus 2010[50]. The Surrogate classification is restricted to results based on current published human clinical trials only. These designations could be further refined by a consideration of preclinical results and unpublished results if a repository of this knowledge existed as called for in the recommendations of this document.

Although type II collagen is an attractive candidate marker of cartilage degradation, it can be difficult to precisely identify the principle tissue sources of a biomarker and the source within a tissue such as articular cartilage which is composed of both calcified (adjacent to subchondral bone) and non-calcified regions. A case in point is represented by the biomarker CTX-II, the most widely tested OA-related biomarker to date. The CTX-II assay exists in two forms: a sandwich ELISA used for animal serum samples that likely recognizes a dimeric form of the EKGPDP epitope; and a competitive ELISA used for human and animal urine samples that likely recognizes monomeric and dimeric forms of the EKGPDP collagen II telopeptide[67]. Unlike the collagen epitope urinary TIINE[40], the exact nature of the immunoreactive cleavage products in urine has not been reported for CTX-II. EKGPDP is released from denatured human type II collagen upon enzymatic digestion with matrilysin, and MMPs-3, -8, and -13[102], and in another study from cartilage sections by enzymatic digestions with MMPs-1, -3, -7, -9, and -13 and cathepsin B[103]. CTX-II immunoreactive epitope can also be released *in vitro* from non-mineralized bovine articular cartilage treated with oncostatin M and TNFalpha and its release can be blocked by estrogen[104]. In young animals and skeletally immature humans, a significant amount of this epitope originates from growth plate cartilage[105–107]. In adult human osteoarthritic cartilage CTX-II immunostaining is in uncalcified fibrillated cartilage as well as calcified articular cartilage[108].

Further complicating the interpretation of collagen type II fragment origins, are the many sites where type II collagen is found in skeletally mature adults, including: articular cartilage, fibrocartilage (intervertebral disc, menisci), respiratory tract cartilage, rib cartilage, insertion sites of tendons and ligaments into bone, and to a small extent, in the ear and eye[63]. However, as pointed out by Lohmander and Eyre, type II collagen makes up only ~1% of all collagen in the body but the normal turnover is low suggesting that pathological turnover from a single joint might be expected to raise the systemic level of fragments significantly[63].

Finally, CTX-II urine levels are very low in individuals with pycnodysostosis compared with age-matched controls[63].

Table IV
Other OA-related biomarkers qualified for various OA outcomes

Biomarker	Process (preliminary)	Tissues of origin (see discussion below Table)	BIPEDS classifications	Surrogacy based on human clinical trials (preliminary)	ELISA assay type
Serum KS	Cartilage catabolism, aggrecan	Cartilage	Knee: BPED Hip: none	Demonstration: changed significantly in one pharmacologic trial meeting primary clinical endpoints[93]	Competitive-inhibition (not commercially available)
Serum YKL-40	Catabolic; macrophages, cartilage, synovium, cells of epithelial origin	Macrophages, cartilage, synovium, cells of epithelial origin	Knee: BE Hip: D	Demonstration: changed significantly in one pharmacologic trial meeting primary clinical endpoints[99]	(Not commercially available)
Urinary TIINE	Cartilage catabolism type II collagen	Cartilage	Knee: BP Hip: none	Exploration: paradoxical response[69]	(Not commercially available)
Serum OC	Anabolic bone turnover	Bone	Knee: BPED Hip: none	Demonstration: changed significantly in one pharmacologic trial meeting primary clinical endpoints[93]	ELISA
Urinary Glc-Gal-PYD	Catabolic synovium	Synovium	Knee: BD Hip: none	Exploration: insignificant change in one pharmacologic trial meeting primary clinical endpoints[70]	HPLC
Urinary PYD	Catabolic bone turnover	Bone	Knee: BED Hip: none	Demonstration: changed significantly in one pharmacologic trial meeting primary clinical endpoints[93]	HPLC
Urinary DPD	Catabolic bone turnover	Bone	Knee: BED Hip: none	Demonstration: changed significantly in one pharmacologic trial meeting primary clinical endpoints[93]	HPLC
MMP-13	Protease	Synovium, cartilage	Knee: E Hip: none	Demonstration: changed significantly in one pharmacologic trial meeting primary clinical endpoints[73]	Sandwich for total MMP-13 assay

Table II abbreviations: uTIINE (mAbs 9A4/5109) = urinary type II collagen collagenase-generated neoepitope; MMP = matrix metalloproteinases:-13 (collagenase-3).

Pycnodysostosis is a lysosomal storage disease of the bone caused by mutation of the gene encoding the enzyme cathepsin K, a cysteine protease expressed by osteoclasts and a major protease involved in bone resorption. In pycnodysostosis (OMIM #265800), osteoclasts function normally in demineralizing bone, but do not adequately degrade the organic matrix. This finding has suggested that a major source of CTX-II is the breakdown and remodeling of mineralized cartilage collagen by osteoclasts[63,104]. In fact, by immunohistochemistry, the EKGPDP epitope is localized in calcified articular cartilage, at the interface between the calcified cartilage and bone, and to some extent at the surface of non-mineralized cartilage lesions, as well as subchondral bone (in a rat model of OA)[104,108] Osteophyte formation and remodeling may thus also be a significant source of CTX-II since, like the growth plate, this also involves endochondral ossification and is a fundamental feature of joint degeneration in OA. Urinary CTX-II has in fact been shown to correlate with total body burden of osteophyte[50].

In summary, and as illustrated here for the most reported OA biomarker CTX-II, the complexities in structure, the paucity of evidence on tissue origins, and the incompletely understood catabolic, clearance, and regulatory pathways currently make it difficult to be certain of the principal sites of origin of OA-related biomarkers. This serves to illustrate how critically important it is to understand as much as we can about each of these biomarkers from *in vivo* and *in vitro* analyses in order to be able to more precisely and correctly interpret biomarker data in preclinical and clinical drug development and assessment.

Summary related to use of biomarkers in clinical trials

There have been few published clinical trials reporting biomarker results. The lack of medications with established chondroprotective activity has limited the availability of clinical trial samples in which to test the utility of biomarkers.

In many cases, especially involving preclinical and clinical trials, biomarker results may not be reported or are not reported in a systematic and standardized manner. So it is difficult to utilize published data from trials to power future trials or to draw conclusions by comparing across studies. Recommendations regarding standardization and access to body fluids can be found at the end of this document.

Of those clinical trials reporting biomarker results, relatively few biomarkers have been tested, often using different methodologies, and very few trials and studies have tested multiple biomarkers in the same samples. Only recently have a variety of biomarkers started to be examined head to head in the same studies[15].

Many promising OA-related biomarkers have never or rarely been tested in clinical trial samples. Existing clinical trials have not used standardized methods of sample collection and assay methods differ among studies for many of the biomarkers tested.

Pathways for biomarker qualification

The increased use of biomarkers is viewed as a critical component in improving the traditional inefficiency of the OA drug development process. Biomarkers can be used in a variety of ways from drug target development in preclinical studies to surrogate endpoints for regulatory approval. How biomarkers are used also defines the level of qualification required.

As described in the Introduction-Definition of biomarkers, a biomarker may be defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention"[2]. The definition has two key components; the measurement of the biomarker and its evaluation as an indicator of some biological process(es). Consequently, any discussion of biomarker qualification must include both characterization of the source of the biomarker, the analytical capabilities of the test used to quantify the biomarker, as well as the evaluation (i.e., qualification) of the association between the biomarker and the pathobiological state and/or clinical outcome.

In general, companies are struggling with defining and developing a process for what *in vitro* (soluble) biomarker data to include in regulatory submissions[39]. Results intended to influence the course of the clinical development process would be considered part of the safety and efficacy evaluation and would need to be part of the regulatory submission[39]. Biomarker data that do not have such a regulatory impact would not need to be part of the regulatory submission. This section describes some of the considerations related to the biomarker qualification process and pathway.

Assessment of analytical capabilities of a biomarker test

Analytical validation of a biomarker follows a different pathway from that of validation of a drug. Validation of analytical methods related to the drug itself is a well-defined process. Regulatory agencies require that critical parameters of tests performed to assess the material conform with current Good Manufacturing Practice[109]. The ICH has published detailed guidelines on the validation of analytical procedures included as part of a registration application[110]. In the case of a drug or biologic development process, analytical methods are considered acceptably validated if the assays perform in a manner that demonstrates that the drug substance or drug product has the appropriate identity, strength, quality, and purity. There is an expectation by regulatory authorities that the analytical capabilities of critical test methods will be enhanced during the drug development process and that the methods are fully validated at the time of the market application.

The contrasting process of biomarker validation was described in the Qualification of Known Biomarkers-Validation, and entails assessment of the accuracy, precision, specificity [what process is it measuring and in which tissue(s)], detection limit, quantification limit, linearity and range. In 1988, Congress passed the Clinical Laboratory Improvement Amendments (CLIA) establishing quality standards for all laboratory testing to ensure the accuracy, reliability and timeliness of patient test results regardless of where the test is performed.

Biomarker qualification — association of a biomarker and a clinical outcome

As mentioned earlier, qualification and validation have different meanings; biomarker qualification consists of the process of associating a biomarker with a clinical outcome or biological parameter. Biomarker qualification processes are in a state of evolution but a recent document outlines the current pilot pathway and regulatory agencies involved[44].

Levels of qualification of biomarkers for drug development use

Biomarkers for drug development use can be divided into four categories according to the degree or level that the biomarker can be shown to be associated with the pathobiological state or clinical outcome. An exploration level biomarker has some evidence of an *in vitro* or preclinical association that may predict a clinical outcome. This type of biomarker is often an outcome of scientific research describing a pathway that may impact a clinical state. The consistency of the data or correlation is low and would be used primarily to support hypothesis generation.

A demonstration level biomarker has higher level of evidence of a correlation between the biomarker and the clinical outcome. This correlative data derive from limited clinical studies. This category is

equivalent to the "probable valid" biomarker defined in the guidance on the submission of pharmacogenomic data to FDA[6] and the process map proposal for validating genomic biomarkers by Goodsaid and Frueh[111]. Often the association of the biomarker and the clinical outcome is the result of a *post-hoc* analysis. While the data showing the association between the biomarker and the biological state may be promising, it is limited until further independent verification can be performed.

Characterization level biomarkers are reproducibly linked to clinical outcomes in more than one prospective clinical study in humans and have been independently verified. This category corresponds to the "known valid" biomarkers in nomenclature suggested in the FDA pharmacogenomic guidance and process map proposal of Goodsaid and Frueh referenced earlier. These biomarkers have been shown to be associated with clinical outcomes as prospectively defined endpoints and are appropriate for making a range of decisions regarding the therapeutic being studied. The biomarkers can be used to identify responders vs non-responders, individuals that may be at risk for toxicity, or assist in defining the appropriate dose for an individual.

The final category of biomarker qualification is when the biomarker can be used as a surrogate for a clinical outcome, and thus can be used as the basis of a regulatory decision. Surrogate level biomarkers should be considered a subset of characterization level biomarkers. The use of a surrogate endpoint as the basis for approval of a new drug requires prior agreement with the regulatory agency, and is also restricted to drugs that are intended to treat serious and life-threatening illnesses[112].

Biomarker qualification can occur both during development of a therapeutic[113] and independent of a therapeutic. As described in Biomarker Applications in Development of Therapeutics for Osteoarthritis, there are many uses for biomarkers that are independent of a therapeutic; in addition they may assess characteristics related to safety or toxicity, such as biomarkers that are correlated with stress or damage to critical organs.

Biomarker qualification independent of a therapeutic

Biomarkers that correlate with disease progression that are developed independent of a therapeutic may be perfect tools for identification of promising new therapeutics. This type of biomarker would be considered an exploration level biomarker. Clinical studies would be required to develop the necessary data to show if it can be categorized as a demonstration or characterization level biomarker. Agencies involved include CLIA, the FDA and the Center for Devices and Radiological Health (CDRH).

For the qualification of biomarkers independent of a therapeutic, a process map for the validation/qualification of genomic biomarkers of drug safety has been proposed[111]. The proposed process could encompass biomarkers other than safety, such as biomarkers of disease progression. As the authors state, the process can be considered intuitive as it follows well-established processes. Following the identification of a potential safety or disease progression biomarker and the development of an appropriate analytical method, a qualification protocol can be proposed and discussed with the regulatory agency. Once approved, the qualification protocol could be executed and the report submitted for review. If the data support the correlation between the biomarker and the safety signal or disease progression, then the biomarker could be considered qualified. The level of qualification could be dependent upon whether the protocol included independent or cross validation of the biomarker.

Biomarker qualification in conjunction with a therapeutic

Biomarkers have been used in drug development for some time, and this practice is expected to expand with the trend toward personalized medicine. Because qualification of a biomarker in conjunction with development of a therapeutic is usually done within a single company, independent verification is rarely feasible. However, the process for qualification would be comparable to the process described earlier. Lesko and Atkinson describe in detail a strategy for biomarker qualification, and note that the criteria used in the qualification of any biomarker are dependent upon the regulatory role a biomarker is expected to play[114]. In 2005, the FDA published a concept paper on drug-diagnostic co-development[115]. This draft document addresses issues related to the development of a test that would be mandatory in the therapeutic use of a drug. Due to the critical role the test would assume, the FDA recommends that the co-development pathway should be determined early in development and that the sponsor should consult with the appropriate drug/biologic/device reviewing centers. The approval of a drug that utilized the analysis of a biomarker as integral in the use of the drug would require the parallel review and approval of the diagnostic. Agencies involved including CLIA, the FDA, and CDRH.

Examples of biomarkers used for regulatory approval of a therapeutic

There are no examples of biomarkers used for OA drug registration. However, examples are emerging in other fields of the successful application of biomarkers in the development of drugs. To date, the primary biomarkers qualified for use with a drug are genomic. In fact, pharmacogenomic information is contained in about 10% of labels for drugs approved by the FDA. The FDA has published a list of valid genomic biomarkers in the context of these FDA-approved drugs[113]. This list, containing approximately 30 drugs, provides the regulatory context in which the biomarker was approved. Currently, only a few drugs recommend or require an assessment of the biomarker in the context of prescribing the drug or arriving at a therapeutic decision.

In summary, the pathway for qualification of a biomarker is defined by how the biomarker will be used, the questions that are addressed, and how closely the biomarker is associated with a clinical outcome. The qualification process can be viewed somewhat as a continuum, with a relatively low bar required of an exploration level biomarker and the highest level required of a surrogate level biomarker.

Conclusions and recommendations

General overview

This guidance document is being prepared at a time of rapid biomarker evolution in this and other fields when studies are revealing many promising and important contributions that could be made by biomarkers to the development of new treatments for OA. The advantages and potential opportunities offered by the use of biomarkers can be traced from preclinical work involving laboratory-based studies, through work with animal models of OA extending into clinical trials and eventually into the treatment of patients. The use and assessment of the value of these biomarkers is seen as very much a work in progress, building on the lessons learned to date and on the ongoing advances in the clinical and imaging biomarker outcomes that form the basis for the qualifying endpoints for non-imaging biomarkers. At the present time biomarker usage will not provide primary outcome measures in OA clinical trials; this in large part stems from the lack of an appropriate gold standard, which allows robust biomarker qualification with regard to symptomatic and structural outcomes. Because OA is a whole organ disease with different tissues and biological processes involved, a combination of a panel of biochemical markers will probably be more powerful for the investigation of

joint damage than assessment of a single biomarker[15,25]. The potential for the effective clinical use of biomarkers may therefore be more readily realized as biomarkers start to be included in OA clinical trials, used in combination rather than individually, and used in combination with imaging such as MRI[35]. It may, in the not so distant future, become possible to use selected individual markers or combinations thereof to inform decisions in clinical trials and patient diagnosis, treatment and monitoring.

Summary of issues related to the application of biomarkers in the development of drugs for OA

Preclinical studies

Biomarkers have already proven their relevance in preclinical studies of arthritis onset, progression, treatment and outcomes. This work includes studies of OA development in mice, rats, guinea pigs, rabbits, horses and dogs, both induced and naturally occurring. Studies with surgically induced joint instability can produce significant biomarker changes in peripheral blood and urine within 2–8 weeks of onset that parallel histologically demonstrated cartilage degeneration. Therapeutic interventions during this period are reflected by biomarker changes in these models. Preclinical studies can be used to link changes in specific biomarker parameters (i.e., magnitude of change with intervention and time to measure change in biomarker from first dose) to histological benefit and therefore inform regarding the use of these biomarkers in clinical studies. Routine use of biomarkers for dose selection will require establishing a link to structural and clinical outcomes. Preclinical model studies of cartilage collagen biomarkers of degradation and synthesis and COMP have proven to be of special value. Such studies should provide valuable insights in human clinical investigations. If biomarkers reflecting structural and/or symptomatic changes can be identified in preclinical studies these can then be considered for use in a clinical trial.

Clinical studies

- There are currently no recognized and approved “disease modifying” therapeutic agents, therefore there is no valid means by which to test the ability of biomarkers to change with therapy.
- Biomarkers may serve as titration tools, facilitating dose setting in early clinical studies.
- Although systemic biomarkers (serum and urine) potentially reflect generalized OA (analogous to a global outcome measure) and local (intra-articular) biomarkers reflect local OA, in general, therapeutic studies are focused on one joint, often knee or hip; data are not routinely collected on symptoms or structure in other joints that may also be affected as part of generalized OA and that may impact systemic biomarkers.
- Rescue medication and *placebo* effects confound trial results but biomarkers provide objective outcomes with the potential to overcome some of the inherent limitations of subjective outcomes.
- Therapeutic trials include patients whose disease is likely to respond to treatment based on symptomatic and imaging criteria, but not biological criteria reflecting tissue metabolic activities; whereas biomarkers and biomarker profiles have the potential to identify molecular and/or metabolic subsets of disease activity and progression that may reflect different responses to a particular intervention.
- Biomarkers provide the only current potential means of identifying the early molecular stages of OA as defined in Fig. 1. These early changes, having been identified by a biomarker, may be most susceptible to disease modification, and also measurable by that same biomarker, based on experiences with biologic therapy in inflammatory arthritis[17].
- A major reason for failure of OA clinical trials to date has been lack of study power due to insufficient numbers of progressors with regard to imaging outcome.
- Biomarkers should offer both sensitive detection of patients with active disease for inclusion in trials and monitoring of effects on tissues.
- Biomarkers provide potential means of increasing trial power with a specified sample size through enrichment of a predominantly disease progressing patient population.
- Biomarkers provide potential to decrease the length of a trial or facilitate early decision-making regarding the therapeutic value of a treatment if early biomarker changes are predictive of later clinical or structural outcomes; this has been exemplified to date by several biomarkers including CTX-II[18,19], MMP-3[88], and considering the combination of collagen degradation and synthesis[24,25].
- Although correlations of biomarkers to symptoms will be informative, very short symptomatic trials may be too short to reflect cartilage or bone biomarker level modifications.
- We lack information on the impact of therapy on biomarkers in generalized OA.
- One shortcoming of most biomarker studies is the failure to account for total body burden of disease.
- Proof of concept studies with serum COMP have shown that systemic concentrations in the serum report on burden of (systemic) disease while intra-articular concentrations report on local disease features[47].
- Little to date is known about markers specific for a particular joint site.
- The use of systemic biomarkers to report on local disease at a specific joint site tends to be confounded by high background from turnover in other cartilage tissues including the spine[48].
- There is a validated measure to evaluate spine OA structural changes[116] that could serve as an endpoint on which to qualify a biomarker for spine OA as exemplified by one past study[48].
- There is no definitive “gold standard” for assessing structural changes in all joint tissues with imaging techniques thus hampering the ability to qualify a biomarker for structural endpoints; sampling of fluid from a given joint will circumvent this problem in a trial.
- Statistically significant biomarker differences may not correlate with clinically meaningful differences in symptomatic or imaging endpoints.
- The interpretation of the biomarker values in urine and blood must take into account the possible confounders such as age, gender, body mass index, ethnicity, diurnal changes, food intake, physical activity and post-menopausal status.
- These confounders require that the biological fluids be collected at well defined times, with standardized procedures, accounting for all known confounders.
- Levels of biomarkers measured in blood and urine provide information on systemic skeletal tissue turnover and are not necessarily specific for the alterations occurring in a single affected joint.
- The clearance of the biomarker may also be affected to different extents by physical activity, time of day, and liver and kidney function. At the joint level, biochemical marker clearance may also vary with synovial inflammation.
- The use of multiple biomarkers that represent various components of the complex OA disease pathway, such as tissue synthesis, destruction and inflammation, may yield intermediate endpoints that offer a more comprehensive assessment of treatment effects such as impact on catabolism and anabolism.

- Because clinical decisions can depend on the quality of biomarker data, appropriate analytical validation of biomarker assays is essential to ensure high-quality data to maximize the value of such decisions[39].
- Many promising OA-related biomarkers have never been tested in appropriate clinical trial samples, often because of lack of access to samples by those developing assays, so it is premature to finalize the choice of the optimal biomarker(s) for OA trials.
- No single biomarker will be representative of all aspects of the biological changes in the complex organ represented by the joint.
- To encourage the application of biochemical and genomic biomarkers in drug development, a consensus on how to interpret results from these measurements is needed for regulatory submissions.

Difficulties encountered

Historically, much work on biomarkers has suffered from a number of limitations and obstacles. First among these include difficulties encountered in translating new biomarker assays developed in the laboratory into preclinical animal models and human clinical trials. Often scientists working independently with animal models and in clinics have had difficulties accessing appropriate collaborative opportunities for biomarker application and assessment. Researchers and companies developing biomarker assays continue to have serious problems evaluating assays due to inability to gain access to clinical samples, especially those from clinical trials. Also many assays do not cross-react between human and other species requiring the development of multiple assays. Second, although early events of the OA-process should be optimal for intervention, clinical studies focused on early disease have been very limited (the Cibere *et al.* 2009 study being a notable example[15]). Diagnosis of OA is typically made late in the disease process and no DMOADs are currently available for treatment, patients are often missed during the early phases of OA. A third obstacle hampering the application of biomarkers has been the lack of understanding of how the processes leading to tissue destruction also lead to symptoms and other clinical parameters and whether there are molecular indicators that correlate with these parameters. Another unknown is whether or how the processes vary over the progression of the disease. Soluble biomarkers are potentially as complex and varied as the biology they model but have the challenge of being qualified based on relatively generic symptomatic and structural outcomes. These obstacles form the basis of a research agenda for the study of OA biomarkers informed by the recommendations below. Fourth, investigators tend to study a single biomarker or a limited set of biomarkers at the exclusion of others. This trend is beginning to change with increased understanding of the need to evaluate many different biomarkers together within a given study. Information can be learned from these biomarkers both individually and in combination as well as in combination with imaging markers.

Critical needs

- To develop better structural endpoints for biomarker qualification;
- To develop biomarkers for various stages of disease;
- To develop biomarkers reporting on specific joint sites and to elucidate the specific joint site contributions to the systemic concentrations of existing biomarkers;
- To determine the clearance of biomarkers from the joint, from the lymphatics, and from the blood as well as the renal processing and elimination *via* the urine and the effect on their correlation with disease progression;
- To assess if there is a circadian rhythm in the level of a biomarker in a particular matrix to better design the sample collection schedule and the interpretation of the results;
- To assess if there are covariates that affect the concentration of a biomarker in the selected matrix such as age, gender, BMI, concomitant diseases/medications, or joint site involvement;
- To study a wide-variety of patient types with varied clinical characteristics and joint-site involvement;
- To develop biomarkers fit-to-purpose;
- To establish an ongoing critical assessment of the value of existing biomarkers in clinical trials;
- To establish minimal clinically important differences in biomarkers once the minimal clinical important differences are defined for the qualifying endpoints for biomarkers, namely in symptomatic and structural endpoints;
- To be able to gain easier access to body fluids from past, present and future clinical trials to enable more comprehensive and critical head to head evaluations of existing and new biomarkers for use in clinical trials;
- To develop multiplex assays incorporating existing promising biomarkers to provide efficient, cost-effective assays informing on multiple domains of joint biology and response to therapy while minimizing demands for sample;
- To increase the available knowledge of biomarker responses in clinical trials for biomarker qualification and clearance by FDA through public release by companies, of information related to use of biomarkers in their preclinical and clinical trials.

Recommendations to advance the science of biomarkers

The availability of an expanding number of biomarkers provides increasing opportunities to combine biomarkers to study disease-subsets and to correlate these to clinical parameters and disease outcome. We recommend measurement of a broad set of biomarkers in available and future sample sets, and analysis of biomarkers singly and in combination, to provide a more comprehensive assessment of ongoing disease and efficacy of treatment. We recommend that a panel of biomarkers be used to examine the same samples and preferably in multiple past and future clinical trials. The most appropriate biomarkers would be those related to the proposed mechanism of drug action. The following commercially available biomarkers, some often studied and others less frequently, are nevertheless recommended for inclusion to provide comparative data and biological insights from which to continue to assess the utility and relevance of an array of established OA-related biomarkers: urinary CTX-II, serum COMP, serum hyaluronan, serum and urine C1, 2C, serum and urine C2C, serum and urine Coll2-1 and Coll2-1NO2, serum CPII, Serum PIIANP, urine/serum NTX-1, urine/serum CTX-1, serum CS846, and serum MMP-3. This panel is considered an initial starting point for a process in evolution. As knowledge is gained and additional OA-qualified biomarkers become either commercially or readily available to the OA community of investigators, it is anticipated that this will be revised.

- Recommendations should be developed for biomarker data presentation in publications from research studies and clinical trials. For clinical trials, this should include, at a minimum, reporting of the mean and SDs (in all groups before and after treatment) of biomarker concentrations and inter- and intra-assay variation.
- Minimal meaningful differences for biomarkers need to be defined and established and this can be done even in the absence of a treatment study in a longitudinal trial. A critical component for the success of this aim will be the establishment of clinically meaningful endpoints related to imaging and

symptom-related outcomes which serve as the qualifying endpoints for biomarker studies.
- For clinical trials, consideration should be given to listing intended biomarker analyses at ClinicalTrials.gov in addition to primary clinical endpoints; alternatively, a separate website could be considered to serve the purpose of tracking and reporting this information, results (both positive and negative apropos of next recommendation), and stimulating advances in the field.
- Biomarker data, both positive and negative, ideally should be released in a timely manner into the public domain, preferably by peer-reviewed publication. This will ensure the optimal development and use of important biomarker tools as exemplified in this guidance document. It will also serve to maintain the momentum generated by a recent increase in collaborative research on biomarkers of OA, ensuring that this continues as a concerted effort to serve the broader stakeholder community to solve common problems. This information could and should be summarized and included in a public database that is managed and regularly updated on a monthly basis.
- Resources should be made available to encourage, through a carefully controlled peer review process, access to body fluids from cohorts such as those harvested from studies of OA onset, progression and OA clinical trials. Many such cohorts are presently available for study (see proceedings of OARSI Biomarker Workshop, Bethesda, MD, 2009) (see http://www.oarsi.org/index2.cfm?section=Meetings_Events&content=OABiomarker). In addition, an effort also needs to be made to obtain cohorts depicting early events, including sample sets for investigation of risk groups after joint trauma, and past and future clinical trial sample sets.
- We note that in existing clinical trials, there has been no standardized method of sample collection. We call for a consensus regarding collection methods and recommend practices in Appendix A.
- We recommend body fluid collection and sample banking in future human (in particular all future prospective OA clinical trials) and animal studies to include serum and plasma, RNA and DNA isolated from whole blood, urine, and where possible, SF. SF is included since it represents the most proximal fluid to the joint and can provide the most direct insight into joint metabolism in the case of biochemical and molecular biomarkers. Peripheral white blood cells exhibit changes in gene expression in OA that are detectable by microarray and polymerase chain reaction (PCR) analyses[117,118]. The process of cell isolation may be associated with artifactual gene expression changes so the collection of whole blood (*via* PaxGene or Tempus tubes), in lieu of cell isolation, may be preferable for studies of gene expression. Just as the FDA has encouraged voluntary submission of pharmacogenomic data in an effort to increase the knowledge base for therapeutic candidates (see http://www.fda.gov/oc/initiatives/criticalpath/Lesko/Lesko.html)[39], and in view of encouraging successful biomarker developments of this kind in other fields (described in Examples of Biomarkers Used for Regulatory Approval of a Therapeutic), we recommend collection of whole blood for future genomic analyses of gene expression in OA clinical trials.
- Since patterns of fragments may vary in different body fluids due to processing in the kidney, we recommend that both urine and serum samples be collected and analyzed when biomarker assays are available for use with both these body fluids.
- Protocols enrolling patients with knee or hip OA (the so-called signal joints) have made measuring and interpreting treatment effects easier, and the development of specific OA measurements has paralleled, and in some ways guided, this signal joint approach. However, exclusive focus on the signal joint will miss what is happening at other OA sites that could affect systemic biomarker concentrations. For this reason it is recommended that clinical trials for OA that include systemic (serum, urine) biomarkers, collect information about other joints in addition to the target joint, such as by using a patient global assessment, or taking specific non-signal-joint measurements. Future developments may demonstrate that the status of particular joints can be distinguished even in the setting of generalized OA.
- Immunoassays based on monoclonal antibodies are preferred (or similar highly specific reactive agents such as those produced by phage libraries). The ability to accurately and quantitatively measure the concentration of epitopes in body fluids is a primary requisite for all assays. Competition immunoassays using a single antibody are often subject to higher assay variability than sandwich assays in which intra- and inter-assay variability can be minimized by use of, ideally, two monoclonal antibodies with different epitope specificities. Sandwich assays however may be problematic with small fragments when these do not span two epitopes. The reliance on polyclonal antisera makes it difficult to ensure continuing assay standardization when new antisera must be raised to replace depleted supplies. The incorporation of an appropriate standard is also an essential requirement for all immunoassays.
- In cases where multiple assays are available for the same analyte, these assays should be compared against each other as different information may be generated according to epitope recognition.
- Although technically challenging, for all existing and future assays, validation of assay specificity should include epitope identification of protein epitopes consisting of sequence verification of the epitope(s) being measured by their isolation and characterization from the sample under investigation using the antibodies that constitute the assay in combination with methods such as mass spectrometry. An example is provided by Nemirovskiy *et al.*[40] who examined the peptides in urine generated by collagenase cleavage of type II collagen and bound by the uTIINE antibody.
- For an improved understanding of a biomarker, the principle tissue source(s) of a given biomarker should be identified as accurately as possible, so that the origin(s) of the epitope(s) is/are clearly understood. These requirements are essential for a clear understanding of what the assay results represent and for the interpretation of data when biochemical and molecular biomarkers are used in preclinical or clinical studies.
- Assays developed in independent laboratories should be made available either commercially or through collaborative agreements.
- For the most effective assessment of existing and new biomarkers, strong collaborations involving both the academic and commercial sectors are essential so that accessibility to body fluids and different biomarker assays in past, present and future clinical trials is ensured. It is possible to envision a time when an expert advisory group could manage this and that one or more central reference laboratories perform assays in a standardized manner in both biomarker assessment/validation and in preclinical and clinical trials.
- Data on epitope stability with storage, and freezing and thawing, should be standardized and available in the public domain.

Declaration of funding and role of funding source

The OARSI FDA OA Initiative received financial support from the following professional organization: ACR.

Additionally the OARSI FDA OA Initiative received financial support from the following companies: Amgen, ArthroLab,

AstraZeneca, Bayer Healthcare, Chondrometrics, CombinatoRx, Cypress BioScience, DePuy Mitek, Expanscience, 4QImaging, Genevrier/IBSA, Genzyme, King (Alpharma), Merck, Merck Serono, NicOx, Pfizer, Rottapharm, Smith & Nephew, Wyeth.

While individuals from pharmaceutical, biotechnology and device companies actively participated in ongoing working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the FDA.

Author contributions

All authors contributed to the writing and revision of the manuscript and approved the final version.

Conflict of interest

Virginia Byers Kraus was funded by 5P30AG028716 and AR50245 from the NIA/NIH and NIAMS/NIH.

Bruce K Burnett was funded by CTSA grant 1 UL1 RR024128 from NCRR/NIH.

JH, DE, JCR: the authors have nothing to disclose.

RM: has received research support or compensation from sanofi-aventis, Pfizer, Bioiberica, Ferring, Endo, Adolor, Merck.

SL: has received research support or compensation from AstraZeneca, Carbylan, Merck Serono, Pfizer, Tigenix, Wyeth.

YH: has received research support or compensation from Danone, Kitozyme, Nestec, Expanscience, Pierre Fabre, Bioiberica, BioXtract, Intervet, Pfizer; Ownership interest or patent in Wellcare product, Medic all design, Zentech.

DH: patent and stock ownership Anamar Medical; research support or compensation from Anamar Medical, Novo Nordisk.

PG: patent with Synarc.

RP: has received research support or compensation from sanofi-aventis, Xceed, XOMA, DePuy-Mitek, Centocor, IBEX, AO Foundation, Ethicon; ownership interest or patent Transition Therapeutics, Biosyntech, GeneNews.

JC: employee of Novartis Pharmaceuticals Corp.

SC: employee of Amgen.

MG: employee of Cypress Biosciences.

JG: previous employee of Wyeth.

AK, JM: employee of Merck & Company, Inc.

GM: employee of Genzyme Corp.

SP: employee of Rottapharm-Madaus.

MT: employee of Smith & Nephew UK, Ltd.

Appendix A. Sample acquisition and handling

Introduction to sample handing and considerations relevant to all samples: the measurement of biomarkers in biological samples has the potential to provide information on diagnosis, evaluation of risk, assessment of prognosis, monitoring treatment, prediction of response to treatment and as a surrogate response marker[119]. Biomarkers can be evaluated in a wide array of fluids or tissues depending on the pathology to be monitored. In this Appendix, we focus on the collection and storage of blood, urine, and SF for the assessment of onset and disease activity in OA. For all samples intended for biomarker analysis, the sample quality is dependent on two major factors: the pre-analytical parameters (methods used for sample collection, handling, processing) and the storage conditions (duration of storage, storage temperature, number of thaws)[119]. The time from body fluid sampling to storage should be reduced to a minimum to avoid degradation. There is a consensus that a temperature below −70°C is required for long-term stability of protein epitopes, although good prospective data on stability in frozen samples are missing for most assays or have not been published. Likewise, it is ideal to perform biomarker analyses as soon after sampling as possible. However, to avoid assay batch effects, it is best, when possible, to run all samples from a particular study at one time and to examine samples from the same patient on the same plate. For a clinical trial, the storage time prior to analyses may be kept to a minimum but this is less readily implemented for routine use. Epitope stability for each assay should be clearly established with respect to duration of storage and effects of freezing and thawing to ensure validity of the measurements. Samples for immunoassays should be aliquoted on isolation into volumes suitable for at least a single immunoassay in triplicate. A volume of 175 μl is generally recommended to be sufficient to accommodate most assays run in triplicate. With technical improvements in assay design in future, much smaller volumes should be able to be accommodated. The time of day of sample collection should be standardized and noted (recommend AM or PM at least 2 h after rising and/or any meals for blood and second morning void for urines). The body fluid collection for human and animal studies should include serum, plasma (to avoid the proteolysis that may be activated in blood coagulation), urine, and where possible, SF samples. Although collection of SF presents some unique challenges in both patients and animal studies, this sample represents the most proximal fluid to the joint and can provide the most direct insight into joint metabolism in the case of biochemical and molecular biomarkers. Whole blood should be collected in appropriate tubes to permit biomarker studies of gene expression and genetic polymorphisms. Robust standardized protocols for sample collection, handling, and storage should be developed and adhered to for high quality biomarker analyses.

Blood collection, handling, and recommendations: blood should be collected and stored separately as serum, plasma, and whole blood. Some assays work better in serum and some only in plasma so the acquisition of both provides for maximal possible assays, as exemplified by assays for the MMPs[120]. The specific needs of the assay should be carefully checked in advance. For instance, consideration must be given to the potential for altered protein conformations and immunoreactivity in an assay upon chelation of divalent cations. Patients can be fasted overnight prior to blood collection but this is often found not necessary. Plasma should be collected into a commercial collection tube with anticoagulant added [commonly ethylenediaminetetraacetic acid (EDTA), citrate, or heparin] followed by centrifugation. The effect of different anticoagulants on the analyte should be examined[119] as the requirements differ for different assays. For instance, EDTA and citrate plasma are unsuitable for MMP activity assays as these anticoagulants chelate calcium required for MMP activity. For serum, blood should be collected in a red top tube without additives. Serum separator (SST) tubes are particularly easy to use and minimize contamination of serum by clot; they have been successfully used for several years by some researchers (VBK). Upon blood collection, the plasma or serum tube is immediately gently inverted 3–5 times, and allowed to clot at room temperature for at least 30 min (maximum 60 min to avoid subsequent fibrinolysis), followed by centrifugation at approximately 1300*g* (~3500 rpm) for 10 min to separate the plasma from the buffy coat and red blood cells (anticoagulant tube), or to separate the serum from the clot (tube without additives). The supernatant from both plasma and serum collection tubes should be aliquoted into small fresh cryotubes (recommended 100 or 175 μl aliquots according to the assay noting the volume to monitor for potential subsequent dessication of sample) and frozen below −70°C. Depending on the intended use of the sample, mixed protease inhibitors can be added to blood sample collections to avoid degradation of specific analytes of interest. The material composition of the tube can affect measurement of analytes so it is recommended to use identical tubes for all samples within a study.

For total RNA and genomic DNA isolation from whole blood: the PaxGene blood collection tube (Becton Dickinson) can be used to obtain RNA and DNA from whole blood. RNA extraction can be performed using Qiagen's PaxGene 96 blood RNA kit. RNA amplification can be achieved using Ambion Illumina AMIL1791 Total Prep RNA amplification kit. DNA isolation can also be achieved from these tubes as described[121]. An alternative but similar system is provided by Tempus tubes (Applied Biosystems). A successful example of blood-derived gene expression analysis in OA is provided by Marshall 2005[117] although *in vitro* manipulation of cells is ideally to be avoided in favor of direct RNA isolation with PAXGene or Tempus tubes.

Urine collection, handing and recommendations: a second morning void urine specimen is recommended as the standard for biomarker assays. Prior to aliquoting (1 ml aliquots recommended for urine), samples should be centrifuged at approximately 1300*g* for 10 min to remove any debris. As with blood samples, collection of urine samples should use a standardized tube and aliquoted supernatants should be stored, as for serum and plasma, in cryotubes below −70°C until measurements are made. Biomarker levels in urine are subject to dilutional variances due to varying hydration level and urine flow rate (volume produced/time) or total volume. This requires adjustment for differences in flow rate or volume to allow comparison of samples collected from different patients or from the same patient over time and is most commonly achieved through normalization of urine biomarker values with urinary creatinine, although urinary creatinine is influenced by age, diet, exercise, muscle mass, medications, tubular secretion and glomerular filtration rate[122].

SF collection, handling, and recommendations: SFs can be aspirated directly in many cases, but if necessary, a small volume (10 ml) of sterile saline can be injected into the knee followed by aspiration of all obtainable fluid[47]. Using this technique, only one needle insertion is required for human studies. For animal studies (usually performed under anesthesia except in rabbits), it is recommended that the needle be withdrawn after saline injection, and the knee flexed and extended 10 times to ensure mixing; this procedure can also potentially increase the yield of fluid aspirated. To obtain a total white blood cell count in the sample, 25 μl of SF can be mixed with 25 μl of trypan blue and the cell count performed with a hemocytometer. SF samples should be cleared by centrifugation (approximately 1300*g* for 10 min) and the remaining supernatant fluid aliquoted (100 μl) and stored in cryotubes at −80°C for future assays. In cases where 10 ml lavage samples were obtained, a nearly simultaneous serum sample should also be obtained in order to determine the dilutional factor of the SF for subsequent correction of biomarker concentrations for this dilutional effect. The dilution factor can be determined as described by Kraus *et al.*[123] based on measuring urea concentrations in the SF and serum. SF up to 2.5-fold diluted shows a similar mass spectroscopic profile as SF aspirated directly (V. Kraus unpublished data); beyond this level of dilution there may be some specimen heterogeneity introduced by lack of mixing.

In the case of small animals, such as mice, a published methodology is available for obtaining SF at the time of sacrifice[124]. This utilizes an alginate product with high absorbancy that wicks the fluid from the joint. The method has to be tested for each biomarker or analyte of interest to insure that the alginate or the buffer components do not interfere with the assay but to date has been shown to be compatible with SF COMP[124] and IL-1 (VBK unpublished). For rabbits, Poole *et al.* 1978[125] have used a 1–2 ml saline injection, containing the tissue culture dye neutral red, into the stifle (knee) joint. Dilution and hence original SF volume is determined by spectrophotometric examination of dye concentration. Intra-articular injection volumes, determined by the relative size of the animal, should be used for other species in relation to the rabbit. However, this weakly cationic dye penetrates cell membranes by non-ionic diffusion and binds intracellularly to sites of the lysosomal matrix[126]; these dye properties may confound the determinations of dilution factor by this method. In the past, Evans blue and indocyanine green dyes were shown to be inappropriate for monitoring dilutional effects of lavage because of their absorption and metabolism by intra-articular cells and precipitation upon exposure to SF[123]. Another useful approach for small animals (rats, guinea pigs, rabbits) at sacrifice is to blot the surfaces of the opened joint with a pre-weighed filter paper and then immediately record the weight with the SF blotted. This will provide a measure of the amount of non-diluted fluid. Biomarkers can be readily eluted from the paper as described and validated previously[124]. This method works well in the guinea pig[127,128]. Biomarker ratios can also be calculated in joint fluids where dilutions cannot be determined. These are independent of dilution and provide useful data (RP unpublished).

Appendix B. Definitions of biomarker terms

Exploratory biomarker: research and development tools accompanied by *in vitro* and/or preclinical evidence, but there is no consistent information linking the biomarker to clinical outcomes in humans. Used for hypothesis generation. First level of surrogacy based on Wagner *et al.*[1]

Demonstration biomarker: associated with adequate preclinical sensitivity and specificity and linked with clinical outcomes, but has not been reproducibly demonstrated in clinical studies. This category corresponds to "probable valid biomarkers" in nomenclature suggested in draft guidance from FDA. Used in decision-making; provides supporting evidence for primary clinical evidence. Second level of surrogacy based on Wagner *et al.*[1]

Characterization biomarker: associated with adequate preclinical sensitivity and specificity and reproducibly linked to clinical outcomes in more than one prospective clinical study in humans. This category corresponds to "known valid biomarkers" in nomenclature suggested in guidance by FDA. Used in decision-making, and dose finding, for secondary/tertiary claims. Third level of surrogacy based on Wagner *et al.*[1]

Surrogate biomarker: a holistic evaluation of the available data demonstrates that the biomarker can substitute for a clinical endpoint. The designation of "surrogate end point" requires agreement with regulatory authorities. Used for drug registration. Fourth level of surrogacy based on Wagner *et al.*[1]

Valid biomarker: has been defined in the "Guidance for Industry: Pharmacogenomic Data Submissions". Therein, a valid biomarker is described as a "biomarker that is measured in an analytical test system with well established performance characteristics and for which there is an established scientific framework or body of evidence that elucidates the physiologic, toxicologic, pharmacologic, or clinical significance of the test results." The classification of biomarkers is context specific. (http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm).

Analytic validity: test's ability to accurately and reliably detect the epitope of interest.

Format: commercial availability, single or multiplex, type of assay (ELISA or mass spectroscopy, etc).

Qualification endpoints: symptoms; structure: radiographic OA, pre-radiographic OA; molecular OA.

The following are summarized from the ICH of Pharmaceuticals for Human Use[110].

Accuracy: expresses the closeness of agreement between the value that is accepted either as a conventional true value or an accepted reference value and the value found.

Detection limit: is the lowest amount of analyte in a sample that can be detected but not necessarily quantified as an exact value.

Linearity: is the ability of an analytical procedure (within a given range) to obtain test results which are directly proportional to the concentration (amount) of analyte in the sample.

Precision: expresses the closeness of agreement (degree of scatter) between a series of measurements obtained from multiple sampling of the same homogeneous sample under the prescribed conditions. Precision may be considered at three levels: repeatability, intermediate precision and reproducibility.

Repeatability: expresses the precision under the same operating conditions over a short interval of time and is also termed intra-assay precision.

Intermediate precision: expresses within-laboratories variations: different days, different analysts, different equipment, etc.

Reproducibility: expresses the precision between laboratories (collaborative studies, usually applied to standardization of methodology).

Quantitation limit: is the lowest amount of analyte in a sample that can be quantitatively determined with suitable precision and accuracy.

Range: the interval between the upper and lower concentration (amounts) of analyte in the sample (including these concentrations) for which it has been demonstrated that the analytical procedure has a suitable level of precision, accuracy and linearity.

Robustness: a measure of its capacity to remain unaffected by small, but deliberate variations in method parameters and provides an indication of its reliability during normal usage.

Specificity: the ability to assess unequivocally the analyte in the presence of components that may be expected to be present.

References

1. Wagner JA, Williams SA, Webster CJ. Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. Clin Pharmacol Ther 2007;81:104–7.
2. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001;69:89–95.
3. FDA FaDA. Guidance for Industry: E15 Definitions for Genomic Biomarkers, Pharmacogenomics, Pharmacogenetics, Genomic Data and Sample Coding Categories 2008
4. Goodsaid F, Frueh F. Biomarker qualification pilot process at the US Food and Drug Administration. AAPS J 2007;9:E105–8.
5. Bauer DC, Hunter DJ, Abramson SB, Attur M, Corr M, Felson D, *et al.* Classification of osteoarthritis biomarkers: a proposed approach. Osteoarthritis Cartilage 2006;14:723–7.
6. FDA FaDA. Guidance for Industry: Pharmacogenomic Data Submissions 2005.
7. FDA FDA. E16 Genomic Biomarkers Related to Drug Response: Context, Structure, and Format of Qualification Submissions 2009.
8. Colman EG, Food and Drug Administration. The Food and Drug Administration's Osteoporosis Guidance Document: past, present, and future. J Bone Miner Res 2003;18:1125–8.
9. Information OoFo. Guidelines for the clinical evaluation of drugs used in the treatment of osteoporosis. In: DoHaH, Ed. Services 1979. Rockville, MD, USA.
10. 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 hand. Arthritis Rheum 1990;33:1601–10.
11. Altman R, Alarcon G, Appelrouth D, Bloch D, Borenstein D, Brandt K, *et al.* The American College of Rheumatology criteria for the classification and reporting of osteoarthritis of the hip. Arthritis Rheum 1991;34:505–14.
12. 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.
13. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis 1957;16:494–502.
14. Hart DJ, Spector TD. Kellgren & Lawrence grade 1 osteophytes in the knee-doubtful or definite? Osteoarthritis Cartilage 2003;11:149–50.
15. Cibere J, Zhang H, Garnero P, Poole AR, Lobanok T, Saxne T, *et al.* Association of biomarkers with pre-radiographically defined and radiographically defined knee osteoarthritis in a population-based study. Arthritis Rheum 2009;60:1372–80.
16. Ling SM, Patel DD, Garnero P, Zhan M, Vaduganathan M, Muller D, *et al.* Serum protein signatures detect early radiographic osteoarthritis. Osteoarthritis Cartilage 2009;17:43–8.
17. Mullan RH, Matthews C, Bresnihan B, FitzGerald O, King L, Poole AR, *et al.* Early changes in serum type II collagen biomarkers predict radiographic progression at one year in inflammatory arthritis patients after biologic therapy. Arthritis Rheum 2007;56:2919–28.
18. Christgau S, Henrotin Y, Tanko LB, Rovati LC, Collette J, Bruyere O, *et al.* Osteoarthritic patients with high cartilage turnover show increased responsiveness to the cartilage protecting effects of glucosamine sulphate. Clin Exp Rheumatol 2004;22:36–42.
19. Garnero P, Aronstein WS, Cohen SB, Conaghan PG, Cline GA, Christiansen C, *et al.* Relationships between biochemical markers of bone and cartilage degradation with radiological progression in patients with knee osteoarthritis receiving risedronate: the Knee Osteoarthritis Structural Arthritis randomized clinical trial. Osteoarthritis Cartilage 2008;16:660–6.
20. Kraus VB, Feng S, Wang S, White S, Ainslie M, Brett A, *et al.* Trabecular morphometry by fractal signature analysis is a novel marker of osteoarthritis progression. Arthritis Rheum 2009;60:3711–22.
21. Eckstein F, Maschek S, Wirth W, Hudelmaier M, Hitzl W, Wyman B, *et al.* One year change of knee cartilage morphology in the first release of participants from the Osteoarthritis Initiative progression subcohort: association with sex, body mass index, symptoms and radiographic osteoarthritis status. Ann Rheum Dis 2009;68:674–9.
22. Qvist P, Bay-Jensen AC, Christiansen C, Dam EB, Pastoureau P, Karsdal MA. The disease modifying osteoarthritis drug (DMOAD): is it in the horizon? Pharm Res 2008;58:1–7.
23. Sharif M, Kirwan JR, Elson CJ, Granell R, Clarke S. Suggestion of nonlinear or phasic progression of knee osteoarthritis based on measurements of serum cartilage oligomeric matrix protein levels over five years. Arthritis Rheum 2004;50:2479–88.
24. Garnero P, Ayral X, Rousseau JC, Christgau S, Sandell LJ, Dougados M, *et al.* Uncoupling of type II collagen synthesis and degradation predicts progression of joint damage in patients with knee osteoarthritis. Arthritis Rheum 2002;46:2613–24.

25. Cahue S, Sharma L, Dunlop D, Ionescu M, Song J, Lobanok T, *et al*. The ratio of type II collagen breakdown to synthesis and its relationship with the progression of knee osteoarthritis. Osteoarthritis Cartilage 2007;15:819–23.
26. Bellamy N, Kirwan J, Boers M, Brooks P, Strand V, Tugwell P, *et al*. Recommendations for a core set of outcome measures for future phase III clinical trials in knee, hip, and hand osteoarthritis. Consensus development at OMERACT III. J Rheumatol 1997;24:799–802.
27. 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.
28. Abadie E, Ethgen D, Avouac B, Bouvenot G, Branco J, Bruyere O, *et al*. Recommendations for the use of new methods to assess the efficacy of disease-modifying drugs in the treatment of osteoarthritis. Osteoarthritis Cartilage 2004;12:263–8.
29. Hunter DJ, Zhang YQ, Tu X, Lavalley M, Niu JB, Amin S, *et al*. Change in joint space width: hyaline articular cartilage loss or alteration in meniscus? Arthritis Rheum 2006;54:2488–95.
30. Emrani PS, Katz JN, Kessler CL, Reichmann WM, Wright EA, McAlindon TE, *et al*. Joint space narrowing and Kellgren–Lawrence progression in knee osteoarthritis: an analytic literature synthesis. Osteoarthritis Cartilage 2008;16:873–82.
31. Charles HC, Kraus VB, Ainslie M, Hellio Le Graverand-Gastineau MP. Optimization of the fixed-flexion knee radiograph. Osteoarthritis Cartilage 2007;15:1221–4.
32. King KB, Lindsey CT, Dunn TC, Ries MD, Steinbach LS, Majumdar S. A study of the relationship between molecular biomarkers of joint degeneration and the magnetic resonance-measured characteristics of cartilage in 16 symptomatic knees. Mag Reson Imaging 2004;22:1117–23.
33. Dam EB, Loog M, Christiansen C, Byrjalsen I, Folkesson J, Nielsen M, *et al*. Identification of progressors in osteoarthritis by combining biochemical and MRI-based markers. Arthritis Res Ther 2009;11. R115.
34. Lukaszkiewicz J, Karczmarewicz E, Pludowski P, Jaworski M, Czerwinski E, Lewinski A, *et al*. Feasibility of simultaneous measurement of bone formation and bone resorption markers to assess bone turnover rate in postmenopausal women: an EPOLOS study. Med Sci Monit 2008;14:PH65–70.
35. Karsdal MA, Henriksen K, Leeming DJ, Mitchell P, Duffin K, Barascuk N, *et al*. Biochemical markers and the FDA Critical Path: how biomarkers may contribute to the understanding of pathophysiology and provide unique and necessary tools for drug development. Biomarkers 2009;14:181–202.
36. Halasz K, Kassner A, Morgelin M, Heinegard D. Comp acts as a catalyst in collagen fibrillogenesis. J Biol Chem 2007.
37. Milan AM, Sugars RV, Embery G, Waddington RJ. Modulation of collagen fibrillogenesis by dentinal proteoglycans. Calcif Tissue Int 2005;76:127–35.
38. Brune K, Katus HA, Moecks J, Spanuth E, Jaffe AS, Giannitsis E. N-terminal pro-B-type natriuretic peptide concentrations predict the risk of cardiovascular adverse events from antiinflammatory drugs: a pilot trial. Clin Chem 2008;54:1149–57.
39. Lee JW, Weiner RS, Sailstad JM, Bowsher RR, Knuth DW, O'Brien PJ, *et al*. Method validation and measurement of biomarkers in nonclinical and clinical samples in drug development: a conference report. Pharm Res 2005;22:499–511.
40. Nemirovskiy OV, Dufield DR, Sunyer T, Aggarwal P, Welsch DJ, Mathews WR. Discovery and development of a type II collagen neoepitope (TIINE) biomarker for matrix metalloproteinase activity: from in vitro to in vivo. Anal Biochem 2007;361:93–101.
41. Eyre DR, Weis MA. The Helix-II epitope: a cautionary tale from a cartilage biomarker based on an invalid collagen sequence. Osteoarthritis Cartilage 2009;17:423–6.
42. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, *et al*. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Ann Intern Med 2003;138:40–4.
43. Felson DT, Lohmander LS. Whither osteoarthritis biomarkers? Osteoarthritis Cartilage 2009;17:419–22.
44. Goodsaid FM, Frueh FW, Mattes W. Strategic paths for biomarker qualification. Toxicology 2008;245:219–23.
45. Lassere MN, Johnson KR, Boers M, Tugwell P, Brooks P, Simon L, *et al*. Definitions and validation criteria for biomarkers and surrogate endpoints: development and testing of a quantitative hierarchical levels of evidence schema. J Rheumatol 2007;34:607–15.
46. Aurich M, Squires GR, Reiner A, Mollenhauer JA, Kuettner KE, Poole AR, *et al*. Differential matrix degradation and turnover in early cartilage lesions of human knee and ankle joints. Arthritis Rheum 2005;52:112–9.
47. Addison S, Coleman RE, Feng S, McDaniel G, Kraus VB. Whole-body bone scintigraphy provides a measure of the total-body burden of osteoarthritis for the purpose of systemic biomarker validation. Arthritis Rheum 2009;60:3366–73.
48. Garnero P, Sornay-Rendu E, Arlot M, Christiansen C, Delmas PD. Association between spine disc degeneration and type II collagen degradation in postmenopausal women: the OFELY study. Arthritis Rheum 2004;50:3137–44.
49. Meulenbelt I, Kloppenburg M, Kroon HM, Houwing-Duistermaat JJ, Garnero P, Hellio Le Graverand MP, *et al*. Urinary CTX-II levels are associated with radiographic subtypes of osteoarthritis in hip, knee, hand, and facet joints in subject with familial osteoarthritis at multiple sites: the GARP study. Ann Rheum Dis 2006;65:360–5.
50. Kraus VB, Kepler TB, Stabler T, Renner J, Jordan J. First qualification study of serum biomarkers as indicators of total body burden of osteoarthritis. PLoS ONE 2010;5:e9739.
51. Clark AG, Jordan JM, Vilim V, Renner JB, Dragomir AD, Luta G, *et al*. Serum cartilage oligomeric matrix protein reflects osteoarthritis presence and severity: the Johnston County Osteoarthritis Project. Arthritis Rheum 1999;42:2356–64.
52. Jordan JM, Luta G, Stabler T, Renner JB, Dragomir AD, Vilim V, *et al*. Ethnic and sex differences in serum levels of cartilage oligomeric matrix protein: the Johnston County Osteoarthritis Project. Arthritis Rheum 2003;48:675–81.
53. Kobayashi T, Yoshihara Y, Samura A, Yamada H, Shinmei M, Roos H, *et al*. Synovial fluid concentrations of the C-propeptide of type II collagen correlate with body mass index in primary knee osteoarthritis. Ann Rheum Dis 1997;56:500–3.
54. Heinegard D, Saxne T. Molecular markers of processes in cartilage in joint disease. Br J Rheumatol 1991;(30 Suppl 1): 21–4.
55. Young-Min SA, Cawston TE, Griffiths ID. Markers of joint destruction: principles, problems, and potential. Ann Rheum Dis 2001;60:545–8.
56. Kong SY, Stabler TV, Criscione LG, Elliott AL, Jordan JM, Kraus VB. Diurnal variation of serum and urine biomarkers in patients with radiographic knee osteoarthritis. Arthritis Rheum 2006;54:2496–504.
57. Gordon CD, Stabler TV, Kraus VB. Variation in osteoarthritis biomarkers from activity not food consumption. Clin Chim Acta 2008;398:21–6.
58. Quintana DJ, Garnero P, Huebner JL, Charni-Ben Tabassi N, Kraus VB. PIIANP and HELIXII diurnal variation. Osteoarthritis Cartilage 2008;16:1192–5.

59. Garnero P, Gourley I, Mareau E, Durn B, Hickey L, Cohen S. Biological variability of biochemical markers of bone, cartilage, and synovial metabolism in patients with knee osteoarthritis: effect of food intake. Osteoarthritis Cartilage 2007;15. C68.
60. Rossler A, Laszlo Z, Kvas E, Hinghofer-Szalkay HG. Plasma hyaluronan concentration: no circadian rhythm but large effect of food intake in humans. Eur J Appl Physiol Occup Physiol 1998;78:573–7.
61. Christgau S. Circadian variation in serum CrossLaps concentration is reduced in fasting individuals. Clin Chem 2000;46:431 [Comment].
62. van Spil WE, Degroot J, Lems WF, Oostveen JC, Lafeber FP. Serum and urinary biochemical markers for knee and hip-osteoarthritis: a systematic review applying the consensus BIPED criteria. Osteoarthritis Cartilage 2010;18:605–12.
63. Lohmander L, Eyre D. Biochemical markers as surrogate end points of joint disease. In: Reid D, Miller C, Eds. Clinical Trials in Rheumatoid Arthritis and Osteoarthritis. New York: Springer; 2008:249–74.
64. Charni-Ben Tabassi N, Garnero P. Monitoring cartilage turnover. Curr Rheumatol Rep 2007;9:16–24.
65. Rousseau JC, Delmas PD. Biological markers in osteoarthritis. Nat Clin Pract Rheumatol 2007;3:346–56.
66. Henrotin Y, Addison S, Kraus V, Deberg M. Type II collagen markers in osteoarthritis: what do they indicate? Curr Opin Rheumatol 2007;19:444–50.
67. Birmingham J, Vilim V, Kraus V. Collagen biomarkers for arthritis applications. Biomarker Insights 2006;2:61–76.
68. Lohmander L, Poole A. Defining and validating the clinical role of molecular markers in osteoarthritis. In: Brandt K, Doherty M, Lohmander L, Eds. Osteoarthritis. Oxford: Oxford University Press; 2003:468–77.
69. Hellio Le Graverand MP, Brandt KD, Mazzuca SA, Katz BP, Buck R, Lane KA, *et al*. Association between concentrations of urinary type II collagen neoepitope (uTIINE) and joint space narrowing in patients with knee osteoarthritis. Osteoarthritis Cartilage 2006;14:1189–95.
70. Gineyts E, Mo JA, Ko A, Henriksen DB, Curtis SP, Gertz BJ, *et al*. Effects of ibuprofen on molecular markers of cartilage and synovium turnover in patients with knee osteoarthritis. Ann Rheum Dis 2004;63:857–61.
71. Christgau S, Garnero P, Fledelius C, Moniz C, Ensig M, Gineyts E, *et al*. Collagen type II C-telopeptide fragments as an index of cartilage degradation. Bone 2001;29:209–15.
72. Gineyts E, Garnero P, Delmas PD. Urinary excretion of glucosylgalactosyl pyridinoline: a specific biochemical marker of synovium degradation. Rheumatology (Oxford) 2001;40:315–23.
73. Manicourt DH, Azria M, Mindeholm L, Thonar EJ, Devogelaer JP. Oral salmon calcitonin reduces Lequesne's algofunctional index scores and decreases urinary and serum levels of biomarkers of joint metabolism in knee osteoarthritis. Arthritis Rheum 2006;54:3205–11.
74. Spector TD, Conaghan PG, Buckland-Wright JC, Garnero P, Cline GA, Beary JF, *et al*. Effect of risedronate on joint structure and symptoms of knee osteoarthritis: results of the BRISK randomized, controlled trial [ISRCTN01928173]. Arthritis Res Ther 2005;7:R625–33.
75. Bingham 3rd CO, Buckland-Wright JC, Garnero P, Cohen SB, Dougados M, Adami S, *et al*. Risedronate decreases biochemical markers of cartilage degradation but does not decrease symptoms or slow radiographic progression in patients with medial compartment osteoarthritis of the knee: results of the two-year multinational knee osteoarthritis structural arthritis study. Arthritis Rheum 2006;54:3494–507.
76. Mazieres B, Hucher M, Zaim M, Garnero P. Effect of chondroitin sulphate in symptomatic knee osteoarthritis: a multicentre, randomised, double-blind, placebo-controlled study. Ann Rheum Dis 2007;66:639–45.
77. Andersson ML, Thorstensson CA, Roos EM, Petersson IF, Heinegard D, Saxne T. Serum levels of cartilage oligomeric matrix protein (COMP) increase temporarily after physical exercise in patients with knee osteoarthritis. BMC Musculoskelet Disord 2006;7:98.
78. Chua Jr SD, Messier SP, Legault C, Lenz ME, Thonar EJ, Loeser RF. Effect of an exercise and dietary intervention on serum biomarkers in overweight and obese adults with osteoarthritis of the knee. Osteoarthritis Cartilage 2008
79. Li XQ, Thonar EJ, Knudson W. Accumulation of hyaluronate in human lung carcinoma as measured by a new hyaluronate ELISA. Connect Tissue Res 1989;19:243–53.
80. Thonar EJ, Lenz ME, Klintworth GK, Caterson B, Pachman LM, Glickman P, *et al*. Quantification of keratan sulfate in blood as a marker of cartilage catabolism. Arthritis Rheum 1985;28:1367–76.
81. Cibere J, Thorne A, Kopec JA, Singer J, Canvin J, Robinson DB, *et al*. Glucosamine sulfate and cartilage type II collagen degradation in patients with knee osteoarthritis: randomized discontinuation trial results employing biomarkers. J Rheumatol 2005;32:896–902.
82. Poole AR, Ionescu M, Fitzcharles MA, Billinghurst RC. The assessment of cartilage degradation in vivo: development of an immunoassay for the measurement in body fluids of type II collagen cleaved by collagenases. J Immunol Methods 2004;294:145–53.
83. Billinghurst RC, Dahlberg L, Ionescu M, Reiner A, Bourne R, Rorabeck C, *et al*. Enhanced cleavage of type II collagen by collagenases in osteoarthritic articular cartilage. J Clin Invest 1997;99:1534–45.
84. Leff RL, Elias I, Ionescu M, Reiner A, Poole AR. Molecular changes in human osteoarthritic cartilage after 3 weeks of oral administration of BAY 12-9566, a matrix metalloproteinase inhibitor. J Rheumatol 2003;30:544–9.
85. Nelson F, Dahlberg L, Laverty S, Reiner A, Pidoux I, Ionescu M, *et al*. Evidence for altered synthesis of type II collagen in patients with osteoarthritis. J Clin Invest 1998;102:2115–25.
86. Hollander AP, Heathfield TF, Webber C, Iwata Y, Bourne R, Rorabeck C, *et al*. Increased damage to type II collagen in osteoarthritic articular cartilage detected by a new immunoassay. J Clin Invest 1994;93:1722–32.
87. Rizkalla G, Reiner A, Bogoch E, Poole AR. Studies of the articular cartilage proteoglycan aggrecan in health and osteoarthritis. Evidence for molecular heterogeneity and extensive molecular changes in disease. J Clin Invest 1992;90:2268–77.
88. Lohmander LS, Brandt KD, Mazzuca SA, Katz BP, Larsson S, Struglics A, *et al*. Use of the plasma stromelysin (matrix metalloproteinase 3) concentration to predict joint space narrowing in knee osteoarthritis. Arthritis Rheum 2005;52:3160–7.
89. Walakovits LA, Moore VL, Bhardwaj N, Gallick GS, Lark MW. Detection of stromelysin and collagenase in synovial fluid from patients with rheumatoid arthritis and posttraumatic knee injury. Arthritis Rheum 1992;35:35–42.
90. Mazzuca SA, Brandt KD, Eyre DR, Katz BP, Askew J, Lane KA. Urinary levels of type II collagen C-telopeptide crosslink are unrelated to joint space narrowing in patients with knee osteoarthritis. Ann Rheum Dis 2006;65:1055–9.
91. Mazzuca SA, Poole AR, Brandt KD, Katz BP, Lane KA, Lobanok T. Associations between joint space narrowing and

molecular markers of collagen and proteoglycan turnover in patients with knee osteoarthritis. J Rheumatol 2006;33:1147–51.
92. Otterness IG, Brandt KD, Le Graverand MP, Mazzuca SA. Urinary TIINE concentrations in a randomized controlled trial of doxycycline in knee osteoarthritis: implications of the lack of association between TIINE levels and joint space narrowing. Arthritis Rheum 2007;56:3644–9.
93. Uebelhart D, Thonar EJ, Delmas PD, Chantraine A, Vignon E. Effects of oral chondroitin sulfate on the progression of knee osteoarthritis: a pilot study. Osteoarthritis Cartilage 1998;(6 Suppl A):39–46.
94. Uebelhart D, Gineyts E, Chapuy MC, Delmas PD. Urinary excretion of pyridinium crosslinks: a new marker of bone resorption in metabolic bone disease. Bone Miner 1990;8:87–96.
95. Hasegawa M, Nakoshi Y, Tsujii M, Sudo A, Masuda H, Yoshida T, *et al*. Changes in biochemical markers and prediction of effectiveness of intra-articular hyaluronan in patients with knee osteoarthritis. Osteoarthritis Cartilage 2008;16:526–9.
96. Yamada H, Miyauchi S, Morita M, Yoshida Y, Yoshihara Y, Kikuchi T, *et al*. Content and sulfation pattern of keratan sulfate in hip osteoarthritis using high performance liquid chromatography. J Rheumatol 2000;27:1721–4.
97. Yoshida K, Miyauchi S, Kikuchi H, Tawada A, Tokuyasu K. Analysis of unsaturated disaccharides from glycosaminoglycuronan by high-performance liquid chromatography. Anal Biochem 1989;177:327–32.
98. Shinmei M, Miyauchi S, Machida A, Miyazaki K. Quantitation of chondroitin 4-sulfate and chondroitin 6-sulfate in pathologic joint fluid. Arthritis Rheum 1992;35:1304–8.
99. Arjmandi BH, Khalil DA, Lucas EA, Smith BJ, Sinichi N, Hodges SB, *et al*. Soy protein may alleviate osteoarthritis symptoms. Phytomedicine 2004;11:567–75.
100. Buckland-Wright JC, Messent EA, Bingham 3rd CO, Ward RJ, Tonkin C. A 2 yr longitudinal radiographic study examining the effect of a bisphosphonate (risedronate) upon subchondral bone loss in osteoarthritic knee patients. Rheumatology (Oxford) 2007;46:257–64.
101. Conrozier T, Poole AR, Ferrand F, Mathieu P, Vincent F, Piperno M, *et al*. Serum concentrations of type II collagen biomarkers (C2C, C1, 2C and CPII) suggest different pathophysiologies in patients with hip osteoarthritis. Clin Exp Rheumatol 2008;26:430–5.
102. Atley L, Shao P, Shaffer K, Eyre D. Matrix metalloproteinase-mediated release of immunoreactive telopeptides from cartilage type II collagen. Trans Orthop Res Soc 1998;23:850 [Abstract].
103. Charni-Ben Tabassi N, Desmarais S, Bay-Jensen AC, Delaisse JM, Percival MD, Garnero P. The type II collagen fragments Helix-II and CTX-II reveal different enzymatic pathways of human cartilage collagen degradation. Osteoarthritis Cartilage 2008.
104. Oestergaard S, Sondergaard BC, Hoegh-Andersen P, Henriksen K, Qvist P, Christiansen C, *et al*. Effects of ovariectomy and estrogen therapy on type II collagen degradation and structural integrity of articular cartilage in rats: implications of the time of initiation. Arthritis Rheum 2006;54:2441–51.
105. Norlund L, Shao P, Yoshihara P, Eyre D. Markers of bone type I and cartilage type II collagen degradation in the Hartley guinea pig model of osteoarthritis. Trans Orthop Res 1997;22:313.
106. Eyre D, Atley L, Wu J-J. Collagen cross-links as markers of bone and cartilage degradation. In: Hascall V, Kuettner K, Eds. The Many Faces of Osteoarthritis. Basel: Birkhauser Verlag; 2002:275–84.
107. Huebner JL, Williams JM, Deberg M, Henrotin Y, Kraus VB. Collagen fibril disruption occurs early in primary guinea pig knee osteoarthritis. Osteoarthritis Cartilage 2009.
108. Bay-Jensen AC, Andersen TL, Charni-Ben Tabassi N, Kristensen PW, Kjaersgaard-Andersen P, Sandell L, *et al*. Biochemical markers of type II collagen breakdown and synthesis are positioned at specific sites in human osteoarthritic knee cartilage. Osteoarthritis Cartilage 2008;16:615–23.
109. FDA UFD. 21 CFR Part 211: Current Good Manufacturing Practice for Finished Pharmaceuticals 2009; http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=211.
110. ICH EWG. Validation of analytical procedures: text and methodology Q2(R1), ICH Harmonised Tripartite Guideline. In: International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. 2005.
111. Goodsaid F, Frueh F. Process map proposal for the validation of genomic biomarkers. Pharmacogenomics 2006;7:773–82.
112. 314.510 C. Approval based upon a surrogate endpoint 2010; http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=314&showFR=1&subpartNode=21:5.0.1.1.4.8
113. FDA UFDAIn: Table of Valid Genomic Biomarkers in the Context of Approved Drug Labels. US Dept Health Human Services; 2009. http://www.fda.gov/Drugs/ScienceResearch/ResearchAreas/Pharmacogenetics/ucm083378.htm.
114. Lesko LJ, Atkinson Jr AJ. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. Annu Rev Pharmacol Toxicol 2001;41:347–66.
115. FDA UFDA. Drug-diagnostic co-development concept paper. 2005.
116. Lane NE, Nevitt MC, Genant HK, Hochberg MC. Reliability of new indices of radiographic osteoarthritis of the hand and hip and lumbar disc degeneration. J Rheumatol 1993;20:1911–8.
117. Marshall KW, Zhang H, Yager TD, Nossova N, Dempsey A, Zheng R, *et al*. Blood-based biomarkers for detecting mild osteoarthritis in the human knee. Osteoarthritis Cartilage 2005;13:861–71.
118. Krasnokutsky S, Attur M, Belitskaya-Levy I, Greenberg J, Samuels J, Smiles S, *et al*. Peripheral blood leukocyte gene expression as biomarkers of disease in knee osteoarthritis (OA). Arthritis Rheum 2007;56. S355:846.
119. Schrohl AS, Wurtz S, Kohn E, Banks RE, Nielsen HJ, Sweep FC, *et al*. Banking of biological fluids for studies of disease-associated protein biomarkers. Mol Cell Proteomics 2008;7:2061–6.
120. Gerlach RF, Uzuelli JA, Souza-Tarla CD, Tanus-Santos JE. Effect of anticoagulants on the determination of plasma matrix metalloproteinase (MMP)-2 and MMP-9 activities. Anal Biochem 2005;344:147–9.
121. Kruhoffer M, Dyrskjot L, Voss T, Lindberg RL, Wyrich R, Thykjaer T, *et al*. Isolation of microarray-grade total RNA, microRNA, and DNA from a single PAXgene blood RNA tube. J Mol Diagn 2007;9:452–8.
122. Jantos-Siwy J, Schiffer E, Brand K, Schumann G, Rossing K, Delles C, *et al*. Quantitative urinary proteome analysis for biomarker evaluation in chronic kidney disease. J Proteome Res 2009;8:268–81.
123. Kraus VB, Huebner JL, Fink C, King JB, Brown S, Vail TP, *et al*. Urea as a passive transport marker for arthritis biomarker studies. Arthritis Rheum 2002;46:420–7.
124. Seifer DR, Furman BD, Guilak F, Olson SA, Brooks 3rd SC, Kraus VB. Novel synovial fluid recovery method allows for quantification of a marker of arthritis in mice. Osteoarthritis Cartilage 2008;16:1532–8.

125. Poole AR, Oldham G, Coombs RR. Early rheumatoid-like lesions in rabbits injected with foreign serum: relationship to localization of immune complexes in the lining tissues of joints and cellular content of synovial fluid. Int Arch Allergy Appl Immunol 1978;57:135–45.
126. Babich H, Borenfreund E. Cytotoxicity of T-2 toxin and its metabolites determined with the neutral red cell viability assay. Appl Environ Microbiol 1991;57:2101–3.
127. Huebner JL, Hanes MA, Beekman B, TeKoppele JM, Kraus VB. A comparative analysis of bone and cartilage metabolism in two strains of guinea-pig with varying degrees of naturally occurring osteoarthritis. Osteoarthritis Cartilage 2002; 10:758–67.
128. Huebner JL, Kraus VB. Assessment of the utility of biomarkers of osteoarthritis in the guinea pig. Osteoarthritis Cartilage 2006;14:923–30.

Osteoarthritis and Cartilage 19 (2011) 543–549

Osteoarthritis and Cartilage

Validity and responsiveness of radiographic joint space width metric measurement in hip osteoarthritis: a systematic review

D. Chu Miow Lin †, W.M. Reichmann ‡§, L. Gossec ||, E. Losina ‡§, P.G. Conaghan ¶, J.F. Maillefert #††‡‡*

† Department of Rheumatology, Tours University Hospital, Tours, France
‡ Department of Orthopedic Surgery, Brigham and Women's Hospital, Boston, MA, USA
§ Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA
|| Paris Descartes University, Medicine Faculty, UPRES-EA 4058, APHP, Rheumatology B Department, Cochin Hospital, Paris France
¶ Section of Musculoskeletal Disease, University of Leeds & NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK
# Department of Rheumatology, Dijon University Hospital, Dijon F-21078, France
†† University of Burgondy, Dijon F-21079, France
‡‡ INSERM U887, Dijon F-21079, France

ARTICLE INFO

Article history:
Received 16 June 2010
Accepted 2 December 2010

Keywords:
Hip osteoarthritis
Joint space metric measurement
X-rays
Concurrent validity
Predictive validity
Responsiveness

SUMMARY

Aim: To perform a systematic review of the literature on the concurrent validity, predictive validity and responsiveness of radiographic metric measurement of femoro-acetabular joint space width (JSW) in hip osteoarthritis (OA).
Methods: Eligibility criteria: studies reporting any data on (1) JSW on X-rays in hip OA patients and (2) concurrent validity (correlations with clinical symptoms), predictive validity (correlations with future symptomatic state, joint space loss or joint replacement), and/or responsiveness (JSW change over time evaluated using the standardized response mean (SRM)). *Search strategy*: Medline PUBMED and Embase databases. *Statistical analysis*: Random-effects models were constructed to obtain pooled SRMs.
Results: Of 448 articles, 79 met the abstract inclusion criteria and were read for further screening. Of these, 15 reported measures of validity and 11 reported measures of responsiveness. *Concurrent validity*: Five studies suggested an association between JSW and symptoms in the general population. Two evaluated the correlations between JSW and symptoms in hip OA patients, with conflicting results. Five demonstrated that JSW is predictive of future hip joint replacement. *Responsiveness* was moderate (SRM = 0.66; 95% confidential interval (95%CI): 0.41, 0.91), but tended to be lower in randomized clinical trials than in cohort studies (0.35 vs 0.83), using an intention to treat rather than a completer analysis (0.30 vs 0.80), and using manual rather than computer-based measurement (0.47 vs 1.12).
Conclusion: There is evidence of a weak association between JSW and symptoms, of predictive validity for subsequent joint replacement, and of moderate responsiveness of metric measurement of JSW.

Osteoarthritis (OA) is a major cause of disability worldwide. For many years, there has been a major interest among the scientific community, pharmaceutical companies, and regulatory agencies in the development of drugs that might influence the natural history of OA by preventing, retarding, or reversing cartilage breakdown. These disease-modifying OA drugs (DMOADs) need to be evaluated in trials using outcomes measures that reflect the natural history of OA. Radiographic variables, particularly metric measurement of minimal joint space width (JSW), are considered the most appropriate structural outcome measure[1]. However, the clinical relevance of this outcome remains doubtful, since there is a debate on whether an association with clinical symptoms exists. Moreover, the responsiveness is questionable since the progression of disease is frequently slow and variable from one patient to another.

Recently, international working groups were created under the auspices of the Food and Drug Administration (FDA) and the Osteoarthritis Research Society International (OARSI) in order to revisit and discuss the outcomes used in OA trials; one of these groups examined the assessment of structural change (ASC). The members of this group agreed that the first stage of their work was to assess the current knowledge on the properties of the instruments used to evaluate structural variables in OA. To assess a potential outcome measure, it is necessary to assess its psychometric properties, as defined by the Outcome Measures in

* Address correspondence and reprint requests to: J.F. Maillefert, Department of Rheumatology, Hôpital Général, 3 rue du Fb Raines, 21000 Dijon, France. Tel: 33-3-80-29-37-45; Fax: 33-3-80-29-36-78.
E-mail address: jean-francis.maillefert@chu-dijon.fr (J.F. Maillefert).

1063-4584/$ – see front matter © 2011 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.joca.2010.12.014

Rheumatology Clinical Trials (OMERACT) filter[2]. The OMERACT filter checks that a potential outcome measure is truthful, reliable, and sensitive to change over time and between different severity stages. This report presents a systematic analysis of the literature performed on the concurrent validity, predictive validity and responsiveness of radiographic metric measurement of hip JSW in hip OA.

Methods

The draft strategy for the literature review was written in December 2008, sent to all members of the ASC working group, underwent iterative revision, and a final version of the protocol was approved in February 2009. The protocol is available and can be obtained from the corresponding author of the present article.

Eligibility criteria

Studies were eligible for analysis when reporting data on hip OA patients (regardless of the definition employed) and including

1 metric measurement of the hip joint JSW on X-rays, irrespective of the measurement technique (manually or computer-based method, evaluation of minimal, mean joint space, or joint area), the study design (cross-sectional or longitudinal), the presence of an intervention or not, or the presence of controls
2 concurrent validity of JSW (correlations with clinical symptoms, in particular pain and function) and/or predictive validity (correlation with future symptoms, joint space loss, or joint replacement), and/or responsiveness (JSW change over time) using either the reported standardized response mean (SRM) or where data allowing calculation of the SRM was available.

Search strategy

A systematic search of the literature was performed in March 2009 and updated in July 2009, using the Medline PUBMED and the Embase databases. The following search terms were used: ((Osteoarthritis[MeSH]) and (hip)) AND (X-ray OR radiography OR diagnostic imaging OR radiology OR disease progression) AND (joint space OR JSW OR disease progression). We limited the search to research conducted in humans and published in English, French, German or Spanish languages.

A quality control of the search terms was performed in January 2009: 30 relevant articles were selected at random from one investigator's personal library. All were found to include the search terms. In addition, a manual search of the references of all screened full-text articles was performed. The abstracts of all potential relevant citations referenced were also screened.

Screening and extraction

All abstracts were read by one reviewer (JFM). Full-text articles were obtained if likely to be relevant or where relevance could not be determined from the abstract.

Criteria for exclusion were: studies reporting results on OA joints other than hip, or combined results on hip and other joint OA which did not present hip results separately, no radiographic evaluation or radiographic data not reported, radiographic assessment not evaluated by metric measurement of JSW (thus excluding studies in which joint space was evaluated using an atlas), secondary OA, and case reports. Reviews, editorials, comments, and systematic literature reviews were not included.

A full-text review of the articles was performed by one reviewer (DCML) using a predetermined data abstraction form approved by the ASC group. The data extracted included the year of publication, name of the first author, study design, X-ray acquisition and measurement technique, evaluated population or patients, demographics, baseline and when available follow-up clinical status (pain, function), baseline and when available follow-up JSW metric measurement, change in JSW (mean and standard deviation), SRM, cross-sectional and longitudinal relationship between JSW metric measurement and clinical status, relationship between JSW and further joint space loss and/or total joint replacement.

After data extraction, a second reviewer (JFM) read all the articles to ensure quality control of data extraction.

Statistical analysis

Responsiveness was assessed by the SRM, defined as the mean change in minimum JSW divided by the standard deviation of change. Articles reporting the SRM or its components were included in the analysis. For randomized clinical trials (RCTs), only the placebo arm was entered to ensure a measure of the natural

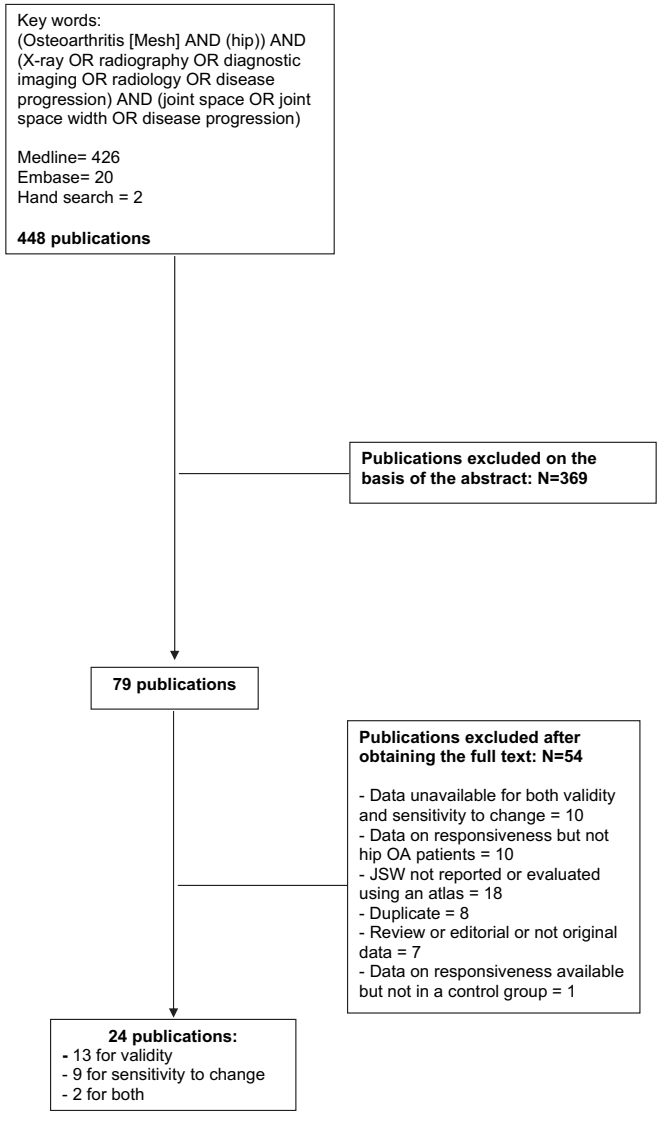

Fig. 1. Flow-chart of the screening process for articles included in the systematic review.

history of disease progression. Pooled estimates of the SRM were performed using random-effects models. We calculated the overall pooled SRM along with the pooled SRM by study design (cohort vs RCT), analysis type (intention to treat (ITT) vs completers), and measurement method (computer vs manual).

Results

We identified 448 articles. Seventy-nine (18%) articles met the initial inclusion criteria and were read for further screening. Of these, 15 (19%) articles reported validity results and 11 (14%) articles reported responsiveness results (Fig. 1).

Concurrent validity

Cross-sectional relationship between JSW and symptoms

Five studies evaluated the correlation between JSW and symptoms in the general population (Table I). In a population-based study (3595 participants), the presence of hip pain, of moderate and severe disability and, to a lesser extent, stiffness, were associated with minimal JSW[3]. In another population-based study (3208 participants), a minimal JSW ≤ 2 mm was significantly associated with self-reported pain in or around the hip joint during the previous 12 months[4]. In 735 participants from the Johnston County Osteoarthritis Project, who had JSW measured at the first follow-up, categorized minimal JSW was not related to pain, but a minimal JSW < 2.5 mm was associated with functional impairment[5]. In a sample of 195 patients presenting with new episodes of pain, there was a negative correlation between JSW and duration of hip pain[6]. In a sample of 220 patients consulting for hip pain, pain duration ≥ 3 months was associated with a minimal JSW ≤ 2.5 mm[7]. In 759 men aged 60–75 years, the prevalence of hip pain was associated with a reduced minimal JSW[8].

There were very few data on the relationship between JSW and symptoms in hip OA patients (Table II). In a sample of 41 hip OA patients, the functional impairment correlated with minimal and sum JSW, in the operated and the contralateral hips[9]. Beside the participants from the Johnston County Osteoarthritis Project, Ref. 5 also provided data from patients included in a 3-year RCT. The baseline clinical parameters explained only 0.4% of the variability of the baseline minimal JSW ($P = 0.44$).[10] In the same sample, categorical JSW was not related to pain nor functional impairment[5].

Longitudinal relationship between JSW and symptoms (Table II)

We did not find any studies that evaluated the relationship between change in symptoms and change in JSW. Two studies evaluated the relationship between baseline symptoms and subsequent joint space loss. In 458 patients included in a 3-year RCT, baseline

Table I
Concurrent validity: correlations between symptoms and hip joint space metric measurement (JSW) in the general population and in patients with hip pain

First author (reference)	Design	Number of subjects/ patients	Mean age: yrs (SD) and % males	Type of JSW	Results
Reijman[3]	Community-based cohort, cross-sectional	3595 subjects aged ≥55 years	66.0 ± 6.9 years, 41.8%	Minimal JSW 7.5% participants with minimal JSW ≤ 2.5 mm, 3.0% with minimal JSW ≤ 2.0 mm, 1.4% with minimal JSW ≤ 1.5 mm	Hip pain associated with minimal JSW ≤ 2.5 mm (OR = 2.4, 95%CI = 1.7–3.4), ≤2.0 mm (OR = 4.5, 95%CI = 2.9–7.0) and ≤1.5 mm (OR = 6.6, 95%CI = 3.6–12.2) Moderate disability associated with minimal JSW ≤ 2.5 mm (OR = 2.7, 95%CI = 2.0–3.7), ≤2.0 mm (OR = 3.7, 95%CI = 2.4–5.9) and ≤1.5 mm (OR = 5.3, 95%CI = 2.9–9.8) Severe disability associated with minimal JSW ≤ 2.5 mm (OR = 3.0, 95%CI = 2.0–4.4), ≤2.0 mm (OR = 4.1, 95%CI = 2.5–7.0) and ≤1.5 mm (OR = 6.1, 95%CI = 3.1–12.1)
Jacobsen[4]	Community-based, cross-sectional	3208	Men: 62.5 (NA) Women: 65.0 (NA), 37.8%	Minimal JSW, 6.0% men and 5.7% women with minimal JSW ≤ 2.0 mm	MJSW ≥ 2 mm significantly associated to self-reported hip pain (OR = 3.5, 95%CI = 2.1–5.7 in men; 1.7, 95%CI = 1.1–2.5 in females), groin pain (OR = 2.3, 95%CI = 1.3–4.1 in men; 2.0, 95%CI = 1.3–3.2 in females), and thigh pain (OR = 1.9, 95%CI = 1.1–3.3 in men; 1.5, 95%CI = 1.0–2.3 in females) during the previous 12 months
Gossec[5]	Subjects from a community-based cohort, cross-sectional	735	67.2 (9.5), 34.3%	Categorical minimal JSW	JSW not related to pain, JSW < 2.5 mm associated with functional impairment, categorized in quartiles (OR = 1.67, 95%CI = 1.0–2.78 compared to JSW > 3 mm)
Birell[6]	Cross-sectional, patients with new episode of hip pain in primary care	195	Median age = 63, 33.3%	Dichotomized minimal JSW, cut off: ≤2.5 mm or 1.5 mm	Pain duration associated with JSW Pain duration < 3 months, 28% with JSW ≤ 2.5 and 7% with JSW ≤ 1.5 mm; Pain duration = 3–12 months, 25% with JSW ≤ 2.5 and 13% with JSW ≤ 1.5 mm; Pain duration > 12 months, 43% with JSW ≤ 2.5 mm and 26% with JSW ≤ 1.5 mm, $P = 0.02$
Bierma-Zienstra[7]	Descriptive, cross-sectional	220	66 (9.6), 27%	Dichotomized minimal JSW, ≤2.5 mm and ≤1.5 mm	JSW ≤ 2.5 mm correlated with pain duration ≥ 3 months (OR = 2.34, 95%CI = 1.26–4.32) and with morning stiffness (OR = 2.0, 95% CI = 1.15–3.62) JSW ≤ 1.5 mm correlated with morning stiffness (OR = 2.6, 95%CI = 1.12–6.06)
Croft[8]	Cross-sectional, men who underwent intravenous urogram	759	Age between 60 and 75 years	Minimal JSW	Pain in 20.4% of hips, 28.3% of hips with JSW ≤ 2.5 mm, and in 56% of hips with JSW ≤ 1.5 mm

OR: Odds Ratio.
95%CI: 95% Confidential Interval.
mm: millimetre.

Table II
Concurrent validity: cross-sectional and longitudinal correlations between symptoms and JSW metric measurement in hip OA patients

Reference	Design	Number of patients	Age, years, mean (SD) and % males	Type of JSW	Results
Amaro[9]	Descriptive, cross-sectional Hip OA patients prior to joint replacement	41	68.4 (9.4), 41%	JSW continuous: minimal and sum (lateral + superior + axial) JSW	Lequesne's index correlated with minimal JSW, r = −0.57, $P < 0.05$ for operated hip and r = −0.70, $P < 0.05$ for non-operated hip Lequesne's index correlated with sum JSW, r = −0.63, $P < 0.05$ for operated hip and r = −0.71, $P < 0.05$ for non-operated hip
Dougados[10]	RCT, cross-sectional and 1-year follow-up	458	63.0 (7.0), 40.4%	JSW continuous: dichotomized change in minimal JSW (≥0.6 mm or not)	Baseline clinical parameters explained only 0.4% of the variability of the baseline JSW ($P = 0.44$) Baseline Lequesne's index > 10 related to 12 months changes in JSW ≥ 0.6 mm (OR = 2.66, 95%CI = 1.46–4.83, $P < 0.0001$)
Gossec[5]	Same RCT as above, cross-sectional	507	63.0 (7.0), 40.4%	Categorical minimal JSW, cut-offs of 1.5, 2.5, and 3.0 mm	JSW not related to pain or functional impairment
Lane[11]	Cohort of women with fractures, aged over 65 years 8-year follow-up	745	71.8 (5.2), 0%	Change in minimal JSW, continuous and dichotomized (> or ≤0.5 mm)	Mean decrease in JSW = 0.5 ± 0.63 and 0.35 ± 0.55 mm in hips with and without baseline pain, respectively ($P = 0.034$) Decrease ≥ 0.5 mm: 53.7% and 30.7% of hips with and without baseline pain, respectively OR = 1.9, 95%CI = 1.4–2.6, $P < 0.001$)

OR: Odds Ratio.
95%CI: 95% Confidential Interval.
mm: millimetre.

Lequesne's index > 10 was an independent predictor of subsequent 1-year change in minimal JSW ≥ 0.6 mm[10]. In a study of 745 women aged over 65 with radiographic hip OA (936 hips), the joint space loss during follow-up was increased in subjects with baseline hip pain[11].

Predictive validity

Prediction of future joint space loss (Table III)

In a retrospective study of 69 patients with hip OA who had undergone total hip replacement (THR), the mean of mean JSW at entry was not related to subsequent annual joint space loss (mean follow-up = 81.2 ± 59.9 months)[12]. In 458 patients included in a 3-year RCT, a baseline minimal JSW < 2.0 mm was an independent predictor of 12 month radiological progression[10].

Prediction of future joint space loss or future joint replacement

In a prospective cohort (mean follow-up = 6.6 ± 0.5 years), a baseline minimal JSW ≤ 2.5 mm was a predictor of a joint space loss ≥ 1.0 mm or a THR on a multivariate analysis performed on all included subjects, but was not on an analysis restricted to the 411 patients with hip pain at baseline[13].

Prediction of total hip joint replacement (Table IV)

A relationship between baseline JSW and later hip replacement was observed in five studies (two of them evaluating the same sample). In a population-based study, a minimal JSW ≤ 2.5 mm was associated with subsequent THR (mean follow-up = 6.6 ± 0.5 years)[3]. In a cohort of 195 patients with a new episode of hip pain, the baseline minimal JSW was predictive of being put on a waiting list for joint replacement (median duration follow-up = 36 months)[14]. In a cohort of 224 subjects aged > 50 years with hip pain followed-up for a mean 2.7 ± 0.25 years then 5.8 ± 0.3 years, a baseline joint space < 2.5 mm was predictive of future joint replacement on unadjusted analysis[15]. In 506 patients included in a 3-year RCT, a baseline minimal JSW < 2 mm and the first year change in minimal JSW were associated with THR during the 2-year follow-up[16]. Patients included in the same RCT were followed-up for an additional 2 years. A decrease of minimal JSW of at least 0.2 mm during the first year predicted joint replacement during the 4 following years and a decrease of minimal JSW of at least 0.4 mm during the first two years predicted joint replacement during the 3 following years[17].

Responsiveness

Data on minimal JSW were extracted from 11 articles (seven cohorts, four RCTs)[11,12,18–26]. Structural assessment analysis was performed as an ITT analysis in three RCTs, and as a completer analysis in the last RCT and in the cohorts. The assessment of minimal JSW was performed using a manual technique in four studies, and a computer-based technique in seven. The mean

Table III
Predictive validity: correlations between JSW metric measurement and future joint space loss in hip OA patients

Reference	Design and follow-up	Number of patients	Age, years, mean (SD) and % males	Type of JSW	Results
Conrozier[12]	Retrospective study from a case registry of patients who had undergone THR for OA, mean radiological follow-up of 81.2 ± 59.9 months	61 patients, 69 hips	Men: 62.0 (10.4) Women: 61.8 (10.4), 44.2%	JSW continuous: mean JSW	The mean JSW at entry was not related to further annual joint space loss
Dougados[10]	3-year RCT	458	63 (7), 40.4%	JSW continuous	Baseline JSW < 2.0 mm was an independent predictor of a further 0–1 year radiological progression, defined as a 1-year JSW loss of at least 0.6 mm (OR = 2.11, 95% CI = 1.30–3.44)

OR: Odds Ratio.
95%CI: 95% Confidential Interval.
mm: millimetre.

Table IV
Predictive validity: correlation between hip joint space metric measurement (JSW) and future THR

Reference	Design and follow-up	Number of subjects/patients	Age, years, mean (SD) and % males	Type of JSW	Results
Reijman[3]	Community-based cohort, mean follow-up = 6.6 ± 0.5 years	3561	67.1 (7.98)	Mean JSW	Baseline JSW ≤ 2.5 mm predicts future THR OR right hip = 18.6, 95%CI = 10.7–32.3 OR left hip = 22.6, 95%CI = 11.8–43.0
Birrell[14]	Cohort of patients with a new episode of hip pain recruited by GPs, median duration follow-up = 36 months	195	63 (11), 32%	Minimal JSW	JSW predictive of future THR In a 0–6 composite score for prediction of THR, the weight of JSW is 2 (joint space > 2.5 = 0, JSW 1.5–2.5 = 1, joint space < 1.5 = 2)
Lievense[15]	Patients aged >50 years with hip pain, followed-up for a mean 2.7 ± 0.25 years then 5.8 ± 0.3 years	193 (mean follow-up 2.7 years) and 163 subjects (mean follow-up = 5.8 years)	65.6 (9.6), 26.9%	Minimal JSW	Baseline JSW < 2.5 mm predictor of future THR on univariate (OR for future 3 years THR = 6.6, $P < 0.01$; OR for future 6 years THR = 7.1, $P < 0.01$), but not on multivariate analysis
Dougados[16]	3-year RCT	506		Minimal JSW: 1-year change in JSW categorized in four grades (no change, worsening < 25%, worsening between 25% and 50%, worsening > 50%)	Baseline JSW < 2 mm associated with a THR during the 3 following years (relative risk = 1.85, 95%CI = 1.18–2.90) First year change in JSW associated with THR during the 2 following years, relative risk of being operated = 2.89; $P < 0.01$ (grade 1 vs 2); 2.09, $P = 0.07$ (grade 2 vs 3); and 5.3, $P < 0.0001$ (grade 3 vs 4)
Maillefert[17]	3-year RCT + 2 years of additional follow-up after end of the trial	422 (first analysis) and 384 (second analysis)	63.0 (6.8), 41.7% (first analysis) and 43.7% (second analysis)	Minimal JSW	A 1-year decrease in JSW ≥ 0.2 mm or 15% predicted THR during the next 4 years (sensibility and specificity of 75% and 68%; 74% and 78%, respectively) Similar results for 0–2 years changes in JSW

OR: Odds Ratio.
95%CI: 95% Confidential Interval.
mm: millimetre.

sample size was 164. Results are shown in Table V. The overall SRM was 0.66 (95% confidential interval (95%CI) = 0.41–0.91). The responsiveness tended to be higher in cohorts (SRM = 0.83; 95%CI: 0.49, 1.16) than in RCTs (SRM = 0.35; 95%CI: 0.12, 0.57). Responsiveness was also higher in analyses of completers (SRM = 0.80; 95%CI: 0.50, 1.10) compared to ITT analyses (SRM = 0.30; 95%CI: 0.06, 0.55). Responsiveness varied by method of measurement, with greater responsiveness seen in studies using computer-based measurement (SRM = 1.12; 95%CI: 0.64, 1.59) compared to manual measurement (SRM = 0.47; 95%CI: 0.31, 0.62).

The data on mean JSW and joint space area were too sparse to allow any pooled analysis. Some studies suggested responsiveness to be comparable to that observed for minimal JSW[12,18,20,27].

Discussion

The present study focused on metric measurement of JSW since it is currently the most frequently used method evaluating structural changes on X-rays in clinical trials[21,22,25,26] and has been demonstrated to be more responsive than other methods, such as the Kellgren and Lawrence or the OARSI grading systems[5]. The main limitation is the heterogeneity of the included studies in their design, inclusion and exclusion criteria, sample size, outcomes.

Table V
Summary of hip responsiveness from radiographs using random-effects pooling of the SRM of the minimum JSW

Analysis	Number of studies	Mean sample size	SRM	95% Confidence Interval
Overall	11	164	0.66	0.41, 0.91
Study design				
RCT	4	111	0.35	0.12, 0.57
Cohort	7	194	0.83	0.49, 1.16
Analysis				
Completers	8	176	0.80	0.50, 1.10
ITT	3	132	0.30	0.06, 0.55
Measurement technique				
Computer	4	40	1.12	0.64, 1.59
Manual	7	234	0.47	0.31, 0.62

The results suggest that, in the general population as well as in the subjects with hip pain, there is an association between minimal JSW and the presence of hip symptoms. Surprisingly, the relationship between JSW and symptoms has rarely been evaluated in hip OA patients. In this review, the results of cross-sectional correlations were too sparse and heterogeneous to allow any conclusion, while longitudinal studies suggested that baseline joint symptoms are moderately correlated to subsequent joint space loss.

Several factors must be taken into account when interpreting these results. First, joint pain is influenced by numerous factors, including patient-related factors. A recent study showed that the relationship between pain and joint space (non-metric measurement) is increased when the patients are their own controls, at least for the knee[28]. It would be interesting to conduct such a study, using JSW metric measurement, in hip OA patients. Second, OA is symptomatically a disease with fluctuating symptoms, which makes it difficult to interpret the correlations between structural data and symptomatic data obtained at only one point in time. Again, additional studies evaluating the relationship between JSW and symptoms obtained at several points of time would be of interest. Third, most studies did not adjust for analgesic and non-steroidal anti-inflammatory drug consumption when evaluating the association between JSW and symptoms. This might alter the associations, at least with respect to pain.

Taken together, the results of this analysis suggest that there is some evidence of a weak association between JSW and symptoms in hip OA. However, additional studies are needed to clarify the association.

Results on predictive validity suggest that absolute levels of JSW might be predictive of later joint space loss, though these data are

heterogeneous; there is more data to suggest that loss of JSW is predictive of subsequent THR. One can question the relevance of joint replacement as an end-point to evaluate the validity of JSW. While arthroplasty is usually performed in patients with advanced symptomatic and structural disease, surgeons have reported that they are weakly or moderately influenced by X-rays when deciding whether joint replacement is indicated or not[29,30]. It has also been shown that in clinical practice, JSW is a major predictive factor of the decision to perform hip replacement[31]. Thus, JSW and joint replacement might not be truly independent. However, the reasons why JSW influences the surgeons' decision remain unclear. If these reasons are differential diagnosis (some surgeons might consider that pain and functional impairment are certainly due to OA in patients with severe joint space narrowing, but might be due, at least in part, to another disease in those with mild joint space narrowing), optional treatments (the surgeons might consider that an additional or complementary medical treatment is less likely to be efficient in patients with severe joint narrowing), and/or disease's potential evolution (surgeons might consider that a spontaneous clinical improvement is less likely to be observed in patients with severe joint loss), joint replacement might be considered as a valid outcome.

The present results suggest good evidence for a moderate responsiveness of JSW in hip OA. It must be pointed out however that the responsiveness tended to be lower in RCTs than in cohort studies, and lower using an ITT rather than a completer analysis (which might explain the higher responsiveness in cohort studies). Potential DMOADs are evaluated using RCTs and an ITT analysis, so the responsiveness of JSW in such studies should be considered as mild, rather than moderate.

Declaration of funding and role of funding source

The OARSI FDA OA Initiative received financial support from the following professional organization: American College of Rheumatology.

Additionally the OARSI FDA OA Initiative received financial support from the following companies: Amgen, ArthroLab, AstraZeneca, Bayer Healthcare, Chondrometrics, CombinatoRx, Cypress BioScience, DePuy Mitek, Expanscience, 4QImaging, Genevrier/IBSA, Genzyme, King (Alpharma), Merck, Merck Serono, NicOx, Pfizer, Rottapharm, Smith & Nephew and Wyeth.

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the FDA.

Author contributions

Delphine Chu Miow Lin: conception and design, data extraction and analysis, redaction of the manuscript.

William Reichmann: conception and design, statistical analysis, data analysis, redaction of the manuscript.

Laure Gossec: conception and design, data extraction, redaction of the manuscript.

Elena Losina: conception and design, statistical analysis, data analysis, redaction of the manuscript.

Philip Conaghan: conception and design, data analysis, redaction of the manuscript.

Jean Francis Maillefert: conception and design, data extraction and analysis, redaction of the manuscript.

Conflict of interest

The authors do not have any conflict of interest to declare.

Acknowledgements

Acknowledgements to the other members of the OARSI/FDA ASC group: B Dardzinski, D Dreher, J Duryea, F Eckstein, D Felson, G Gold, A Guermazi, MP Hellio Le Graverand-Gastineau, J Huckle, D Hunter, H Keen, S Kingsbury, J Kraines, G Lester, R Maciewicz, S Mazzuca, E Morris, T Mosher, JP Pelletier, H Potter, MA Preston, J Randle, L Rovati, E Schneider, B Story, S Totterman, W Tsuji and M Wahba.

References

1. FDA. Clinical Development Programs for Drugs, Devices and Biological Products Intended for the Treatment of OA. FDA; 1999.
2. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25:198–9.
3. Reijman M, Hazes JM, Pols HA, Bernsen RM, Koes BW, Bierma-Zeinstra SM. Validity and reliability of three definitions of hip osteoarthritis: cross sectional and longitudinal approach. Ann Rheum Dis 2004;63:1427–33.
4. Jacobsen S, Sonne-Holm S, Soballe K, Gebuhr P, Lund B. The relationship of hip joint space to self reported hip pain. A survey of 4.151 subjects of the Copenhagen city heart study: the osteoarthritis substudy. Osteoarthritis Cartilage 2004;12:692–7.
5. Gossec L, Jordan JM, Lam MA, Fang FF, Renner JB, Davis A, *et al*. Comparative evaluation of 3 semi-quantitative radiographic grading techniques for hip osteoarthritis in terms of validity and reproducibility in 1404 radiographs: report of the OARSI-OMERACT task force. Osteoarthritis Cartilage 2009;17:182–7.
6. Birrell F, Croft P, Cooper C, Hosie G, Macfarlane GL, Silman A. Radiographic change is common in new presenters in primary care with hip pain. PCR Hip Study Group. Rheumatology 2000;39:772–5.
7. Bierma-Zeinstra SMA, Oster JD, Bernsen RMD, Verhaar JAN, Ginai AZ, Bohnen AM. Joint space narrowing and relationship with symptoms and signs in adults consulting for hip pain in primary care. J Rheumatol 2002;29:1713–8.
8. Croft P, Cooper C, Wickham C, Coggen D. Defining osteoarthritis of the hip for epidemiologic studies. Am J Epidemiol 1990;132:514–22.
9. Amaro A, Amado F, Duarte JA, Appell HJ. Gluteus medius muscle atrophy is related to contralateral and ipsilateral hip joint osteoarthritis. Int J Sports Med 2007;28:1035–9.
10. Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, *et al*. Radiological progression of hip osteoarthritis: definition, risk factors and correlations with clinical status. Ann Rheum Dis 1996;55:356–62.
11. Lane NE, Nevitt MC, Hochberg MC, Hung YY, Palermo L. Progression of radiographic hip osteoarthritis over eight years in a community sample of elderly white women. Arthritis Rheum 2004;50:1477–86.
12. Conrozier T, Jousseaume CA, Mathieu P, Tron AM, Caton J, Bejui J, *et al*. Quantitative measurement of joint space narrowing progression in hip osteoarthritis: a longitudinal retrospective study of patients treated by total hip arthroplasty. Br J Rheumatol 1998;37:961–8.
13. Reijman M, Hazes JM, Pols HA, Bernsen RM, Koes BW, Bierma-Zeinstra SM. Role of radiography in predicting progression of osteoarthritis of the hip: prospective cohort study. BMJ 2005;330:1183.
14. Birrell F, Afzal C, Nahit E, Lunt M, Macfarlane GJ, Cooper C, *et al*. Predictors of hip joint replacement in new attenders in primary care with hip pain. Br J Gen Pract 2003;53:26–30.

15. Lievense AM, Koes BW, Verhaar JA, Bohnen AM, Bierma-Zeinstra SM. Prognosis of hip pain in general practice: a prospective followup study. Arthritis Rheum 2007;57: 1368–74.
16. Dougados M, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, *et al*. Requirement for total hip arthroplasty: an outcome measure of hip osteoarthritis? J Rheumatol 1999;26: 855–61.
17. Maillefert JF, Gueguen A, Nguyen M, Berdah L, Lequesne M, Mazieres B, *et al*. Relevant change in radiological progression in patients with hip osteoarthritis. I. Determination using predictive validity for total hip arthroplasty. Rheumatology 2002;41:142–7.
18. Chevalier X, Conrozier T, Gehrmann M, Claudepierre P, Mathieu P, Unger S, *et al*. Tissue inhibitor of metalloprotease-1 (TIMP-1) serum level may predict progression of hip osteoarthritis. Osteoarthritis Cartilage 2001;9:300–7.
19. Conrozier T, Saxne T, Shan Sei Fan C, Mathieu P, Tron AM, Heinegard D, *et al*. Serum concentration of cartilage oligomeric matrix protein and bone sialoprotein in hip osteoarthritis: a one year prospective study. Ann Rheum Dis 1998;57: 527–32.
20. Dougados M, Villers C, Amor B. Sensitivity to change of various roentgenological severity scoring systems for osteoarthritis of the hip. Rev Rhum 1995;62:169–73.
21. Dougados M, Nguyen M, Berdah L, Mazieres B, Vignon E, Lequesne M. Evaluation of the structure-modifying effects of diacerein in hip osteoarthritis. Arthritis Rheum 2001;44: 2539–47.
22. Lequesne M, Maheu E, Cadet C, Dreiser RL. Structural effect of avocado/soybean unsaponifiables on joint space loss in osteoarthritis of the hip. Arthritis Care Res 2002;47:50–8.
23. Maheu E, Cadet C, Marty M, Dougados M, Ghabri S, Kerloch I, *et al*. Reproducibility and sensitivity to change of various methods to measure joint space width in osteoarthritis of the hip: a double reading of three different radiographic views taken with a three-year interval. Arthritis Res Ther 2005;7: R1375–85.
24. Papacoulas CD, Ward RJ, Tonkin CJ, Buckland-Wright C. Cancellous bone changes in hip osteoarthritis: a short-term longitudinal study using fractal signature analysis. Osteoarthritis Cartilage 2005;13:998–1003.
25. Pavelka K, Gatterova J, Gollerova V, Urbanova Z, Sedlackova M, Altman RD. A 5-year randomized controlled, double-blind study of glycosaminoglycan polysulphuric acid complex (Rumalon®) as a structure modifying therapy in osteoarthritis of the hip and knee. Osteoarthritis Cartilage 2000;8:335–42.
26. Rozendaal RM, Koes BW, van Osch GJVM, Uitterlinden EJ, Garling EH, Willemsen SP, *et al*. Effect of glucosamine sulfate on hip osteoarthritis. Ann Intern Med 2008;148:268–77.
27. Maillefert JF, Sharp JT, Aho LS, Dougados M. Comparison of a computer-based method and the classical manual method for radiographic joint space width assessment in hip osteoarthritis. J Rheumatol 2002;29:2592–6.
28. Neogi T, Felson D, Niu J, Nevitt M, Lewis CE, Aliabadi P, *et al*. Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies. BMJ 2009; 339:b2844.
29. Mancuso CA, Ranawat CS, Esdaile JM, Johanson NA, Charlson ME. Indications for total hip and total knee arthroplasties. Results of orthopaedic surveys. J Arthroplasty 1996; 11:34–46.
30. Dreinhöfer KE, Dieppe P, Stürmer T, Gröber-Grätz D, Flören M, Günther KP, *et al*. Indication for total hip replacement: comparison of assessments of orthopaedic surgeons and referring physicians. Ann Rheum Dis 2006;65:1346–50.
31. Maillefert JF, Roy C, Cadet C, Nizard R, Cohen P, Ravaud P. Factors influencing surgeons' decisions in the indication for total joint replacement in hip osteoarthritis in real life. Arthritis Care Res 2008;59:255–62.

Osteoarthritis and Cartilage 19 (2011) 550–556

Osteoarthritis and Cartilage

Responsiveness to change and reliability of measurement of radiographic joint space width in osteoarthritis of the knee: a systematic review

W.M. Reichmann †‡*, J.F. Maillefert §‖¶, D.J. Hunter #††, J.N. Katz †‡‡§§, P.G. Conaghan |||, E. Losina †‡

† Department of Orthopedic Surgery, Brigham and Women's Hospital, Boston, MA, USA
‡ Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA
§ Department of Rheumatology, Dijon University Hospital, Dijon F-21078, France
‖ University of Burgundy, Dijon F-21079, France
¶ INSERM U887, Dijon F-21079, France
# Rheumatology Department, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Sydney, NSW Australia
†† Division of Research, New England Baptist Hospital, Boston, MA, USA
‡‡ Division of Rheumatology, Immunology and Allergy, Brigham and Women's Hospital, Boston, MA, USA
§§ Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA
||| Section of Musculoskeletal Disease, University of Leeds & NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK

ARTICLE INFO

Article history:
Received 15 June 2010
Accepted 9 January 2011

Keywords:
Knee osteoarthritis
X-ray
Radiograph
Responsiveness
Reliability
Standardized response mean

SUMMARY

Objective: The goal of this systematic review was to report the responsiveness to change and reliability of conventional radiographic joint space width (JSW) measurement.
Method: We searched the PubMed and Embase databases using the following search criteria: [osteoarthritis (OA) (MeSH)] AND (knee) AND (X-ray OR radiography OR diagnostic imaging OR radiology OR disease progression) AND (joint space OR JSW or disease progression). We assessed responsiveness by calculating the standardized response mean (SRM). We assessed reliability using intra- and inter-reader intra-class correlation (ICC) and coefficient of variation (CV). Random-effects models were used to pool results from multiple studies. Results were stratified by study duration, design, techniques of obtaining radiographs, and measurement method.
Results: We identified 998 articles using the search terms. Of these, 32 articles (43 estimates) reported data on responsiveness of JSW measurement and 24 (50 estimates) articles reported data on measures of reliability. The overall pooled SRM was 0.33 [95% confidence interval (CI): 0.26, 0.41]. Responsiveness of change in JSW measurement was improved substantially in studies of greater than 2 years duration (0.57). Further stratifying this result in studies of greater than 2 years duration, radiographs obtained with the knee in a flexed position yielded an SRM of 0.71. Pooled intra-reader ICC was estimated at 0.97 (95% CI: 0.92, 1.00) and the intra-reader CV estimated at 3.0 (95% CI: 2.0, 4.0). Pooled inter-reader ICC was estimated at 0.93 (95% CI: 0.86, 0.99) and the inter-reader CV estimated at 3.4% (95% CI: 1.3%, 5.5%).
Conclusions: Measurement of JSW obtained from radiographs in persons with knee is reliable. These data will be useful to clinicians who are planning RCTs where the change in minimum JSW is the outcome of interest.

Introduction

Knee osteoarthritis (OA) is a painful and disabling disease for many with 12% of adults 60 years of age or older having symptomatic knee OA[1]. As the population ages, the prevalence of knee OA continues to rise. Currently, available pharmacologic regimens for knee OA focus on alleviating pain, but do not slow the structural progression of disease[2]. Disease modifying OA drugs (DMOADS) are in the early developmental stages, and thus it is important to quantify the expected rate of structural progression to facilitate trial planning.

Minimum joint space width (JSW) is commonly used to assess knee OA progression[3]. It has been shown to be sensitive to change[4,5] and change in the minimum JSW has been the primary outcome for previous DMOAD trials[4–7]. An analytic literature

* Address correspondence and reprint requests to: William M. Reichmann, Orthopedic and Arthritis Center for Outcomes Research, Brigham and Women's Hospital, 75 Francis Street, BC 4-4016, Boston, MA 02115, USA. Tel: 1-617-732-5081; Fax: 1-617-525-7900.
E-mail address: wreichmann@partners.org (W.M. Reichmann).

1063-4584/$ – see front matter © 2011 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.joca.2011.01.023

synthesis by Emrani *et al.* in 2008 showed an interaction between study design and radiographic technique was associated with annual change in minimum JSW. The greatest annual change was seen in observational studies that used a semi-flexed technique without fluoroscopy, while the smallest annual change was see in randomized controlled trials (RCTs) with the same technique[5].

The objective of this paper was to update results of Emrani *et al.* by adding the most recent studies and report responsiveness of JSW in terms of standardized response mean (SRM). The SRM is defined as the mean change divided by the standard deviation (SD) of change and can be interpreted as the number of SDs of change, which will be useful for planning future DMOAD trials. We also report pooled estimates of reliability, which include inter- and intra-reader intra-class correlations (ICCs) and coefficients of variation (CVs).

Method

Eligibility criteria

Studies were eligible for our analyses if they satisfied all four requirements of the PICO (Patients Interventions Controls Outcomes). To be included in the review, the study population had to include patients with knee OA followed over time with radiograph-based measures of JSW. We included studies that reported responsiveness (mean change/SD of change or SRM) or reliability measures (inter- or intra-reader ICC or CV). If the study was a RCT then we used data from the control group. This was done to ensure quantification of the natural history of responsiveness of radiographs in those with knee OA. Studies were not limited by publication date (latest search: April 2009) and we included studies that were published in English, French, Spanish, and German.

Information sources and search

We searched the PubMed and Embase databases using the following search criteria: (osteoarthritis [MeSH]) AND (knee) AND (X-ray OR radiography OR diagnostic imaging OR radiology OR disease progression) AND (joint space OR JSW or disease progression).

Study selection

All abstracts were read by one reviewer. The reviewer obtained full-length articles of all abstracts that were considered as probably relevant or of unknown relevance. These articles were subsequently reviewed and data extracted into a data abstraction form. Abstracts of all potentially relevant references in the full-text review were obtained if probably relevant or of unknown relevance.

Studies were excluded if they did not report change in minimum JSW in the knee or if they did not provide a measure of reliability in measuring minimum JSW.

Data items

We abstracted the following study characteristics from each article: study design, radiographic technique, use of fluoroscopy, method of measurement, follow-up time, whether readers were blinded to the order of the radiographic studies, and sample size. Study design was classified as RCT or observational and radiographic technique was categorized as extended view or flexed (includes semi-flexed). Method of measuring minimum JSW was performed manually or using a computer. Follow-up time was categorized as 1-year or less, 1–2 years, or greater than 2 years.

Summary measures

The principal summary measure for our review is the SRM. In articles that reported the SRM directly, we abstracted the reported value. In articles that only reported mean change and SD of change, we calculated the SRM from the two reported measures. Inter- and intra-reader reliability measures (ICC, CV) were also abstracted from the articles.

Synthesis of results

Random-effects models were built to obtain pooled estimates for the SRM and reliability measures across studies adjusting for variability across the studies. Heterogeneity in the estimates was assessed using I-squared, which assesses the percentage of variation across studies that was due to between study variation.

Analyses were performed for all studies that reported these measures and by study characteristics, including study design, radiographic approach, radiographic technique, use of fluoroscopy, method of measurement, and follow-up time. Ninety-five percent CIs were derived for all estimates.

Results

Study selection

We identified 866 articles using our electronic search and another 132 were identified manually for a total of 998 articles. Two-hundred eighty-five articles met the initial abstract screening inclusion criteria and the full-text article was obtained and read for further screening. Of these, 32 articles reported responsiveness results (43 estimates) and 24 articles reported reliability results. Of the 24 articles reporting reliability results, the inter-reader ICC was reported eight times, the intra-reader ICC 17 times, the inter-reader CV six times, and the intra-reader CV 19 times (Fig. 1).

Study characteristics

Of the 43 estimates on responsiveness, 21 (49%) estimates were obtained from studies with follow-up of 1 year or less, 10 (23%) estimates were derived from studies with follow-up of 1–2 years, and 12 (28%) came from studies with greater than 2 years of follow-up. The mean sample size was 100 (SD = 86). Sixteen estimates (37%) were obtained from studies that used a radiographic approach with the knee fully extended and 27 (63%) from studies

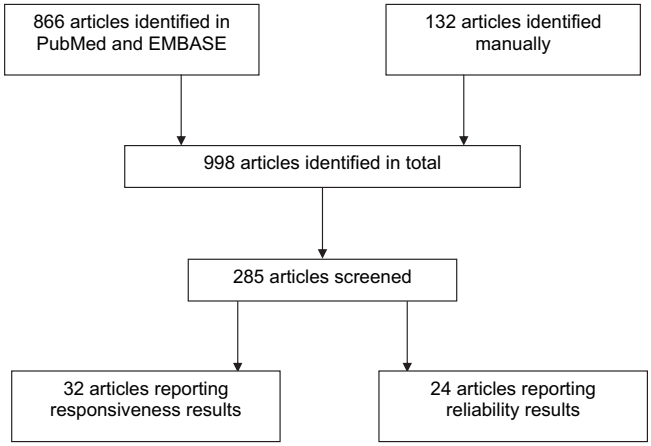

Fig. 1. Flow chart of the screening process for articles included in the systematic review.

that had the knee in flexion. Fluoroscopy was used for 23 (53%) of the estimates and computerized methods of measuring the minimum JSW was used for 24 of the estimates (56%). Nineteen (44%) of the estimates came from RCTs. Of the 43 estimates, only 21 (49%) disclosed whether the readers were blinded to the sequence of the radiographs. Of these 21 estimates, 19 (90%) came from studies that used blinded readers. Study characteristics for all 32 studies are shown in Table I.

Of the eight estimates evaluating the inter-reader ICC, four (50%) used a fully extended radiographic approach, four (50%) used fluoroscopy, and seven (88%) measured the joint space manually. The mean sample size in these studies was 110 (SD = 110).

Of the 17 estimates evaluating the intra-reader ICC, six (35%) used a fully extended radiographic approach, eight (47%) used fluoroscopy, and nine (53%) measured the joint space manually. The mean sample size in these studies was 80 (SD = 88).

Of the six estimates evaluating the inter-reader CV, three (50%) used a fully extended radiographic approach, six (100%) used fluoroscopy, and six (100%) measured the joint space manually. The mean sample size in these studies was 120 (SD = 99).

Of the 19 estimates evaluating the intra-reader CV, six (32%) used a fully extended radiographic approach, 14 (74%) used fluoroscopy, and 11 (58%) measured the joint space manually. The mean sample size was 43 (SD = 38). Study characteristics for all 24 studies are shown in Table II.

Synthesis of responsiveness results

The I-squared value for the 43 estimates was 0.82 [95% confidence interval (CI): 0.76, 0.86] indicating substantial between study variation. The I-squared values are shown in Table III.

The random-effects analysis yielded an overall pooled SRM for the 43 estimates of 0.33 (95% CI: 0.26, 0.41). The pooled SRM was similar when the analysis was stratified by radiographic approach, the use of fluoroscopy, measurement method, and study type. Follow-up time was related to the magnitude of the SRM. Estimates derived from studies with 1 year or less and 1–2 years of follow-up had similar responsiveness (0.24 and 0.25 respectively), while estimates coming from studies with greater than 2 years follow-up had an SRM of 0.57 (95% CI: 0.39, 0.75). Similar effects of follow-up time are shown when use of fluoroscopy, measurement method, and study type were stratified by follow-up time. However, when radiographic approach was stratified by follow-up time, estimates derived from studies that used a flexion-based radiographic

Table I
Study Characteristics of the manuscripts reviewed for responsiveness

Author, year (Ref.)	Study type	Sample size	Follow-up months	Radiographic approach	Method of measurement	Delta (SD)
Ayral *et al.* 1996[8]	Cohort	41	12	Extension without fluoroscopy	Manual	0.40 (1.00)
Ravaud *et al.* 1996[9]	Cohort	55	12	Extension without fluoroscopy	Manual	0.42 (1.11)
Listrat *et al.* 1997[10]	RCT	17	12	Extension without fluoroscopy	Manual	0.70 (1.20)
Pavelka *et al.* 2000[4]	RCT	139	60	Extension with fluoroscopy	Manual	0.42 (0.94)
Mazzuca *et al.* 2001[6]	Cohort	402	31.60	Extension without fluoroscopy	Manual	0.37 (1.25)
Reginster *et al.* 2001[7]	RCT	106	36	Extension with fluoroscopy	Computerized	0.40 (0.92)
Gandy *et al.* 2002[11]	Cohort	11	37	Extension without fluoroscopy	Manual	0.21 (0.37)
Miyazaki *et al.* 2002[12]	Cohort	74	72	Flexion without fluoroscopy	Manual	1.40 (1.20)
Boegard *et al.* 2003[13]	Cohort	50	25	Flexion with fluoroscopy	Manual	0.06 (0.45)
Mazzuca *et al.* 2003[14]	Cohort	52	14	Flexion with fluoroscopy	Computerized	0.09 (0.31)
		52	14	Flexion without fluoroscopy	Manual	−0.09 (0.66)
Pessis *et al.* 2003[15]	Cohort	20	12	Flexion with fluoroscopy	Manual	0.00 (0.60)
		20	12	Extension with fluoroscopy	Manual	0.10 (0.90)
Sugiyama *et al.* 2003[16]	Cohort	110	48	Flexion with fluoroscopy	Computerized	0.53 (0.43)
Vignon *et al.* 2003[17]	Cohort	58	24	Extension with fluoroscopy	Computerized	0.17 (0.75)
		58	24	Flexion with fluoroscopy	Computerized	0.24 (0.50)
Pavelka *et al.* 2004[18]	RCT	89	24	Extension with fluoroscopy	Manual	0.40 (0.79)
Pham *et al.* 2004[19]	RCT	79	12	Extension without fluoroscopy	Manual	0.21 (0.59)
		69	12	Extension without fluoroscopy	Manual	0.12 (0.32)
Pham *et al.* 2004[20]	RCT	277	12	Extension without fluoroscopy	Manual	0.09 (0.55)
Uebelhart *et al.* 2004[21]	RCT	76	12	Extension without fluoroscopy	Computerized	0.32 (1.11)
Brandt *et al.* 2005[22]	RCT	180	30	Flexion with fluoroscopy	Manual	0.45 (0.70)
Conrozier *et al.* 2005[23]	Cohort	96	12	Flexion with fluoroscopy	Computerized	0.19 (0.48)
Michel *et al.* 2005[24]	RCT	150	24	Flexion without fluoroscopy	Computerized	0.07 (0.56)
Spector *et al.* 2005[25]	RCT	98	12	Flexion with fluoroscopy	Computerized	0.12 (0.42)
Bingham *et al.* 2006[26]	RCT	269	24	Flexion with fluoroscopy	Computerized	0.13 (1.08)
		280	24	Flexion with fluoroscopy	Computerized	0.09 (1.31)
Cline *et al.* 2006[27]	RCT	112	9.84	Flexion without fluoroscopy	Computerized	0.00 (0.53)
		85	11.76	Flexion with fluoroscopy	Computerized	0.12 (0.42)
		99	8.16	Flexion without fluoroscopy	Computerized	−0.07 (0.63)
Mikesky *et al.* 2006[28]	RCT	60	30	Flexion with fluoroscopy	Manual	0.54 (0.70)
Botha-Scheepers *et al.* 2007[29]	Cohort	122	24	Flexion without fluoroscopy	Computerized	0.21 (0.52)
Krzeski *et al.* 2007[30]	RCT	71	12	Extension with fluoroscopy	N/A	0.14 (0.53)
Nevitt *et al.* 2007[31]	Cohort	53	37	Flexion without fluoroscopy	Computerized	0.43 (0.66)
Sharif *et al.* 2007[32]	Cohort	115	60	Extension without fluoroscopy	Manual	0.18 (0.93)
Le Graverand *et al.* 2008[33]	Cohort	62	12	Flexion with fluoroscopy	Computerized	0.22 (0.41)
		62	12	Flexion without fluoroscopy	Computerized	−0.01 (0.46)
Mazzuca *et al.* 2008[34]	Cohort	27	12	Flexion without fluoroscopy	Computerized	0.25 (0.54)
		27	12	Flexion without fluoroscopy	Computerized	0.02 (0.40)
		47	12	Flexion with fluoroscopy	Computerized	0.16 (0.37)
		47	12	Flexion with fluoroscopy	Computerized	−0.01 (0.51)
Gensburger *et al.* 2009[35]	Cohort	81	48	Flexion with fluoroscopy	Computerized	0.32 (0.76)
Kahan *et al.* 2009[36]	RCT	313	12	Flexion with fluoroscopy	Computerized	0.31 (0.71)

Delta: change in minimum JSW from baseline to follow-up (measured in millimeters).

Table II
Study characteristics of the manuscripts reviewed for reliability

Author, year (Ref.)	Sample size	Radiographic approach	Method of measurement	Reliability estimator	Observer	Value
Buckland-Wright *et al.* 1995[37]	5	Flexion with fluoroscopy	Computer	CV	Intra	3.8%
	5	Flexion with fluoroscopy	Computer	CV	Intra	1.2%
	7	Flexion with fluoroscopy	Manual	CV	Intra	3.6%
	7	Flexion with fluoroscopy	Manual	CV	Intra	0.6%
Ravaud *et al.* 1996[9]	55	Extension without fluoroscopy	Manual	ICC	Intra	0.95
	55	Extension without fluoroscopy	Manual	ICC	Inter	0.85
Pavelka *et al.* 2000[4]	10	Extension with fluoroscopy	Manual	CV	Intra	2.0%
	10	Extension with fluoroscopy	Manual	ICC	Intra	0.99
	280	Extension with fluoroscopy	Manual	CV	Inter	6.6%
	280	Extension with fluoroscopy	Manual	ICC	Inter	0.97
Mazzuca *et al.* 2001[6]	20	Extension without fluoroscopy	Manual	CV	Intra	4.4%
Myazaki *et al.* 2002[12]	10	Flexion without fluoroscopy	Manual	ICC	Intra	0.92
Pavelka *et al.* 2002[38]	40	Extension with fluoroscopy	Manual	CV	Intra	1.9%
	202	Extension with fluoroscopy	Manual	CV	Inter	2.6%
Boegard *et al.* 2003[13]	51	Flexion with fluoroscopy	Manual	CV	Intra	2.3%
	51	Flexion with fluoroscopy	Manual	CV	Intra	1.0%
	51	Flexion with fluoroscopy	Manual	CV	Inter	2.7%
	51	Flexion with fluoroscopy	Manual	CV	Inter	1.1%
Mazzuca *et al.* 2003[14]	71	Flexion without fluoroscopy	Manual	CV	Intra	5.8%
Sugiyama *et al.* 2003[16]	10	Flexion with fluoroscopy	Computer	CV	Intra	1.5%
Vignon *et al.* 2003[17]	20	Extension with fluoroscopy	Computer	ICC	Intra	0.98
	36	Flexion with fluoroscopy	Computer	ICC	Intra	0.98
Mazzuca *et al.* 2004[39]	30	Flexion with fluoroscopy	Manual	ICC	Intra	0.996
	30	Flexion with fluoroscopy	Manual	ICC	Inter	0.956
Pavelka *et al.* 2004[18]	89	Extension with fluoroscopy	Manual	CV	Intra	3.6%
	89	Extension with fluoroscopy	Manual	CV	Inter	6.5%
Pham *et al.* 2004[19]	156	Extension without fluoroscopy	Manual	ICC	Intra	0.993
Pham *et al.* 2004[20]	292	Extension without fluoroscopy	Manual	ICC	Intra	0.996
	292	Extension without fluoroscopy	Manual	ICC	Inter	0.912
Sharif *et al.* 2004[40]	20	Extension without fluoroscopy	Manual	CV	Intra	11.3%
Cicuttini *et al.* 2005[41]	123	Extension without fluoroscopy	Computer	CV	Intra	4.8%
Conrozier *et al.* 2005[23]	106	Flexion with fluoroscopy	Computer	CV	Intra	1.15%
	106	Flexion with fluoroscopy	Computer	ICC	Intra	0.99
Michel *et al.* 2005[24]	284	Flexion without fluoroscopy	Computer	ICC	Intra	0.98
Szebenyi *et al.* 2006[42]	60	Extension without fluoroscopy	Manual	ICC	Intra	0.895
	60	Extension without fluoroscopy	Manual	ICC	Inter	0.868
Nevitt *et al.* 2007[31]	80	Flexion without fluoroscopy	Manual	ICC	Intra	0.90
	80	Flexion without fluoroscopy	Manual	ICC	Inter	0.98
	25	Flexion without fluoroscopy	Computer	ICC	Intra	0.96
	25	Flexion without fluoroscopy	Computer	CV	Intra	2.9%
Le Graverand *et al.* 2008[33]	36	Flexion with fluoroscopy	Computer	ICC	Intra	0.99
	36	Flexion with fluoroscopy	Computer	ICC	Intra	0.99
	18	Flexion without fluoroscopy	Computer	ICC	Intra	0.99
Mazzuca *et al.* 2008[34]	39	Flexion with fluoroscopy	Computer	CV	Intra	0.80
Gensburger *et al.* 2009[35]	42	Flexion with fluoroscopy	Manual	ICC	Intra	0.89
	42	Flexion with fluoroscopy	Manual	CV	Intra	2.9%
	44	Flexion with fluoroscopy	Manual	ICC	Inter	0.80
	44	Flexion with fluoroscopy	Manual	CV	Inter	0.8%
Kahan *et al.* 2009[36]	100	Flexion with fluoroscopy	Computer	CV	Intra	1.2%
	100	Flexion with fluoroscopy	Computer	ICC	Intra	0.99

approach and had greater than 2 years of follow-up time had a higher SRM of 0.71 (95% CI: 0.44, 0.98).

Synthesis of reliability results

Results of random-effects pooling of the reliability estimates showed good inter- and intra-reader reliability for measuring the minimum JSW. The eight estimates of inter-reader ICC produced an estimate of 0.93 (95% CI: 0.86, 0.99), while the 17 estimates of intra-reader ICC produced an estimate of 0.97 (95% CI: 0.92, 1.00). Additional results stratified by study characteristics are shown in Table IV. Six estimates for the inter-reader CV produced an estimate of 3.4% (95% CI: 1.3%, 5.5%) and 19 estimates for the intra-reader CV produced an estimate of 3.0% (95% CI: 2.0%, 4.0%). Additional results stratified by study characteristics are shown in Table V.

Discussion

We performed an analytic systematic review of the responsiveness and reliability of knee radiographs when measuring the minimum JSW. We analyzed responsiveness using the SRM. This measure can be interpreted as the number of SDs of change. The overall SRM was 0.33 (95% CI: 0.26, 0.41). Follow-up time was the main study characteristic that was related to responsiveness. Studies with follow-up times greater than 2 years showed greater responsiveness (SRM = 0.57; 95% CI: 0.39, 0.75). It is critical to note that

Table III
Results of random-effects pooling for studies that reported estimates of responsiveness by different study characteristics

	Number of estimates	I-squared (95% CI)	SRM (95% CI)
Overall	43	0.82 (0.76, 0.86)	0.33 (0.26, 0.41)
Knee flexion			
Extended	16	0.19 (0.00, 0.55)	0.32 (0.26, 0.37)
Flexed	27	0.88 (0.84, 0.91)	0.34 (0.22, 0.45)
Fluoroscopy			
Fluoro	23	0.83 (0.76,0.88)	0.38 (0.27, 0.48)
No fluoro	20	0.79 (0.69 0.86)	0.28 (0.17, 0.39)
Measurement method			
Manual	18	0.80 (0.70, 0.87)	0.38 (0.26, 0.50)
Computerized	24	0.84 (0.77, 0.89)	0.31 (0.20, 0.41)
Study type			
RCT	19	0.82 (0.73, 0.88)	0.30 (0.20, 0.40)
Cohort	24	0.82 (0.74, 0.87)	0.36 (0.24, 0.49)
Follow-up time			
1-year or less	21	0.56 (0.27, 0.73)	0.24 (0.15, 0.32)
1–2 years	10	0.80 (0.63, 0.89)	0.25 (0.13, 0.37)
Greater than 2 years	12	0.88 (0.81, 0.93)	0.57 (0.39, 0.75)
Reader blinded to order of radiographs			
Yes	19	0.76 (0.63, 0.85)	0.30 (0.19, 0.40)
No	2	0.59 (0.00, 0.90)	0.55 (0.33, 0.76)
Unknown	22	0.85 (0.78, 0.89)	0.35 (0.23, 0.46)
Knee flexion by follow-up time			
Extended/1-year or less	9	0.00 (0.00, 0.63)	0.26 (0.19, 0.34)
Extended/1–2 years	2	0.61 (0.00, 0.91)	0.38 (0.10, 0.65)
Extended/greater than 2 years	5	0.32 (0.00, 0.74)	0.34 (0.24, 0.44)
Flexed/1-year or less	12	0.68 (0.42, 0.83)	0.19 (0.06, 0.32)
Flexed/1–2 years	8	0.82 (0.65, 0.90)	0.22 (0.08, 0.36)
Flexed/greater than 2 years	7	0.88 (0.78, 0.94)	0.71 (0.44, 0.98)
Fluoroscopy by follow-up time			
Fluoro/1-year or less	9	0.33 (0.00, 0.69)	0.29 (0.18, 0.39)
Fluoro/1–2 years	7	0.81 (0.62, 0.91)	0.29 (0.14, 0.44)
Fluoro/greater than 2 years	7	0.87 (0.75, 0.93)	0.58 (0.36, 0.80)
No fluoro/1-year or less	12	0.61 (0.28, 0.79)	0.21 (0.10, 0.32)
No fluoro/1–2 years	3	0.82 (0.45, 0.94)	0.15 (−0.13, 0.42)
No fluoro/greater than 2 years	5	0.89 (0.78, 0.95)	0.56 (0.24, 0.87)
Measurement method by follow-up time			
Manual/1-year or less	8	0.20 (0.00, 0.63)	0.28 (0.17, 0.38)
Manual/1–2 years	2	0.92 (0.73, 0.98)	0.19 (−0.44, 0.82)
Manual/greater than 2 years	8	0.87 (0.77, 0.93)	0.51 (0.31, 0.71)
Computerized/1-year or less	12	0.68 (0.42, 0.83)	0.21 (0.08, 0.33)
Computerized/1–2 years	8	0.78 (0.56, 0.89)	0.26 (0.13, 0.38)
Computerized/greater than 2 years	4	0.90 (0.77, 0.96)	0.68 (0.31, 1.06)
Study type by follow-up time			
RCT/1-year or less	10	0.60 (0.19, 0.80)	0.21 (0.11, 0.32)
RCT/1–2 years	5	0.87 (0.72, 0.94)	0.24 (0.07, 0.41)
RCT/greater than 2 years	4	0.51 (0.00, 0.84)	0.56 (0.41, 0.70)
Cohort/1-year or less	11	0.51 (0.03, 0.75)	0.26 (0.13, 0.40)
Cohort/1–2 years	5	0.69 (0.20, 0.88)	0.26 (0.06, 0.46)
Cohort/greater than 2 years	8	0.92 (0.86, 0.95)	0.57 (0.30, 0.85)

Table IV
Results of random-effects pooling for studies that reported estimates of intra-ICC by different study characteristics

	Number of estimates	Inter-reader ICC (95% CI)	Number of estimates	Intra-reader ICC (95% CI)
Overall	8	0.93 (0.86, 0.99)	17	0.97 (0.92, 1.00)
Knee flexion				
Extended	4	0.93 (0.85, 1.00)	6	0.98 (0.90, 1.00)
Flexed	4	0.94 (0.79, 1.00)	11	0.97 (0.90, 1.00)
Fluoroscopy				
Fluoro	4	0.95 (0.85, 1.00)	8	0.98 (0.88, 1.00)
No fluoro	4	0.91 (0.82, 1.00)	9	0.97 (0.91, 1.00)
Measurement method				
Manual	7	0.93 (0.86, 0.99)	9	0.97 (0.89, 1.00)
Computerized	1	0.99 (N/A)	8	0.98 (0.90, 1.00)

Table V
Results of random-effects pooling for studies that reported estimates of CV by different study characteristics

	Number of estimates	Inter-reader CV (95% CI)	Number of estimates	Intra-reader CV (95% CI)
Overall	6	3.4% (1.3, 5.5)	19	3.0% (2.0, 4.0)
Knee flexion				
Extended	3	5.2% (2.5, 8.0)	6	4.7% (2.7, 6.7)
Flexed	3	1.5% (0.3, 2.7)	13	2.2% (1.3, 3.2)
Fluoroscopy				
Fluoro	6	3.4% (1.3, 5.5)	14	2.0% (1.4, 2.5)
No fluoro	0	N/A	5	5.8% (3.8, 7.9)
Measurement method				
Manual	6	3.4% (1.3, 5.5)	11	3.6% (2.1, 5.1)
Computerized	0	N/A	8	2.2% (0.8, 3.5)

studies with a follow-up of 1 year or shorter showed a responsiveness of 0.24. This limitation of the radiographic technique means that to adequately power a study to demonstrate change over this short interval will require much larger sample sizes. Studies that used a flexed view and had greater than 2 years of follow-up showed the greatest responsiveness (SRM = 0.71; 95% CI: 0.44, 0.98). Based upon this literature there does appear to be some advantage to standardized positioning and fluoroscopy with slight improvements in responsiveness. Despite what one may have expected there does not appear to be any advantage in computerized measurement of JSW over manual measures. In studies with greater than 2 years of follow-up, the responsiveness was higher for those that used computerized methods (0.68) compared to those that used manual methods (0.51). However, the 95% CIs substantially overlap due to substantial variability in these estimates (Table III).

The reliability of measuring minimum JSW provided to be excellent with pooled ICCs ranging from 0.91 to 0.99 and pooled CVs ranging from 1.5 to 5.8. Radiographic method, use of fluoroscopy, and measurement method did not affect reliability albeit the majority of the estimates come from different studies with no direct study comparison.

Our findings complement the work of Emrani *et al.* who published a systematic review in 2008 on the change in minimum JSW. While they found effects of radiographic approach and study type, they also analyzed the crude change in minimum JSW rather than the SRM. They also found that increased follow-up time was inversely associated with change in minimum JSW, while we found that increasing the follow-up time increased the responsiveness of radiographs to change. This difference may be due to differences in definition of primary outcomes and additional assumption of linearity of change that Emrani *et al.* used in their analysis[5].

A major strength of this study is that it is the first literature synthesis to summarize responsiveness in terms of the SRM. These data will be useful to clinicians who are planning studies where the change in the minimum JSW is the outcome of interest. The results of this analysis suggest that studies using JSW as primary outcome measure based on radiographs should plan to have a follow-up period that is greater than 2 years and have the knee in a flexed position when performing the radiographs to ensure the greatest possible responsiveness. While the pooled SRM was higher for studies that did not blind the reader to the sequence of the radiographs (0.55), it is unlikely that blinding of the readers of the radiographs substantially influenced our results since only two estimates came from studies that did not blind their readers. Also, the pooled SRM for estimates coming from studies that did blind the readers was similar to those that did not report this information (0.30 vs 0.35 respectively).

Also, this is the first known literature synthesis that pools reliability data on measuring minimum JSW. In general, these measurements can be considered to be reliable as the intra- and inter-reader ICCs were large and the CVs were low.

A major limitation of our review is that we did not report our results by risk factors for knee OA progression (body mass index, knee alignment, age, concurrent OA in other joints, synovitis, etc.) since they were not uniformly reported. The fact that we were not able to account for these factors may have contributed to the heterogeneity in the SRMs. It is important for future studies that report results on quantitative changes of knee OA progression to report these risk factors. Also, we did not collect data on the number of readers and the time interval between reads for our reliability data. It would be interesting to examine how these factors affected our estimates of reliability.

We found that radiographs provide moderate responsiveness and good reliability measures for measuring the minimum JSW in persons with knee OA. These data will be useful to clinicians who wish to plan future RCTs in which change in minimum JSW is their primary outcome.

Declaration of funding and role of funding source

The Osteoarthritis Research Society International (OARSI) Food and Drug Administration (FDA) OA Initiative received financial support from the following professional organization: American College of Rheumatology.

Additionally the OARSI FDA OA Initiative received financial support from the following companies: Amgen, ArthroLab, AstraZeneca, Bayer Healthcare, Chondrometrics, CombinatoRx, Cypress BioScience, DePuy Mitek, Expanscience, 4QImaging, Genevrier/IBSA, Genzyme, King (Alpharma), Merck, Merck Serono, NicOx, Pfizer, Rottapharm, Smith & Nephew, Wyeth.

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the FDA. This work was also supported in part by the National Institutes of Health (K24 AR057827 and T32 AR055885) and they had no role in the content being published.

Author contributions

- Conception and design (WMR, JFM, EL).
- Analysis and interpretation of the data (WMR, JFM, DJH, PGC, JNK, EL).
- Drafting of the article (WMR, JFM, DJH, PGC, JNK, EL).
- Critical revision of the article for important intellectual content (WMR, JFM, DJH, PGC, JNK, EL).
- Final approval of the article (WMR, JFM, DJH, PGC, JNK, EL).
- Statistical expertise (WMR, EL).
- Collection and assembly of data (WMR, JFM).

The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflict of interest

WR, JFM, JK, PC, EL: no conflict of interest to declare.
DH: receives research or institutional support from DonJoy, NIH, and Stryker.

Acknowledgments

We recognize the invaluable support of Valorie Thompson for administrative and editorial support, the OA community for their feedback and OARSI for their invaluable support of this activity. This analysis and literature review was undertaken to facilitate discussions and development of recommendations by the Assessment of Structural Change Working group for the OARSI FDA Initiative.

References

1. Dillon CF, Rasch EK, Gu Q, Hirsch R. Prevalence of knee osteoarthritis in the United States: arthritis data from the Third National Health and Nutrition Examination Survey 1991-94. J Rheumatol 2006;33:2271–9.
2. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, *et al*. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis Cartilage 2007;15:981–1000.
3. Lequesne M. Quantitative measurements of joint space during progression of osteoarthritis: chondrometry. In: Kuettner KE, Goldberg VM, Eds. Osteoarthritic Disorders. Rosemont: American Academy of Orthopaedic Surgeons; 1995:427–44.
4. Pavelka K, Gatterova J, Gollerova V, Urbanova Z, Sedlackova M, Altman RD. A 5-year randomized controlled, double-blind study of glycosaminoglycan polysulphuric acid complex (Rumalon) as a structure modifying therapy in osteoarthritis of the hip and knee. Osteoarthritis Cartilage 2000;8:335–42.
5. Emrani PS, Katz JN, Kessler CL, Reichmann WM, Wright EA, McAlindon TE, *et al*. Joint space narrowing and Kellgren–Lawrence progression in knee osteoarthritis: an analytic literature synthesis. Osteoarthritis Cartilage 2008;16:873–82.
6. Mazzuca SA, Brandt KD, Dieppe PA, Doherty M, Katz BP, Lane KA. Effect of alignment of the medial tibial plateau and X-ray beam on apparent progression of osteoarthritis in the standing anteroposterior knee radiograph. Arthritis Rheum 2001;44:1786–94.
7. Reginster JY, Deroisy R, Rovati LC, Lee RL, Lejeune E, Bruyere O, *et al*. Long-term effects of glucosamine sulphate on osteoarthritis progression: a randomised, placebo-controlled clinical trial. Lancet 2001;357:251–6.
8. Ayral X, Dougados M, Listrat V, Bonvarlet JP, Simonnet J, Amor B. Arthroscopic evaluation of chondropathy in osteoarthritis of the knee. J Rheumatol 1996;23:698–706.
9. Ravaud P, Giraudeau B, Auleley GR, Chastang C, Poiraudeau S, Ayral X, *et al*. Radiographic assessment of knee osteoarthritis: reproducibility and sensitivity to change. J Rheumatol 1996;23:1756–64.
10. Listrat V, Ayral X, Patarnello F, Bonvarlet JP, Simonnet J, Amor B, *et al*. Arthroscopic evaluation of potential structure modifying activity of hyaluronan (Hyalgan) in osteoarthritis of the knee. Osteoarthritis Cartilage 1997;5:153–60.
11. Gandy SJ, Dieppe PA, Keen MC, Maciewicz RA, Watt I, Waterton JC. No loss of cartilage volume over three years in patients with knee osteoarthritis as assessed by magnetic resonance imaging. Osteoarthritis Cartilage 2002;10:929–37.
12. Miyazaki T, Wada M, Kawahara H, Sato M, Baba H, Shimada S. Dynamic load at baseline can predict radiographic disease progression in medial compartment knee osteoarthritis. Ann Rheum Dis 2002;61:617–22.
13. Boegard TL, Rudling O, Petersson IF, Jonsson K. Joint space width of the tibiofemoral and of the patellofemoral joint in chronic knee pain with or without radiographic osteoarthritis: a 2-year follow-up. Osteoarthritis Cartilage 2003;11:370–6.
14. Mazzuca SA, Brandt KD, Buckwalter KA. Detection of radiographic joint space narrowing in subjects with knee osteoarthritis: longitudinal comparison of the metatarsophalangeal and semiflexed anteroposterior views. Arthritis Rheum 2003;48:385–90.

15. Pessis E, Drape JL, Ravaud P, Chevrot A, Dougados M, Ayral X. Assessment of progression in knee osteoarthritis: results of a 1 year study comparing arthroscopy and MRI. Osteoarthritis Cartilage 2003;11:361–9.
16. Sugiyama S, Itokazu M, Suzuki Y, Shimizu K. Procollagen II C propeptide level in the synovial fluid as a predictor of radiographic progression in early knee osteoarthritis. Ann Rheum Dis 2003;62:27–32.
17. Vignon E, Piperno M, Le Graverand MP, Mazzuca SA, Brandt KD, Mathieu P, *et al*. Measurement of radiographic joint space width in the tibiofemoral compartment of the osteoarthritic knee: comparison of standing anteroposterior and Lyon Schuss views. Arthritis Rheum 2003;48:378–84.
18. Pavelka K, Forejtova S, Olejarova M, Gatterova J, Senolt L, Spacek P, *et al*. Hyaluronic acid levels may have predictive value for the progression of knee osteoarthritis. Osteoarthritis Cartilage 2004;12:277–83.
19. Pham T, Maillefert JF, Hudry C, Kieffert P, Bourgeois P, Lechevalier D, *et al*. Laterally elevated wedged insoles in the treatment of medial knee osteoarthritis. A two-year prospective randomized controlled study. Osteoarthritis Cartilage 2004;12:46–55.
20. Pham T, Le Henanff A, Ravaud P, Dieppe P, Paolozzi L, Dougados M. Evaluation of the symptomatic and structural efficacy of a new hyaluronic acid compound, NRD101, in comparison with diacerein and placebo in a 1 year randomised controlled study in symptomatic knee osteoarthritis. Ann Rheum Dis 2004;63:1611–7.
21. Uebelhart D, Malaise M, Marcolongo R, de Vathaire F, Piperno M, Mailleux E, *et al*. Intermittent treatment of knee osteoarthritis with oral chondroitin sulfate: a one-year, randomized, double-blind, multicenter study versus placebo. Osteoarthritis Cartilage 2004;12:269–76.
22. Brandt KD, Mazzuca SA, Katz BP, Lane KA, Buckwalter KA, Yocum DE, *et al*. Effects of doxycycline on progression of osteoarthritis: results of a randomized, placebo-controlled, double-blind trial. Arthritis Rheum 2005;52:2015–25.
23. Conrozier T, Mathieu P, Piperno M, Favret H, Colson F, Vignon M, *et al*. Selection of knee radiographs for trials of structure-modifying drugs in patients with knee osteoarthritis: a prospective, longitudinal study of Lyon Schuss knee radiographs with the definition of adequate alignment of the medial tibial plateau. Arthritis Rheum 2005;52:1411–7.
24. Michel BA, Stucki G, Frey D, De Vathaire F, Vignon E, Bruehlmann P, *et al*. Chondroitins 4 and 6 sulfate in osteoarthritis of the knee: a randomized, controlled trial. Arthritis Rheum 2005;52:779–86.
25. Spector TD, Conaghan PG, Buckland-Wright JC, Garnero P, Cline GA, Beary JF, *et al*. Effect of risedronate on joint structure and symptoms of knee osteoarthritis: results of the BRISK randomized, controlled trial [ISRCTN01928173]. Arthritis Res Ther 2005;7:R625–33.
26. Bingham 3rd CO, Buckland-Wright JC, Garnero P, Cohen SB, Dougados M, Adami S, *et al*. Risedronate decreases biochemical markers of cartilage degradation but does not decrease symptoms or slow radiographic progression in patients with medial compartment osteoarthritis of the knee: results of the two-year multinational knee osteoarthritis structural arthritis study. Arthritis Rheum 2006;54:3494–507.
27. Cline GA, Meyer JM, Stevens R, Buckland-Wright C, Peterfy C, Beary JF. Comparison of fixed flexion, fluoroscopic semi-flexed and MTP radiographic methods for obtaining the minimum medial joint space width of the knee in longitudinal osteoarthritis trials. Osteoarthritis Cartilage 2006;14(Suppl A):A32–6.
28. Mikesky AE, Mazzuca SA, Brandt KD, Perkins SM, Damush T, Lane KA. Effects of strength training on the incidence and progression of knee osteoarthritis. Arthritis Rheum 2006;55: 690–9.
29. Botha-Scheepers S, Kloppenburg M, Kroon HM, Hellio Le Graverand MP, Breedveld FC, Ravaud P, *et al*. Fixed-flexion knee radiography: the sensitivity to detect knee joint space narrowing in osteoarthritis. Osteoarthritis Cartilage 2007;15:350–3.
30. Krzeski P, Buckland-Wright C, Balint G, Cline GA, Stoner K, Lyon R, *et al*. Development of musculoskeletal toxicity without clear benefit after administration of PG-116800, a matrix metalloproteinase inhibitor, to patients with knee osteoarthritis: a randomized, 12-month, double-blind, placebo-controlled study. Arthritis Res Ther 2007;9. R109.
31. Nevitt MC, Peterfy C, Guermazi A, Felson DT, Duryea J, Woodworth T, *et al*. Longitudinal performance evaluation and validation of fixed-flexion radiography of the knee for detection of joint space loss. Arthritis Rheum 2007;56:1512–20.
32. Sharif M, Kirwan J, Charni N, Sandell LJ, Whittles C, Garnero P. A 5-yr longitudinal study of type IIA collagen synthesis and total type II collagen degradation in patients with knee osteoarthritis–association with disease progression. Rheumatology (Oxford) 2007;46:938–43.
33. Le Graverand MP, Vignon EP, Brandt KD, Mazzuca SA, Piperno M, Buck R, *et al*. Head-to-head comparison of the Lyon Schuss and fixed flexion radiographic techniques. Long-term reproducibility in normal knees and sensitivity to change in osteoarthritic knees. Ann Rheum Dis 2008;67:1562–6.
34. Mazzuca SA, Hellio Le Graverand MP, Vignon E, Hunter DJ, Jackson CG, Kraus VB, *et al*. Performance of a non-fluoroscopically assisted substitute for the Lyon Schuss knee radiograph: quality and reproducibility of positioning and sensitivity to joint space narrowing in osteoarthritic knees. Osteoarthritis Cartilage 2008;16:1555–9.
35. Gensburger D, Arlot M, Sornay-Rendu E, Roux JP, Delmas P. Radiologic assessment of age-related knee joint space changes in women: a 4-year longitudinal study. Arthritis Rheum 2009; 61:336–43.
36. Kahan A, Uebelhart D, De Vathaire F, Delmas PD, Reginster JY. Long-term effects of chondroitins 4 and 6 sulfate on knee osteoarthritis: the study on osteoarthritis progression prevention, a two-year, randomized, double-blind, placebo-controlled trial. Arthritis Rheum 2009;60:524–33.
37. Buckland-Wright JC, MacFarlane DG, Lynch JA, Jasani MK. Quantitative microfocal radiography detects changes in OA knee joint space width in patients in placebo controlled trial of NSAID therapy. J Rheumatol 1995;22:937–43.
38. Pavelka K, Gatterova J, Olejarova M, Machacek S, Giacovelli G, Rovati LC. Glucosamine sulfate use and delay of progression of knee osteoarthritis: a 3-year, randomized, placebo-controlled, double-blind study. Arch Intern Med 2002;162:2113–23.
39. Mazzuca SA, Brandt KD, Buckwalter KA, Lequesne M. Pitfalls in the accurate measurement of joint space narrowing in semi-flexed, anteroposterior radiographic imaging of the knee. Arthritis Rheum 2004;50:2508–15.
40. Sharif M, Kirwan JR, Elson CJ, Granell R, Clarke S. Suggestion of nonlinear or phasic progression of knee osteoarthritis based on measurements of serum cartilage oligomeric matrix protein levels over five years. Arthritis Rheum 2004;50:2479–88.
41. Cicuttini F, Hankin J, Jones G, Wluka A. Comparison of conventional standing knee radiographs and magnetic resonance imaging in assessing progression of tibiofemoral joint osteoarthritis. Osteoarthritis Cartilage 2005;13:722–7.
42. Szebenyi B, Hollander AP, Dieppe P, Quilty B, Duddy J, Clarke S, *et al*. Associations between pain, function, and radiographic features in osteoarthritis of the knee. Arthritis Rheum 2006; 54:230–5.

Osteoarthritis and Cartilage 19 (2011) 557–588

Osteoarthritis and Cartilage

Systematic review of the concurrent and predictive validity of MRI biomarkers in OA

D.J. Hunter †‡*, W. Zhang §, Philip G. Conaghan ||, K. Hirko ‡, L. Menashe ‡, L. Li ‡, W.M. Reichmann ¶, E. Losina ¶

† Rheumatology Department, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Sydney, NSW, Australia
‡ Division of Research, New England Baptist Hospital, Boston, MA, USA
§ Academic Rheumatology, The University of Nottingham, Nottingham, UK
|| Section of Musculoskeletal Disease, University of Leeds & NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK
¶ Department of Orthopedic Surgery, Brigham and Women's Hospital, Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

ARTICLE INFO

Article history:
Received 15 June 2010
Accepted 17 October 2010

Keywords:
Osteoarthritis
Magnetic resonance imaging
Validity

SUMMARY

Objective: To summarize literature on the concurrent and predictive validity of MRI-based measures of osteoarthritis (OA) structural change.
Methods: An online literature search was conducted of the OVID, EMBASE, CINAHL, PsychInfo and Cochrane databases of articles published up to the time of the search, April 2009. 1338 abstracts obtained with this search were preliminarily screened for relevance by two reviewers. Of these, 243 were selected for data extraction for this analysis on validity as well as separate reviews on discriminate validity and diagnostic performance. Of these 142 manuscripts included data pertinent to concurrent validity and 61 manuscripts for the predictive validity review. For this analysis we extracted data on criterion (concurrent and predictive) validity from both longitudinal and cross-sectional studies for all synovial joint tissues as it relates to MRI measurement in OA.
Results: Concurrent validity of MRI in OA has been examined compared to symptoms, radiography, histology/pathology, arthroscopy, CT, and alignment. The relation of bone marrow lesions, synovitis and effusion to pain was moderate to strong. There was a weak or no relation of cartilage morphology or meniscal tears to pain. The relation of cartilage morphology to radiographic OA and radiographic joint space was inconsistent. There was a higher frequency of meniscal tears, synovitis and other features in persons with radiographic OA. The relation of cartilage to other constructs including histology and arthroscopy was stronger. Predictive validity of MRI in OA has been examined for ability to predict total knee replacement (TKR), change in symptoms, radiographic progression as well as MRI progression. Quantitative cartilage volume change and presence of cartilage defects or bone marrow lesions are potential predictors of TKR.
Conclusion: MRI has inherent strengths and unique advantages in its ability to visualize multiple individual tissue pathologies relating to pain and also predict clinical outcome. The complex disease of OA which involves an array of tissue abnormalities is best imaged using this imaging tool.

Introduction

Magnetic Resonance Imaging (MRI) is being developed as a method to assess joint morphology in osteoarthritis (OA), with the goal of providing a sensitive non-invasive tool for the study of healthy and diseased states, and a means of assessing the effectiveness of interventions for osteoarthritis. Traditionally structural assessment of OA has relied upon the plain radiograph which has capacity to image the joint space and osteophytes[1]. MRI has many advantages in visualizing the joint, and recent efforts are yielding a variety of approaches that offer the potential for monitoring this prevalent synovial joint disease[2]. Because OA is a disease of the whole synovial joint, not just the cartilage, measurements of *structure* need to be seen broadly and capture important anatomic features, including osteophytes, effusions, meniscal tears,

* Address correspondence and reprint requests to: David J. Hunter, Rheumatology Department, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Sydney, NSW, Australia. Tel.: 61-2-9926-7379; Fax: 61-2-9906-1859.
E-mail address: David.Hunter@sydney.edu.au (D.J. Hunter).

1063-4584/$ – see front matter
doi:10.1016/j.joca.2010.10.029

subchondral bone architectural changes or ligamentous instability, in addition to cartilage loss[2]. There is an abundant literature describing the concurrent validity of MRI as it relates to comparable constructs such as histology and radiography but little if any effort has been made to systematically summarize this literature.

Similarly the merits of any OA structural assessment will undoubtedly be assessed for their clinical relevance. There are multiple determinants of pain and functional limitation in OA and there may be many more unknown[3]. Many studies have examined whether the loss of structural integrity is in some way the physical correlate of these symptoms. Traditionally most epidemiologic studies have relied upon plain radiography to define disease. The major limitation of this method is that measures of symptoms correlate poorly with x-ray features. Less than 50% of people with evidence of OA on plain radiographs have symptoms related to these findings[4]. Uncertainty as to whether measurements of MRI structure alone will adequately reflect what structure connotes, or whether other metrics of structure should also be considered, need to be systematically evaluated. The relationships between structure and pain and/or function and between structure and future outcomes (e.g., arthroplasty) are critical in determining the clinical relevance of MRI.

In psychometrics, validity refers to the degree to which a study accurately reflects or assesses the specific concept that the researcher is attempting to measure. There are many types of validity of which one, criterion validity, is used to demonstrate the accuracy of a measure or procedure by comparing it with another measure or procedure which has been demonstrated to be valid. There is a contention in the OA field about the validity of a number of biomarkers and clinical endpoints and their inclusion here is in an effort to be comprehensive and does not diminish the credible concerns about the lack of well validated clinical endpoints[5]. If the test data and criterion data are collected at the same time, this is referred to as concurrent validity evidence. If the test data is collected first in order to predict criterion data collected at a later point in time, then this is referred to as predictive validity evidence. The purpose of this systematic review was to summarize the OA MRI literature with regards to both concurrent and predictive validity.

Material and methods

Systematic literature search details

An online literature search was conducted using the OVID MEDLINE (1945–), EMBASE (1980–) and Cochrane databases (1998–) to identify the articles published up to April 2009, with the search entries "MRI", and "osteoarthritis", "osteoarthritides", "osteoarthrosis", "osteoarthroses", "degenerative arthritis", "degenerative arthritides", or "osteoarthritis deformans". The abstracts of the 1330 citations received with this search were then preliminarily screened for relevance by two reviewers (KH and DJH). For this preliminary search, all articles which used MRI, in some form, on patients with osteoarthritis of the knee, hip, or hand were included. Although review articles were not included (see Inclusion/exclusion criteria), citations found in any review articles which were not already included in our preliminary search were screened for possible inclusion in this study. This added 7 more relevant studies to our search. One further article was added, before publication, by one of authors of this meta-analysis bringing the preliminary total to 1338.

Inclusion/exclusion criteria

Only studies published in English were included. Studies presenting non-original data were excluded, such as reviews, editorials, opinion papers, or letters to the editor. Studies with questionable clinical relevance and those using non-human subjects or specimens were excluded. Studies in which rheumatoid, inflammatory, or other forms of arthritis were included in the OA datasets were excluded, as well as general joint-pertinent MRI studies not focused on OA. Studies with no extractable, numerical data were excluded. Only those articles which had some measure of diagnostic performance were included. Any duplicates which came up in the preliminary search were excluded. Of the preliminary 1338 abstracts, 243 were selected for data extraction (Fig. 1).

Data abstraction

Two reviewers (KH and LM) independently abstracted the following data: (1) patient demographics; (2) MRI make, sequences and techniques used, tissue types viewed; (3) study type and funding source; (4) details on rigor of study design to construct the Downs methodological quality score[6]; (5) MRI reliability/reproducibility data; (6) MRI diagnostic measures and performance; (7) gold standard measures against which the MRI measure was evaluated; (8) treatment and MRI measures (when appropriate).

The Downs methodological quality score[6] collects a profile of scores for both randomized trials and observational studies in terms of quality of reporting, internal validity (bias and confounding), power, external validity so that the overall study quality score reflects all of these elements. Answers were scored 0 (No) or 1 (Yes), except for one item in the Reporting subscale, which scored 0–2 and the single item on power, which was scored 0–5. The possible range is from 0 to 27 where 0 represents poor quality and 27 optimal quality.

We used a data abstraction tool constructed in EpiData (Entry version 2.0 Odense, Denmark) and more than one reviewer undertook the data abstraction. The data collection forms were designed to target the objectives of the review, and were piloted prior to conducting the study.

The outcomes for psychometric properties on MRI were examined using the OMERACT filter[7,8]. The specific focus of this review is upon the truth domain: is the measure truthful, does it measure what it intends to measure? More specifically we were interested in criterion validity; for both the concurrent [Does it agree (by independent and blind comparison) with a measure that reflects the same concept] and predictive [Does it predict (by independent and blind comparison) a future 'gold standard'] validity of MRI in OA. If

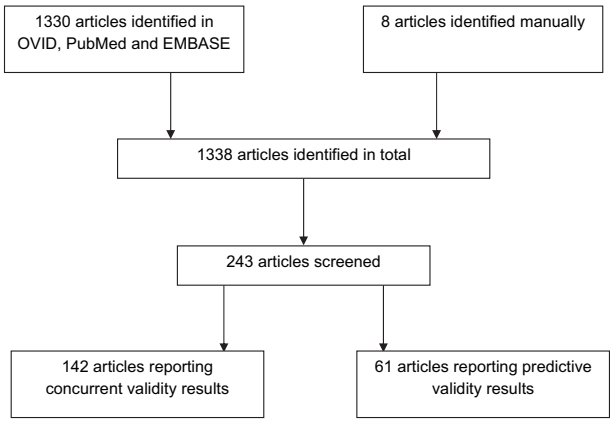

Fig. 1. Flow chart of the screening process for articles included in the systematic review.

the test data and criterion data are collected at the same time, this is referred to as concurrent validity evidence. If the test data is collected first in order to predict criterion data collected at a later point in time, then this is referred to as predictive validity evidence.

It is critical to delineate what we mean by the various terms used, as current usage is often incorrect, and this ambiguity may stem from an incorrect understanding of appropriate definitions. Whilst there are several definitions that have been proposed[9–13], the brief synthesis of some working definitions is as follows:

1. *biological marker* (biomarker)— a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes or pharmacologic responses to a therapeutic agent.
2. *clinical endpoint*—a clinically meaningful measure of how a patient feels, functions, or survives.

For the purposes of this analysis, MRI (the biomarker) is directly compared both to clinical endpoints (symptoms, total knee replacement (TKR)) as well as other biomarkers (including radiography, CT, histology, arthroscopy, alignment). The presentation of the data in the results reflects presentation of clinical endpoints before comparison with other biomarkers.

There is some overlap in the manuscripts for which data is extracted for these two types of validity. The large majority of studies for concurrent validity were cross-sectional studies although some longitudinal studies reported cross-sectional results and thus are included in the concurrent validity data. There is no attempt made to create summary estimates as the validity effect measures [i.e., odds ratio (OR), Beta coefficient, *r*, *P*-value of difference] used in this literature are very heterogeneous.

Results

Concurrent validity (Table I)

The analysis included data from 142 manuscripts. The mean Downs criteria score for these manuscripts was 8.3 (range 3–17). What follows below are important excerpts from this data pertaining to different aspects of concurrent validity. The data is further summarized in Table II to discretely identify the associations examined and those where a significant association was found.

Relation to symptoms

21 studies examined the concurrent relation of MRI findings in OA to symptoms. Of these, 62% demonstrated a statistically significant association, defined as $P < 0.05$. Bone marrow lesions were found in 272 of 351 (77.5%) persons with painful knees compared with 15 of 50 (30%) persons with no knee pain ($P < 0.001$). Large lesions were present almost exclusively in persons with knee pain (35.9% vs 2%; $P < 0.001$). After adjustment for severity of radiographic disease, effusion, age, and sex, lesions and large lesions remained associated with the occurrence of knee pain [odds ratio, 3.31 (95% confidence interval (CI), 1.54–7.41)]. Using the same analytical approach, large lesions were also strongly associated with the presence of pain [odds ratio, 5.78 (CI, 1.04–111.11)]. Among persons with knee pain, bone marrow lesions were not associated with pain severity[14].

- After adjusting for the severity of radiographic OA, there was a difference between those with and without knee pain in prevalence of moderate or larger effusions ($P < 0.001$) and synovial thickening, independent of effusion ($P < 0.001$). Among those with small (grade 1) or no knee (grade 0) effusion, those with knee pain had a prevalence of synovial thickening of 73.6% compared to 21.4% of those without knee pain ($P < 0.001$). There was a significant difference in visual analogue scale (VAS) pain scores in those with synovial thickening compared to those without synovial thickening, after adjustment for radiographic severity, size of effusion, age, sex, and BMI. The mean pain score in those with synovial thickening after adjustment for radiographic severity and size of effusion was 47.2 mm [standard error (SE) 6.0], compared to 28.2 mm (SE 2.8) in those without synovial thickening ($P = 0.006$)[15].
- A medial or lateral meniscal tear was a very common finding in the asymptomatic subjects (prevalence, 76%) but was more common in the patients with symptomatic osteoarthritis (91%) ($P < 0.005$). There was no significant difference with regard to the pain or WOMAC score between the patients with and those without a medial or lateral meniscal tear in the osteoarthritic group ($P = 0.8$ to 0.9 for all comparisons)[16].
- Significant differences between WOMAC scores were found for the grades of cartilage lesions ($P < 0.05$) but not bone marrow edema pattern, and ligamentous and meniscal lesions[17].
- Bone marrow lesions >1 cm were more frequent (OR = 5.0; 95% CI = 1.4, 10.5) in the painful knee OA group than all other groups. While the frequency of BME lesions was similar in the painless OA and painful OA groups, there were more lesions, >1 cm, in the painful OA group. Full-thickness cartilage defects occurred frequently in painful OA. Women with radiographic OA, full-thickness articular cartilage defects, and adjacent subchondral cortical bone defects were significantly more likely to have painful knee OA than other groups (OR = 3.2; 95% CI = 1.3, 7.6)[18].
- Peripatellar lesions (prepatellar or superficial infrapatellar) were present in 12.1% of the patients with knee pain and ROA, in 20.5% of the patients with ROA and no knee pain, and in 0% of subjects with neither ROA nor knee pain ($P = 0.116$). However, other periarticular lesions (including bursitis and iliotibial band syndrome) were present in 14.9% of patients with both ROA and knee pain, in only 3.9% of patients with ROA but no knee pain, and in 0% of the group with no knee pain and no ROA ($P = 0.004$)[19].
- More severe symptoms relating to knee OA (pain, stiffness, and function) are weakly inversely related to tibial cartilage volume. Patients with lower cartilage volume had more severe symptoms of knee OA than those with higher cartilage volume[20].
- The increase in median pain from median quantile regression, adjusting for age and BMI, was significant for bone attrition (1.91, 95% CI 0.68, 3.13), bone marrow lesions (3.72, 95% CI 1.76, 5.68), meniscal tears (1.99, 95% CI 0.60, 3.38), and grade 2 or 3 synovitis/effusion vs grade 0 (9.82, 95% CI 0.38, 19.27). The relationship with pain severity was of borderline significance for osteophytes and cartilage morphology and was not significant for bone cysts or meniscal subluxation. When compared to the pain severity in knees with high scores for both bone attrition and bone marrow lesions (median pain severity 40 mm), knees with high attrition alone (30 mm) were not significantly different, but knees with high bone marrow lesion without high attrition scores (15 mm) were significantly less painful[21].
- A large joint effusion was associated with pain (OR, 9.99; 99% CI: 1.28, 149) and stiffness (OR, 4.67; 99% CI: 1.26, 26.1). The presence of an osteophyte in the patellofemoral compartment (OR, 2.25; 99% CI: 1.06, 4.77) was associated with pain. All other imaging findings, including focal or diffuse cartilaginous abnormalities, subchondral cysts, bone marrow edema,

Table I
Summary table of studies reporting data on concurrent validity of MRI in OA

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, mean(SD), range	No. (%) of females	Quantitative cartilage	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Score of methodological quality
Chan WP; American Journal of Roentgenology; 1991; 1892040[36]	20	20	0	58(Range: 42–73)	11	No	No	Yes	Yes	No	No	No	Yes	Yes	Cross-sectional	6
McAlindon TE; Annals of the Rheumatic Diseases; 1991; 1994861[90]	12					No	No	Yes	Yes	Yes	Yes	No	Yes	No	Case control	3
Li KC; Magnetic Resonance Imaging; 1988; 3398728[91]	10	10	0	(Range: 33–78)	9(90%)	No	No	Yes	Yes	No	No	No	No	No	Longitudinal Prospective	4
Fernandez-Madrid F; Magnetic Resonance Imaging; 1994; 7934656[92]	92	52	40	Controls: 49(15), (Rang: 22–78); OA patients: 55(14), (Range: 25–86)	60	No	No	Yes	Yes	Yes	No	No	Yes	No	Cross-sectional	11
Karvonen RL; Journal of Rheumatology; 1994; 7966075[27]	92	52	40	Reference: 49(15), (Range: 22–78); All OA patients: 55(14), (Range: 25–86); Bilateral OA: 53(13), (Range: 25–73)	60	Yes	No	No	Yes	No	Yes	No	No	No	Case control	11
Peterfy CG; Radiology; 1994; 8029420[93]	8	5	3	62(Range: 45–82)	4(50%)	Yes	No	No	Yes	No	No	No	No	No	Cross-sectional	4
Blackburn WD Jr; Journal of Rheumatology; 1994; 8035392[37]	33	33	0	62.7(9.1), (Range: 44–79)	17	No	No	Yes	Yes	No	No	No	No	No	Cross-sectional	6
Broderick LS; American Journal of Roentgenology; 1994; 8273700[61]	23	13	10			No	No	Yes	Yes	No	No	No	No	No	Cross-sectional	4
Miller TT; Radiology; 1996; 8816552[94]	384			47(Range: 14–88)		No	No	Yes	Yes	No	No	No	Yes	Yes	Cross-sectional	8
Dupuy DE; Academic Radiology; 1996; 8959181[57]	7			TKA patients: (Range: 64–75); Asymptomatic: 35(Range: 25–35)	3	Yes	No	No	Yes	No	No	No	No	No	Other	6
Kenny C; Clinical Orthopaedics & Related Research; 1997; 9186215[95]	136					No	No	Yes	No	No	No	No	Yes	No	Case control	6
Breitenseher MJ; Acta Radiologica; 1997; 9332248[96]	60	12	48	37(14.3), (Range: 15–68)	30(50%)	No	No	Yes	No	No	No	No	Yes	No	Cross-sectional	5

Ostergaard M; British Journal of Rheumatology; 1997; 9402860[97]	46	14	47	70(Range: 24–85)		No	No	No	No	Yes	No	No	No	No	Cross-sectional	7
Trattnig S; Journal of Computer Assisted Tomography; 1998; 9448754[98]	20	20	0	72.2(Range: 62–82)	18	No	No	Yes	Yes	No	No	No	No	No	Other	8
Kawahara Y; Acta Radiologica; 1998; 9529440[62]	72			58(Range: 41–74)	46	No	No	Yes	Yes	No	No	No	No	No	Other	6
Drape JL; Radiology; 1998; 9646792[63]	43	43	0	63(Range: 53–78)	30	No	No	Yes	Yes	No	No	No	No	No	Cross-sectional	5
Eckstein F; Clinical Orthopaedics & Related Research; 1998; 9678042[56]	8	0	8	50.6(Range: 39–64)		Yes	No	No	Yes	No	No	No	No	No	Other	7
Uhl M; European Radiology; 1998; 9724423[58]	22			(Range: 50–72)		No	No	Yes	Yes	No	No	No	No	No	Cross-sectional	5
Boegard T; Acta Radiologica - Supplementum; 1998; 9759121[99]	61					No	No	Yes	Yes	No	No	No	Yes	No	Longitudinal Prospective	5
Bachmann GF; European Radiology; 1999; 9933399[64]	320			29.3(8.7), (Range: 13–56)	122	No	No	Yes	Yes	No	No	No	Yes	No	Cross-sectional	7
Cicuttini F; Osteoarthritis & Cartilage; 1999; 10329301[100]	28			Males: 40.4(Range: 42–58); Females: 31.2(8.6);	11	Yes	No	No	Yes	No	No	No	No	No	Cross-sectional	7
Boegard T; Annals of the Rheumatic Diseases; 1999; 10343536[101]	58			Women: 40.4(Range: 42–58); Men: 57(49.5), (Range: 41–57)	29	No	No	Yes	Yes	No	No	No	No	No	Cross-sectional	6
Adams JG; Clinical Radiology; 1999; 10484216[44]	62	32	30			No	No	Yes	Yes	No	No	No	Yes	No	Case control	8
Pham XV; Revue du Rhumatisme; 1999; 10526380[102]	10	10	10	67.2(7.34), (Range: 57–80)	6	No	No	Yes	No	No	No	No	No	Yes	Cross-sectional	13
Gale DR; Osteoarthritis & Cartilage; 1999; 10558850[43]	291	233	58			No	No	No	No	No	No	No	Yes	No	Case control	10
Kladny B; International Orthopaedics; 1999; 10653290[59]	26					Yes	No	No	Yes	No	No	No	No	No	Cross-sectional	6
Zanetti M; Radiology; 2000; 10831707[103]	16	16	0	67(Range: 43–79)	15	Yes	No	No	Yes	No	Yes	Yes	No	No	Cross-sectional	6

(*continued on next page*)

Table I (*continued*)

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, mean(SD), range	No. (%) of females	Quantitative cartilage	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Score of methodological quality
Jones G; Arthritis & Rheumatism; 2000; 11083279 [104]	92	92	0	Boys: 12.8(2.7); Girls: 12.6(2.9)	43	Yes	No	No	Yes	No	Yes	No	No	No	Cross-sectional	13
McCauley TR; American Journal of Roentgenology; 2001; 11159074[105]	193			40(Range: 11–86)	83	No	No	Yes	Yes	No	No	No	Yes	Yes	Cross-sectional	8
Wluka AE; Annals of the Rheumatic Diseases; 2001; 11247861[106]	81	42	39	Cases: 58(6.1); Controls: 56(5.4)	81(100%)	Yes	No	Yes	Yes	No	No	No	No	No	Case control	16
Felson DT; Annals of Internal Medicine; 2001; 11281736[14]	401	401	0	66.8		No	No	Yes	No	No	No	Yes	No	No	Cross-sectional	13
Hill CL; Journal of Rheumatology; 2001; 11409127[15]	458	433	25	67	(34%)	No	No	Yes	No	Yes	No	No	No	No	Case control	13
Kawahara Y; Journal of Computer Assisted Tomography; 2001; 11584226[107]	35			57(Range: 33–70)	23	No	No	Yes	Yes	No	No	No	Yes	No	Cross-sectional	8
Arokoski JP; Annals of the Rheumatic Diseases; 2002; 11796401[108]	57	27	30	Cases: 56.2(4.9), (Range: 47–64); Controls: 56.3(4.5), (Range: 47–64)	0	Yes	No	No	No	No	No	No	No	No	Case control	8
Bergin D; Skeletal Radiology; 2002; 11807587[109]	60	30	30	Cases: 50; Controls: 57		No	No	Yes	No	No	No	No	Yes	Yes	Case control	9
Beuf O; Arthritis & Rheumatism; 2002; 11840441[110]	46	18	28	Mild OA: 68(9.1); Severe OA: 70(6.3)	17	Yes	No	No	No	No	No	No	No	No	Case control	5
Arokoski MH; Journal of Rheumatology; 2002; 12375331[111]	57	27	30	Cases: 56.2(4.9), (Range: 47–64); Controls: 56.3(4.5), (Range: 47–64)	0	Yes	No	No	Yes	No	No	No	No	No	Case control	8
Bhattacharyya T; Journal of Bone & Joint Surgery - American Volume; 2003; 12533565[16]	203	154	49	Cases: 65; Controls: 67		No	No	Yes	No	No	No	No	Yes	No	Case control	9
Link TM; Radiology; 2003; 12563128[17]	50	50	0	63.7(11.5), (Range: 43–81)	30	No	No	Yes	Yes	No	No	No	Yes	Yes	Cross-sectional	6

Tiderius CJ; Magnetic Resonance in Medicine; 2003; 12594751[112]	17			50(Range: 35–70)	4	No	Yes	No	Yes	No	No	No	No	No	Cross-sectional	6
Cicuttini FM; Arthritis & Rheumatism; 2003; 12632421[28]	252			60.2(10)	157962%)	Yes	No	No	Yes	No	Yes	No	No	No	Cross-sectional	9
Cicuttini FM; Clinical & Experimental Rheumatology; 2003; 12673893[113]	81	42	39	ERT: 58(6.1); Controls: 56(5.4)	81(100%)	Yes	No	No	Yes	No	Yes	No	No	No	Case control	12
Sowers MF; Osteoarthritis & Cartilage; 2003; 12801478[18]	120	60	60	no OAK, no Pain: 45(0.8); OAK, no Pain: 46(0.6); No OAK, Pain: 47(0.8); OAK and Pain: 47(0.7)	(100%)	No	No	Yes	Yes	No	No	Yes	No	No	Case control	11
McGibbon CA; Osteoarthritis & Cartilage; 2003; 12814611[60]	4					No	No	Yes	Yes	No	No	No	No	No	Other	5
Cicuttini FM; Clinical & Experimental Rheumatology; 2003; 12846050[46]	157	157	0	62(10)	(62%)	Yes	No	No	Yes	No	No	No	No	No	Cross-sectional	10
Felson DT; Annals of Internal Medicine; 2003; 12965941[51]	256	256	0	Followed: 66.2(9.4); Not followed: 67.8(9.6)	(38.3%)	No	No	Yes	No	No	No	Yes	No	No	Longitudinal prospective	11
Tarhan S; Clinical Rheumatology; 2003; 14505208[114]	74	58	16	OA Patients: 57.4(8.5), (Range: 45–75); Healthy controls: 59.1(5.8), (Range: 46–77)	60	Yes	No	Yes	Yes	Yes	No	No	No	No	Case control	8
Hill CL; Arthritis & Rheumatism; 2003; 14558089[19]	451	427		Knee pain/ROA/Male: 68.3; Knee pain/ROA/ Female: 65; No knee pain/ROA/Male: 66.8; No knee pain/ROA/ Female: 66.1		No	No	Yes	No	No	No	Yes	No	No	Cross-sectional	10
Kim YJ; Journal of Bone & Joint Surgery - American Volume; 2003; 14563809[115]	43			30(Range: 11–47); Median = 31	40	No	Yes	No	Yes	No	No	No	No	No	Other	5

(*continued on next page*)

Table I (*continued*)

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, mean(SD), range	No. (%) of females	Quantitative cartilage	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Score of methodological quality
Lindsey CT; Osteoarthritis & Cartilage; 2004; 14723868[29]	74	33	21	Controls: 34.2(12.5); OA1 (KL1/2): 62.7(10.9); OA2(KL3/4): 66.6(11.6)	39	Yes	No	No	Yes	No	Yes	No	No	No	Case control	8
Jones G; Osteoarthritis & Cartilage; 2004; 14723876[47]	372	186	186	45(Range: 26–61)		Yes	No	No	Yes	No	Yes	No	No	No	Case control	9
Raynauld JP; Arthritis & Rheumatism; 2004; 14872490[48]	32	32	0	62.9(8.2)	(74%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	10
Wluka AE; Annals of the Rheumatic Diseases; 2004; 14962960[20]	132	132	0	63.1(Range: 41–86)	71(54%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	10
Cicuttini F; Rheumatology; 2004; 14963201[52]	117	117	0	67(10.6)	(58%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	12
Peterfy CG; Osteoarthritis & Cartilage; 2004; 14972335[116]	19	19	0	61(8)	4	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Other	5
Graichen H; Arthritis & Rheumatism; 2004; 15022323[117]	21	21	0	70.6(7.7), (Range: 58–86)	17	Yes	No	No	Yes	No	Yes	No	No	No	Cross-sectional	6
Dashti M; Scandinavian Journal of Rheumatology; 2004; 15163109[118]	174	117	57	61.6(9.5)	123(70.7%)	Yes	No	No	Yes	No	No	No	No	No	Case control	11
Arokoski JP; Journal of Clinical Densitometry; 2004; 15181262[119]	57	27	30	Cases: 56.2(4.9), Range: (47–64); Controls: 56.3(4.5), (Range: 47–64)	0	No	Yes	No	No	No	No	No	No	No	Case control	9
Dunn TC; Radiology; 2004; 15215540[120]	55	48	7	Healthy: 38(Range: 22–71); Mild OA: 63(Range: 46–81); Severe OA: 67 (Range: 43–88)	30	No	Yes	No	Yes	No	No	No	No	No	Case control	8
Regatte RR; Academic Radiology; 2004; 15217591[121]	14	6	8	Asymptomatic: 33.5(Range: 22–45); Symptomatic: 45.5(Range: 28–63)	2	No	Yes	No	Yes	No	No	No	No	No	Case control	7

Baysal O; Swiss Medical Weekly; 2004; 15243849[122]	65	65	0	53.1(7), (Range: 45–75)	65(100%)	Yes	No	Yes	Yes	No	Yes	No	No	No	Cross-sectional	7
Lerer DB; Skeletal Radiology; 2004; 15316679[123]	205			46.5(Range: 15–88); Median = 46	113	No	No	Yes	Yes	No	No	No	Yes	No	Cross-sectional	6
Berthiaume MJ; Annals of the Rheumatic Diseases; 2005; 15374855[78]	32					Yes	No	Yes	Yes	No	No	No	Yes	No	Longitudinal Prospective	10
King KB; Magnetic Resonance Imaging; 2004; 15527998[124]	16	16	0	Males: Median = 58.5, (11.3), (Range: 43–76); Females: Median = 70 (14.4), (Range: 46–88)	8(50%)	Yes	Yes	No	Yes	No	No	No	No	No	Cross-sectional	7
Carbone LD; Arthritis & Rheumatism; 2004; 15529367[125]	818			Non-users: 74.8(2.94); Antiresportive users: 74.8(2.9)	818(100%)	No	No	Yes	Yes	Yes	Yes	No	No	No	Cross-sectional	11
Cicuttini F; Journal of Rheumatology; 2004; 15570649[126]	123					Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	6
Wluka AE; Annals of the Rheumatic Diseases; 2005; 15601742[38]	149	68	81	Normal: 57(5.8); OA: 63(10.3)	1499(100%)	No	No	No	No	No	Yes	No	No	No	Longitudinal Prospective	13
Ding C; Osteoarthritis & Cartilage; 2005; 15727885[49]	372	162	210	No cartilage defects: 43.6(7.1); Any cartilage defect: 47(6.1)	(56.5%)	Yes	No	Yes	Yes	No	No	No	No	No	Case control	9
Hill CL; Arthritis & Rheumatism; 2005; 15751064[45]	433	360	73	Cases males: 68.2; Cases females: 65; Control males: 66.8; Control females: 65.8	143	No	No	Yes	No	No	No	No	No	Yes	Case control	12
Kornaat PR; European Radiology; 2005; 15754163[127]	205	205	0	Median = 60; (Range: 43–77)	163(80%)	No	No	Yes	Yes	No	No	Yes	Yes	No	Cross-sectional	8
Zhai G; Arthritis & Rheumatism; 2005; 15818695[128]	151	23	128	Men: 64(8.1); Women: 62(7.7)	72	Yes	No	No	Yes	No	No	No	No	No	Cross-sectional	8
Cicuttini F; Osteoarthritis & Cartilage; 2005; 15922634[50]	28	28	0	62.8(9.8)	(57%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	10
Blankenbaker DG; Skeletal Radiology; 2005; 15940487[129]	247	74	173	44	126	No	No	Yes	Yes	No	No	No	Yes	Yes	Cross-sectional	6

(*continued on next page*)

Table I (*continued*)

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, mean(SD), range	No. (%) of females	Quantitative cartilage	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Score of methodological quality
Huh YM; Korean Journal of Radiology; 2005; 15968151[130]	94	73	21	RA group: 49.2 (Range: 37–76), Median = 48; OA group: 57.8 (Range: 40–80), Median = 58	73	No	No	Yes	No	Yes	No	No	No	No	Longitudinal Retrospective	7
von Eisenhart-Roth; Annals of the Rheumatic Diseases; 2006; 15975965[131]	26	26	0	70.4(7.6), (Range: 58–86)	20	Yes	No	No	Yes	No	No	No	No	No	Cross-sectional	5
Tan AL; Arthritis & Rheumatism; 2005; 16052535[132]	58	40	18	Early OA: 56 (Range: 49–69); Chronic OA: 60 (Range: 51–68); Hand OA: 60 (Range: 46–72);	44	No	No	Yes	Yes	Yes	Yes	Yes	No	Yes	Cross-sectional	7
Lo GH; Arthritis & Rheumatism; 2005; 16145676[133]	268	80	188	No BMLs: 64.8(8.5); Medial BMLs: 68.3(7); Lateral BMLs: 66.6(9.5)	(59%)	No	No	Yes	No	No	No	Yes	No	No	Cross-sectional	10
Li X; Magnetic Resonance in Medicine; 2005; 16155867[134]	19	9	10	Cases: Median = 52, (Range: 18–72); Controls: Median = 30, (Range: 22–74)	8	No	Yes	No	Yes	No	No	No	No	No	Case control	7
Rhodes LA; Rheumatology; 2005; 16188949[135]	35	35	0	Median = 63; (Range: 49–77)	23	No	No	Yes	No	Yes	No	No	No	No	Cross-sectional	9
Williams A; Arthritis & Rheumatism; 2005; 16255024[32]	31	31	0	67(10.4), 9 (Range: 45–86)	24(77%)	No	Yes	No	Yes	No	No	No	No	No	Cross-sectional	9
Loeuille D; Arthritis & Rheumatism; 2005; 16255041[136]	39	39	0	56.4(12.71)	(56.4%)	No	No	Yes	No	Yes	No	No	No	No	Cross-sectional	10
Roos EM; Arthritis & Rheumatism; 2005; 16258919[137]	30			45.8(3.3)	10(33.3%)	No	Yes	No	Yes	No	No	No	No	No	Randomized controlled trial	17
Hunter DJ; Journal of Rheumatology; 2005; 16265702[53]	132	162	0	33.5(9.7)	(44.2%)	No	No	Yes	Yes	No	Yes	No	Yes	Yes	Cross-sectional	8

Nojiri T; Knee Surgery, Sports Traumatology, Arthroscopy; 2006; 16395564[33]	28	9	21	40.3(Range: 16–74)	17	No	Yes	No	Yes	No	No	No	No	No	Cross-sectional	7
Kimelman T; Invest Radiol; 2006; 16428993[138]	7	4	3	Healthy controls: 23; OA cases: 56	4	No	Yes	No	Yes	No	No	No	No	No	Other	6
Sengupta M; Osteoarthritis & Cartilage; 2006; 16442316[139]	217	217	0	67.3(9.1)	(30%)	No	No	Yes	Yes	Yes	Yes	Yes	No	No	Cross-sectional	7
Hunter DJ; Arthritis & Rheumatism; 2006; 16508930[81]	257	257	0	66.6(9.2), (Range: 47–93)	(41.6%)	No	No	Yes	Yes	No	No	No	Yes	No	Longitudinal Prospective	10
Hunter DJ; Arthritis & Rheumatism; 2006; 16646037[83]	217	217	0	66.4(9.4)	(44%)	No	No	Yes	Yes	No	No	Yes	No	No	Longitudinal Prospective	10
Grainger AJ; European Radiology; 2007; 16685505[140]	43	43	0	64(Range: 48–75)	19	No	No	Yes	No	Yes	No	No	Yes	No	Cross-sectional	8
Cashman PM; IEEE Transactions on Nanobioscience; 2002; 16689221[141]	27	10	17	OA patients: (Range: 45–73); Similar age controls: (Range: 50–65); Young healthy controls: (Range: 21–32);	8(29.6%)	Yes	No	No	Yes	No	No	No	No	No	Other	6
Torres L; Osteoarthritis & Cartilage; 2006; 16713310[21]	143	143	0	70(10)	(78%)	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Cross-sectional	9
Kornaat PR; Radiology; 2006; 16714463[22]	205	97	103	60 (Range: 43–77)	163(80%)	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Cross-sectional	9
Bamac B; Saudi Medical Journal; 2006; 16758050[142]	46	36	10	Cases: 41.9 (Range: 20–67); Controls: 39.7 (Range: 21–66)	25	No	No	No	No	No	No	No	Yes	No	Case control	8
Boks SS; American Journal of Sports Medicine; 2006; 16861575[143]	134	136	132	40.8(Range: 18.8–63.8)		No	No	Yes	Yes	No	No	No	Yes	Yes	Cross-sectional	7
Koff MF; Osteoarthritis & Cartilage; 2007; 16949313[34]	113	113	0	56(11), (Range: 33–82)	84	No	Yes	No	Yes	No	No	No	No	No	Cross-sectional	8
Nakamura M; Magnetic Resonance Imaging; 2006; 17071336[39]	63			51.8 (Range: 40–59)	42	No	No	Yes	No	No	No	No	Yes	No	Cross-sectional	6

(*continued on next page*)

Table I (*continued*)

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, mean(SD), range	No. (%) of females	Quantitative cartilage	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Score of methodological quality
Folkesson J; IEEE Transactions on Medical Imaging; 2007; 17243589[144]	139			56(Range: 22–79)	(59%)	Yes	No	No	Yes	No	No	No	No	No	Other	7
Li X; Osteoarthritis & Cartilage; 2007; 17307365[35]	26	10	16	Healthy: 41.3 (Range: 22–74); OA patients: 55.9 (37–72)	11	No	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Case control	7
Iwasaki J; Clinical Rheumatology; 2007; 17322963[145]	26	26	0	63.8(Rang: 49–82)	18	No	No	No	No	No	No	No	No	No	Cross-sectional	5
Dam EB; Osteoarthritis & Cartilage; 2007; 17353132[31]	139			Evaluation set: 55(Range: 21–78); Scan-rescan set: 61 (Range: 26–75)	(54.5%)	Yes	No	No	Yes	No	No	No	No	No	Other	9
Tiderius CJ; Magnetic Resonance in Medicine; 2007; 17390362[146]	18	10	8	Controls: 28(Range: 20–47); Cases: 39 (Range: 25–58)		No	Yes	No	Yes	No	No	No	No	No	Case control	6
Baranyay FJ; Seminars in Arthritis & Rheumatism; 2007; 17391738[147]	297		297	58(5.5)	(63%)	Yes	No	No	Yes	No	No	Yes	No	No	Cross-sectional	16
Issa SN; Arthritis & Rheumatism; 2007; 17394225[54]	146	146	0	70	109	No	No	Yes	Yes	No	Yes	Yes	Yes	No	Cross-sectional	8
Hanna F; Menopause; 2007; 17413649[148]	176	0	176	52.3(6.6), (Range: 40–67)	176(100%)	Yes	No	No	Yes	No	No	No	No	No	Cross-sectional	13
Hunter DJ; Annals of the Rheumatic Diseases; 2008; 17472995[23]	71			67.9(9.3)	(28.2%)	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Other	8
Hill CL; Annals of the Rheumatic Diseases; 2007; 17491096[24]	270	270	0	66.7(9.2)	112	No	No	Yes	Yes	Yes	No	No	No	No	Longitudinal Prospective	9
Qazi AA; Osteoarthritis & Cartilage; 2007; 17493841[149]	71					Yes	No	No	Yes	No	No	No	No	No	Cross-sectional	8
Lammentausta E; Osteoarthritis & Cartilage; 2007; 17502160[150]	14			55(18)	2	No	Yes	No	Yes	No	Yes	No	No	No	Other	5

Guymer E; Osteoarthritis & Cartilage; 2007; 17560134[151]	176	0	176	52.3(6.6)	176(100%)	Yes	No	Yes	Yes	No	Yes	Yes	No	No	Cross-sectional	11
Nishii T; Osteoarthritis & Cartilage; 2008; 17644363[152]	33	23	10	Volunteers: 34(Range: 23–51); Patients: 40(Range: 22–69)	33(1005)	No	Yes	No	Yes	No	No	No	No	No	Case control	8
Janakiramanan N; Journal of Orthopaedic Research; 2008; 17763451[55]	202	74	128	61(9)	(73%)	No	No	Yes	Yes	No	No	No	No	No	Cross-sectional	11
Lo GH; Osteoarthritis & Cartilage; 2008; 17825586[153]	845	170		63.6(8.8)	(58%)	No	No	Yes	No	No	No	No	Yes	No	Cross-sectional	10
Davies-Tuck M; Osteoarthritis & Cartilage; 2008; 17869546[154]	100	100	0	63.3(10.2)	61(61%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	11
Qazi AA; Academic Radiology; 2007; 17889338[155]	159			(Range: 21–81)		No	No	Yes	Yes	No	No	No	No	No	Other	8
Folkesson J; Academic Radiology; 2007; 17889339[156]	71			56(Range: 22–79)	(59%)	No	No	No	No	No	No	No	No	No	Other	7
Englund M; Arthritis & Rheumatism; 2007; 18050201[40]	310	102	208	Cases: 62.9(8.3); Controls: 61.2(8.3)	211(68%)	No	No	Yes	No	No	No	No	Yes	No	Case control	15
Kamei G; Magnetic Resonance Imaging; 2008; 18083319[157]	37	27	0	Cartilage defect: 51.6(Range: 42–61); No cartilage defect: 54.5(Range: 45–61)	20	No	No	Yes	Yes	No	No	No	Yes	No	Case control	7
Li W; Journal of Magnetic Resonance Imaging; 2008; 18183573[158]	29	19	10	OA subjects: 61.7(Range: 40–86); Controls: 31 (Range: 18–40)	19	No	Yes	No	Yes	No	No	No	No	No	Cross-sectional	5
Amin S; Osteoarthritis & Cartilage; 2008; 18203629[86]	265	265		67(9)	(43%)	No	No	Yes	Yes	No	No	No	Yes	Yes	Longitudinal Prospective	11
Taljanovic MS; Skeletal Radiology; 2008; 18274742[159]	19	19	0	66	8	No	Yes	No	No	No	No	No	No	No	Case control	8
Oda H; Journal of Orthopaedic Science; 2008; 18274849[160]	161			58.5(Range: 11–85)	98	No	No	Yes	No	Yes	No	No	Yes	Yes	Cross-sectional	8

(*continued on next page*)

Table I (*continued*)

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, mean(SD), range	No. (%) of females	Quantitative cartilage	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Score of methodological quality
Hanna FS; Arthritis Research & Therapy; 2008; 18312679[161]	176			52.3(6.6)	(100%)	Yes	No	No	Yes	No	No	No	No	No	Cross-sectional	10
Reichenbach S; Osteoarthritis & Cartilage; 2008; 18367415[41]	964	217	747	63.3	(57%)	No	No	Yes	Yes	No	Yes	No	No	No	Cross-sectional	8
Petterson SC; Medicine & Science in Sports & Exercise; 2008; 18379202[162]	123	123	0	64.9(8.5)	67	No	No	No	No	No	No	No	No	No	Case control	11
Bolbos RI; Osteoarthritis & Cartilage; 2008; 18387828[163]	32	16	16	Cases: 47.2(11.54), (Range: 29–72); Controls: 36.3(10.54), (Range: 27–56)	14	Yes	Yes	No	Yes	No	Yes	No	No	No	Case control	7
Quaia E; Skeletal Radiology; 2008; 18404267[164]	35	35	0	42(17), (Range: 22–67)	14	No	Yes	No	Yes	No	No	No	No	No	Other	6
Folkesson J; Magnetic Resonance in Medicine; 2008; 18506845[42]	245		143	KL0: 48(Range: 21–78); KL1: 62(Range: 37–81); KL2: 67(Range: 47–78); KL3&4: 68(Range: 58–78)		No	No	No	Yes	No	No	No	No	No	Other	12
Mills PM; Osteoarthritis & Cartilage; 2008; 18515157[165]	49	25	24	APMM: 46.8(5.3); Controls: 43.6(6.6)	18(36.7%)	Yes	No	Yes	Yes	No	No	No	No	No	Case control	12
Dore D; Osteoarthritis & Cartilage; 2008; 18515160[166]	50	50		64.5(7.1)	23	Yes	No	Yes	Yes	No	Yes	No	No	No	Cross-sectional	9
Mutimer J; Journal of Hand Surgery; 2008; 18562375[167]	20	20	0	47 (Range: 26–69)	9	No	No	Yes	Yes	No	No	No	No	No	Cross-sectional	6
Amin S; Journal of Rheumatology; 2008; 18597397[168]	192	192		69(9)	0.	No	No	Yes	Yes	No	No	No	No	No	Cross-sectional	10
Li X; Journal of Magnetic Resonance Imaging; 2008; 18666183[169]	38	13	25	Healthy: 28.5 (Range: 20–34); Knee OA or injury: 37.4 (Range: 20–66)	10	Yes	No	No	Yes	No	No	Yes	No	No	Other	7

Pelletier JP; Osteoarthritis & Cartilage; 2008; 18672386[25]	27	1		64.1(9.6)	14	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Other	9
Stahl R; European Radiology; 2009; 18709373[170]	37	17	20	Mild OA: 54(9.98); Healthy control: 33.6(9.44)	19	No	Yes	Yes	Yes	No	No	No	No	No	Case control	10
Brem MH; Acta Radiologica; 2008; 18720084[171]	23	23	0	55.5(10.3)	8	No	No	Yes	No	No	No	Yes	No	No	Other	6
Lancianese SL; Bone; 2008; 18755303[172]	4			80(14)	3	No	No	No	No	No	Yes	No	No	No	Cross-sectional	5
Englund M; New England Journal of Medicine; 2008; 18784100[26]	991	171		62.3(8.6), (Range: 50.1–90.5)	565(57%)	No	No	Yes	No	No	No	No	Yes	No	Cross-sectional	10
Mamisch TC; Magnetic Resonance in Medicine; 2008; 18816842[173]	26					No	Yes	No	Yes	No	No	No	No	No	Cross-sectional	7
Rauscher I; Radiology; 2008; 18936315[174]	60	37	23	Healthy controls: 34.1(10); Mild OA: 52.5(10); Severe OA: 61.6(11.6)	32	No	Yes	No	Yes	No	No	No	Yes	No	Case control	9
Li W; Journal of Magnetic Resonance Imaging; 2009; 19161210[175]	31	17	14	OA patients: 61.8(Range: 40–86); Healthy controls: 29.2(Range: 18–40)	21	No	Yes	No	Yes	No	No	No	No	No	Case control	7
Choi JW; Journal of Computer Assisted Tomography; 2009; 19188805[176]	36			39.7(Range: 8–69)	21	No	No	Yes	Yes	No	No	No	Yes	No	Longitudinal Retrospective	7
Chen YH; Journal of Computer Assisted Tomography; 2008; 19204464[177]	96	25	71	OA patients: 56; Non-OA: 46		No	No	Yes	Yes	No	No	No	Yes	No	Case control	8

Table II
Summary of Concurrent Validity of MRI in OA

Outcome of interest	Number of studies examining this outcome	Number of studies finding significant associations ($P < .05$)
Symptoms	21 studies	13 of 21 (62%)
Radiographic features	43 studies	39 of 43 (90%)
Radiographic joint space	9 studies	9 of 9 (100%)
Alignment	10 studies	9 of 10 (90%)
CT	4 studies	4 of 4 (100%)
Histology/Pathology	5 studies	3 of 5 (60%)
Arthroscopy	7 studies	5 of 7 (71%)

subluxation of the meniscus, meniscal tears, or Baker cysts, were not associated with symptoms[22].

- Maximal bone marrow lesion (BML) size on the Boston Leeds Osteoarthritis Score (BLOKS) scale had a positive linear relation with VAS pain (P for linear trend $= 0.04$)[23].
- No correlation of baseline synovitis with baseline pain score ($r = 0.09$, $P = 0.17$)[24].
- No relation between baseline synovitis score and VAS pain score ($r = 0.11$, $P = 0.60$)[25].
- In the group of persons with radiographic evidence of osteoarthritis (Kellgren–Lawrence grade 2 or higher), the prevalence of a meniscal tear was 63% among those with knee pain, aching, or stiffness on most days and 60% among those without these symptoms ($P = 0.75$); the corresponding prevalences in the group without radiographic evidence of osteoarthritis were 32% and 23% ($P = 0.02$). The majority of the meniscal tears – 180 of 297 (61%) were in subjects who had not had any pain, aching, or stiffness in the previous month[26].

Relation to radiographic features

43 studies examined the concurrent relation of MRI findings in OA to radiographic features. Of these, 90% demonstrated a statistically significant association, defined as $P < 0.05$.

Relation of quantitative cartilage morphometry measures to radiographic abnormalities.

- Significant differences in lateral and medial femorotibial cartilage thickness were found between those with and without radiographic OA. Significant cartilage thinning could be detected by MRI in patients with OA, even when the joint space was normal radiographically[27].
- For every increase in grade of lateral tibiofemoral osteophytes the lateral tibial cartilage volume was significantly reduced by 255 mm^3, after adjustment. There was a reduction of 77 mm^3 in medial tibial cartilage volume for every increase in grade of medial tibiofemoral osteophytes, but this finding was only of borderline statistical significance[28].
- Cartilage volume and thickness were less in patients with OA compared to normal controls ($P < 0.1$)[29].
- Kellgren and Lawrence (KLG)2 participants displayed, on average, thicker cartilage than healthy controls in the medial femorotibial compartment [particularly anterior subregion of the medial tibia (MT) and peripheral (external, internal) subregions of the medial femur], and in the lateral femur. KLG3 participants displayed significantly thinner cartilage than KLG0 participants in the medial weight-bearing femur (central subregion), in the external subregion of the MT, and in the internal subregion of the lateral tibia[30].
- Mean cartilage signal intensity provided a clear separation of healthy from KLG1 ($P = 0.0009$). Quantification of cartilage homogeneity by entropy was able to clearly 11separate healthy from OA subjects ($P = 0.0003$). Furthermo121re, entropy was also able to separate healthy from KL 1 subjects ($P = 0.0004$)[31].

Relation of other MRI measures to radiographic abnormalities.

- Significant difference ($P = 0.002$) in the average T(1rho) within patellar and femoral cartilage between controls (45.04 ± 2.59 ms) and osteoarthritis patients (53.06 ± 4.60 ms). A significant correlation was found between T(1rho) and T(2); however, the difference of T(2) was not statistically significant between controls and osteoarthritis patients[31].
- Trend toward a lower dGEMRIC index with increasing KLG; the spared compartments of knees with a KLG grade 2 had a higher dGEMRIC index than those of knees with a KLG grade 4 (mean 425 msec vs 371 msec; $P < 0.05$)[32].
- All cases demonstrating decreased T1 values on dGEMRIC, showed abnormal arthroscopic or direct viewing findings. The diagnosis of damage in articular cartilage was possible in all 16 cases with radiographic KLG 1 on dGEMRIC, while the intensity changes were not found in 10 of 16 cases on Proton density Weighted Image (PDWI)[33].
- No differences of T2 values were found across the stages of OA ($P = 0.25$), but the factor of BMI did have a significant effect $P < 0.0001$) on T2 value[34].
- Average T(1rho) and T(2) values were significantly increased in OA patients compared with controls [52.04 ± 2.97 ms vs 45.53 ± 3.28 ms with $P = 0.0002$ for T(1rho), and 39.63 ± 2.69 ms vs 34.74 ± 2.48 ms with $P = 0.001$ for T(2)]. Increased T(1rho) and T(2) values were correlated with increased severity in radiographic and MR grading of OA. T(1rho) has a larger range and higher effect size than T(2), 3.7 vs 3.0[35].
- Statistically significant correlation between radiography and MR cartilage loss in the medial (r 0.7142, P .0001) and lateral compartments ($r = 0.4004$, P .0136). Significant correlations also found between radiographic assessment of sclerosis and osteophytes and those found on MRI[36].
- Patients in whom plain radiographs, MRI, and arthroscopy were compared, the plain radiographs and MRI significantly underestimated the extent of cartilage abnormalities[37].
- Presence of synovial thickening was more likely with increasing KLG, from 24.0% in those with KLG 0–78.3% in those with KLG 3/4 ($P < 0.001$)[15].
- Higher KLG was correlated with a higher frequency of meniscal tears ($r = 0.26$, $P < 0.001$)[16].
- KLG correlated significantly ($P < 0.05$) with the grade of cartilage lesions, and a substantially higher percentage of bone marrow and meniscal lesions with higher KLG found on MR images[17].
- Women with osteoarthritis had larger medial and lateral tibial plateau bone area [mean (SD): 1850 (240) mm^2 and 1279 (220) mm^2, respectively] than healthy women [1670 (200) mm^2 and

Table III
Summary table of studies reporting data on predictive validity of MRI in knee OA

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, Mean(SD), Range	No. (%) of females	Quantitative cartilage	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Score of methodological quality
Boegard TL; Osteoarthritis & Cartilage; 2001; 11467896[178]	47			Women: Median = 50, (Range: 42–57); Men: Median = 50, (Range: 41–57)	25(53.2%)	No	No	Yes	Yes	No	No	No	Yes	No	Longitudinal Prospective	9
Wluka AE; Arthritis & Rheumatism; 2002; 12209510[179]	123	123	0	63.1(10.6)	71	Yes	No	No	Yes	No	Yes	No	No	No	Longitudinal Prospective	14
Cicuttini FM; Journal of Rheumatology; 2002; 12233892[180]	21	8	13	Case: 41.3(13.2); Controls: 49.2(17.8)	14(66.7%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Retrospective	13
Biswal S; Arthritis & Rheumatism; 2002; 12428228[76]	43	4	39	54.4(Range: 17–65)	21	No	No	Yes	Yes	No	No	Yes	Yes	Yes	Longitudinal Retrospective	8
Cicuttini F; Journal of Rheumatology; 2002; 12465162[181]	110	110	0	63.2(10.2)	66	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	12
Pessis E; Osteoarthritis & Cartilage; 2003; 12744942[182]	20	20		63.9(9)	13	Yes	No	Yes	Yes	No	Yes	Yes	No	No	Longitudinal Prospective	12
Felson DT; Annals of Internal Medicine; 2003; 12965941[51]	256	156	0	Followed: 66.2(9.4); Not followed: 67.8(9.6)	(38.3%)	No	No	Yes	No	No	No	Yes	No	No	Longitudinal Prospective	11
Cicuttini FM; Arthritis & Rheumatism; 2004; 14730604[77]	117	117		63.7(10.2)	(58.1%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	9
Wluka AE; Annals of the Rheumatic Diseases; 2004; 14962960[20]	132	132	0	63.1(Range: 41–86)	71(54%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	10
Cicuttini F; Rheumatology; 2004; 14963201[52]	117	117	0	67(10.6)	(58%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	12
Cicuttini FM; Ann Rheum Dis; 2004; 15115714[65]	123	123	0	Joint replacement: 64.1(9.3); No joint replacement: 63.1(10.3)	65(52.8%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	11
Dashti M; Scandinavian Journal of Rheumatology; 2004; 15163109[118]	174	117	57	61.6(9.5)	123(70.7%)	Yes	No	No	Yes	No	No	No	No	No	Case control	11
Cicuttini FM; Journal of Rheumatology; 2004; 15229959[183]	102	102	0	63.8(10.1)	(63%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	10
Berthiaume MJ; Annals of the Rheumatic Diseases; 2005; 15374855[78]	32					Yes	No	Yes	Yes	No	No	No	Yes	No	Longitudinal Prospective	10
Cicuttini F; Journal of Rheumatology; 2004; 15570649[126]	123					Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	6

(*continued on next page*)

Table III (*continued*)

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, Mean(SD), Range	No. (%) of females	Quantitative cartilage	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Score of methodological quality
Cubukcu D; Clinical Rheumatology; 2005; 15599642 [184]	40	40		HA group: 52.6(7.16); Saline group: 57.6(2.77)	24(60%)	No	No	Yes	Yes	No	Yes	No	Yes	Yes	Randomized controlled trial	15
Ozturk C; Rheumatol Int; 2006;15703953[185]	47	47	0	HA-only group: 58(7.7); HA&Cortico group: 58.1(10.3)	39(97.5%)	No	No	Yes	Yes	No	No	Yes	No	No	Randomized controlled trial	17
Wang Y; Arthritis Res Ther; 2005; 15899054[186]	126	126		63.6(10.1)	68	No	No	No	No	No	Yes	No	No	No	Longitudinal Prospective	12
Cicuttini F; Osteoarthritis & Cartilage; 2005; 15922634[50]	28	28	0	62.8(9.8)	(57%)	Yes	No	No	Yes	No	No	No	No	No	Longitudinal Prospective	10
Wluka AE; Rheumatology; 2005; 16030084[66]	126	126	0	63.6(10.1)	68(54%)	Yes	No	Yes	Yes	No	Yes	No	No	No	Longitudinal Prospective	14
Garnero P; Arthritis & Rheumatism; 2005; 16145678[187]	377	377	0	62.5(8.1)	(76%)	No	No	Yes	Yes	No	No	No	No	No	Longitudinal Prospective	11
Wang Y; Rheumatology; 2006; 16188947[79]	124	124	0	Females: 57.1(5.8); Males: 52.5(13.2)	81(65.3%)	No	No	Yes	Yes	No	Yes	No	No	No	Longitudinal Prospective	11
Phan CM; European Radiology; 2006; 16222533[68]	40	34	6	57.7(15.6), (Range: 28–81)	16	No	No	Yes	Yes	No	Yes	Yes	Yes	Yes	Longitudinal Prospective	7
Hayes CW; Radiology; 2005; 16251398[188]	117	117	115	No OA, No Pain: 44.6(10.7); OA, No Pain: 16.2(0.8); No OA, Pain: 47(0.7); OA&Pain: 47.1(0.8)	(100%)	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Longitudinal Prospective	13
Wang Y; Journal of Rheumatology; 2005; 16265703[189]	40	0	40	52.3(13)	0	Yes	No	No	Yes	No	Yes	No	No	No	Longitudinal Prospective	11
Ding C; Arthritis & Rheumatism; 2005; 16320339[80]	325			45.2(6.5)	190	Yes	No	Yes	Yes	No	No	No	No	No	Longitudinal Prospective	10
Bruyere O; Annals of the Rheumatic Diseases; 2006; 16396978[190]	62	62	0	64.9(10.3)	49	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Longitudinal Prospective	10
Katz JN; Osteoarthritis & Cartilage; 2006; 16413210[69]	83			61(11), (Range: 45–89)	50(60%)	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Longitudinal Prospective	9
Raynauld JP; Arthritis Research & Therapy; 2006; 16507119[72]	110	110	0	62.4(7.5)	(64%)	Yes	No	Yes	Yes	No	No	Yes	Yes	No	Longitudinal Prospective	11
Hunter DJ; Arthritis & Rheumatism; 2006; 16508930[81]	257	257	0	66.6(9.2), (Range: 47–93)	(41.6%)	No	No	Yes	Yes	No	No	No	Yes	No	Longitudinal Prospective	10
Ding C; Archives of Internal Medicine; 2006; 16567605[82]	325			Decrease defects: 45.4(6.4); Stable defects: 44.2(7.1); Increase defects: 46.1(5.9)	(58.1%)	No	No	Yes	Yes	No	No	No	No	No	Longitudinal Prospective	14

Brandt KD; Rheumatology; 2006; 16606655[191]	30	20	10	62	29	No	No	No	No	Yes	No	No	No	No	Other	10
Hunter DJ; Arthritis & Rheumatism; 2006; 16646037[83]	217	217	0	66.4(9.4)	(44%)	No	No	Yes	Yes	No	No	Yes	No	No	Longitudinal Prospective	10
Wluka AE; Arthritis Research & Therapy; 2006; 16704746[192]	105	105	0	All eligible: 62.5 (10.7); MRI at FU: 63.8(10.6); Lost to FU: 61.6(11.3)	59(53%)	Yes	No	No	Yes	No	Yes	No	No	No	Longitudinal Prospective	17
Hunter DJ; Osteoarthritis & Cartilage; 2007; 16857393[193]	127	127		67(9.05)	(46.7%)	No	Yes	No	Yes	No	No	No	No	No	Cross-sectional	12
Bruyere O; Osteoarthritis & Cartilage; 2007; 16890461[73]	62	62	0	64.9(10.3)	46	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Longitudinal Prospective	10
Amin S; Annals of the Rheumatic Diseases; 2007; 17158140[194]	196	196	0	68(9)	0	No	No	Yes	Yes	No	No	No	No	No	Longitudinal Prospective	13
Nevitt MC; Arthritis & Rheumatism; 2007; 17469126[74]	80	39	0	73.5(3.1)	(63.6%)	No	No	Yes	Yes	No	No	No	No	No	Longitudinal Prospective	10
Hill CL; Annals of the Rheumatic Diseases; 2007; 17491096[24]	270	270	0	66.7(9.2)	112	No	No	Yes	Yes	Yes	No	No	No	No	Longitudinal Prospective	9
Pelletier JP; Arthritis Research & Therapy; 2007; 17672891[71]	110	110	0	Q1 greatest loss global: 63.7(7.2); Q4 least loss gobal: 61.3(7.5); Q1 greatest loss_medial: 64.1 (7.4); Q1 least loss_medial: 61.6(7.8)	(68.3%)	No	No	Yes	Yes	No	No	Yes	Yes	No	Longitudinal Prospective	15
Davies-Tuck ML; Osteoarthritis & Cartilage; 2008; 17698376[195]	117	117	0	63.7(10.2)	68(58%)	Yes	No	Yes	Yes	No	Yes	No	No	No	Longitudinal Prospective	14
Raynauld JP; Annals of the Rheumatic Diseases; 2008; 17728333[84]	107	107	0	62.4(7.5)	(64%)	Yes	No	Yes	Yes	No	No	No	Yes	No	Longitudinal Retrospective	15
Felson DT; Arthritis & Rheumatism; 2007; 17763427[70]	330	110	220	Cases: 62.9(8.3); Controls: 61.2(8.4)	211(63.9%)	No	No	Yes	No	No	Yes	Yes	No	No	Case control	12
Kornaat PR; European Radiology; 2007; 17823802 [196]	182	71		59(Range: 43–76)	157(80%)	No	No	Yes	No	No	No	Yes	No	No	Longitudinal Prospective	8
Hunter DJ; Arthritis Research & Therapy; 2007; 17958892[197]	160	80	80	67(9)	(46%)	No	No	Yes	Yes	No	No	No	No	No	Case control	11

(*continued on next page*)

Table III (*continued*)

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, Mean(SD), Range	No. (%) of females	Quantitative cartilage	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Score of methodological quality
Englund M; Arthritis & Rheumatism; 2007; 18050201[40]	310	102	208	Cases: 62.9(8.3)	211(68.1%)	No	No	Yes	No	No	No	No	Yes	No	Case control	15
Davies-Tuck ML; Osteoarthritis & Cartilage; 2008; 18093847[85]	74	0	74	Meniscal tear: 58.8(6); No meniscal tear: 55.5(4.3)	74(100%)	No	Yes	Yes	Yes	No	Yes	No	Yes	No	Longitudinal Prospective	13
Hernandez-Molina G; Arthritis & Rheumatism; 2008; 18163483[198]	258	258	0	66.6(9.2)	(42.6%)	No	No	Yes	Yes	No	No	Yes	No	Yes	Longitudinal Prospective	11
Teichtahl AJ; Osteoarthritis & Cartilage; 2009; 18194873[199]	99	99	0	63 (10)	(60%)	Yes	No	No	Yes	No	Yes	No	No	No	Longitudinal Prospective	14
Amin S; Osteoarthritis & Cartilage; 2008; 18203629[86]	265	265		67(9)	(43%)	No	No	Yes	Yes	No	No	No	Yes	Yes	Longitudinal Prospective	11
Teichtahl AJ; Obesity; 2008; 18239654[200]	297		297	58(5.5)	186	Yes	No	Yes	Yes	No	Yes	No	No	No	Longitudinal Prospective	14
Blumenkrantz G; Osteoarthritis & Cartilage; 2008; 18337129[201]	18	8	10	Cases: 55.7(7.3); Controls: 57.6(6.2)	18(100%)	No	Yes	Yes	Yes	No	No	No	No	No	Case control	12
Song IH; Annals of the Rheumatic Diseases; 2009; 18375537[202]	41	41		65(6.7)	26	No	No	Yes	No	No	No	No	No	Yes	Randomized controlled trial	14
Scher C; Skeletal Radiology; 2008; 18463865[67]	65	65	0	OA-only: 49.3 (Range: 28–75); OA&BME group: 53.5(35–82)		No	No	Yes	Yes	No	No	Yes	No	No	Longitudinal Retrospective	10
Sharma L; Arthritis & Rheumatism; 2008; 18512777[87]	153	153	0	66.4(11)		Yes	No	Yes	Yes	No	No	No	Yes	No	Longitudinal Prospective	11
Owman H; Arthritis & Rheumatism; 2008; 18512778[203]	15	9	7	50(Range: 35–70)		No	Yes	No	Yes	No	No	No	No	No	Longitudinal Prospective	10
Madan-Sharma R; Skeletal Radiology; 2008; 18566813[75]	186	74	112	60.2(Range: 43–76)	150	No	No	Yes	Yes	No	No	Yes	Yes	No	Longitudinal Prospective	11
Amin S; Journal of Rheumatology; 2008; 18597397[168]	192	192		69(9)		No	No	Yes	Yes	No	No	No	No	No	Cross-sectional	10
Pelletier JP; Osteoarthritis & Cartilage; 2008; 18672386[25]	27	1		64.1(9.6)	14	No	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Other	9
Amin S; Arthritis & Rheumatism; 2009; 19116936[204]	265	265	0	67(9)		No	No	Yes	Yes	No	No	No	No	No	Longitudinal Prospective	16

1050 (130) mm^2] ($P < 0.001$ for both differences). For each increase in grade of osteophyte, an increase in bone area was seen of 146 mm^2 in the medial compartment and 102 mm^2 in the lateral compartment[38].

- Statistically significant correlations were observed between the medial tibial spur classification on X-ray, the medial meniscal displacement rate on MRI and the medial meniscal signal change classification on MRI[39].
- Meniscal damage was mostly present in knees with OA and demonstrates a relation to KLG[40].
- Bone attrition of the tibiofemoral joint, scored >1, was found in 228 MRIs (23.6%) and in 55 radiographs (5.7%). Moderate to strong correlation between MRIs and radiographs for bone attrition of the tibiofemoral joint ($r = 0.50$, $P < 0.001$)[41].
- Surface curvature of articular cartilage for both the fine- and coarse-scale estimates were significantly higher in the OA population compared with the healthy population, with $P < 0.001$ and $P < 0.001$, respectively[42].
- The prevalence of meniscal damage was significantly higher among subjects with radiographic evidence of tibiofemoral osteoarthritis (KLG 2 or higher) than among those without such evidence (82% vs 25%, $P < 0.001$), and the prevalence increased with a higher KLG ($P < 0.001$ for trend). Among persons with radiographic evidence of severe osteoarthritis (KLG 3 or 4 in their right knee), 95% had meniscal damage[26].

Relation to radiographic joint space width

Nine studies examined the concurrent relation of MRI findings in OA to radiographic joint space. Of these, 100% demonstrated a statistically significant association, defined as $P < 0.05$.

- Strong correlation between the degree of medial meniscal subluxation and the severity of medial joint space narrowing (JSN) ($r = 0.56$, $P = 0.0001$)[43].
- Meniscal extrusion identified in all 32 patients with JSN (KLG 1–4). Definite thinning or loss of articular cartilage was identified in only 15 of the 32 cases. In 17 patients with radiographic JSN (KLG 1–3) and meniscal extrusion, no loss of articular cartilage was observed. A statistically significant correlation ($P < 0.001$) was observed between KLG and degree of meniscal extrusion and cartilage thinning on MRI[44].
- For each increase in grade of JSN, tibial plateau bone area increased by 160 mm^2 in the medial compartment and 131 mm^2 in the lateral compartment (significance of regression coefficients all $P < 0.001$)[38].
- Persons with symptomatic knee OA with ACL rupture had more severe radiologic OA ($P < 0.0001$) and were more likely to have medial JSN ($P < 0.0001$) than a control sample[45].
- Compartments of the knee joint without JSN had a higher dGEMRIC index than those with any level of narrowing (mean 408 msec vs 365 msec; $P = 0.001$). In knees with 1 unnarrowed (spared) and 1 narrowed (diseased) compartment, the dGEMRIC index was greater in the spared vs the diseased compartment (mean 395 msec vs 369 msec; $P = 0.001$)[32].
- Grade of JSN as measured on skyline and lateral patellofemoral radiographs was inversely associated with patella cartilage volume. After adjusting for age, gender and body mass index, for every increase in grade of skyline JSN (0–3), the patella cartilage volume was reduced by 411 mm^3. For every increase in lateral patellofemoral JSN grade (0–3), the adjusted patella cartilage volume was reduced by 125 mm^3. The relationship was stronger for patella cartilage volume and skyline JSN ($r = -0.54$, $P < 0.001$) than for lateral patellofemoral JSN ($r = -0.16$, $P = 0.015$)[46].
- Grade one medial JSN was associated with substantial reductions in cartilage volume at both the medial and lateral tibial and patellar sites within the knee (adjusted mean difference 11–13%, all $P < 0.001$)[47].
- Cartilage volume in the medial compartment and the narrowest JSW obtained by radiography at baseline in 31 knee OA patients, revealed that some level of correlation exists between these two measurements ($r = 0.46$, $P < 0.007$)[48].
- Knee cartilage defects are inconsistently associated with JSN after adjustment for osteophytes but consistently with knee cartilage volume (beta: −0.27 to −0.70/ml; OR: 0.16–0.56/ml, all $P < 0.01$ except for OR at lateral tibial cartilage site $P = 0.06$)[49].
- Moderate, but statistically significant, correlation between JSW and femoral and tibial cartilage volumes in the medial tibiofemoral joint, which was strengthened by adjusting for medial tibial bone size ($R = 0.58$–0.66, $P = 0.001$)[50].
- JSN seen on both medial and lateral radiographs of the tibiofemoral joint was inversely associated with the respective tibial cartilage volume. This inverse relationship was strengthened with adjustment for age, sex, body mass index (BMI), and bone size. After adjustment for these confounders, for every increase in JSN grade (0–3), the medial tibial cartilage volume was reduced by 257 mm^3 (95% CI 193–321) and the lateral tibial cartilage volume by 396 mm^3 (95% CI 283–509). The relationship between mean cartilage volume and radiologic grade of JSN was linear[28].

Relation to alignment

10 studies examined the concurrent relation of MRI findings in OA to alignment. Of these, 90% demonstrated a statistically significant association, defined as $P < 0.05$.

- Valgus-aligned knees tended to have lower dGEMRIC values laterally, and varus-aligned knees tended to have lower dGEMRIC values medially; as a continuous variable, alignment correlated with the lateral: medial dGEMRIC ratio (Pearson's $R = 0.43$, $P = 0.02$)[32].
- Limbs with varus alignment, especially if marked (≥7 degrees), had a remarkably high prevalence of medial lesions compared with limbs that were neutral or valgus (74.3% vs 16.4%; $P < 0.001$ for relation between alignment and medial lesions). Conversely, limbs that were neutral or valgus had a much higher prevalence of lateral lesions than limbs that were in the most varus group (29.5% vs 8.6%; $P = 0.002$ for alignment and lateral lesions)[51].
- Medial tibial and femoral cartilage volumes increased as the angle decreased (i.e., was less varus). Similarly, in the lateral compartment there was an inverse association at baseline between tibial and femoral cartilage volumes and the measured knee angle[52].
- The main univariate determinants of varus alignment in decreasing order of influence were medial bone attrition, medial meniscal degeneration, medial meniscal subluxation, and medial tibiofemoral cartilage loss. Multivariable analysis revealed that medial bone attrition and medial tibiofemoral cartilage loss explained more of the variance in varus malalignment than other variables. The main univariate determinants of valgus malalignment in decreasing order of influence

were lateral tibiofemoral cartilage loss, lateral osteophyte score, and lateral meniscal degeneration[53].

- Correlation between medial meniscal displacement rate on MRI and the femorotibial angle ($r = 0.398$)[39].
- Worsening in the status of each medial lesion cartilage morphology, subarticular bone marrow lesions, meniscal tear, meniscal subluxation, and bone attrition was associated with greater varus malalignment[54].
- For every one degree increase in a valgus direction, there was an associated reduced risk of the presence of cartilage defects in the medial compartment of subjects with knee OA ($P = 0.02$). Moreover, for every one degree increase in a valgus direction, there was an associated increased risk of the presence of lateral cartilage defects in the OA group ($P = 0.006$)[55].

Relation to CT

Four studies examined the concurrent relation of MRI findings in OA to CT. Of these, 100% demonstrated a statistically significant association, defined as $P < 0.05$. MR frequently showed tri-compartmental cartilage loss when radiography and CT showed only bicompartmental involvement in the medial and patellofemoral compartments. In the lateral compartment, MR showed a higher prevalence of cartilage loss (60%) than radiography (35%) and CT (25%) did. In the medial compartment, CT and MR showed osteophytes in 100% of the knees, whereas radiography showed osteophytes in only 60%. Notably, radiography often failed to show osteophytes in the posterior medial femoral condyle. On MR images, meniscal degeneration or tears were found in all 20 knees studied. Partial and complete tears of the anterior cruciate ligament were found in three and seven patients, respectively. MR is more sensitive than radiography and CT for assessing the extent and severity of osteoarthritic changes and frequently shows tri-compartmental disease in patients in whom radiography and CT show only bicompartmental involvement. MR imaging is unique for evaluating meniscal and ligamentous disease related to osteoarthritis[36].

- Strong linear relationship ($r = 0.998$) between MRI imaging and CT arthrography. The mean absolute volume deviation between magnetic resonance imaging and computed tomography arthrography was 3.3%[56].

Relation to histology/pathology

Five studies examined the concurrent relation of MRI findings in OA to histology/pathology. Of these, 60% demonstrated a statistically significant association, defined as $P < 0.05$.Observed measurements of MRI volume of articular cartilage correlated with actual weight and volume displacement measurements with an accuracy of 82%–99% and linear correlation coefficients of 0.99 ($P = 2.5e\text{-}15$) and 0.99 ($P = 4.4e\text{-}15$)[57].

- The signal behavior of hyaline articular cartilage does not reflect the laminar histologic structure. Osteoarthrosis and cartilage degeneration are visible on MR images as intracartilaginous signal changes, superficial erosions, diffuse cartilage thinning, and cartilage ulceration[58].
- Comparison of data on cartilage thickness measurements with MRI with corresponding histological sections in the middle of each sector revealed a very good magnetic resonance/anatomic correlation ($r = 0.88$)[59].
- Correlation between MRI Noyes grading scores and Mankin grading scores of natural lesions was moderately high ($r = 0.7$) and statistically significant ($P = 0.001$)[60].

Relation to arthroscopy

Seven studies examined the concurrent relation of MRI findings in OA to arthroscopy. Of these, 71% demonstrated a statistically significant association, defined as $P < 0.05$.

- Moderate correlation between imaged cartilage scores and the arthroscopy scores (Pearson correlation coefficient = 0.40)[37].
- Spearman rank linear correlation between arthroscopic and MR cartilage grading was highly significant ($P < 0.002$) for each of the six articular regions evaluated. The MR and arthroscopic grades were the same in 93 (68%) of 137 joint surfaces, they were the same or differed by one grade in 123 surfaces (90%), and they were the same or differed by one or two grades in 129 surfaces (94%)[61].
- The overall sensitivity and specificity of MR in detecting chondral abnormalities were 60.5% (158/261) and 93.7% (89/95) respectively. MR imaging was more sensitive to the higher grade lesions: 31.8% (34/107) in grade 1; 72.4% (71/98) in grade 2; 93.5% (43/46) in grade 3; and 100% (10/10) in grade 4. The MR and arthroscopic grades were the same in 46.9% (167/356), and differed by no more than 1 grade in 90.2% (321/356) and 2 grades in 99.2% (353/356). The correlation between arthroscopic and MR grading scores was highly significant with a correlation coefficient of 0.705 ($P < 0.0001$)[62].
- Statistically significant correlation between the SFA-arthroscopic score and the SFA-MR score ($r = 0.83$) and between the SFA-arthroscopic grade and the SFA-MR grade (weighted kappa = 0.84). The deepest cartilage lesions graded with arthroscopy and MR imaging showed correlation in the medial femoral condyle (weighted kappa = 0.83) and in the medial tibial plateau (weighted kappa = 0.84)[63].
- Magnetic resonance imaging was in agreement with arthroscopy in 81% showing more degeneration but less tears of menisci than arthroscopy. Using a global system for grading the total damage of the knee joint into none, mild, moderate, or severe changes, agreement between arthroscopy and MRI was found in 82%[64].

Predictive validity (Table III)

The analysis included data from 61 manuscripts of which 1 pertains to the hip and the remainder to the knee. The mean Downs criteria score for these manuscripts was 11.5 (range 6–17). What follows below are important excerpts from this data pertaining to different aspects of predictive validity. The data is further summarized in Table IV to discretely identify the associations examined and those where a significant association was found.

Prediction of joint replacement

Three studies examined the predictive relation of MRI findings to joint replacement. Of these, 100% demonstrated a statistically significant association, defined as $P < 0.05$.

- One study investigated the relation of change in quantitative cartilage volume to risk of knee replacement. For every 1% increase in the rate of tibial cartilage loss there was a 20% increase risk of undergoing a knee replacement at four years (95% CI, 10%–30%). Those in the highest tertile of tibial cartilage loss had 7.1 (1.4–36.5) higher odds of undergoing a knee replacement than those in the lowest tertile. Change in bone area also predicted risk of TKR OR 12 (95% CI 1–14)[65].
- Higher total cartilage defect scores (8–15) were associated with a 6.0-fold increased risk of joint replacement over 4 yr compared with those with lower scores (2–7) (95% CI 1.6, 22.3), independently of potential confounders[66].

Table IV
Summary of Predictive Validity of MRI in OA

Outcome of interest	Number of studies examining this outcome	Number of studies finding significant associations ($P < .05$)
Joint replacement	3 studies	3 of 3 (100%)
Change in symptoms	6 studies	5 of 6 (83%)
Radiographic progression	8 studies	5 of 8 (63%)
MRI progression	19 studies	16 of 19 (84%)

- A separate smaller study investigated the relation of bone marrow lesions (assessed semi-quantitatively) to need for TKR. Subjects who had a bone marrow lesion were 8.95 times as likely to progress rapidly to a TKA when compared to subjects with no BME ($P = 0.016$). There was no relation of TKR with meniscal tear or cartilage loss[67].

Prediction of change in symptoms

Six studies examined the predictive relation of MRI findings to change in symptoms. Of these, 83% demonstrated a statistically significant association, defined as $P < 0.05$.

- Weak associations between worsening of symptoms of OA and increased cartilage loss: pain [$r(s) = 0.28$, $P = 0.002$], stiffness [$r(s) = 0.17$, $P = 0.07$], and deterioration in function [$r(s) = 0.21$, $P = 0.02$][20].
- Small study did not find a significant relation between changes in WOMAC scores with the amount of cartilage loss and the change in BME ($P > 0.05$)[68].
- Multivariate analyses of knee pain 1 year following arthroscopic partial meniscectomy demonstrated that medial tibial cartilage damage accounting for 13% of the variability in pain scores[69].
- The BOKS study examined the relationship between longitudinal fluctuations in synovitis with change in pain and cartilage in knee osteoarthritis. Change in summary synovitis score was correlated with the change in pain ($r = 0.21$, $P = 0.0003$). An increase of one unit in summary synovitis score resulted in a 3.15-mm increase in VAS pain score (0–100 scale). Effusion change was not associated with pain change. Of the three locations for synovitis, changes in the infrapatellar fat pad were most strongly related to pain change[24].
- A nested case-control study examined if enlarging BMLs are associated with new knee pain. Case knee was defined as absence of knee pain at baseline but presence of knee pain both times at follow-up. Controls were selected randomly from among knees with absence of pain at baseline. Among case knees, 54 of 110 (49.1%) showed an increase in BML score within a compartment, whereas only 59 of 220 control knees (26.8%) showed an increase ($P < 0.001$ by chi-square test). A BML score increase of at least 2 units was much more common in case knees than in control knees (27.5% vs 8.6%; adjusted odds ratio 3.2, 95% CI 1.5–6.8)[70].
- Increases in Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) pain and patient global scores over time are associated with change in cartilage volume of the medial tibial plateau and medial femoral condyle[71].
- Weak association of cartilage volume loss with less knee pain. Medial cartilage volume loss and simultaneous pain change at 24 months (beta coefficient −0.45, $P = 0.03$) and SF-36 physical components (beta coefficient 0.22, $P = 0.04$)[72].

Prediction of radiographic progression

Eight studies examined the predictive relation of MRI findings to radiographic progression. Of these, 63% demonstrated a statistically significant association, defined as $P < 0.05$.

- No significant association between reduction in JSW and cartilage volume ($R < 0.13$). Trend toward a significant association between change in medial tibiofemoral cartilage volume and joint replacement at 4 years (OR = 9.0, $P = 0.07$) but not change in medial tibiofemoral JSW (OR = 1.1, $P = 0.92$)[50].
- No correlation between the cartilage volume loss changes (either by using absolute or percentage values) and the JSW changes at 24 months (global cartilage volume, $r = 0.11$; medial compartment cartilage volume, $r = 0.19$)[72].
- Medial femorotibial JSN after 1 year, assessed by radiography, was significantly correlated with a loss of medial tibial cartilage volume ($r = 0.25$, $P = 0.046$) and medial tibial cartilage thickness ($r = 0.28$, $P = 0.025$), over the same period[73].
- Higher baseline composite cartilage scores and increases in composite cartilage scores during follow-up were moderately correlated with greater joint space loss ($r = 0.33$, $P = 0.0002$ and $r = 0.26$, $P = 0.01$, respectively)[74].
- Loss in JSW correlated with the loss of cartilage volume on the central weight-bearing area of the condyles and the plateaus as well as on the medial compartment[71].
- Study examined the relation of MRI features at baseline with radiographically determined JSN in the medial compartment of the knee after 2 years in a group of patients with symptomatic osteoarthritis. A significant association was observed for meniscal tears (RR 3.57; CI 1.08–10.0) and meniscal subluxation (RR 2.73; CI 1.20–5.41), between KL < 2 and meniscal subluxation (RR 11.3; CI 2.49–29.49) and KL ≥ 2 and meniscus tears (RR 8.91; CI 1.13–22.84) and radiographic JSN 2 years later[75].

Prediction of MRI progression

Nineteen studies examined the predictive relation of MRI findings to MRI progression. Of these, 84% demonstrated a statistically significant association, defined as $P < 0.05$.

- Patients who had sustained meniscal tears showed a higher average rate of progression of cartilage loss (22%) than that seen in those who had intact menisci (14.9%) ($P \leq 0.018$). Anterior cruciate ligament (ACL) tears had a borderline significant influence $P \leq 0.06$) on the progression of cartilage pathology. Lesions located in the central region of the medial compartment were more likely to progress to more advanced cartilage pathology (progression rate 28%; $P \leq 0.003$) than lesions in the anterior (19%; $P \leq 0.564$) and posterior (17%; $P \leq 0.957$) regions or lesions located in the lateral compartment (average progression rate 15%; $P \leq 0.707$). Lesions located in the anterior region of the lateral compartment showed less progression of cartilage degradation (6%; $P \leq 0.001$). No specific grade of lesion identified at baseline had a predilection for more rapid cartilage loss ($P \leq 0.93$)[76].
- There was a significant correlation between the degree of loss of tibial cartilage and the degree of loss of femoral cartilage, in both tibiofemoral joints ($r = 0.81$, $P < 0.001$ at the medial tibiofemoral joint; $r = 0.71$, $P < 0.001$ at the lateral tibiofemoral joint)[77].
- A highly significant difference in global cartilage volume loss was observed between severe medial meniscal tear and absence of tear [mean (SD), −10.1 (2.1)% v −5.1 (2.4)%, $P = 0.002$]. An even greater difference was found between the medial meniscal changes and medial compartment cartilage volume loss [−14.3 (3.0)% in the presence of severe tear v −6.3 (2.7)% in the absence of tear; $P < 0.0001$]. Similarly, a major difference was found between the presence of a medial meniscal extrusion and loss of medial compartment cartilage

volume [−15.4 (4.1)% in the presence of extrusion v −4.5 (1.7)% with no extrusion; $P < 0.001$][78].

- Annual patellar cartilage loss was highest in those with defects compared with no defects (5.5% vs 3.2%, $P = 0.01$). Tibial cartilage loss was not associated with defects in the medial (4.6% vs 5.8%, $P = 0.42$) or lateral (4.7% vs 6.5%, $P = 0.21$) tibial cartilages[66].
- Baseline cartilage defect score was negatively associated with the progression of cartilage defects in each compartment (all $P < 0.001$)[79].
- Baseline cartilage defect scores at the medial tibia, lateral tibia, and patella had a dose-response association with the annual rate of change in knee cartilage volume at the corresponding site (beta = −1.3% to −1.2% per grade; $P < 0.05$ for all comparisons). In addition, an increase in knee cartilage defect score (change of more than or equal to 1) was associated with higher rates of knee cartilage volume loss at all sites (beta = −1.9% to −1.7% per year; $P < 0.01$ for all comparisons). Furthermore, a decrease in the knee cartilage defect score (change of less than or equal to −1) was associated with an increase in knee cartilage volume at all sites (beta = 1.0%–2.7% per year; $P < 0.05$ for all comparisons)[80].
- Predictors of fast progression included the presence of severe meniscal extrusion ($P = 0.001$), severe medial tear ($P = 0.005$), medial and/or lateral bone edema ($P = 0.03$), high body mass index ($P < 0.05$, fast vs slow), weight ($P < 0.05$, fast vs slow) and age ($P < 0.05$ fast vs slow)[72].
- In the medial tibiofemoral joint, each measure of meniscal malposition was associated with an increased risk of cartilage loss. There was also a strong association between meniscal damage and cartilage loss[81].
- A worsening in cartilage defect score was significantly associated with tibiofemoral osteophytes (OR, 6.22 and 6.04 per grade), tibial bone area (OR, 1.24 and 2.07 per square centimeter), and cartilage volume (OR, 2.91 and 1.71 per ml in the medial tibiofemoral and patellar compartments)[82].
- Knee compartments with a higher baseline BML score had greater cartilage loss. An increase in BMLs was strongly associated with further worsening of the cartilage score[83].
- Despite cartilage loss occurring in over 50% of knees, synovitis was not associated with cartilage loss in either tibiofemoral or patellofemoral compartment[24].
- Significant correlations were seen between the loss of cartilage volume and edema size change in the medial condyle (−0.40, $P = 0.0001$) and the medial tibial plateau (−0.23, $P = 0.03$), and the changes in cyst size in the medial condyle (−0.29, $P = 0.01$). A multivariate analysis showed that the edema size change was strongly and independently associated with medial cartilage volume loss (−0.31, $P = 0.0004$)[84].
- Medial meniscal tear was associated with 103 mm(2) greater tibial plateau bone area within the medial (95% CI 6.2, 200.3; $P = 0.04$) and a lateral meniscal tear with a 120 mm(2) greater area within the lateral compartment (95% CI 45.5, 195.2; $P = 0.002$)[85].
- Adjusting for age, body mass index, gender and baseline cartilage scores, complete ACL tear increased the risk for cartilage loss at the medial tibiofemoral compartment (OR: 1.8, 95% CI: 1.1, 3.2). However, following adjustment for the presence of medial meniscal tears, no increased risk for cartilage loss was further seen (OR: 1.1, 95% CI: 0.6, 1.8)[86].
- Medial meniscal damage predicted medial tibial cartilage volume loss and tibial and femoral denuded bone increase, while varus malalignment predicted medial tibial cartilage volume and thickness loss and tibial and femoral denuded bone increase. Lateral meniscal damage predicted every lateral outcome[87].
- A positive correlation was found between the global severity of synovitis at baseline and the loss of cartilage volume at 60 days ($P < 0.03$)[25].

Discussion

The performance of MRI as an outcome measure in OA has been extensively studied providing strong support for both its concurrent and predictive validity.

As outlined in this review numerous studies have examined the relation of MRI to related constructs such as symptom measures, plain radiography, histology and arthroscopy. These studies demonstrate the following:

1. Inconsistent relation of structural features to symptoms with 13 of 21 studies finding a significant relation. Generally strong relation of large bone marrow lesions, moderate relation of synovitis and effusion and weak relation of cartilage volume/thickness to presence of pain. No relation of meniscal tears to presence of pain.
2. In general there was an inconsistent relation of cartilage volume and thickness and compositional measures to presence of radiographic OA. Higher frequency of meniscal tears, synovitis, increased bone area, increased bone attrition/curvature in persons with radiographic OA. Radiographic change insensitive to early changes found on MRI. 39 of 43 studies found significant associations between MRI and radiographic features.
3. There was a strong relation of meniscal subluxation and increased subchondral bone area to reduced radiographic joint space. Inconsistent (but generally moderate) relation of reduced cartilage volume and thickness to reduced radiographic joint space. Nine of nine studies found significant associations between MRI and radiographic joint space.
4. In general there was a strong correlation of cartilage volume measures to histologic findings. Three of five studies found significant relation of MRI to histology/pathology.
5. Moderate to strong relation of arthroscopic findings to cartilage and meniscal findings on MRI with five of seven studies finding a significant association
6. Strong relation of CT arthrography to MRI cartilage volume with all four studies examining this relation finding a significant association.

An important obstacle to biomarker validation and qualification is the adequate delineation of a gold standard. Unlike other diseases where surrogate endpoints exist, OA does not have a clear gold standard clinical endpoint and further is a remarkably heterogeneous disease. Therefore, the 'clinical endpoint' is more difficult to establish. A number of experts in the field have advocated that joint replacement be the clinical outcome of interest but due to constraints over comorbidities, insurance status and a number of other factors that influence determining if a person receives a joint replacement, alternate suggestions have been recommended including the use of virtual TKR (vTKR)[88]. This is a composite endpoint that includes domains of pain, physical function and joint structure on X-rays[89]. At this point it remains to be validated and as a consequence the constituent literature in this review does not include this endpoint to establish the predictive validity of MRI.

This work may be susceptible to publication bias as there was no effort made to search either clinical trial registries or meeting abstracts for potential unpublished studies that might tend to invalidate the MRI biomarkers examined.

The literature on the predictive validity of MRI in OA demonstrated the following:

1. Quantitative cartilage volume change and presence of cartilage defects or bone marrow lesions are potential predictors of TKR. Three of three studies found a significant relation.
2. Inconsistent but generally weak relation of cartilage loss to symptom change. Moderate relation of BML change to incident symptoms and pain change. Weak relation of change in synovitis to change in pain. Five of six studies found significant association between MRI and change in symptoms.
3. At best a weak relation between change in cartilage thickness and change in joint space. Five of eight studies found a significant relation.
4. Presence of meniscal damage, cartilage defects and BMLs predicts MRI progression. 16 of 19 studies found a significant relation.

Some MRI biomarkers correlate with some other biomarkers. Moreover in a limited number of studies some MRI biomarkers correlate with clinical endpoints and/or predict clinical outcomes Future research should be directed toward improving the predictive validity of current structural measures as they relate to important clinical outcomes so their role as surrogate outcomes can be substantiated. In addition, studies to improve the precision of assessment of structural features more closely related to symptom change such as BMLs and synovitis are warranted.

Role of funding source

The OARSI FDA OA Initiative received financial support from the following professional organization:

American College of Rheumatology

Additionally the OARSI FDA OA Initiative received financial support from the following companies:

Amgen
ArthroLab
AstraZeneca
Bayer Healthcare
Chondrometrics
CombinatoRx
Cypress BioScience
DePuy Mitek
Expanscience 4QImaging
Genevrier/IBSA
Genzyme
King (Alpharma) Merck
Merck Serono NicOx
Pfizer
Rottapharm
Smith & Nephew Wyeth

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the Food and Drug Administration.

Author contributions

DJH conceived and designed the study, drafted the manuscript and takes responsibility for the integrity of the work as a whole, from inception to finished article. EL and WZ were also involved in the design of the study. All authors contributed to acquisition of the data. All authors critically revised the manuscript and gave final approval of the article for submission.

The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflict of interest statement

David Hunter receives research or institutional support from DonJoy, NIH, and Stryker.

Other authors declared no conflict of interest

Acknowledgements

We recognize the invaluable support of Valorie Thompson for administrative and editorial support and OARSI for their invaluable support of this activity. This analysis and literature review was undertaken to facilitate discussions and development of recommendations by the Assessment of Structural Change Working group for the OARSI FDA Initiative. The members of the working group were: Philip Conaghan, MB, BS, PhD (Chair), David Hunter, MBBS, PhD, Jeffrey Duryea, PhD, Garry Gold, MD, Steven Mazzuca, PhD, Jean Francis Maillefert, MD, Timothy Mosher, MD, Hollis Potter, MD, David Felson, MD, Ali Guermazi, MD, Helen Keen, MD, Gayle Lester, PhD, Wayne Tsuji, MD, John Randle, PhD, Felix Eckstein, MD, Erika Schneider, PhD, Elena Losina, PhD, Sarah Kingsbury, PhD, William Reichman, PhD, Jean Pierre Pelletier, MD, Saara Totterman, MD, PhD, Rose Maciewicz, PhD, Bernard Dardzinski, PhD, Mona Wahba, MD, Marie Pierre Hellio Le Graverand-Gastineau, MD, PhD, DSc, Elisabeth Morris, DVM, Jeffrey Kraines, MD, Lucio Rovati, MD, Don Dreher, MD, PhD, James Huckle, PhD, Mary-Ann Preston, PhD, Brooks Story, PhD.

References

1. Kellgren JH, Lawrence JS. Atlas of standard radiographs. Oxford: Blackwell Scientific; 1963.
2. Guermazi A, Burstein D, Conaghan P, Eckstein F, Hellio Le Graverand-Gastineau MP, Keen H, *et al*. Imaging in osteoarthritis. Rheum Dis Clin North Am 2008;34:645–87.
3. Dieppe PA, Lohmander LS. Pathogenesis and management of pain in osteoarthritis. Lancet 2005;365:965–73 [Review] [100 refs].
4. Hannan MT, Felson DT, Pincus T. Analysis of the discordance between radiographic changes and knee pain in osteoarthritis of the knee. J Rheumatol 2000;27:1513–7.
5. Hunter DJ, Losina E, Guermazi A, Burstein D, Lassere MN, Kraus V. A pathway and approach to biomarker validation and qualification for osteoarthritis clinical trials. Curr Drug Targets 2010;11:536–45.
6. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998;52:377–84.
7. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998; 25:198–9.
8. Lassere M. A users guide to measurement in medicine. Osteoarthritis Cartilage 2006;14(Suppl 1):10–4.
9. Goodsaid FM, Frueh FW, Mattes W. Strategic paths for biomarker qualification. Toxicology 2008;245:219–23 [Review] [14 refs].
10. Wagner JA. Overview of biomarkers and surrogate endpoints in drug development. Dis Markers 2002;18:41–6 [Review] [21 refs].

11. NIH-FDA Conference. Biomarkers and surrogate endpoints: advancing clinical research and applications. Abstracts. Dis Markers 1998;14:187–334.
12. Lesko LJ, Atkinson Jr AJ. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. Annu Rev Pharmacol Toxicol 2001;41:347–66 [Review] [61 refs].
13. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin Pharmacol Ther 2001;69:89–95 [Review] [29 refs].
14. Felson DT, Chaisson CE, Hill CL, Totterman SM, Gale ME, Skinner KM, *et al*. The association of bone marrow lesions with pain in knee osteoarthritis. Ann Intern Med 2001;134:541–9 [see comments].
15. Hill CL, Gale DG, Chaisson CE, Skinner K, Kazis L, Gale ME, *et al*. Knee effusions, popliteal cysts, and synovial thickening: association with knee pain in osteoarthritis. J Rheumatol 2001;28:1330–7.
16. Bhattacharyya T, Gale D, Dewire P, Totterman S, Gale ME, McLaughlin S, *et al*. The clinical importance of meniscal tears demonstrated by magnetic resonance imaging in osteoarthritis of the knee. J Bone Joint Surg Am 2003;85-A:4–9 [see comment].
17. Link TM, Steinbach LS, Ghosh S, Ries M, Lu Y, Lane N, *et al*. Osteoarthritis: MR imaging findings in different stages of disease and correlation with clinical findings. Radiology 2003;226:373–81.
18. Sowers MF, Hayes C, Jamadar D, Capul D, Lachance L, Jannausch M, *et al*. Magnetic resonance-detected subchondral bone marrow and cartilage defect characteristics associated with pain and X-ray-defined knee osteoarthritis. Osteoarthritis Cartilage 2003;11:387–93.
19. Hill CL, Gale DR, Chaisson CE, Skinner K, Kazis L, Gale ME, *et al*. Periarticular lesions detected on magnetic resonance imaging: prevalence in knees with and without symptoms. Arthritis Rheum 2003;48:2836–44.
20. Wluka AE, Wolfe R, Stuckey S, Cicuttini FM. How does tibial cartilage volume relate to symptoms in subjects with knee osteoarthritis? Ann Rheum Dis 2004;63:264–8.
21. Torres L, Dunlop DD, Peterfy C, Guermazi A, Prasad P, Hayes KW, *et al*. The relationship between specific tissue lesions and pain severity in persons with knee osteoarthritis. Osteoarthritis Cartilage 2006;14:1033–40.
22. Kornaat PR, Bloem JL, Ceulemans RY, Riyazi N, Rosendaal FR, Nelissen RG, *et al*. Osteoarthritis of the knee: association between clinical features and MR imaging findings. Radiology 2006;239:811–7.
23. Hunter DJ, Lo GH, Gale D, Grainger AJ, Guermazi A, Conaghan PG. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). Ann Rheum Dis 2008;67:206–11 [Review] [28 refs].
24. Hill CL, Hunter DJ, Niu J, Clancy M, Guermazi A, Genant H, *et al*. Synovitis detected on magnetic resonance imaging and its relation to pain and cartilage loss in knee osteoarthritis. Ann Rheum Dis 2007;66:1599–603.
25. 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.
26. Englund M, Guermazi A, Gale D, Hunter DJ, Aliabadi P, Clancy M, *et al*. Incidental meniscal findings on knee MRI in middle-aged and elderly persons. N Engl J Med 2008;359:1108–15 [see comment].
27. Karvonen RL, Negendank WG, Teitge RA, Reed AH, Miller PR, Fernandez-Madrid F. Factors affecting articular cartilage thickness in osteoarthritis and aging. J Rheumatol 1994;21:1310–8.
28. Cicuttini FM, Wluka AE, Forbes A, Wolfe R, Cicuttini FM, Wluka AE, *et al*. Comparison of tibial cartilage volume and radiologic grade of the tibiofemoral joint. Arthritis Rheum 2003;48:682–8.
29. Lindsey CT, Narasimhan A, Adolfo JM, Jin H, Steinbach LS, Link T, *et al*. Magnetic resonance evaluation of the interrelationship between articular cartilage and trabecular bone of the osteoarthritic knee. Osteoarthritis Cartilage 2004;12:86–96.
30. Hellio Le Graverand MP, Buck RJ, Wyman BT, Vignon E, Mazzuca SA, Brandt KD, *et al*. Subregional femorotibial cartilage morphology in women–comparison between healthy controls and participants with different grades of radiographic knee osteoarthritis. Osteoarthritis Cartilage 2009;17:1177–85.
31. Dam EB, Folkesson J, Pettersen PC, Christiansen C. Automatic morphometric cartilage quantification in the medial tibial plateau from MRI for osteoarthritis grading. Osteoarthritis Cartilage 2007;15:808–18.
32. Williams A, Sharma L, McKenzie CA, Prasad PV, Burstein D. Delayed gadolinium-enhanced magnetic resonance imaging of cartilage in knee osteoarthritis: findings at different radiographic stages of disease and relationship to malalignment. Arthritis Rheum 2005;52:3528–35.
33. Nojiri T, Watanabe N, Namura T, Narita W, Ikoma K, Suginoshita T, *et al*. Utility of delayed gadolinium-enhanced MRI (dGEMRIC) for qualitative evaluation of articular cartilage of patellofemoral joint. Knee Surg Sports Traumatol Arthrosc 2006;14:718–23.
34. Koff MF, Amrami KK, Kaufman KR. Clinical evaluation of T2 values of patellar cartilage in patients with osteoarthritis. Osteoarthritis Cartilage 2007;15:198–204.
35. Li X, Benjamin MC, Link TM, Castillo DD, Blumenkrantz G, Lozano J, *et al*. In vivo T(1rho) and T(2) mapping of articular cartilage in osteoarthritis of the knee using 3 T MRI. Osteoarthritis Cartilage 2007;15:789–97.
36. Chan WP, Lang P, Stevens MP, Sack K, Majumdar S, Stoller DW, *et al*. Osteoarthritis of the knee: comparison of radiography, CT, and MR imaging to assess extent and severity. AJR Am J Roentgenol 1991;157:799–806.
37. Blackburn Jr WD, Bernreuter WK, Rominger M, Loose LL. Arthroscopic evaluation of knee articular cartilage: a comparison with plain radiographs and magnetic resonance imaging. J Rheumatol 1994;21:675–9.
38. Wluka AE, Wang Y, Davis SR, Cicuttini FM. Tibial plateau size is related to grade of joint space narrowing and osteophytes in healthy women and in women with osteoarthritis. Ann Rheum Dis 2005;64:1033–7.
39. Nakamura M, Sumen Y, Sakaridani K, Exham H, Ochi M. Relationship between the shape of tibial spurs on X-ray and meniscal changes on MRI in early osteoarthritis of the knee. Magn Reson Imaging 2006;24:1143–8.
40. Englund M, Niu J, Guermazi A, Roemer FW, Hunter DJ, Lynch JA, *et al*. Effect of meniscal damage on the development of frequent knee pain, aching, or stiffness. Arthritis Rheum 2007;56:4048–54.
41. Reichenbach S, Guermazi A, Niu J, Neogi T, Hunter DJ, Roemer FW, *et al*. Prevalence of bone attrition on knee radiographs and MRI in a community-based cohort. Osteoarthritis Cartilage 2008;16:1005–10.
42. Folkesson J, Dam EB, Olsen OF, Karsdal MA, Pettersen PC, Christiansen C. Automatic quantification of local and global

articular cartilage surface curvature: biomarkers for osteoarthritis? Magn Reson Med 2008;59:1340–6.
43. Gale DR, Chaisson CE, Totterman SM, Schwartz RK, Gale ME, Felson D. Meniscal subluxation: association with osteoarthritis and joint space narrowing. Osteoarthritis Cartilage 1999;7:526–32.
44. Adams JG, McAlindon T, Dimasi M, Carey J, Eustace S. Contribution of meniscal extrusion and cartilage loss to joint space narrowing in osteoarthritis. Clin Radiol 1999;54:502–6.
45. Hill CL, Seo GS, Gale D, Totterman S, Gale ME, Felson DT. Cruciate ligament integrity in osteoarthritis of the knee. Arthritis Rheum 2005;52:794–9.
46. Cicuttini FM, Wang YY, Forbes A, Wluka AE, Glisson M. Comparison between patella cartilage volume and radiological assessment of the patellofemoral joint. Clin Exp Rheumatol 2003;21:321–6.
47. Jones G, Ding C, Scott F, Glisson M, Cicuttini F. Early radiographic osteoarthritis is associated with substantial changes in cartilage volume and tibial bone surface area in both males and females. Osteoarthritis Cartilage 2004;12: 169–74.
48. Raynauld JP, Martel-Pelletier J, Berthiaume MJ, Labonte F, Beaudoin G, de Guise JA, *et al*. Quantitative magnetic resonance imaging evaluation of knee osteoarthritis progression over two years and correlation with clinical symptoms and radiologic changes. Arthritis Rheum 2004;50:476–87.
49. Ding C, Garnero P, Cicuttini F, Scott F, Cooley H, Jones G, *et al*. Knee cartilage defects: association with early radiographic osteoarthritis, decreased cartilage volume, increased joint surface area and type II collagen breakdown. Osteoarthritis Cartilage 2005;13:198–205.
50. Cicuttini F, Hankin J, Jones G, Wluka A. Comparison of conventional standing knee radiographs and magnetic resonance imaging in assessing progression of tibiofemoral joint osteoarthritis. Osteoarthritis Cartilage 2005;13: 722–7.
51. Felson DT, McLaughlin S, Goggins J, Lavalley MP, Gale ME, Totterman S, *et al*. Bone marrow edema and its relation to progression of knee osteoarthritis. Ann Intern Med 2003;139:330–6.
52. Cicuttini F, Wluka A, Hankin J, Wang Y. Longitudinal study of the relationship between knee angle and tibiofemoral cartilage volume in subjects with knee osteoarthritis. Rheumatology (Oxford) 2004;43:321–4.
53. Hunter DJ, Zhang Y, Niu J, Tu X, Amin S, Goggins J, *et al*. Structural factors associated with malalignment in knee osteoarthritis: the Boston osteoarthritis knee study. J Rheumatol 2005;32:2192–9 [see comment].
54. Issa SN, Dunlop D, Chang A, Song J, Prasad PV, Guermazi A, *et al*. Full-limb and knee radiography assessments of varus-valgus alignment and their relationship to osteoarthritis disease features by magnetic resonance imaging. Arthritis Rheum 2007;57:398–406.
55. Janakiramanan N, Teichtahl AJ, Wluka AE, Ding C, Jones G, Davis SR, *et al*. Static knee alignment is associated with the risk of unicompartmental knee cartilage defects. J Orthop Res 2008;26:225–30.
56. Eckstein F, Schnier M, Haubner M, Priebsch J, Glaser C, Englmeier KH, *et al*. Accuracy of cartilage volume and thickness measurements with magnetic resonance imaging. Clin Orthop Relat Res 1998;(352):137–48.
57. Dupuy DE, Spillane RM, Rosol MS, Rosenthal DI, Palmer WE, Burke DW, *et al*. Quantification of articular cartilage in the knee with three-dimensional MR imaging. Acad Radiol 1996;3:919–24.
58. Uhl M, Ihling C, Allmann KH, Laubenberger J, Tauer U, Adler CP, *et al*. Human articular cartilage: in vitro correlation of MRI and histologic findings. Eur Radiol 1998;8:1123–9.
59. Kladny B, Martus P, Schiwy-Bochat KH, Weseloh G, Swoboda B. Measurement of cartilage thickness in the human knee-joint by magnetic resonance imaging using a three-dimensional gradient-echo sequence. Int Orthop 1999;23:264–7.
60. McGibbon CA, Trahan CA. Measurement accuracy of focal cartilage defects from MRI and correlation of MRI graded lesions with histology: a preliminary study. Osteoarthritis Cartilage 2003;11:483–93.
61. Broderick LS, Turner DA, Renfrew DL, Schnitzer TJ, Huff JP, Harris C. Severity of articular cartilage abnormality in patients with osteoarthritis: evaluation with fast spin-echo MR vs arthroscopy. AJR Am J Roentgenol 1994;162:99–103.
62. Kawahara Y, Uetani M, Nakahara N, Doiguchi Y, Nishiguchi M, Futagawa S, *et al*. Fast spin-echo MR of the articular cartilage in the osteoarthrotic knee. Correlation of MR and arthroscopic findings. Acta Radiol 1998;39:120–5.
63. Drape JL, Pessis E, Auleley GR, Chevrot A, Dougados M, Ayral X. Quantitative MR imaging evaluation of chondropathy in osteoarthritic knees. Radiology 1998;208:49–55.
64. Bachmann GF, Basad E, Rauber K, Damian MS, Rau WS. Degenerative joint disease on MRI and physical activity: a clinical study of the knee joint in 320 patients. Eur Radiol 1999;9:145–52.
65. Cicuttini FM, Jones G, Forbes A, Wluka AE. Rate of cartilage loss at two years predicts subsequent total knee arthroplasty: a prospective study. Ann Rheum Dis 2004;63:1124–7.
66. Wluka AE, Ding C, Jones G, Cicuttini FM. The clinical correlates of articular cartilage defects in symptomatic knee osteoarthritis: a prospective study. Rheumatology (Oxford) 2005;44:1311–6.
67. Scher C, Craig J, Nelson F. Bone marrow edema in the knee in osteoarthrosis and association with total knee arthroplasty within a three-year follow-up. Skeletal Radiol 2008;37:609–17.
68. Phan CM, Link TM, Blumenkrantz G, Dunn TC, Ries MD, Steinbach LS, *et al*. MR imaging findings in the follow-up of patients with different stages of knee osteoarthritis and the correlation with clinical symptoms. Eur Radiol 2006;16:608–18.
69. Katz JN, Meredith DS, Lang P, Creel AH, Yoshioka H, Neumann G, *et al*. Associations among preoperative MRI features and functional status following arthroscopic partial meniscectomy. Osteoarthritis Cartilage 2006;14:418–22.
70. Felson DT, Niu J, Guermazi A, Roemer F, Aliabadi P, Clancy M, *et al*. Correlation of the development of knee pain with enlarging bone marrow lesions on magnetic resonance imaging. Arthritis Rheum 2007;56:2986–92.
71. Pelletier JP, Raynauld JP, Berthiaume MJ, Abram F, Choquette D, Haraoui B, *et al*. Risk factors associated with the loss of cartilage volume on weight-bearing areas in knee osteoarthritis patients assessed by quantitative magnetic resonance imaging: a longitudinal study. Arthritis Res Ther 2007;9:R74.
72. Raynauld JP, Martel-Pelletier J, Berthiaume MJ, Beaudoin G, Choquette D, Haraoui B, *et al*. Long term evaluation of disease progression through the quantitative magnetic resonance imaging of symptomatic knee osteoarthritis patients: correlation with clinical symptoms and radiographic changes. Arthritis Res Ther 2006;8:R21.
73. Bruyere O, Genant H, Kothari M, Zaim S, White D, Peterfy C, *et al*. Longitudinal study of magnetic resonance imaging and

standard X-rays to assess disease progression in osteoarthritis. Osteoarthritis Cartilage 2007;15:98–103.
74. Nevitt MC, Peterfy C, Guermazi A, Felson DT, Duryea J, Woodworth T, *et al*. Longitudinal performance evaluation and validation of fixed-flexion radiography of the knee for detection of joint space loss. Arthritis Rheum 2007;56:1512–20.
75. Madan-Sharma R, Kloppenburg M, Kornaat PR, Botha-Scheepers SA, Le Graverand MP, Bloem JL, *et al*. Do MRI features at baseline predict radiographic joint space narrowing in the medial compartment of the osteoarthritic knee 2 years later? Skeletal Radiol 2008;37:805–11.
76. Biswal S, Hastie T, Andriacchi TP, Bergman GA, Dillingham MF, Lang P. Risk factors for progressive cartilage loss in the knee: a longitudinal magnetic resonance imaging study in forty-three patients. Arthritis Rheum 2002;46: 2884–92.
77. Cicuttini FM, Wluka AE, Wang Y, Stuckey SL, Cicuttini FM, Wluka AE, *et al*. Longitudinal study of changes in tibial and femoral cartilage in knee osteoarthritis. Arthritis Rheum 2004;50:94–7 [see comment].
78. Berthiaume MJ, Raynauld JP, Martel-Pelletier J, Labonte F, Beaudoin G, Bloch DA, *et al*. Meniscal tear and extrusion are strongly associated with progression of symptomatic knee osteoarthritis as assessed by quantitative magnetic resonance imaging. Ann Rheum Dis 2005;64:556–63.
79. Wang Y, Ding C, Wluka AE, Davis S, Ebeling PR, Jones G, *et al*. Factors affecting progression of knee cartilage defects in normal subjects over 2 years. Rheumatology (Oxford) 2006;45:79–84.
80. Ding C, Cicuttini F, Scott F, Boon C, Jones G. Association of prevalent and incident knee cartilage defects with loss of tibial and patellar cartilage: a longitudinal study. Arthritis Rheum 2005;52:3918–27.
81. Hunter DJ, Zhang YQ, Niu JB, Tu X, Amin S, Clancy M, *et al*. The association of meniscal pathologic changes with cartilage loss in symptomatic knee osteoarthritis. Arthritis Rheum 2006;54:795–801.
82. Ding C, Cicuttini F, Scott F, Cooley H, Boon C, Jones G. Natural history of knee cartilage defects and factors affecting change. Arch Intern Med 2006;166:651–8.
83. Hunter DJ, Zhang Y, Niu J, Goggins J, Amin S, LaValley MP, *et al*. Increase in bone marrow lesions associated with cartilage loss: a longitudinal magnetic resonance imaging study of knee osteoarthritis. Arthritis Rheum 2006;54:1529–35.
84. Raynauld JP, Martel-Pelletier J, Berthiaume MJ, Abram F, Choquette D, Haraoui B, *et al*. Correlation between bone lesion changes and cartilage volume loss in patients with osteoarthritis of the knee as assessed by quantitative magnetic resonance imaging over a 24-month period. Ann Rheum Dis 2008;67:683–8.
85. Davies-Tuck ML, Martel-Pelletier J, Wluka AE, Pelletier JP, Ding C, Jones G, *et al*. Meniscal tear and increased tibial plateau bone area in healthy post-menopausal women. Osteoarthritis Cartilage 2008;16:268–71.
86. Amin S, Guermazi A, Lavalley MP, Niu J, Clancy M, Hunter DJ, *et al*. Complete anterior cruciate ligament tear and the risk for cartilage loss and progression of symptoms in men and women with knee osteoarthritis. Osteoarthritis Cartilage 2008;16:897–902.
87. Sharma L, Eckstein F, Song J, Guermazi A, Prasad P, Kapoor D, *et al*. Relationship of meniscal damage, meniscal extrusion, malalignment, and joint laxity to subsequent cartilage loss in osteoarthritic knees. Arthritis Rheum 2008;58:1716–26.
88. Gossec L, Hawker G, Davis AM, Maillefert JF, Lohmander LS, Altman R, *et al*. OMERACT/OARSI initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis. J Rheumatol 2007;34:1432–5 [12 refs].
89. Ornetti P, Brandt K, Hellio Graverand M, Hochberg M, Hunter D, Kloppenburg M, *et al*. OARSI-OMERACT definition of relevant radiological progression in hip/knee osteoarthritis. Osteoarthritis Cartilage 2009 Jul;17(7):856–63.
90. McAlindon TE, Watt I, McCrae F, Goddard P, Dieppe PA. Magnetic resonance imaging in osteoarthritis of the knee: correlation with radiographic and scintigraphic findings. Ann Rheum Dis 1991;50:14–9.
91. Li KC, Higgs J, Aisen AM, Buckwalter KA, Martel W, McCune WJ. MRI in osteoarthritis of the hip: gradations of severity. Magn Reson Imaging 1988;6:229–36.
92. Fernandez-Madrid F, Karvonen RL, Teitge RA, Miller PR, Negendank WG. MR features of osteoarthritis of the knee. Magn Reson Imaging 1994;12:703–9.
93. Peterfy C, van Dijke C, Janzen D, Gluer C, Namba R, Majumdar S, *et al*. Quantification of articular cartilage in the knee with pulsed saturation transfer subtraction and fat-suppressed MR imaging: optimization and validation. Radiology 1994;192:485–91.
94. Miller TT, Staron RB, Koenigsberg T, Levin TL, Feldman F. MR imaging of Baker cysts: association with internal derangement, effusion, and degenerative arthropathy. Radiology 1996;201:247–50 [see comment].
95. Kenny C. Radial displacement of the medial meniscus and Fairbank's signs. Clin Orthop Relat Res 1997;(339):163–73.
96. Breitenseher MJ, Trattnig S, Dobrocky I, Kukla C, Nehrer S, Steiner E, *et al*. MR imaging of meniscal subluxation in the knee. Acta Radiol 1997;38:876–9.
97. Ostergaard M. Different approaches to synovial membrane volume determination by magnetic resonance imaging: manual versus automated segmentation. Br J Rheumatol 1997;36:1166–77.
98. Trattnig S, Huber M, Breitenseher MJ, Trnka HJ, Rand T, Kaider A, *et al*. Imaging articular cartilage defects with 3D fat-suppressed echo planar imaging: comparison with conventional 3D fat-suppressed gradient echo sequence and correlation with histology. J Comput Assist Tomogr 1998;22:8–14.
99. Boegard T. Radiography and bone scintigraphy in osteoarthritis of the knee—comparison with MR imaging. Acta Radiol 1998;418:7–37. Supplementum.
100. Cicuttini F, Forbes A, Morris K, Darling S, Bailey M, Stuckey S. Gender differences in knee cartilage volume as measured by magnetic resonance imaging. Osteoarthritis Cartilage 1999;7:265–71.
101. Boegard T, Rudling O, Dahlstrom J, Dirksen H, Petersson IF, Jonsson K. Bone scintigraphy in chronic knee pain: comparison with magnetic resonance imaging. Ann Rheum Dis 1999;58:20–6.
102. Pham XV, Monteiro I, Judet O, Sissakian JF, Plantin P, Aegerter P, *et al*. Magnetic resonance imaging changes in periarticular soft tissues during flares of medial compartment knee osteoarthritis. Preliminary study in 10 p.tients. Rev Rhum Engl Ed 1999;66:398–403.
103. Zanetti M, Bruder E, Romero J, Hodler J. Bone marrow edema pattern in osteoarthritic knees: correlation between MR imaging and histologic findings. Radiology 2000;215: 835–40.
104. Jones G, Glisson M, Hynes K, Cicuttini F. Sex and site differences in cartilage development: a possible explanation for variations in knee osteoarthritis in later life. Arthritis Rheum 2000;43:2543–9.

105. McCauley TR, Kornaat PR, Jee WH. Central osteophytes in the knee: prevalence and association with cartilage defects on MR imaging. AJR Am J Roentgenol 2001;176:359–64.
106. Wluka AE, Davis SR, Bailey M, Stuckey SL, Cicuttini FM. Users of oestrogen replacement therapy have more knee cartilage than non-users. Ann Rheum Dis 2001;60:332–6.
107. Kawahara Y, Uetani M, Fuchi K, Eguchi H, Hashmi R, Hayashi K. MR assessment of meniscal movement during knee flexion: correlation with the severity of cartilage abnormality in the femorotibial joint. J Comput Assist Tomogr 2001;25:683–90.
108. Arokoski JP, Arokoski MH, Jurvelin JS, Helminen HJ, Niemitukia LH, Kroger H. Increased bone mineral content and bone size in the femoral neck of men with hip osteoarthritis. Ann Rheum Dis 2002;61:145–50.
109. Bergin D, Keogh C, O'Connell M, Rowe D, Shah B, Zoga A, *et al*. Atraumatic medial collateral ligament oedema in medial compartment knee osteoarthritis. Skeletal Radiol 2002;31: 14–8.
110. Beuf O, Ghosh S, Newitt DC, Link TM, Steinbach L, Ries M, *et al*. Magnetic resonance imaging of normal and osteoarthritic trabecular bone structure in the human knee. Arthritis Rheum 2002;46:385–93.
111. Arokoski MH, Arokoski JP, Haara M, Kankaanpaa M, Vesterinen M, Niemitukia LH, *et al*. Hip muscle strength and muscle cross sectional area in men with and without hip osteoarthritis. J Rheumatol 2002;29:2185–95.
112. Tiderius CJ, Olsson LE, Leander P, Ekberg O, Dahlberg L. Delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) in early knee osteoarthritis. Magn Reson Med 2003;49:488–92.
113. Cicuttini FM, Wluka AE, Wang Y, Stuckey SL, Davis SR. Effect of estrogen replacement therapy on patella cartilage in healthy women. Clin Exp Rheumatol 2003;21:79–82.
114. Tarhan S, Unlu Z. Magnetic resonance imaging and ultrasonographic evaluation of the patients with knee osteoarthritis: a comparative study. Clin Rheumatol 2003;22:181–8.
115. Kim YJ, Jaramillo D, Millis MB, Gray ML, Burstein D. Assessment of early osteoarthritis in hip dysplasia with delayed gadolinium-enhanced magnetic resonance imaging of cartilage. J Bone Joint Surg Am 2003;85-A:1987–92.
116. Peterfy CG, Guermazi A, Zaim S, Tirman PF, Miaux Y, White D, *et al*. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthritis Cartilage 2004;12:177–90.
117. Graichen H, Eisenhart-Rothe R, Vogl T, Englmeier KH, Eckstein F. Quantitative assessment of cartilage status in osteoarthritis by quantitative magnetic resonance imaging: technical validation for use in analysis of cartilage volume and further morphologic parameters. Arthritis Rheum 2004;50:811–6.
118. Dashti M, Wluka AE, Geso M, Davis SR, Stuckey S, Cicuttini FM. Relationship between the area of cartilage shown on the magnetic resonance imaging middle-slice image of the medial and lateral tibial cartilages with cartilage volume and grade of osteoarthritis over time. Scand J Rheumatol 2004;33:87–93.
119. Arokoski JP, Arokoski MH, Vainio P, Kroger H, Jurvelin JS. Estimation of femoral head bone density using magnetic resonance imaging: comparison between men with and without hip osteoarthritis. J Clin Densitom 2004;7:183–91.
120. Dunn TC, Lu Y, Jin H, Ries MD, Majumdar S. T2 relaxation time of cartilage at MR imaging: comparison with severity of knee osteoarthritis. Radiology 2004;232:592–8.
121. Regatte RR, Akella SV, Wheaton AJ, Lech G, Borthakur A, Kneeland JB, *et al*. 3D-T1rho-relaxation mapping of articular cartilage: in vivo assessment of early degenerative changes in symptomatic osteoarthritic subjects. Acad Radiol 2004;11:741–9.
122. Baysal O, Baysal T, Alkan A, Altay Z, Yologlu S. Comparison of MRI graded cartilage and MRI based volume measurement in knee osteoarthritis. Swiss Med Wkly 2004;134:283–8.
123. Lerer DB, Umans HR, Hu MX, Jones MH. The role of meniscal root pathology and radial meniscal tear in medial meniscal extrusion. Skeletal Radiol 2004;33:569–74.
124. King KB, Lindsey CT, Dunn TC, Ries MD, Steinbach LS, Majumdar S. A study of the relationship between molecular biomarkers of joint degeneration and the magnetic resonance-measured characteristics of cartilage in 16 symptomatic knees. Magn Reson Imaging 2004;22:1117–23.
125. Carbone LD, Nevitt MC, Wildy K, Barrow KD, Harris F, Felson D, *et al*. The relationship of antiresorptive drug use to structural findings and symptoms of knee osteoarthritis. Arthritis Rheum 2004;50:3516–25 [see comment].
126. Cicuttini F, Morris KF, Glisson M, Wluka AE. Slice thickness in the assessment of medial and lateral tibial cartilage volume and accuracy for the measurement of change in a longitudinal study. J Rheumatol 2004;31:2444–8.
127. Kornaat PR, Watt I, Riyazi N, Kloppenburg M, Bloem JL. The relationship between the MRI features of mild osteoarthritis in the patellofemoral and tibiofemoral compartments of the knee. Eur Radiol 2005;15:1538–43.
128. Zhai G, Cicuttini F, Srikanth V, Cooley H, Ding C, Jones G. Factors associated with hip cartilage volume measured by magnetic resonance imaging: the Tasmanian older adult cohort study. Arthritis Rheum 2005;52:1069–76.
129. Blankenbaker DG, De Smet AA, Fine JP. Is intra-articular pathology associated with MCL edema on MR imaging of the non-traumatic knee? Skeletal Radiol 2005;34:462–7.
130. Huh YM, Kim S, Suh JS, Song H, Song K, Shin KH. The role of popliteal lymph nodes in differentiating rheumatoid arthritis from osteoarthritis by using CE 3D FSPGR MR imaging: relationship of the inflamed synovial volume. Korean J Radiol 2005;6:117–24.
131. von Eisenhart-Rothe R, Graichen H, Hudelmaier M, Vogl T, Sharma L, Eckstein F. Femorotibial and patellar cartilage loss in patients prior to total knee arthroplasty, heterogeneity, and correlation with alignment of the knee. Ann Rheum Dis 2006;65:69–73.
132. Tan AL, Grainger AJ, Tanner SF, Shelley DM, Pease C, Emery P, *et al*. High-resolution magnetic resonance imaging for the assessment of hand osteoarthritis. Arthritis Rheum 2005;52: 2355–65.
133. Lo GH, Hunter DJ, Zhang Y, McLennan CE, LaValley MP, Kiel DP, *et al*. Bone marrow lesions in the knee are associated with increased local bone density. Arthritis Rheum 2005;52: 2814–21.
134. Li X, Han ET, Ma CB, Link TM, Newitt DC, Majumdar S. In vivo 3T spiral imaging based multi-slice T(1rho) mapping of knee cartilage in osteoarthritis. Magn Reson Med 2005;54:929–36.
135. Rhodes LA, Grainger AJ, Keenan AM, Thomas C, Emery P, Conaghan PG. The validation of simple scoring methods for evaluating compartment-specific synovitis detected by MRI in knee osteoarthritis. Rheumatology 2005;44:1569–73.
136. Loeuille D, Chary-Valckenaere I, Champigneulle J, Rat AC, Toussaint F, Pinzano-Watrin A, *et al*. Macroscopic and microscopic features of synovial membrane inflammation in the osteoarthritic knee: correlating magnetic resonance imaging findings with disease severity. Arthritis Rheum 2005;52:3492–501.
137. Roos EM, Dahlberg L, Roos EM, Dahlberg L. Positive effects of moderate exercise on glycosaminoglycan content in knee

cartilage: a four-month, randomized, controlled trial in patients at risk of osteoarthritis. Arthritis Rheum 2005;52:3507–14.
138. Kimelman T, Vu A, Storey P, McKenzie C, Burstein D, Prasad P. Three-dimensional T1 mapping for dGEMRIC at 3.0 T using the look locker method. Invest Radiol 2006;41:198–203.
139. Sengupta M, Zhang YQ, Niu JB, Guermazi A, Grigorian M, Gale D, *et al*. High signal in knee osteophytes is not associated with knee pain. Osteoarthritis Cartilage 2006;14:413–7.
140. Grainger AJ, Rhodes LA, Keenan AM, Emery P, Conaghan PG. Quantifying peri-meniscal synovitis and its relationship to meniscal pathology in osteoarthritis of the knee. Eur Radiol 2007;17:119–24.
141. Cashman PM, Kitney RI, Gariba MA, Carter ME. Automated techniques for visualization and mapping of articular cartilage in MR images of the osteoarthritic knee: a base technique for the assessment of microdamage and submicro damage. IEEE Trans Nanobioscience 2002;1:42–51.
142. Bamac B, Ozdemir S, Sarisoy HT, Colak T, Ozbek A, Akansel G. Evaluation of medial and lateral meniscus thicknesses in early osteoarthritis of the knee with magnetic resonance imaging. Saudi Med J 2006;27:854–7.
143. Boks SS, Vroegindeweij D, Koes BW, Hunink MM, Bierma-Zeinstra SM. Magnetic resonance imaging abnormalities in symptomatic and contralateral knees: prevalence and associations with traumatic history in general practice. Am J Sports Med 2006;34:1984–91.
144. Folkesson J, Dam EB, Olsen OF, Pettersen PC, Christiansen C. Segmenting articular cartilage automatically using a voxel classification approach. IEEE Trans Med Imaging 2007;26:106–15.
145. Iwasaki J, Sasho T, Nakagawa K, Ogino S, Ochiai N, Moriya H. Irregularity of medial femoral condyle on MR imaging serves as a possible indicator of objective severity of medial-type osteoarthritic knee-a pilot study. Clin Rheumatol 2007;26:1705–8.
146. Tiderius CJ, Jessel R, Kim YJ, Burstein D. Hip dGEMRIC in asymptomatic volunteers and patients with early osteoarthritis: the influence of timing after contrast injection. Magn Reson Med 2007;57:803–5.
147. Baranyay FJ, Wang Y, Wluka AE, English DR, Giles GG, Sullivan RO, *et al*. Association of bone marrow lesions with knee structures and risk factors for bone marrow lesions in the knees of clinically healthy, community-based adults. Semin Arthritis Rheum 2007;37:112–8.
148. Hanna F, Teichtahl AJ, Bell R, Davis SR, Wluka AE, O'Sullivan R, *et al*. The cross-sectional relationship between fortnightly exercise and knee cartilage properties in healthy adult women in midlife. Menopause 2007;14:830–4 [see comment].
149. Qazi AA, Folkesson J, Pettersen PC, Karsdal MA, Christiansen C, Dam EB. Separation of healthy and early osteoarthritis by automatic quantification of cartilage homogeneity. Osteoarthritis Cartilage 2007;15:1199–206.
150. Lammentausta E, Kiviranta P, Toyras J, Hyttinen s, Kiviranta I, Nieminen MT, *et al*. Quantitative MRI of parallel changes of articular cartilage and underlying trabecular bone in degeneration. Osteoarthritis Cartilage 2007;15:1149–57.
151. Guymer E, Baranyay F, Wluka AE, Hanna F, Bell RJ, Davis SR, *et al*. A study of the prevalence and associations of subchondral bone marrow lesions in the knees of healthy, middle-aged women. Osteoarthritis Cartilage 2007;15:1437–42.
152. Nishii T, Tanaka H, Sugano N, Sakai T, Hananouchi T, Yoshikawa H. Evaluation of cartilage matrix disorders by T2 relaxation time in patients with hip dysplasia. Osteoarthritis Cartilage 2008;16:227–33.
153. Lo GH, Niu J, McLennan CE, Kiel DP, McLean RR, Guermazi A, *et al*. Meniscal damage associated with increased local subchondral bone mineral density: a Framingham study. Osteoarthritis Cartilage 2008;16:261–7.
154. Davies-Tuck M, Teichtahl AJ, Wluka AE, Wang Y, Urquhart DM, Cui J, *et al*. Femoral sulcus angle and increased patella facet cartilage volume in an osteoarthritic population. Osteoarthritis Cartilage 2008;16:131–5.
155. Qazi AA, Dam EB, Nielsen M, Karsdal MA, Pettersen PC, Christiansen C. Osteoarthritic cartilage is more homogeneous than healthy cartilage: identification of a superior region of interest colocalized with a major risk factor for osteoarthritis. Acad Radiol 2007;14:1209–20.
156. Folkesson J, Dam EB, Olsen OF, Christiansen C. Accuracy evaluation of automatic quantification of the articular cartilage surface curvature from MRI. Acad Radiol 2007;14:1221–8.
157. Kamei G, Sumen Y, Sakaridani K. Evaluation of cartilage defect at medial femoral condyle in early osteoarthritis of the knee. Magn Reson Imaging 2008;26:567–71.
158. Li W, Scheidegger R, Wu Y, Vu A, Prasad PV. Accuracy of T1 measurement with 3-D look-locker technique for dGEMRIC. J Magn Reson Imaging 2008;27:678–82.
159. Taljanovic MS, Graham AR, Benjamin JB, Gmitro AF, Krupinski EA, Schwartz SA, *et al*. Bone marrow edema pattern in advanced hip osteoarthritis: quantitative assessment with magnetic resonance imaging and correlation with clinical examination, radiographic findings, and histopathology. Skeletal Radiol 2008;37:423–31.
160. Oda H, Igarashi M, Sase H, Sase T, Yamamoto S. Bone bruise in magnetic resonance imaging strongly correlates with the production of joint effusion and with knee osteoarthritis. J Orthop Sci 2008;13:7–15.
161. Hanna FS, Bell RJ, Cicuttini FM, Davison SL, Wluka AE, Davis SR. High sensitivity C-reactive protein is associated with lower tibial cartilage volume but not lower patella cartilage volume in healthy women at mid-life. Arthritis Res Ther 2008;10:R27.
162. Petterson SC, Barrance P, Buchanan T, Binder-Macleod S, Snyder-Mackler L. Mechanisms underlying quadriceps weakness in knee osteoarthritis. Med Sci Sports Exerc 2008;40:422–7.
163. Bolbos RI, Zuo J, Banerjee S, Link TM, Ma CB, Li X, *et al*. Relationship between trabecular bone structure and articular cartilage morphology and relaxation times in early OA of the knee joint using parallel MRI at 3 T. Osteoarthritis Cartilage 2008;16:1150–9.
164. Quaia E, Toffanin R, Guglielmi G, Ukmar M, Rossi A, Martinelli B, *et al*. Fast T2 mapping of the patellar articular cartilage with gradient and spin-echo magnetic resonance imaging at 1.5 T: validation and initial clinical experience in patients with osteoarthritis. Skeletal Radiol 2008;37:511–7.
165. Mills PM, Wang Y, Cicuttini FM, Stoffel K, Stachowiak GW, Podsiadlo P, *et al*. Tibio-femoral cartilage defects 3-5 years following arthroscopic partial medial meniscectomy. Osteoarthritis Cartilage 2008;16:1526–31.
166. Dore D, Ding C, Jones G. A pilot study of the reproducibility and validity of measuring knee subchondral bone density in the tibia. Osteoarthritis Cartilage 2008;16:1539–44.
167. Mutimer J, Green J, Field J. Comparison of MRI and wrist arthroscopy for assessment of wrist cartilage. J Hand Surg Eur Vol 2008;33:380–2.

168. Amin S, Goggins J, Niu J, Guermazi A, Grigoryan M, Hunter DJ, *et al*. Occupation-related squatting, kneeling, and heavy lifting and the knee joint: a magnetic resonance imaging-based study in men. J Rheumatol 2008;35:1645–9.
169. Li X, Ma BC, Bolbos RI, Stahl R, Lozano J, Zuo J, *et al*. Quantitative assessment of bone marrow edema-like lesion and overlying cartilage in knees with osteoarthritis and anterior cruciate ligament tear using MR imaging and spectroscopic imaging at 3 Tesla. J Magn Reson Imaging 2008;28:453–61.
170. Stahl R, Luke A, Li X, Carballido-Gamio J, Ma CB, Majumdar S, *et al*. T1rho, T2 and focal knee cartilage abnormalities in physically active and sedentary healthy subjects versus early OA patients–a 3.0-Tesla MRI study. Eur Radiol 2009;19: 132–43.
171. Brem MH, Schlechtweg PM, Bhagwat J, Genovese M, Dillingham MF, Yoshioka H, *et al*. Longitudinal evaluation of the occurrence of MRI-detectable bone marrow edema in osteoarthritis of the knee. Acta Radiol 2008;49:1031–7.
172. Lancianese SL, Kwok E, Beck CA, Lerner AL. Predicting regional variations in trabecular bone mechanical properties within the human proximal tibia using MR imaging. Bone 2008;43:1039–46.
173. Mamisch TC, Dudda M, Hughes T, Burstein D, Kim YJ. Comparison of delayed gadolinium enhanced MRI of cartilage (dGEMRIC) using inversion recovery and fast T1 mapping sequences. Magn Reson Med 2008;60:768–73.
174. Rauscher I, Stahl R, Cheng J, Li X, Huber MB, Luke A, *et al*. Meniscal measurements of T1rho and T2 at MR imaging in healthy subjects and patients with osteoarthritis. Radiology 2008;249:591–600.
175. Li W, Du H, Scheidegger R, Wu Y, Prasad PV. Value of precontrast T(1) for dGEMRIC of native articular cartilage. J Magn Reson Imaging 2009;29:494–7.
176. Choi JW, Chung HW, Ahn JH, Yoon YC. Central hole tear of the discoid meniscus of the knee in magnetic resonance imaging: mimicking the bucket-handle tear. J Comput Assist Tomogr 2009;33:155–9.
177. Chen YH, Carrino JA, Raman SP, Morrison WB, Fayad LM. Atraumatic lateral collateral ligament complex signal abnormalities by magnetic resonance imaging in patients with osteoarthrosis of the knee. J Comput Assist Tomogr 2008;32: 982–6.
178. Boegard TL, Rudling O, Petersson IF, Jonsson K. Magnetic resonance imaging of the knee in chronic knee pain. A 2-year follow-up. Osteoarthritis Cartilage 2001;9:473–80.
179. Wluka AE, Stuckey S, Snaddon J, Cicuttini FM. The determinants of change in tibial cartilage volume in osteoarthritic knees. Arthritis Rheum 2002;46:2065–72.
180. Cicuttini FM, Forbes A, Yuanyuan W, Rush G, Stuckey SL. Rate of knee cartilage loss after partial meniscectomy. J Rheumatol 2002;29:1954–6.
181. Cicuttini F, Wluka A, Wang Y, Stuckey S, Cicuttini F, Wluka A, *et al*. The determinants of change in patella cartilage volume in osteoarthritic knees. J Rheumatol 2002;29:2615–9.
182. Pessis E, Drape JL, Ravaud P, Chevrot A, Dougados M, Ayral X. Assessment of progression in knee osteoarthritis: results of a 1 year study comparing arthroscopy and MRI. Osteoarthritis Cartilage 2003;11:361–9.
183. Cicuttini FM, Wluka AE, Hankin J, Stuckey S. Comparison of patella cartilage volume and radiography in the assessment of longitudinal joint change at the patellofemoral joint. J Rheumatol 2004;31:1369–72.
184. Cubukcu D, Ardic F, Karabulut N, Topuz O, Hylan G- F. 20 efficacy on articular cartilage quality in patients with knee osteoarthritis: clinical and MRI assessment. Clin Rheumatol 2005;24:336–41.
185. Ozturk C, Atamaz F, Hepguler S, Argin M, Arkun R. The safety and efficacy of intraarticular hyaluronan with/without corticosteroid in knee osteoarthritis: 1-year, single-blind, randomized study. Rheumatol Int 2006;26:314–9.
186. Wang Y, Wluka AE, Cicuttini FM, Wang Y, Wluka AE, Cicuttini FM. The determinants of change in tibial plateau bone area in osteoarthritic knees: a cohort study. Arthritis Res Ther 2005;7:R687–93.
187. Garnero P, Peterfy C, Zaim S, Schoenharting M. Bone marrow abnormalities on magnetic resonance imaging are associated with type II collagen degradation in knee osteoarthritis: a three-month longitudinal study. Arthritis Rheum 2005;52:2822–9.
188. Hayes CW, Jamadar DA, Welch GW, Jannausch ML, Lachance LL, Capul DC, *et al*. Osteoarthritis of the knee: comparison of MR imaging findings with radiographic severity measurements and pain in middle-aged women. Radiology 2005;237:998–1007.
189. Wang Y, Ebeling PR, Hanna F, O'Sullivan R, Cicuttini FM. Relationship between bone markers and knee cartilage volume in healthy men. J Rheumatol 2005;32:2200–4.
190. Bruyere O, Collette J, Kothari M, Zaim S, White D, Genant H, *et al*. Osteoarthritis, magnetic resonance imaging, and biochemical markers: a one year prospective study. Ann Rheum Dis 2006;65:1050–4.
191. Brandt KD, Mazzuca SA, Buckwalter KA. Acetaminophen, like conventional NSAIDs, may reduce synovitis in osteoarthritic knees. Rheumatology (Oxford) 2006;45: 1389–94.
192. Wluka AE, Forbes A, Wang Y, Hanna F, Jones G, Cicuttini FM. Knee cartilage loss in symptomatic knee osteoarthritis over 4.5 years. Arthritis Res Ther 2006;8:R90.
193. Hunter DJ, LaValley M, Li J, Zhang Y, Bauer D, Nevitt M, *et al*. Urinary pentosidine does not predict cartilage loss among subjects with symptomatic knee OA: the BOKS Study. Osteoarthritis Cartilage 2007;15:93–7.
194. Amin S, Niu J, Guermazi A, Grigoryan M, Hunter DJ, Clancy M, *et al*. Cigarette smoking and the risk for cartilage loss and knee pain in men with knee osteoarthritis. Ann Rheum Dis 2007;66:18–22.
195. Davies-Tuck ML, Wluka AE, Wang Y, Teichtahl AJ, Jones G, Ding C, *et al*. The natural history of cartilage defects in people with knee osteoarthritis. Osteoarthritis Cartilage 2008;16: 337–42.
196. Kornaat PR, Kloppenburg M, Sharma R, Botha-Scheepers SA, Le Graverand MP, Coene LN, *et al*. Bone marrow edema-like lesions change in volume in the majority of patients with osteoarthritis; associations with clinical features. Eur Radiol 2007;17:3073–8.
197. Hunter DJ, Li J, LaValley M, Bauer DC, Nevitt M, DeGroot J, *et al*. Cartilage markers and their association with cartilage loss on magnetic resonance imaging in knee osteoarthritis: the Boston Osteoarthritis Knee Study. Arthritis Res Ther 2007;9:R108.
198. Hernandez-Molina G, Guermazi A, Niu J, Gale D, Goggins J, Amin S, *et al*. Central bone marrow lesions in symptomatic knee osteoarthritis and their relationship to anterior cruciate ligament tears and cartilage loss. Arthritis Rheum 2008;58:130–6.
199. Teichtahl AJ, Wluka AE, Cicuttini FM. Frontal plane knee alignment is associated with a longitudinal reduction in patella cartilage volume in people with knee osteoarthritis. Osteoarthritis Cartilage 2008;16:851–4.

200. Teichtahl AJ, Wang Y, Wluka AE, Szramka M, English DR, Giles GG, *et al*. The longitudinal relationship between body composition and patella cartilage in healthy adults. Obesity (Silver Spring) 2008;16:421–7.
201. Blumenkrantz G, Stahl R, Carballido-Gamio J, Zhao S, Lu Y, Munoz T, *et al*. The feasibility of characterizing the spatial distribution of cartilage T(2) using texture analysis. Osteoarthritis Cartilage 2008;16:584–90.
202. Song IH, Althoff CE, Hermann KG, Scheel AK, Knetsch T, Burmester GR, *et al*. Contrast-enhanced ultrasound in monitoring the efficacy of a bradykinin receptor 2 antagonist in painful knee osteoarthritis compared with MRI. Ann Rheum Dis 2009;68:75–83.
203. Owman H, Tiderius CJ, Neuman P, Nyquist F, Dahlberg LE. Association between findings on delayed gadolinium-enhanced magnetic resonance imaging of cartilage and future knee osteoarthritis. Arthritis Rheum 2008;58:1727–30.
204. Amin S, Baker K, Niu J, Clancy M, Goggins J, Guermazi A, *et al*. Quadriceps strength and the risk of cartilage loss and symptom progression in knee osteoarthritis. Arthritis Rheum 2009;60:189–98.

Osteoarthritis and Cartilage 19 (2011) 589–605

Osteoarthritis and Cartilage

Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence

D.J. Hunter †‡*, W. Zhang §, P.G. Conaghan ||, K. Hirko ‡, L. Menashe ‡, W.M. Reichmann ¶, E. Losina ¶

† *Rheumatology Department, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Sydney, NSW, Australia*
‡ *Division of Research, New England Baptist Hospital, Boston, MA, USA*
§ *Academic Rheumatology, The University of Nottingham, Nottingham, UK*
|| *Section of Musculoskeletal Disease, University of Leeds & NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK*
¶ *Department of Orthopedic Surgery, Brigham and Women's Hospital, Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA*

ARTICLE INFO

Article history:
Received 15 June 2010
Accepted 17 October 2010

Keywords:
Osteoarthritis
Magnetic resonance imaging
Reliability
Responsiveness

SUMMARY

Objective: To summarize literature on the responsiveness and reliability of MRI-based measures of knee osteoarthritis (OA) structural change.
Methods: A literature search was conducted using articles published up to the time of the search, April 2009. 1338 abstracts obtained with this search were preliminarily screened for relevance and of these, 243 were selected for data extraction. For this analysis we extracted data on reliability and responsiveness for every reported synovial joint tissue as it relates to MRI measurement in knee OA. Reliability was defined by inter- and intra-reader intra-class correlation (ICC), or coefficient of variation, or kappa statistics. Responsiveness was defined as standardized response mean (SRM) - ratio of mean of change over time divided by standard deviation of change. Random-effects models were used to pool data from multiple studies.
Results: The reliability analysis included data from 84 manuscripts. The inter-reader and intra-reader ICC were excellent (range 0.8–0.94) and the inter-reader and intra-reader kappa values for quantitative and semi-quantitative measures were all moderate to excellent (range 0.52–0.88). The lowest value (kappa = 0.52) corresponded to semi-quantitative synovial scoring intra-reader reliability and the highest value (ICC = 0.94) for semi-quantitative cartilage morphology.
The responsiveness analysis included data from 42 manuscripts. The pooled SRM for quantitative measures of cartilage morphometry for the medial tibiofemoral joint was −0.86 (95% confidence intervals (CI) −1.26 to −0.46). The pooled SRM for the semi-quantitative measurement of cartilage morphology for the medial tibiofemoral joint was 0.55 (95% CI 0.47–0.64). For the quantitative analysis, SRMs are negative because the quantitative value, indicating a loss of cartilage, goes down. For the semi-quantitative analysis, SRMs indicating a loss in cartilage are positive (increase in score).
Conclusion: MRI has evolved substantially over the last decade and its strengths include the ability to visualize individual tissue pathologies, which can be measured reliably and with good responsiveness using both quantitative and semi-quantitative techniques.

Introduction

One proposed osteoarthritis (OA) treatment goal is modification of the underlying joint structure. Highly reproducible and responsive measures of the rate of disease progression are a prerequisite for assessing structural change in clinical trials. Conventional radiography (CR) has been the mainstay of assessing structural change in OA clinical trials and is currently part of FDA recommendations on how to conduct trials to assess structural progression. The focus of such evaluations has been on the radiographic joint space as a surrogate for hyaline cartilage assessment.

There has been a growing awareness that symptomatic OA represents a process involving all the tissues in the OA joint. Structure modification should therefore be considered in a broader context than that of cartilage alone. Modern imaging, especially magnetic resonance imaging (MRI), allows unparalleled direct visualization of all the tissues involved in OA joint pathology,

*Address correspondence and reprint requests to: David J. Hunter, Rheumatology Department, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Sydney, NSW, Australia. Tel.: 61-2-9926-7379; Fax: 61-2-9906-1859.
E-mail address: David.Hunter@sydney.edu.au (D.J. Hunter).

1063-4584/$ – see front matter © 2011 Osteoarthritis Research Society International. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.joca.2010.10.030

including cartilage, menisci, subchondral bone and soft tissue. MRI is ideally suited for imaging arthritic joints as is it is free of ionizing radiation, and its tomographic viewing perspective obviates morphological distortion, magnification and superimposition. More importantly, MRI has a rich image contrast variability resulting in an ability to discriminate articular tissues and it therefore holds great potential as a tool for whole-organ imaging of the OA joint. The last 20 years has seen a rapid improvement in imaging technology and in the last decade this has translated into improved understanding of the importance of individual features, their relation to clinical outcome and disease pathogenesis and better data on the quantification of these pathologies[1,2]. There is a wealth of literature on the measurement properties of MRI in the setting of OA including responsiveness and reliability. Prior to considering the merits of MRI in the setting of potential disease modifying trials and trial guidance it is important to review this systematically.

The objective of this review was to summarize the literature on the responsiveness and reliability of MRI-based measures of knee OA structural change.

Material and methods

Systematic literature search details

An online literature search was conducted using the OVID MEDLINE (1945–), EMBASE (1980–) and Cochrane databases (1998–). The search was not limited by publication date and the last search occurred in April 2009, with the search entries "MRI", and "osteoarthritis", "osteoarthritides", "osteoarthrosis", "osteoarthroses", "degenerative arthritis", "degenerative arthritides", or "osteoarthritis deformans". The abstracts of the 1330 citations received with this search were then preliminarily screened for relevance by two reviewers (KH and DJH). Although review articles were not included (see Inclusion/exclusion criteria), citations found in any review articles which were not already included in our preliminary search were screened for possible inclusion in this study. This added seven more relevant studies to our search. One further article was added, before publication, by one of the authors of this meta-analysis bringing the preliminary total to 1338.

Inclusion/exclusion criteria

Only studies published in English were included. Studies presenting non-original data were excluded, such as reviews, editorials, opinion papers, case studies or letters to the editor. Studies with questionable clinical relevance and those using non-human subjects or specimens were excluded. Studies in which rheumatoid, inflammatory, or other forms of arthritis were included in the OA datasets were excluded, as well as general joint-pertinent MRI studies not focused on OA. Studies with no extractable, numerical data were excluded. Any duplicates which came up in the preliminary search were excluded. Of the preliminary 1338 abstracts, 243 were selected for data extraction (Fig. 1).

Data abstraction

We used a data abstraction tool constructed in EpiData (Entry version 2.0 Odense, Denmark). Two reviewers (KH and LM) independently abstracted the following data: (1) patient demographics; (2) MRI make (vendor and field strength), sequences and techniques used (see further description below), tissue types viewed; (3) study type and funding source; (4) details on rigor of study design to construct the Downs methodological quality score (see further description below)[3]; (5) MRI reliability/reproducibility data; (6) MRI diagnostic measures and performance; (7) gold standard measures against which the MRI measure was evaluated; (8) treatment and MRI measures (when appropriate).

Multiple techniques have been used to measure structural abnormality and change on MRI in OA. Broadly speaking these methods are divided into quantitative and semi-quantitative methods[1]. Quantitative measurements using computer-aided image processing to assess whole joint quantification (cartilage morphometry, bone volume, bone marrow lesion volume, meniscal position and volume, synovial volume, etc). The three-dimensional (3D) coverage of an entire cartilaginous region by MRI allows for the direct quantification of volumetric structures. Compositional measures of articular cartilage are also included within the quantitative measures as the measurement provides for a quantitative output. These methods include T2 mapping, dGEMRIC and T1rho and are extensively reviewed elsewhere[4,5].

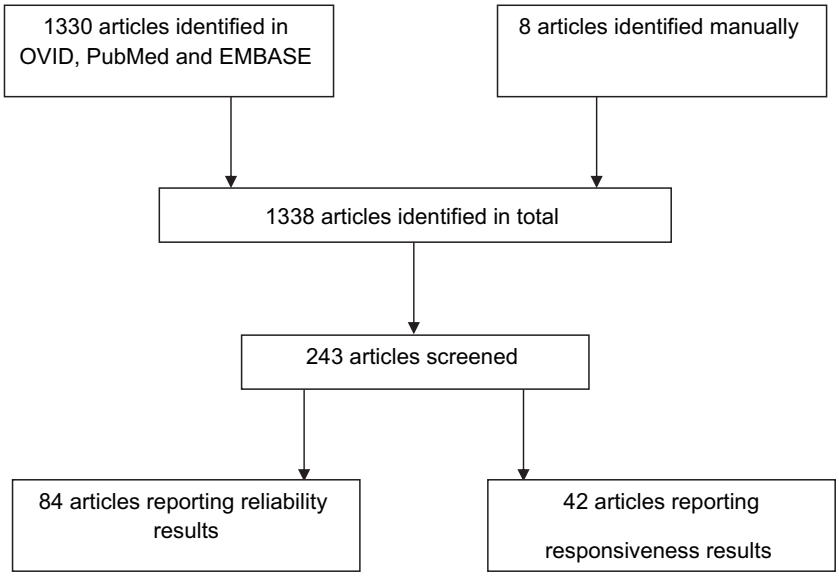

Fig. 1. Flow chart of the screening process for articles included in the systematic review.

Table I
Summary table of studies reporting data on reliability of MRI in knee OA

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, mean (SD), range	No. (%) of females	Quantitative	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Downs criteria score
Karvonen RL; Journal of Rheumatology; 1994; 7966075[13]	92	52	40	All OA Pts: 55(14), (Range: 25–86); Bilateral OA Pts: 53(13) (Range: 25–73); Control: 49(15), (Range: 22–78)	All OA Pts: 35; Bilateral OA Pts: 19; Control: 25	X			X		X				Case control	11
Peterfy CG; Radiology; 1994; 8029420[14]	8	5	3	62 (Range: 45–82)	4(50%)	X			X						Cross-sectional	4
Marshall KW; Journal of Orthopaedic Research; 1995; 8544016[15]	2			31		X			X						Other	6
Disler DG; AJR Am J Roentgel.; 1996; 8659356[16]	114	79	35	36	48			X	X						Cross-sectional	6
Dupuy DE; Academic Radiology; 1996; 8959181[17]	7	2	5	TKA Pts: (Range: 64–75); Asymptomatic Pts: (Range: 25–35)	TKA Pts: 1(50%); Asymptomatic Pts: 2	X			X						Other	6
Trattnig S; Journal of Computer Assisted Tomography; 1998; 9448754[18]	20	20	0	72.2 (Range: 62–82)	18			X	X						Other	8
Drape JL; Radiology; 1998; 9646792[19]	43	43	0	63 (Range: 53–78)	30			X	X						Cross-sectional	5
Cicuttini F; Osteoarthritis & Cartilage; 1999; 10329301[20]	28			Males: 41.4(14.8); Females: 31.2(8.6)	11(39%)	X			X						Cross-sectional	7
Pham XV; Revue du Rhumatisme; 1999; 10526380[21]	10	10	10	67.2(7.34), (Range: 57–80)	6			X						X	Cross-sectional	13
Gale DR; Osteoarthritis & Cartilage; 1999; 10558850[22]	291	233	58	Men cases: 67(10); Men controls 65(10); Women cases: 66(10); Women controls: 66(8)	61(21%)								X		Case control	10
Hyhlik–Durr A; European Radiology; 2000; 10663760[23]	11	3	8	OA group: (Range: 61–75); Healthy group: (Range: 25–36)	5(45.5%)	X			X						Cross-sectional	6
Jones G; Arthritis & Rheumatism; 2000; 11083279[24]	92	0	92	Boys:12.8(2.7); Girls: 12.6(2.9)	43(46.8%)	X			X		X				Cross-sectional	13
Wluka AE; Annals of the Rheumatic Diseases; 2001; 11247861[25]	81	42	39	Cases: 58(6.1); Controls: 56(5.4)	81(100%)	X		X	X						Case control	16
Felson DT; Annals of Internal Medicine; 2001; 11281736[26]	401	401	0	66.8				X				X			Cross-sectional	13
Hill CL; Journal of Rheumatology; 2001; 11409127[27]	458	433	25	67	(34%)			X		X					Case control	13
Bergin D; Skeletal Radiology; 2002; 11807587[28]	60	30	30	Cases: 50; Controls: 57				X					X	X	Case control	9

(*continued on next page*)

Table I (*continued*)

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, mean (SD), range	No. (%) of females	Quantitative	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Downs criteria score
Beuf O; Arthritis & Rheumatism; 2002; 11840441[29]	46	18	28	Mild OA: 56.3(4.5); Severe OA: 70(6.3)	17(37%)	X									Case control	5
Wluka AE; Arthritis & Rheumatism; 2002; 12209510[30]	123	123	0	63.1(10.6)	71	X			X		X				Longitudinal Prospective	14
Gandy SJ; Osteoarthritis & Cartilage; 2002; 12464553[31]	16	16	0		6	X			X						Longitudinal Prospective	8
Bhattacharyya T; Journal of Bone & Joint Surgery - American Volume; 2003; 12533565[32]	203	154	49	Cases: 65; Controls: 67				X					X		Case control	9
Cicuttini FM; Clinical & Experimental Rheumatology; 2003; 12673893[33]	81	42	39	ERT: 58(6.1); Controls: 56(5.4)	81(100%)	X			X		X				Case control	12
Raynauld JP; Osteoarthritis & Cartilage; 2003; 12744941[34]	28	17	11	Healthy subjects: (Range: 25–35); OA Pts: 63.5		X			X						Other	7
Felson DT; Annals of Internal Medicine; 2003; 12965941[35]	256	256	0	Followed: 66.2(9.4); t followed: 67.8(9.6)	Followed: 41.7%; t followed: 15.2%			X				X			Other	11
Hill CL; Arthritis & Rheumatism; 2003; 14558089[36]	451	427		Knee pain/ROA/MALE: 68.3; Knee pain/ROA/ Female: 65; knee pain/ ROA/male: 66.8; knee pain/ROA/ female:66.1				X				X			Cross-s ectional	10
Glaser C; Magnetic Resonance in Medicine; 2003; 14648571[37]	23	7	16	Healthy subjects: (Range: 23–33); OA Pts: 60–85	13(56.5%)	X			X						Cross-sectional	5
Lindsey CT; Osteoarthritis & Cartilage; 2004; 14723868[38]	74	33	21	OA1(KL = 1/2):62.7(10.9); OA2(KL = 3/4):66.6(11.6); Controls: 34.2(12.5)	39(52.7%)	X			X		X				Case control	8
Cicuttini FM; Arthritis & Rheumatism; 2004; 14730604[39]	117	117		63.7(10.2)	(58.1%)	X			X						Longitudinal Prospective	9
Raynauld JP; Arthritis & Rheumatism; 2004; 14872490[40]	32	32	0	62.9(8.2)	(74%)	X			X						Longitudinal Prospective	10
Cicuttini F; Rheumatology; 2004; 14963201[41]	117	117	0	67(10.6)	(58.1%)	X			X						Longitudinal Prospective	12
Peterfy CG; Osteoarthritis & Cartilage; 2004; 14972335[6]	19	19	0	61(8)	4			X	X	X	X	X	X	X	Other	5
Dashti M; Scandinavian Journal of Rheumatology; 2004; 15163109[42]	174	117	57	61.6(9.5)	123(70.7%)	X			X						Case control	11
Cicuttini FM; Journal of Rheumatology; 2004; 15229959[43]	102	102	0	63.8(10.1)	(63%)	X			X						Longitudinal Prospective	10
Baysal O; Swiss Medical Weekly; 2004; 15243849[44]	65	65	0	53.1(7), (Range: 45–75)	(100%)	X		X	X		X				Cross-sectional	7

Kornaat PR; Skeletal Radiology; 2005; 15480649[9]	25	25	0	Median age = 63, (Range: 50–75)				X	X	X	X	X	X		Other	6
Yoshioka H; Journal of Magnetic Resonance Imaging; 2004; 15503323[45]	28	28	0	55.6 (Range: 40–73)	10	X		X	X	X		X	X		Other	5
Ding C; Osteoarthritis & Cartilage; 2005; 15727885[46]	372	162	210	cartilage defects: 43.6(7.1); Any cartilage defects: 47(6.1)	(58%)	X		X	X						Case control	9
Hill CL; Arthritis & Rheumatism; 2005; 15751064[47]	433	360	73	Case males:68.2; Case females:65; Control males:66.8; Control females:65.8	143(33%)			X						X	Case control	12
Maataoui A; European Radiology; 2005; 15856246[48]	12	12	0	median age = 70.5, (Range: 60–86)	9	X			X						Cross-sectional	6
Cicuttini F; Osteoarthritis & Cartilage; 2005; 15922634[49]	28	28	0	62.8(9.8)	(57%)	X			X						Longitudinal Prospective	10
Huh YM; Korean Journal of Radiology; 2005; 15968151[50]	94	73	21	OA group: 57.8, (Range: 40–80), Median = 58; RA group:49.6, (Range: 37–76), Median = 48	73(80%)			X		X					Longitudinal Retropective	7
Wluka AE; Rheumatology; 2005; 16030084[51]	126	126	0	63.6(10.1)	68(54%)	X		X	X		X				Longitudinal Prospective	14
Eckstein, F; Annals of the Rheumatic Diseases; 2006; 16126797[52]	19	10	9	51 (Range: 40–71)	12	X			X		X				Other	8
Eckstein F; Arthritis & Rheumatism; 2005; 16200592[53]	30	15	15	Cases: 49.6(Range: 37–76); Controls: 62.3(11.5)	30(100%)	X			X						Cross-sectional	7
Sengupta M; Osteoarthritis & Cartilage; 2006; 16442316[54]	217	217	0	67.3(9.1)	(30%)			X	X	X	X	X			Cross-sectional	7
Raynauld JP; Arthritis Research & Therapy; 2006; 16507119[55]	110	110	0	62.4(7.5)	(64%)	X		X	X			X	X		Longitudinal Prospective	11
Hunter DJ; Arthritis & Rheumatism; 2006; 16508930[56]	257	257	0	66.6(9.2), (Range: 47–93)	(41.6%)			X	X				X		Longitudinal Prospective	10
Brandt KD; Rheumatology; 2006; 16606655[57]	30	20	10	62	29					X					Other	10
Jaremko JL; Osteoarthritis & Cartilage; 2006; 16644245[58]	12	3	9	OA: (Range: 59–71);) Healthy: 37(8), (Range: 23–48)	4(33.3%)	X			X						Cross-sectional	8
Hunter DJ; Osteoarthritis & Cartilage; 2007; 16857393[59]	127	127		67(9.05)	(46.7%)		X		X						Cross-sectional	12
Boks SS; American Journal of Sports Medicine; 2006; 16861575[60]	134	136	132	40.8 (Range: 18.8–63.8)				X	X				X	X	Cross-sectional	7
Brem MH; Skeletal Radiology; 2007; 17219231[61]	5	5	0	64.3 (Range: 40–73)	2	X			X						Other	6
Folkesson J; IEEE Transactions on Medical Imaging; 2007; 17243589[62]	139			56 (Range: 22–79)	(59%)	X			X						Other	7

(*continued on next page*)

Table I (*continued*)

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, mean (SD), range	No. (%) of females	Quantitative	Compositional techniques	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Downs criteria score
Dam EB; Osteoarthritis & Cartilage; 2007; 17353132[63]	139			Evaluation set: 55(Range: 21–78); Scan-rescan set: 51,(Range: 26–75)	(55%)	X			X						Other	9
Baranyay FJ; Seminars in Arthritis & Rheumatism; 2007; 17391738[64]	297		297	58(5.5)	(63%)	X			X			X			Cross-sectional	16
Hanna F; Mepause; 2007; 17413649[65]	176	0	176	52.3(6.6), (Range: 40–67)	176(100%)	X			X						Cross-sectional	13
Hunter DJ; Annals of the Rheumatic Diseases; 2008; 17472995[8]	71			67.9(9.3)	(28.2%)			X	X	X	X	X	X	X	Other	8
Hill CL; Annals of the Rheumatic Diseases; 2007; 17491096[66]	270	270	0	66.7(9.2)	112			X	X	X					Longitudinal Prospective	9
Qazi AA; Osteoarthritis & Cartilage; 2007; 17493841[67]						X			X						Cross-sectional	8
Guymer E; Osteoarthritis & Cartilage; 2007; 17560134[68]	176	0	176	52.3(6.6)	(100%)	X		X	X		X				Cross-sectional	11
Eckstein F; Osteoarthritis & Cartilage; 2007; 17560813[69]	9	9		52.2(9.3)	5	X			X						Other	9
Akhtar, S; Osteoarthritis & Cartilage; 2007; 17707660[70]	6			(Range: 25–69)	2(33%)	X			X						Other	7
Raynauld JP; Annals of the Rheumatic Diseases; 2008; 17728333[71]	107	107	0	62.4(7.5)	(64%)	X		X	X				X		Longitudinal Retropective	15
Felson DT; Arthritis & Rheumatism; 2007; 17763427[72]	330	110	220	Cases: 62.9(8.3); Controls: 61.2(8.4)	211(64%)			X			X	X			Case control	12
Lo GH; Osteoarthritis & Cartilage; 2008; 17825586[73]	845	170		63.6(8.8)	(58%)			X					X		Cross-sectional	10
Davies-Tuck M; Osteoarthritis & Cartilage; 2008; 17869546[74]	100	100	0	63.6(10.2)	61(61%)	X			X						Longitudinal Prospective	11
Folkesson J; Academic Radiology; 2007; 17889339[75]	.			56 (Range: 22–79)	(59%)										Other	7
Sanz R; Journal of Magnetic Resonance Imaging; 2008; 18022850[76]	22	9		Normal: 43.6(15); Chondromalacia: 33.3(11.8); OA Pts: 58.9(11.5)	14(64%)		X		X						Case control	6
Englund M; Arthritis & Rheumatism; 2007; 18050201[77]	310	102	208	Cases: 62.9(8.3); Controls: 61.2(8.3)	211(68%)			X					X		Case control	15
Hernandez-Molina G; Arthritis & Rheumatism; 2008; 18163483[78]	258	258	0	66.6(9.2)	(42.6%)			X	X			X		X	Longitudinal Prospective	11
Amin S; Osteoarthritis & Cartilage; 2008; 18203629[79]	265	265		67(9)	(43%)			X	X				X	X	Longitudinal Prospective	11
Teichtahl AJ; Obesity; 2008; 18239654[80]	297		297	58(5.5)	186	X		X	X		X				Longitudinal Prospective	14
Anandacoomarasamy; Journal of Rheumatology; 2008; 18278831[81]	32	32		Males: 64(11.5); Females: 66(9.5); Total: 65(Range: 42–87)	17(53%)	X		X	X						Longitudinal Prospective	11

Eckstein F; Annals of the Rheumatic Diseases; 2008; 18283054[82]	158			Mild to moderate OA2: 57.6(8.3); Controls: 56.1(8.7)	158(100%)	X				X						Case control	10
Reichenbach S; Osteoarthritis & Cartilage; 2008; 18367415[83]	964	217	747	63.3	(57%)				X	X		X				Cross-sectional	8
Petterson SC; Medicine & Science in Sports & Exercise; 2008; 18379202[84]	123	123	0	64.9(8.5)	67											Case control	11
Bolbos RI; Osteoarthritis & Cartilage; 2008; 18387828[85]	32			Cases: 47.2(11.5), (Range: 29–72); Controls: 36.3(10.5), Range: (27–56)	14(44%)	X	X			X		X				Case control	7
Pai A; Magnetic Resonance Imaging; 2008; 18502073[86]	10	0	10	27 (Range: 21–31)	4(40%)		X			X						Other	6
Folkesson J; Magnetic Resonance in Medicine; 2008; 18506845[87]			143	Healthy subjects: 48(Range: 21–78); KL1: 62(Range: 37–81); KL2: 67(Range: 47–78); KL3&4: 68(58–78)						X						Other	12
Mills PM; Osteoarthritis & Cartilage; 2008; 18515157[88]	49	25	24	APMM: 46.8(5.3); Controls: 43.6(6.6)	18(36.7%)	X			X	X						Case control	12
Dore D; Osteoarthritis & Cartilage; 2008; 18515160[89]	50	50		64.5(7.1)	23	X			X	X		X				Cross-sectional	9
Pelletier JP; Osteoarthritis & Cartilage; 2008; 18672386[90]	27	1		64.1(9.6)	14				X	X	X	X	X	X		Other	9
Englund M; New England Journal of Medicine; 2008; 18784100[91]	991	171		62.3(8.6), (Range: 50.1–90.5)	565(57%)				X					X		Cross-sectional	10
Rauscher I; Radiology; 2008; 18936315[92]	60	37	23	Healthy controls: 34.1(10); Mild OA: 52.5(10.9); Severe OA: 61.6(11.6)	32(53.3%)		X			X				X		Case control	9
Kijowski R.; Radiology; 2009; 19164121[93]	200		200	1.5T image group: 38.9(Range: 16–63); 3T image group: 39.1(Range: 15–65)	87(43.5%)				X	X						Longitudinal Retropective	10

In contrast to quantitative measures semi-quantitative image analysis is typically much more observer dependent and generates grades or scales rather than truly continuous output. Semi-quantitative scoring of MRI's are a valuable method for performing multi-feature assessment of the knee using conventional MRI acquisitions[6–8,98]. Such approaches score, in an observer dependent semi-quantitative manner, a variety of features that are currently believed to be relevant to the functional integrity of the knee and/or potentially involved in the pathophysiology of OA. These articular features can include articular cartilage morphology, subarticular bone marrow abnormality, subarticular cysts, subarticular bone attrition, marginal and central osteophytes, medial and lateral meniscal integrity, anterior and posterior cruciate ligament integrity, medial and lateral collateral ligament integrity, synovitis/effusion, intra-articular loose bodies, and periarticular cysts/bursitis.

The Downs methodological quality score[3] collects a profile of scores (quality of reporting, internal validity (bias and confounding), power, external validity so that the overall study quality score reflects all of these elements. Answers were scored 0 (No) or 1 (Yes), except for one item in the Reporting subscale, which scored 0–2 and the single item on power, which was scored 0–5. The possible range is from 0–27 where 0 represents poor quality and 27 optimal quality.

The outcomes for psychometric properties on MRI were examined using the OMERACT filter[10,11]. The material pertinent to this manuscript is Discrimination: does the measure discriminate between situations that are of interest? The situations can be states at one time (for classification or prognosis) or states at different times (to measure change). This criterion captures the issues of reliability and responsiveness (sensitivity to change).

Statistical analysis

Reliability was defined by inter- and intra-reader measures of coefficient of variation (CV), or intra-class correlation (ICC), or kappa statistics.

Responsiveness was defined as standardized response mean (SRM) - ratio of mean of change over time divided by standard deviation of change. Whenever possible, both reliability measures and SRMs were stratified by measurement method (quantitative and semi-quantitative), tissue lesion (cartilage, synovium, bone, bone marrow lesions, meniscus and ligament) and plate/region for cartilage divisions.

For the quantitative analysis, a negative SRM expresses cartilage loss whereas a positive SRM would indicate cartilage gain. For the semi-quantitative analysis, positive SRMs indicate a loss in cartilage with higher scores reflecting greater lesions.

Random-effects models were used to summarize data from multiple studies. Since some studies reported more than one measure for each region, to avoid substantial skewness of results influenced by multiple observations from a single study and to ensure that the estimates included in the analysis came from independent studies, we repeated analyses 500 times. We did this by selecting one observation (estimate) from each study at random so that the number of observations in the model reflected the number of studies. We then ran a random-effects model to obtain the pooled summary measure and its standard error. The process was repeated 500 times to obtain the empirical distribution of the summary measure. The final pooled summary measure and its standard error were obtained by averaging the 500 summary measures and the 500 standard errors obtained from the random-effects models respectively. Ninety-five percent confidence intervals (CI) were obtained using a normal approximation for the final pooled summary measure and its standard error.

Table II
Results of random-effects pooling of *intra-reader* CV from MRI studies stratified by measure (quantitative and semi-quantitative) and tissue (cartilage, synovium, bone, bone marrow lesion, meniscus, and ligament)

Stratification	Number of estimates (Studies)	Mean sample size	Pooled CV (%)	95% CI
Quantitative				
Cartilage	32 (10)	60	3	−2, 7
Synovium	2 (1)	94	8	−6, 22
Compositional	6 (1)	60	5	−5, 15

Table III
Results of random-effects pooling of *inter-reader* CV from MRI studies stratified by measure (quantitative and semi-quantitative) and tissue (cartilage, synovium, bone, bone marrow lesion, meniscus, and ligament)

Stratification	Number of estimates (Studies)	Mean sample size	Pooled CV (%)	95% CI
Quantitative				
Cartilage	42 (13)	65	3	−1, 6
Synovium	1 (1)	94	5	−15, 25
Bone	9 (5)	119	2	−4, 8

Results

Reliability

The reliability analysis included data from 84 manuscripts (Table I). The mean Downs criteria score for these manuscripts was 9.4 (range 4–16).

Inter- and intra-reader CV and test-retest measures were confined to quantitative or compositional measures (Tables II and III). The pooled CV for quantitative cartilage was 3% for both inter- and intra-reader reliability.

The inter-reader and intra-reader ICCs for quantitative and semi-quantitative measures were all excellent (range 0.8–0.94)(Tables IV and V). For quantitative measures the intra-reader ICC ranged from 0.87 (0.61–1.00) for synovium to 0.93 (0.82–1.00) for meniscus measurement. For quantitative measures the inter-reader ICC ranged from 0.81 (0.72–0.89) for meniscus to 0.90 (0.86–0.95) for cartilage morphometry measurement.

The inter-reader and intra-reader kappa values for quantitative and semi-quantitative measures were all moderate to excellent (range 0.52–0.88)(Tables VI and VII). For semi-quantitative measures the range for intra-reader kappa values extended from 0.52 (0.28–0.77) for synovium to 0.66 (0.54–0.78) for BML assessment. For semi-quantitative measures the range for inter-reader kappa values extended from 0.57 (0.44–0.71) for cartilage morphology to 0.88 (0.79–0.97) for BML assessment.

Table IV
Results of random-effects pooling of *intra-reader* ICC from MRI studies stratified by measure (quantitative and semi-quantitative) and tissue (cartilage, synovium, bone, bone marrow lesion, meniscus, and ligament)

Stratification	Number of estimates (Studies)	Mean sample size	Pooled ICC	95% CI
Quantitative				
Cartilage	23 (9)	108	0.92	0.88, 0.96
Synovium	2 (1)	30	0.87	0.61, 1.00
Meniscus	1 (1)	291	0.93	0.82, 1.00
Semi-quantitative				
Cartilage	7 (4)	114	0.94	0.87, 1.00
Synovium	3 (2)	26	0.88	0.66, 1.00
Bone Marrow Lesion	2 (2)	178	0.93	0.83, 1.00
Meniscus	2 (1)	25	0.77	0.49, 1.00

Table V
Results of random-effects pooling of *inter-reader* ICC from MRI studies stratified by measure (quantitative and semi-quantitative) and tissue (cartilage, synovium, bone, bone marrow lesion, meniscus, and ligament)

Stratification	Number of estimates (Studies)	Mean sample size	Pooled ICC	95% CI
Quantitative				
Cartilage	10 (4)	196	0.90	0.86, 0.95
Meniscus	2 (1)	291	0.81	0.72, 0.89
Semi-Quantitative				
Cartilage	9 (7)	88	0.85	0.77, 0.94
Synovium	5 (4)	46	0.87	0.74, 1.00
Bone	3 (2)	23	0.90	0.66, 1.00
Bone Marrow Lesion	2 (2)	22	0.84	0.54, 1.00
Meniscus	5 (3)	67	0.93	0.82, 1.00
Ligament	4 (2)	105	0.80	0.56, 1.00

Responsiveness

The responsiveness analysis included data from 42 manuscripts (Table VIII). The mean Downs criteria score for these manuscripts was 11.2 (range 8–21). Table IX includes the summary responsiveness data for both types of measurement methods (quantitative and semi-quantitative). As some studies reported multiple estimates, random-effects model pooling was done to reduce potential bias from studies reporting multiple estimates. The pooled SRM for quantitative measures of cartilage morphometry for the medial tibiofemoral joint was −0.86 (95%CI −1.26 to −0.46), for lateral tibofemoral joint was −1.01 (95%CI −2.04 to 0.02), and for the patella was −0.63 (95%CI −0.90 to −0.37). The quantitative cartilage morphometry pooled SRM ranged from −0.21 (−0.48 to 0.05) for the lateral femoral plate to −1.01 (−2.04 to 0.02) for lateral tibiofemoral plate. The results for the compositional measures are from one study and should be interpreted with caution. The pooled SRM for semi-quantitative measures of cartilage for medial tibiofemoral joint was 0.55 (95%CI 0.47–0.64), for lateral tibofemoral joint was 0.37 (95%CI 0.18–0.57), and for the patella was 0.29 (95% CI 0.03–0.56). The semi-quantitative cartilage morphology SRMs ranged from −0.07 (−0.18 to 0.04) for the medial tibial region to 0.55 (0.47–0.64 for the medial tibiofemoral region. The pooled SRM for semi-quantitative measures of synovium was 0.47 (95%CI 0.18–0.77), and for BMLs was 0.43 (95%CI −0.17 to 1.03).

There has been some concern that some of the earlier literature for quantitative measures of cartilage morphometry was more responsive than more recent estimates. Table X reflects an effort to distil distinct time periods. In general, the earlier estimates demonstrate larger SRMs than more recent studies with the medial tibiofemoral estimates from 2002–2006 being −0.95 (−1.15, −0.76) and from more recent studies (2007–2009) being −0.84 (−1.35, −0.33).

Table XI shows the results of random-effects pooling of SRM from MRI studies evaluating quantitative cartilage stratified by duration of study and plate region for studies published between 2007 and 2009. Studies with multiple estimates had an estimate selected at random and a pooled analysis was performed. In this analysis the pooled SRM for the medial tibiofemoral joint for studies of 1 year or less is −0.80 (−1.27, −0.33) and for studies of 1–2 years is −1.16 (−2.90, 0.58).

Table VI
Results of random-effects pooling of *intra-reader kappa* values from MRI studies stratified by measure (quantitative and semi-quantitative) and tissue (cartilage, synovium, bone, bone marrow lesion, meniscus, and ligament)

Stratification	Number of estimates (Studies)	Mean sample size	Pooled Kappa	95% CI
Quantitative				
Cartilage	1 (1)	158	0.66	0.50, 0.82
Semi-Quantitative				
Synovium	4 (2)	317	0.52	0.28, 0.77
Bone Marrow Lesion	1 (1)	256	0.66	0.54, 0.78

Table VII
Results of random-effects pooling of *inter-reader kappa* values from MRI studies stratified by measure (quantitative and semi-quantitative) and tissue (cartilage, synovium, bone, bone marrow lesion, meniscus, and ligament)

Stratification	Number of estimates (Studies)	Mean sample size	Pooled Kappa	95% CI
Semi-quantitative				
Cartilage	15 (4)	136	0.57	0.44, 0.71
Bone marrow lesion	2 (2)	237	0.88	0.79, 0.97
Meniscus	3 (3)	418	0.73	0.63, 0.84
Ligament	3 (3)	209	0.80	0.69, 0.90

Discussion

The purpose of this study was to summarize the literature on the responsiveness and reliability of MRI-based measures of knee OA structural change. In general, this review provides clear evidence that structural change in OA can be measured both reliably and with good responsiveness on MRI.

The data from this review indicates that quantitative measures of joint structure have excellent reliability (ICC range 0.81–0.94). Similarly agreement for semi-quantitative measures is good to excellent (kappa range 0.52–0.88). Directly comparing the reliability between quantitative and semi-quantitative techniques is not possible given the extracted data comes from different studies and the statistical methods used are frequently distinct but an overarching view would suggest they are broadly comparable with a slight benefit in reliability for quantitative measures. This is not surprising given the continuous nature of these measures, the greater use of technology to automate processes and quality control vigilance in quantitative measures.

The aim of the systematic review is to provide a summary of the best evidence. However, as a result of issues related to the quality of research, findings of studies can sometimes be misleading or incorrect. To minimize these risks, the quality of the studies was critically appraised using Downs checklist[3]. The findings from our review indicate that in general this literature is of adequate quality. No studies were identified in our search prior to 1994.

Several studies have suggested that baseline clinical, biomarker and imaging features are predictive of progression of cartilage loss in the medial compartment of the knee and could be used to provide greater study power by selecting a population at greater risk for more rapid progression. Whilst the estimates included in this analysis reflect these studies we have not explicitly selected for these studies so the pooled estimates reflect all studies not just selected estimates for those at highest risk for progression.

This review does not include results focused upon using MRI to stage OA. Whilst MRI has been extensively used for measuring progression its use in staging OA as a disease is at this point quite limited. In an effort to shorten discovery and development timelines, clinical trial brevity is paramount. As OA is typically a very slowly progressive condition, one can optimize trial efficiency by finding more responsive endpoint/s. The results of the responsiveness data reaffirm the potential benefit of MRI compared to plain radiography that generally has SRMs in the 0.3–0.4 range[12].

Table VIII
Summary table of studies reporting data on responsiveness of MRI in OA

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, Mean(SD), Range	No. (%) of females	Quantitative	Compositional technique	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Downs criteria score
Wluka AE; Arthritis & Rheumatism; 2002; 12209510[30]	123	123	0	63.1(10.6)	71	X			X		X				Longitudinal Prospective	14
Cicuttini FM; Journal of Rheumatology; 2002; 12233892[94]	21	8	13	Case:41.3(13.2); Controls: 49.2(17.8)	14(66.7%)	X			X						Longitudinal Prospective	13
Biswal S; Arthritis & Rheumatism; 2002; 12428228[7]	43	4	39	54.4(Range 17–65)	21			X	X						Longitudinal Prospective	8
Gandy SJ; Osteoarthritis & Cartilage; 2002; 12464553[31]	16	16	0	63.4 (Range 52–70)	6	X			X						Longitudinal Prospective	8
Wluka AE; Journal of Rheumatology; 2002; 12465157[95]	136	136	0	Vitamin E group: 64.3(11); Placebo group: 63.7(10)	75(55%)	X			X						Randomized controlled trial	21
Cicuttini F; Journal of Rheumatology; 2002; 12465162[96]	110	110	0	63.2(10.2)	66	X			X						Longitudinal Prospective	12
Pessis E; Osteoarthritis & Cartilage; 2003; 12744942[97]	20	20		63.9(9)	13	X		X	X						Longitudinal Prospective	12
Cicuttini FM; Arthritis & Rheumatism; 2004; 14730604[39]	117	117		63.7(10.2)	(58.1%)	X			X						Longitudinal Prospective	9
Raynauld JP; Arthritis & Rheumatism; 2004; 14872490[40]	32	32	0	62.9(8.2)	(74%)	X			X						Longitudinal Prospective	10
Wluka AE; Annals of the Rheumatic Diseases; 2004; 14962960[98]	132	132	0	63.1(Range: 41–86)	71(54%)	X			X						Longitudinal Prospective	10
Cicuttini FM; Journal of Rheumatology; 2004; 15229959[43]	102	102	0	63.8(10.1)	(63%)	X			X						Longitudinal Prospective	10
Blumenkrantz G; Osteoarthritis & Cartilage; 2004; 15564067[99]	38	30	8	58(Range: 28–81)	(39.5%)	X	X		X						Longitudinal Prospective	9
Zhai G; BMC Musculoskeletal Disorders; 2005; 15720725[100]	150	80	70	TASOAC dataset: 62.3(7.6); KCV dataset: 42.8(6.1)	79(52.7%)	X			X						Other	9
Wang Y; Arthritis Res Ther; 2005; 15899054[101]	126	126		63.6(10.1)	68						X				Longitudinal Prospective	12
Cicuttini F; Osteoarthritis & Cartilage; 2005; 15922634[49]	28	28	0	62.8(9.8)	(57%)	X			X						Longitudinal Prospective	10
Wluka AE; Rheumatology; 2005; 16030084[51]	126	126	0	63.6(10.1)	68(54%)	X			X						Longitudinal Prospective	14
Ding C; Arthritis & Rheumatism; 2005; 16320339[102]	325			45.2(6.5)	190	X		X	X						Longitudinal Prospective	10
Raynauld JP; Arthritis Research & Therapy; 2006; 16507119[55]	110	110	0	62.4(7.5)	(64%)	X			X						Longitudinal Prospective	11
Hunter DJ; Arthritis & Rheumatism; 2006; 16508930[56]	257	257	0	66.6(Range: 47–93)	(41.6%)			X	X						Longitudinal Prospective	10

Hunter, DJ; Osteoarthritis & Cartilage; 2006; 16678452[103]	150	150	0	58.9(Range: 44–81)	(72%)	X			X	X	X	X	X			Longitudinal Prospective	9
Wluka AE; Arthritis Research & Therapy; 2006; 16704746[104]	105	105	0	All eligible: 62.5(10.7); MRI at FU: 63.8(10.6); Lost to FU: 61.6(11.3)	61(58.1%)	X				X						Longitudinal Prospective	17
Ding C; Rheumatology; 2007; 16861710[105]	325			45.2(6.4)	190	X				X						Longitudinal Prospective	12
Hunter DJ; Arthritis & Rheumatism; 2006; 16868968[106]	264	264	0	66.7(9.2), (Range: 47–93)	(40.9%)				X	X				X		Longitudinal Prospective	9
Bruyere O; Osteoarthritis Cartilage; 2007; 16890461[107]	62			64.9 (10.3)	(74%)	X				X						Longitudinal Prospective	10
Stahl R; Osteoarthritis & Cartilage; 2007; 17561417[108]	18	8	10	OA Pts: 55.7(7.3); Controls: 57.6(6.2)	18(100%)		X			X						Case control	10
Pelletier JP; Arthritis Research & Therapy; 2007; 17672891[109]	110	110		Q1greatestlossglobal: 63.7(7.2); Q4 least loss global: 61.3(7.5); Q1 greatest loss_med: 64.1(7.4); Q1 least loss_medial: 61.6(7.8)	74(67.3%)	X				X						Longitudinal Prospective	15
Raynauld JP; Annals of the Rheumatic Diseases; 2008; 17728333[71]	107	107	0	62.4(7.5)	(64%)	X				X			X			Longitudinal Retrospective	15
Davies-Tuck M; Osteoarthritis & Cartilage; 2008; 17869546[74]	100	100	0	63.3(10.2)	61(61%)	X				X						Longitudinal Prospective	11
Hunter DJ; Arthritis Research & Therapy; 2007; 17958892[110]	160	80	80	67(9)	(46%)				X	X						Case control	11
Teichtahl AJ; Osteoarthritis & Cartilage; 2008; 18194873[111]	99	99	0	63(10)	(60%)	X				X						Longitudinal Prospective	14
Hunter DJ; Annals of the Rheumatic Diseases; 2009; 18408248[112]	150	150		60.9(9.9)	76(51%)	X				X						Longitudinal Prospective	8
Folkesson J; Magnetic Resonance in Medicine; 2008; 18506845[87]	288		143	KL0(Healthy): 48(Range: 21–78); KL1: 62(Range: 37–81); KL2: 67(Range: 47–78); KL3&4: 68(Range: 58–78)	(44%)					X						Other	12
Sharma L; Arthritis & Rheumatism; 2008; 18512777[113]	153	153	0	66.4(11)		X				X						Longitudinal Prospective	11
Teichtahl AJ; Osteoarthritis & Cartilage; 2009; 18590972[114]	78			63 (10.5)	(52%)	X				X						Longitudinal Prospective	14
Raynauld JP; Annals Rheumatic Disease; 2009; 18653484[115]	154			60.3 (8.1)	100 (65%)	X				X						Randomized controlled trial	11
Pelletier JP; Osteoarthritis & Cartilage; 2008; 18672386[90]	27	1		64.1(9.6)	14	X			X	X	X		X			Other	9
Wirth W; Osteoarthritis & Cartilage; 2009; 18789729[116]	79			60.3 (9.5)	79 (100%)	X				X						Longitudinal Prospective	14

(*continued on next page*)

Table VIII (*continued*)

Reference: Author, Journal, Year, PMID	Whole sample size	No. of cases	No. of controls	Age, yrs, Mean(SD), Range	No. (%) of females	Quantitative	Compositional technique	Semi-quantitative	Cartilage	Synovium	Bone	Bone marrow lesions	Meniscus	Ligament	Study design	Downs criteria score
Eckstein F; Arthritis & Rheumatism; 2008; 18975356[117]	174	174	0	66(11.1)	(76%)	X			X						Longitudinal Prospective	8
Hellio Le Graverand MP; Annals Rheumatic Diseases; 2008; 19103634[118]	180				(100%)	X			X						Longitudinal Prospective	19
Eckstein F; Arthritis Research Therapy; 2009; 19534783[119]	79			60.3 (9.5)	79 (100%)	X			X						Longitudinal Prospective	15
Eckstein F; Arthritis Rheum; 2009: 19714595[120]	80			60.9 (9.1)	48 (60%)	X			X						Longitudinal Prospective	14
Hunter DJ; Osteoarthritis & Cartilage; 2009; 19744588[121]	150	150		60.9(9.9)	76(51%)	X			X						Longitudinal Prospective	18

Table IX
Results of random-effects pooling of SRM from MRI studies stratified by measure (quantitative and semi-quantitative) and tissue (cartilage, synovium, bone, bone marrow lesion, meniscus, and ligament). Studies with multiple estimates had an estimate selected at random and a pooled analysis was performed. The process was repeated 500 times to obtain the empirical distribution of pooled SRMs

Stratification	Number of estimates (Studies)	Mean sample size	Pooled SRM	95% CI
Quantitative cartilage*				
Medial femoral	54 (12)	118	−0.51	−0.74, −0.28
Medial tibial	55 (17)	134	−0.48	−0.63, −0.34
Medial tibiofemoral	31 (12)	92	−0.86	−1.26, −0.46
Lateral femoral	32 (8)	151	−0.21	−0.48, 0.05
Lateral tibial	44 (14)	152	−0.56	−0.72, −0.39
Lateral tibiofemoral	14 (5)	110	−1.01	−2.04, 0.02
Patella	13 (9)	131	−0.63	−0.90, −0.37
Global	5 (4)	48	−0.89	−2.59, 0.80
Quantitative other*				
Denuded area	19 (2)	114	−0.20	−0.85, 0.45
Bone	14 (2)	167	0.12	−0.46, 0.70
Bone marrow lesion	4 (1)	107	0.11	−0.08, 0.30
Meniscus	2 (1)	264	−0.24	−0.36, −0.12
Compositional	3 (1)	18	−3.27	−3.73, −2.81
Semi-quantitative cartilage†				
Medial tibial	1 (1)	325	−0.07	−0.18, 0.04
Medial tibiofemoral	3 (3)	224	0.55	0.47, 0.64
Lateral tibial	1 (1)	325	−0.05	−0.15, 0.06
Lateral tibiofemoral	3 (3)	224	0.37	0.18, 0.57
Patella	2 (2)	238	0.29	0.03, 0.56
Semi-quantitative other*				
Synovium	3 (2)	68	0.47	0.18, 0.77
Osteophytes	4 (1)	150	0.36	0.20, 0.52
Bone marrow lesion	6 (2)	130	0.43	−0.17, 1.03
Meniscus	2 (1)	264	0.27	0.15, 0.39

* Analysis used re-sampling techniques.
† Analysis did not use re-sampling techniques.

For MRI there is quite a lot of variability between different regions within the knee, and with different measures of different tissues, yet the SRM of −0.86 (95%CI −1.26 to −0.46) for the medial tibiofemoral joint quantitative cartilage measure provides advantages with regards to adequately powering studies.

Interestingly there have been a number of concerns raised about what appears to be conflicting data from earlier studies that were more responsive than studies conducted more recently. This analysis confirms that more recent studies (2007–2009) have slightly more conservative SRMs than earlier studies (2002–2006). For example the SRM for the medial tibiofemoral joint quantitative cartilage measure is −0.95 (−1.15, −0.76) for studies from 2002–2006 and is −0.84 (−1.35, −0.33) for studies from 2007 to 2009. The CIs for both these periods overlap and while there may be some differences in techniques between the two time periods including routine blinding to sequence in more recent studies that may explain differences, identifying the reasons for these differences was not the focus of this analysis. We have also been able to clearly demonstrate that adequate responsiveness can be attained in periods as short as 12 months.

Semi-quantitative scoring of MRIs is a valuable method for performing multi-feature assessment of the knee using conventional MRI acquisitions[6–8,98]. The responsiveness of the semi-quantitative assessment of medial tibiofemoral cartilage morphology (SRM 0.55) is broadly consistent with quantitative assessment for the medial tibiofemoral joint. Semi-quantitative assessment of synovium also demonstrated good responsiveness (SRM 0.52). In addition the semi-quantitative assessment of BMLs, a structural target with good clinical and predictive validity was also adequately responsive (SRM 0.43).

Table X
Results of random-effects pooling of SRM from MRI studies evaluating quantitative cartilage stratified by year of publication and plate region. Studies with multiple estimates had an estimate selected at random and a pooled analysis was performed. The process was repeated 500 times to obtain the empirical distribution of pooled SRMs

Stratification	Number of estimates (Studies)	Mean sample size	Pooled SRM	95% CI
2002–2006				
Medial femoral	3 (3)	126	−0.59	−1.21, 0.03
Medial tibial	7 (7)	123	−0.58	−0.81, −0.35
Medial tibiofemoral*	4 (3)	51	−0.95	−1.15, −0.76
Lateral femoral	1 (1)	117	−0.01	−0.19, 0.17
Lateral tibial	6 (6)	139	−0.55	−0.82, −0.29
Patella	5 (5)	141	−0.68	−1.04, −0.32
Global	2 (2)	24	−0.58	−1.15, −0.02
2007–2009				
Medial femoral	51 (9)	117	−0.49	−0.75, −0.22
Medial tibial	48 (10)	135	−0.42	−0.62, −0.22
Medial tibiofemoral	27 (9)	98	−0.84	−1.35, −0.33
Lateral femoral	31 (7)	152	−0.24	−0.53, 0.05
Lateral tibial	38 (8)	154	−0.56	−0.79, −0.33
Lateral tibiofemoral	14 (5)	110	−1.01	−2.04, 0.02
Patella	8 (4)	125	−0.58	−0.97, −0.18
Global	3 (2)	63	−1.24	−4.42, 1.94

* Note: All analyses of articles published in 2002–2006 did not use re-sampling techniques except for the medial tibial-femoral component. All analyses of articles published in 2007–2009 did use re-sampling techniques.

In summary, OA changes on MRI can be measured reliably using both quantitative and semi-quantitative techniques. MRI can accurately and feasibly measure change in quantitative cartilage morphometry over 12 months for knee OA. Based upon extant literature these study findings strongly support inclusion of MRI structure in updated regulatory guidance statements for clinical trials of structure modifying agents in OA.

Table XI
Results of random-effects pooling of SRM from MRI studies evaluating quantitative cartilage stratified by duration of study and plate region for studies published between 2007 and 2009. Studies with multiple estimates had an estimate selected at random and a pooled analysis was performed. The process was repeated 500 times to obtain the empirical distribution of pooled SRMs

Stratification	Number of estimates (Studies)	Mean sample size	Pooled SRM	95% CI
Quantitative cartilage				
1 year or less				
Medial femoral	27 (5)	82	−0.49	−0.81, −0.17
Medial tibial	18 (6)	93	−0.33	−0.53, −0.13
Medial tibiofemoral	16 (6)	83	−0.80	−1.27, −0.33
Lateral femoral	7 (3)	137	−0.30	−0.98, 0.38
Lateral tibial	8 (4)	130	−0.56	−0.88, −0.24
Lateral tibiofemoral	3 (2)	79	−1.03	−2.79, 0.73
Patella	7 (3)	129	−0.47	−0.92, −0.02
Global	2 (1)	18	0.45	−0.01, 0.92
1–2 years				
Medial femoral	6 (3)	104	−0.51	−1.15, 0.13
Medial tibial	6 (3)	104	−0.63	−1.14, −0.12
Medial tibiofemoral	5 (2)	53	−1.16	−2.90, 0.58
Lateral femoral	6 (3)	104	−0.21	−0.51, 0.09
Lateral tibial	6 (3)	104	−0.61	−1.14, −0.08
Lateral tibiofemoral	5 (2)	53	−1.28	−3.48, 0.91
Patella	1 (1)	99	−0.90	−1.10, −0.71
Global	1 (1)	154	−2.85	−3.01, −2.70
Greater than 2 years*				
Medial femoral	18 (1)	174	−0.32	−0.47, −0.17
Medial tibial	24 (1)	174	−0.27	−0.42, −0.12
Medial tibiofemoral	6 (1)	174	−0.41	−0.56, −0.26
Lateral femoral	18 (1)	174	−0.22	−0.37, −0.07
Lateral tibial	24 (1)	174	−0.42	−0.57, −0.27
Lateral tibiofemoral	6 (1)	174	−0.43	−0.57, −0.28

* Represents results of one study[117].

Declaration of funding and role of funding source

The OARSI FDA OA Initiative received financial support from the following professional organization:

American College of Rheumatology

Additionally the OARSI FDA OA Initiative received financial support from the following companies:

Amgen
ArthroLab
AstraZeneca
Bayer Healthcare
Chondrometrics
CombinatoRx
Cypress BioScience
DePuy Mitek
Expanscience 4QImaging
Genevrier/IBSA
Genzyme
King (Alpharma)
Merck
Merck Serono
NicOx
Pfizer
Rottapharm
Smith & Nephew
Wyeth

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the Food and Drug Administration.

Author contributions

DJH conceived and designed the study, drafted the manuscript and takes responsibility for the integrity of the work as a whole, from inception to finished article. EL and WZ were also involved in the design of the study. All authors contributed to acquisition of the data. All authors critically revised the manuscript and gave final approval of the article for submission.

The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Conflict of interest

David Hunter receives research or institutional support from DonJoy, NIH, and Stryker.
Other authors declared no conflict of interest.

Acknowledgements

We recognize the invaluable support of Valorie Thompson for administrative and editorial support and OARSI for their invaluable support of this activity. This analysis and literature review was undertaken to facilitate discussions and development of recommendations by the Assessment of Structural Change Working group for the OARSI FDA Initiative. The members of the working group were: Philip Conaghan, MB, BS, PhD (Chair), David Hunter, MBBS,

PhD, Jeffrey Duryea, PhD, Garry Gold, MD, Steven Mazzuca, PhD, Jean Francis Maillefert, MD, Timothy Mosher, MD, Hollis Potter, MD, David Felson, MD, Ali Guermazi, MD, Helen Keen, MD, Gayle Lester, PhD, Wayne Tsuji, MD, John Randle, PhD, Felix Eckstein, MD, Erika Schneider, PhD, Elena Losina, PhD, Sarah Kingsbury, PhD, William Reichman, PhD, Jean Pierre Pelletier, MD, Saara Totterman, MD, PhD, Rose Maciewicz, PhD, Bernard Dardzinski, PhD, Mona Wahba, MD, Marie Pierre Hellio Le Graverand-Gastineau, MD, PhD, DSc, Elisabeth Morris, DVM, Jeffrey Kraines, MD, Lucio Rovati, MD, Don Dreher, MD, PhD, James Huckle, PhD, Mary-Ann Preston, PhD, Brooks Story, PhD.

References

1. Guermazi A, Burstein D, Conaghan P, Eckstein F, Hellio Le Graverand-Gastineau MP, Keen H, *et al*. Imaging in osteoarthritis. [Review] [183 refs]. Rheum Dis Clin North Am 2008;34:645–87.
2. Eckstein F, Burstein D, Link TM. Quantitative MRI of cartilage and bone: degenerative changes in osteoarthritis. [Review] [238 refs]. NMR Biomed 2006;19:822–54.
3. Downs SH, Black N. The feasibility of creating a checklist for the assessment of the methodological quality both of randomised and non-randomised studies of health care interventions. J Epidemiol Community Health 1998;52:377–84.
4. Eckstein F, Mosher T, Hunter D. Imaging of knee osteoarthritis: data beyond the beauty. [Review] [100 refs]. Curr Opin Rheumatol 2007;19:435–43.
5. Gray ML, Burstein D. Molecular (and functional) imaging of articular cartilage. [Review] [69 refs]. J Musculoskelet Neuronal Interact 2004;4:365–8.
6. Peterfy CG, Guermazi A, Zaim S, Tirman PF, Miaux Y, White D, *et al*. Whole-Organ Magnetic Resonance Imaging Score (WORMS) of the knee in osteoarthritis. Osteoarthritis Cartilage 2004;12:177–90.
7. Biswal S, Hastie T, Andriacchi TP, Bergman GA, Dillingham MF, Lang P. Risk factors for progressive cartilage loss in the knee: a longitudinal magnetic resonance imaging study in forty-three patients. Arthritis Rheum 2002;46:2884–92.
8. Hunter D, Gale D, Grainger G, Lo G, Guermazi A, Conaghan P. The reliability of a new scoring system for knee osteoarthritis MRI and the validity of bone marrow lesion assessment: BLOKS (Boston Leeds Osteoarthritis Knee Score). Ann Rheum Dis 2008;67:206–11.
9. Kornaat PR, Ceulemans RY, Kroon HM, Riyazi N, Kloppenburg M, Carter WO, *et al*. MRI assessment of knee osteoarthritis: Knee Osteoarthritis Scoring System (KOSS)—inter-observer and intra-observer reproducibility of a compartment-based scoring system. Skeletal Radiol 2005;34:95–102.
10. Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for outcome measures in rheumatology. J Rheumatol 1998;25:198–9.
11. Lassere M. A users guide to measurement in medicine. Osteoarthritis Cartilage 2006;14(Suppl 1):10–4.
12. Emrani PS, Katz JN, Kessler CL, Reichmann WM, Wright EA, McAlindon TE, *et al*. Joint space narrowing and Kellgren–Lawrence progression in knee osteoarthritis: an analytic literature synthesis. Osteoarthritis Cartilage 2008;16:873–82.
13. Karvonen RL, Negendank WG, Teitge RA, Reed AH, Miller PR, Fernandez-Madrid F. Factors affecting articular cartilage thickness in osteoarthritis and aging. J Rheumatol 1994;21:1310–8.
14. Peterfy CG, van Dijke CF, Janzen DL, Gluer CC, Namba R, Majumdar S, *et al*. Quantification of articular cartilage in the knee with pulsed saturation transfer subtraction and fat-suppressed MR imaging: optimization and validation. Radiology 1994;192:485–91.
15. Marshall KW, Mikulis DJ, Guthrie BM. Quantitation of articular cartilage using magnetic resonance imaging and three-dimensional reconstruction. J Orthop Res 1995;13:814–23.
16. Disler DG, McCauley TR, Kelman CG, Fuchs MD, Ratner LM, Wirth CR, *et al*. Fat-suppressed three-dimensional spoiled gradient-echo MR imaging of hyaline cartilage defects in the knee: comparison with standard MR imaging and arthroscopy. AJR Am J Roentgenol 1996;167:127–32.
17. Dupuy DE, Spillane RM, Rosol MS, Rosenthal DI, Palmer WE, Burke DW, *et al*. Quantification of articular cartilage in the knee with three-dimensional MR imaging. Acad Radiol 1996;3:919–24.
18. Trattnig S, Huber M, Breitenseher MJ, Trnka HJ, Rand T, Kaider A, *et al*. Imaging articular cartilage defects with 3D fat-suppressed echo planar imaging: comparison with conventional 3D fat-suppressed gradient echo sequence and correlation with histology. J Comput Assist Tomogr 1998;22:8–14.
19. Drape JL, Pessis E, Auleley GR, Chevrot A, Dougados M, Ayral X. Quantitative MR imaging evaluation of chondropathy in osteoarthritic knees. Radiology 1998;208:49–55.
20. Cicuttini F, Forbes A, Morris K, Darling S, Bailey M, Stuckey S. Gender differences in knee cartilage volume as measured by magnetic resonance imaging. Osteoarthritis Cartilage 1999;7:265–71.
21. Pham XV, Monteiro I, Judet O, Sissakian JF, Plantin P, Aegerter P, *et al*. Magnetic resonance imaging changes in periarticular soft tissues during flares of medial compartment knee osteoarthritis. Preliminary study in 10 p.tients. Rev Rhum Engl Ed 1999;66:398–403.
22. Gale DR, Chaisson CE, Totterman SM, Schwartz RK, Gale ME, Felson D. Meniscal subluxation: association with osteoarthritis and joint space narrowing. Osteoarthritis Cartilage 1999;7:526–32.
23. Hyhlik-Durr A, Faber S, Burgkart R, Stammberger T, Maag KP, Englmeier KH, *et al*. Precision of tibial cartilage morphometry with a coronal water-excitation MR sequence. Eur Radiol 2000;10:297–303.
24. Jones G, Glisson M, Hynes K, Cicuttini F. Sex and site differences in cartilage development: a possible explanation for variations in knee osteoarthritis in later life. Arthritis Rheum 2000;43:2543–9.
25. Wluka AE, Davis SR, Bailey M, Stuckey SL, Cicuttini FM. Users of oestrogen replacement therapy have more knee cartilage than non-users. Ann Rheum Dis 2001;60:332–6.
26. Felson DT, Chaisson CE, Hill CL, Totterman SM, Gale ME, Skinner KM, *et al*. The association of bone marrow lesions with pain in knee osteoarthritis. [see comments.]. Ann Intern Med 2001;134:541–9.
27. Hill CL, Gale DG, Chaisson CE, Skinner K, Kazis L, Gale ME, *et al*. Knee effusions, popliteal cysts, and synovial thickening: association with knee pain in osteoarthritis. J Rheumatol 2001;28:1330–7.
28. Bergin D, Keogh C, O'Connell M, Rowe D, Shah B, Zoga A, *et al*. Atraumatic medial collateral ligament oedema in medial compartment knee osteoarthritis. Skeletal Radiol 2002;31:14–8.
29. Beuf O, Ghosh S, Newitt DC, Link TM, Steinbach L, Ries M, *et al*. Magnetic resonance imaging of normal and osteoarthritic trabecular bone structure in the human knee. Arthritis Rheum 2002;46:385–93.
30. Wluka AE, Stuckey S, Snaddon J, Cicuttini FM. The determinants of change in tibial cartilage volume in osteoarthritic knees. Arthritis Rheum 2002;46:2065–72.

31. Gandy SJ, Dieppe PA, Keen MC, Maciewicz RA, Watt I, Waterton JC, *et al*. No loss of cartilage volume over three years in patients with knee osteoarthritis as assessed by magnetic resonance imaging. Osteoarthritis Cartilage 2002;10:929–37.
32. Bhattacharyya T, Gale D, Dewire P, Totterman S, Gale ME, McLaughlin S, *et al*. The clinical importance of meniscal tears demonstrated by magnetic resonance imaging in osteoarthritis of the knee. [comment]. J Bone Joint Surg Am 2003; 85-A:4–9.
33. Cicuttini FM, Wluka AE, Wang Y, Stuckey SL, Davis SR. Effect of estrogen replacement therapy on patella cartilage in healthy women. Clin Exp Rheumatol 2003;21:79–82.
34. Raynauld JP, Kauffmann C, Beaudoin G, Berthiaume MJ, de Guise JA, Bloch DA, *et al*. Reliability of a quantification imaging system using magnetic resonance images to measure cartilage thickness and volume in human normal and osteoarthritic knees. Osteoarthritis Cartilage 2003;11: 351–60.
35. Felson DT, McLaughlin S, Goggins J, LaValley MP, Gale ME, Totterman S, *et al*. Bone marrow edema and its relation to progression of knee osteoarthritis. Ann Intern Med 2003;139: 330–6.
36. Hill CL, Gale DR, Chaisson CE, Skinner K, Kazis L, Gale ME, *et al*. Periarticular lesions detected on magnetic resonance imaging: prevalence in knees with and without symptoms. Arthritis Rheum 2003;48:2836–44.
37. Glaser C, Burgkart R, Kutschera A, Englmeier KH, Reiser M, Eckstein F. Femoro-tibial cartilage metrics from coronal MR image data: technique, test-retest reproducibility, and findings in osteoarthritis. Magn Reson Med 2003;50:1229–36.
38. Lindsey CT, Narasimhan A, Adolfo JM, Jin H, Steinbach LS, Link T, *et al*. Magnetic resonance evaluation of the interrelationship between articular cartilage and trabecular bone of the osteoarthritic knee. Osteoarthritis Cartilage 2004;12:86–96.
39. Cicuttini FM, Wluka AE, Wang Y, Stuckey SL, Cicuttini FM, Wluka AE, *et al*. Longitudinal study of changes in tibial and femoral cartilage in knee osteoarthritis.[see comment]. Arthritis Rheum 2004;50:94–7.
40. Raynauld JP, Martel-Pelletier J, Berthiaume MJ, Labonte F, Beaudoin G, de Guise JA, *et al*. Quantitative magnetic resonance imaging evaluation of knee osteoarthritis progression over two years and correlation with clinical symptoms and radiologic changes. Arthritis Rheum 2004;50:476–87.
41. Cicuttini F, Wluka A, Hankin J, Wang Y, Cicuttini F, Wluka A, *et al*. Longitudinal study of the relationship between knee angle and tibiofemoral cartilage volume in subjects with knee osteoarthritis. Rheumatology 2004;43:321–4.
42. Dashti M, Wluka AE, Geso M, Davis SR, Stuckey S, Cicuttini FM. Relationship between the area of cartilage shown on the magnetic resonance imaging middle-slice image of the medial and lateral tibial cartilages with cartilage volume and grade of osteoarthritis over time. Scand J Rheumatol 2004;33:87–93.
43. Cicuttini FM, Wluka AE, Hankin J, Stuckey S. Comparison of patella cartilage volume and radiography in the assessment of longitudinal joint change at the patellofemoral joint. J Rheumatol 2004;31:1369–72.
44. Baysal O, Baysal T, Alkan A, Altay Z, Yologlu S. Comparison of MRI graded cartilage and MRI based volume measurement in knee osteoarthritis. Swiss Med Wkly 2004;134:283–8.
45. Yoshioka H, Stevens K, Hargreaves BA, Steines D, Genovese M, Dillingham MF, *et al*. Magnetic resonance imaging of articular cartilage of the knee: comparison between fat-suppressed three-dimensional SPGR imaging, fat-suppressed FSE imaging, and fat-suppressed three-dimensional DEFT imaging, and correlation with arthroscopy. J Magn Reson Imaging 2004;20: 857–64.
46. Ding C, Garnero P, Cicuttini F, Scott F, Cooley H, Jones G, *et al*. Knee cartilage defects: association with early radiographic osteoarthritis, decreased cartilage volume, increased joint surface area and type II collagen breakdown. Osteoarthritis Cartilage 2005;13:198–205.
47. Hill CL, Seo GS, Gale D, Totterman S, Gale ME, Felson DT. Cruciate ligament integrity in osteoarthritis of the knee. Arthritis Rheum 2005;52:794–9.
48. Maataoui A, Graichen H, Abolmaali ND, Khan MF, Gurung J, Straub R, *et al*. Quantitative cartilage volume measurement using MRI: comparison of different evaluation techniques. Eur Radiol 2005;15:1550–4.
49. Cicuttini F, Hankin J, Jones G, Wluka A. Comparison of conventional standing knee radiographs and magnetic resonance imaging in assessing progression of tibiofemoral joint osteoarthritis. Osteoarthritis Cartilage 2005;13:722–7.
50. Huh YM, Kim S, Suh JS, Song H, Song K, Shin KH. The role of popliteal lymph nodes in differentiating rheumatoid arthritis from osteoarthritis by using CE 3D FSPGR MR imaging: relationship of the inflamed synovial volume. Korean J Radiol 2005;6:117–24.
51. Wluka AE, Ding C, Jones G, Cicuttini FM. The clinical correlates of articular cartilage defects in symptomatic knee osteoarthritis: a prospective study. Rheumatology (Oxford) 2005;44:1311–6.
52. Eckstein F, Hudelmaier M, Wirth W, Kiefer B, Jackson R, Yu J, *et al*. Double echo steady state magnetic resonance imaging of knee articular cartilage at 3 tesla: a pilot study for the Osteoarthritis Initiative. Ann Rheum Dis 2006;65:433–41.
53. Eckstein F, Charles HC, Buck RJ, Kraus VB, Remmers AE, Hudelmaier M, *et al*. Accuracy and precision of quantitative assessment of cartilage morphology by magnetic resonance imaging at 3.0T. Arthritis Rheum 2005;52:3132–6.
54. Sengupta M, Zhang YQ, Niu JB, Guermazi A, Grigorian M, Gale D, *et al*. High signal in knee osteophytes is not associated with knee pain. Osteoarthritis Cartilage 2006;14:413–7.
55. Raynauld JP, Martel-Pelletier J, Berthiaume MJ, Beaudoin G, Choquette D, Haraoui B, *et al*. Long term evaluation of disease progression through the quantitative magnetic resonance imaging of symptomatic knee osteoarthritis patients: correlation with clinical symptoms and radiographic changes. Arthritis Res Ther 2006;8:R21.
56. Hunter DJ, Zhang YQ, Niu JB, Tu X, Amin S, Clancy M, *et al*. The association of meniscal pathologic changes with cartilage loss in symptomatic knee osteoarthritis. Arthritis Rheum 2006;54:795–801.
57. Brandt KD, Mazzuca SA, Buckwalter KA. Acetaminophen, like conventional NSAIDs, may reduce synovitis in osteoarthritic knees. Rheumatology (Oxford) 2006;45:1389–94.
58. Jaremko JL, Cheng RW, Lambert RG, Habib AF, Ronsky JL. Reliability of an efficient MRI-based method for estimation of knee cartilage volume using surface registration. Osteoarthritis Cartilage 2006;14:914–22.
59. Hunter DJ, LaValley M, Li J, Zhang Y, Bauer D, Nevitt M, *et al*. Urinary pentosidine does not predict cartilage loss among subjects with symptomatic knee OA: the BOKS study. Osteoarthritis Cartilage 2007;15:93–7.
60. Boks SS, Vroegindeweij D, Koes BW, Hunink MM, Bierma-Zeinstra SM. Magnetic resonance imaging abnormalities in symptomatic and contralateral knees: prevalence and associations with traumatic history in general practice. Am J Sports Med 2006;34:1984–91.

61. Brem MH, Pauser J, Yoshioka H, Brenning A, Stratmann J, Hennig FF, *et al*. Longitudinal in vivo reproducibility of cartilage volume and surface in osteoarthritis of the knee. Skeletal Radiol 2007;36:315–20.
62. Folkesson J, Dam EB, Olsen OF, Pettersen PC, Christiansen C. Segmenting articular cartilage automatically using a voxel classification approach. IEEE Trans Med Imaging 2007;26: 106–15.
63. Dam EB, Folkesson J, Pettersen PC, Christiansen C. Automatic morphometric cartilage quantification in the medial tibial plateau from MRI for osteoarthritis grading. Osteoarthritis Cartilage 2007;15:808–18.
64. Baranyay FJ, Wang Y, Wluka AE, English DR, Giles GG, Sullivan RO, *et al*. Association of bone marrow lesions with knee structures and risk factors for bone marrow lesions in the knees of clinically healthy, community-based adults. Semin Arthritis Rheum 2007;37:112–8.
65. Hanna F, Teichtahl AJ, Bell R, Davis SR, Wluka AE, O'Sullivan R, *et al*. The cross-sectional relationship between fortnightly exercise and knee cartilage properties in healthy adult women in midlife. Menopause 2007;14: 830–4.
66. Hill CL, Hunter DJ, Niu J, Clancy M, Guermazi A, Genant H, *et al*. Synovitis detected on magnetic resonance imaging and its relation to pain and cartilage loss in knee osteoarthritis. Ann Rheum Dis 2007;66:1599–603.
67. Qazi AA, Folkesson J, Pettersen PC, Karsdal MA, Christiansen C, Dam EB. Separation of healthy and early osteoarthritis by automatic quantification of cartilage homogeneity. Osteoarthritis Cartilage 2007;15:1199–206.
68. Guymer E, Baranyay F, Wluka AE, Hanna F, Bell RJ, Davis SR, *et al*. A study of the prevalence and associations of subchondral bone marrow lesions in the knees of healthy, middle-aged women. Osteoarthritis Cartilage 2007;15: 1437–42.
69. Eckstein F, Kunz M, Schutzer M, Hudelmaier M, Jackson RD, Yu J, *et al*. Two year longitudinal change and test-retest-precision of knee cartilage morphology in a pilot study for the osteoarthritis initiative. Osteoarthritis Cartilage 2007;15:1326–32.
70. Akhtar S, Poh CL, Kitney RI. An MRI derived articular cartilage visualization framework. Osteoarthritis Cartilage 2007;15:1070–85.
71. Raynauld JP, Martel-Pelletier J, Berthiaume MJ, Abram F, Choquette D, Haraoui B, *et al*. Correlation between bone lesion changes and cartilage volume loss in patients with osteoarthritis of the knee as assessed by quantitative magnetic resonance imaging over a 24-month period. Ann Rheum Dis 2008;67:683–8.
72. Felson DT, Niu J, Guermazi A, Roemer F, Aliabadi P, Clancy M, *et al*. Correlation of the development of knee pain with enlarging bone marrow lesions on magnetic resonance imaging. Arthritis Rheum 2007;56:2986–92.
73. Lo GH, Niu J, McLennan CE, Kiel DP, McLean RR, Guermazi A, *et al*. Meniscal damage associated with increased local subchondral bone mineral density: a Framingham study. Osteoarthritis Cartilage 2008;16:261–7.
74. Davies-Tuck M, Teichtahl AJ, Wluka AE, Wang Y, Urquhart DM, Cui J, *et al*. Femoral sulcus angle and increased patella facet cartilage volume in an osteoarthritic population. Osteoarthritis Cartilage 2008;16:131–5.
75. Folkesson J, Dam EB, Olsen OF, Christiansen C. Accuracy evaluation of automatic quantification of the articular cartilage surface curvature from MRI. Acad Radiol 2007;14:1221–8.
76. Sanz R, Marti-Bonmati L, Rodrigo JL, Moratal D. MR pharmacokinetic modeling of the patellar cartilage differentiates normal from pathological conditions. J Magn Reson Imaging 2008;27:171–7.
77. Englund M, Niu J, Guermazi A, Roemer FW, Hunter DJ, Lynch JA, *et al*. Effect of meniscal damage on the development of frequent knee pain, aching, or stiffness. Arthritis Rheum 2007;56:4048–54.
78. Hernandez-Molina G, Guermazi A, Niu J, Gale D, Goggins J, Amin S, *et al*. Central bone marrow lesions in symptomatic knee osteoarthritis and their relationship to anterior cruciate ligament tears and cartilage loss. Arthritis Rheum 2008;58: 130–6.
79. Amin S, Guermazi A, LaValley MP, Niu J, Clancy M, Hunter DJ, *et al*. Complete anterior cruciate ligament tear and the risk for cartilage loss and progression of symptoms in men and women with knee osteoarthritis. Osteoarthritis Cartilage 2008;16:897–902.
80. Teichtahl AJ, Wang Y, Wluka AE, Szramka M, English DR, Giles GG, *et al*. The longitudinal relationship between body composition and patella cartilage in healthy adults. Obesity (Silver Spring) 2008;16:421–7.
81. Anandacoomarasamy A, Bagga H, Ding C, Burkhardt D, Sambrook PN, March LM. Predictors of clinical response to intraarticular Hylan injections – a prospective study using synovial fluid measures, clinical outcomes, and magnetic resonance imaging. J Rheumatol 2008;35:685–90.
82. Eckstein F, Buck RJ, Burstein D, Charles HC, Crim J, Hudelmaier M, *et al*. Precision of 3.0 tesla quantitative magnetic resonance imaging of cartilage morphology in a multicentre clinical trial. Ann Rheum Dis 2008;67:1683–8.
83. Reichenbach S, Guermazi A, Niu J, Neogi T, Hunter DJ, Roemer FW, *et al*. Prevalence of bone attrition on knee radiographs and MRI in a community-based cohort. Osteoarthritis Cartilage 2008;16:1005–10.
84. Petterson SC, Barrance P, Buchanan T, Binder-Macleod S, Snyder-Mackler L. Mechanisms underlying quadriceps weakness in knee osteoarthritis. Med Sci Sports Exerc 2008; 40:422–7.
85. Bolbos RI, Zuo J, Banerjee S, Link TM, Ma CB, Li X, *et al*. Relationship between trabecular bone structure and articular cartilage morphology and relaxation times in early OA of the knee joint using parallel MRI at 3 T. Osteoarthritis Cartilage 2008;16:1150–9.
86. Pai A, Li X, Majumdar S. A comparative study at 3 T of sequence dependence of T2 quantitation in the knee. Magn Reson Imaging 2008;26:1215–20.
87. Folkesson J, Dam EB, Olsen OF, Karsdal MA, Pettersen PC, Christiansen C. Automatic quantification of local and global articular cartilage surface curvature: biomarkers for osteoarthritis? Magn Reson Med 2008;59:1340–6.
88. Mills PM, Wang Y, Cicuttini FM, Stoffel K, Stachowiak GW, Podsiadlo P, *et al*. Tibio-femoral cartilage defects 3-5 years following arthroscopic partial medial meniscectomy. Osteoarthritis Cartilage 2008;16:1526–31.
89. Dore D, Ding C, Jones G. A pilot study of the reproducibility and validity of measuring knee subchondral bone density in the tibia. Osteoarthritis Cartilage 2008;16:1539–44.
90. 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.
91. Englund M, Guermazi A, Gale D, Hunter DJ, Aliabadi P, Clancy M, *et al*. Incidental meniscal findings on knee MRI in middle-aged and elderly persons. N Engl J Med 2008;359: 1108–15.

92. Rauscher I, Stahl R, Cheng J, Li X, Huber MB, Luke A, *et al*. Meniscal measurements of T1rho and T2 at MR imaging in healthy subjects and patients with osteoarthritis. Radiology 2008;249:591–600.
93. Kijowski R, Blankenbaker DG, Davis KW, Shinki K, Kaplan LD, De Smet AA. Comparison of 1.5- and 3.0-T MR imaging for evaluating the articular cartilage of the knee joint. Radiology 2009;250:839–48.
94. Cicuttini FM, Forbes A, Yuanyuan W, Rush G, Stuckey SL. Rate of knee cartilage loss after partial meniscectomy. J Rheumatol 2002;29:1954–6.
95. Wluka AE, Stuckey S, Brand C, Cicuttini FM, Wluka AE, Stuckey S, *et al*. Supplementary vitamin E does not affect the loss of cartilage volume in knee osteoarthritis: a 2 year double blind randomized placebo controlled study. J Rheumatol 2002;29:2585–91.
96. Cicuttini F, Wluka A, Wang Y, Stuckey S, Cicuttini F, Wluka A, *et al*. The determinants of change in patella cartilage volume in osteoarthritic knees. J Rheumatol 2002;29:2615–9.
97. Pessis E, Drape JL, Ravaud P, Chevrot A, Dougados M, Ayral X. Assessment of progression in knee osteoarthritis: results of a 1 year study comparing arthroscopy and MRI. Osteoarthritis Cartilage 2003;11:361–9.
98. Wluka AE, Wolfe R, Stuckey S, Cicuttini FM. How does tibial cartilage volume relate to symptoms in subjects with knee osteoarthritis? Ann Rheum Dis 2004;63:264–8.
99. Blumenkrantz G, Lindsey CT, Dunn TC, Jin H, Ries MD, Link TM, *et al*. A pilot, two-year longitudinal study of the interrelationship between trabecular bone and articular cartilage in the osteoarthritic knee. Osteoarthritis Cartilage 2004;12:997–1005.
100. Zhai G, Ding C, Cicuttini F, Jones G. Optimal sampling of MRI slices for the assessment of knee cartilage volume for cross-sectional and longitudinal studies. BMC Musculoskelet Disord 2005;6:10.
101. Wang Y, Wluka AE, Cicuttini FM, Wang Y, Wluka AE, Cicuttini FM. The determinants of change in tibial plateau bone area in osteoarthritic knees: a cohort study. Arthritis Res Ther 2005;7:R687–93.
102. Ding C, Cicuttini F, Scott F, Boon C, Jones G. Association of prevalent and incident knee cartilage defects with loss of tibial and patellar cartilage: a longitudinal study. Arthritis Rheum 2005;52:3918–27.
103. Hunter DJ, Conaghan PG, Peterfy CG, Bloch D, Guermazi A, Woodworth T, *et al*. Responsiveness, effect size, and smallest detectable difference of Magnetic Resonance Imaging in knee osteoarthritis. Osteoarthritis Cartilage 2006;14(Suppl A):A112–5.
104. Wluka AE, Forbes A, Wang Y, Hanna F, Jones G, Cicuttini FM. Knee cartilage loss in symptomatic knee osteoarthritis over 4.5 years. Arthritis Res Ther 2006;8:R90.
105. Ding C, Cicuttini F, Blizzard L, Scott F, Jones G, Ding C, *et al*. A longitudinal study of the effect of sex and age on rate of change in knee cartilage volume in adults. Rheumatology 2007;46:273–9.
106. Hunter DJ, Zhang YQ, Tu X, LaValley M, Niu JB, Amin S, *et al*. Change in joint space width: hyaline articular cartilage loss or alteration in meniscus? Arthritis Rheum 2006;54:2488–95.
107. Bruyere O, Genant H, Kothari M, Zaim S, White D, Peterfy C, *et al*. Longitudinal study of magnetic resonance imaging and standard X-rays to assess disease progression in osteoarthritis. Osteoarthritis Cartilage 2007;15:98–103.
108. Stahl R, Blumenkrantz G, Carballido-Gamio J, Zhao S, Munoz T, Hellio Le Graverand-Gastineau MP, *et al*. MRI-derived T2 relaxation times and cartilage morphometry of the tibio-femoral joint in subjects with and without osteoarthritis during a 1-year follow-up. Osteoarthritis Cartilage 2007;15:1225–34.
109. Pelletier JP, Raynauld JP, Berthiaume MJ, Abram F, Choquette D, Haraoui B, *et al*. Risk factors associated with the loss of cartilage volume on weight-bearing areas in knee osteoarthritis patients assessed by quantitative magnetic resonance imaging: a longitudinal study. Arthritis Res Ther 2007;9:R74.
110. Hunter DJ, Li J, LaValley M, Bauer DC, Nevitt M, DeGroot J, *et al*. Cartilage markers and their association with cartilage loss on magnetic resonance imaging in knee osteoarthritis: the Boston Osteoarthritis Knee Study. Arthritis Res Ther 2007;9:R108.
111. Teichtahl AJ, Wluka AE, Cicuttini FM. Frontal plane knee alignment is associated with a longitudinal reduction in patella cartilage volume in people with knee osteoarthritis. Osteoarthritis Cartilage 2008;16:851–4.
112. Hunter DJ, Niu J, Zhang Y, Totterman S, Tamez J, Dabrowski C, *et al*. Change in cartilage morphometry: a sample of the progression cohort of the Osteoarthritis Initiative. Ann Rheum Dis 2009;68:349–56.
113. Sharma L, Eckstein F, Song J, Guermazi A, Prasad P, Kapoor D, *et al*. Relationship of meniscal damage, meniscal extrusion, malalignment, and joint laxity to subsequent cartilage loss in osteoarthritic knees. Arthritis Rheum 2008;58:1716–26.
114. Teichtahl AJ, Davies-Tuck ML, Wluka AE, Jones G, Cicuttini FM. Change in knee angle influences the rate of medial tibial cartilage volume loss in knee osteoarthritis. Osteoarthritis Cartilage 2009;17:8–11.
115. Raynauld JP, Martel-Pelletier J, Bias P, Laufer S, Haraoui B, Choquette D, *et al*. Protective effects of licofelone, a 5-lipoxygenase and cyclo-oxygenase inhibitor, versus naproxen on cartilage loss in knee osteoarthritis: a first multicentre clinical trial using quantitative MRI. Ann Rheum Dis 2009;68:938–47.
116. Wirth W, Hellio Le Graverand MP, Wyman BT, Maschek S, Hudelmaier M, Hitzl W, *et al*. Regional analysis of femorotibial cartilage loss in a subsample from the Osteoarthritis Initiative progression subcohort. Osteoarthritis Cartilage 2009;17:291–7.
117. Eckstein F, Wirth W, Hudelmaier M, Stein V, Lengfelder V, Cahue S, *et al*. Patterns of femorotibial cartilage loss in knees with neutral, varus, and valgus alignment. Arthritis Rheum 2008;59:1563–70.
118. Hellio Le Graverand MP, Buck RJ, Wyman BT, Vignon E, Mazzuca SA, Brandt KD, *et al*. Change in regional cartilage morphology and joint space width in osteoarthritis participants versus healthy controls - a multicenter study using 3.0 Tesla MRI and Lyon schuss radiography. Ann Rheum Dis 2010;69:155–62.
119. Eckstein F, Wirth W, Hudelmaier MI, Maschek S, Hitzl W, Wyman BT, *et al*. Relationship of compartment-specific structural knee status at baseline with change in cartilage morphology: a prospective observational study using data from the osteoarthritis initiative. Arthritis Res Ther 2009;11:R90.
120. Eckstein F, Benichou O, Wirth W, Nelson DR, Maschek S, Hudelmaier M, *et al*. Magnetic resonance imaging-based cartilage loss in painful contralateral knees with and without radiographic joint space narrowing: data from the Osteoarthritis Initiative. Arthritis Rheum 2009;61:1218–25.
121. Hunter DJ, Li L, Zhang YQ, Totterman S, Tamez J, Kwoh CK, *et al*. Region of interest analysis: by selecting regions with denuded areas can we detect greater amounts of change? Osteoarthritis Cartilage 2010;18:175–83.

Osteoarthritis and Cartilage 19 (2011) 606–610

Osteoarthritis and Cartilage

Summary and recommendations of the OARSI FDA osteoarthritis Assessment of Structural Change Working Group

P.G. Conaghan †‡*, D.J. Hunter §‖, J.F. Maillefert ¶#††, W.M. Reichmann ‡‡§§, E. Losina ‡‡§§

† Section of Musculoskeletal Disease, University of Leeds, Leeds, UK
‡ NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK
§ Rheumatology Department, Royal North Shore Hospital and Northern Clinical School, University of Sydney, Sydney, NSW, Australia
‖ Division of Research, New England Baptist Hospital, Boston, MA, USA
¶ Department of Rheumatology, Dijon University Hospital, Dijon F-21078, France
# University of Burgondy, Dijon F-21079, France
†† INSERM U887, Dijon F-21079, France
‡‡ Department of Orthopedic Surgery, Brigham and Women's Hospital, Boston, MA, USA
§§ Department of Biostatistics, Boston University School of Public Health, Boston, MA, USA

ARTICLE INFO

Article history:
Received 17 December 2010
Accepted 11 February 2011

Keywords:
Osteoarthritis
Conventional radiography
Magnetic resonance imaging
Validity
Responsiveness
Reliability

SUMMARY

Objective: The Osteoarthritis Research Society International initiated a number of working groups to address a call from the US Food and Drug Administration (FDA) on updating draft guidance on conduct of osteoarthritis (OA) clinical trials. The development of disease-modifying osteoarthritis drugs (DMOADs) remains challenging. The Assessment of Structural Change (ASC) Working Group aimed to provide a state-of-the-art critical update on imaging tools for OA clinical trials.
Methods: The Group focussed on the performance metrics of conventional radiographs (CR) and magnetic resonance imaging (MRI), performing systematic literature reviews for these modalities. After acquiring these reviews, summary and research recommendations were developed through a consensus process.
Results: For CR, there is some evidence for construct and predictive validity, with good evidence for reliability and responsiveness of metric measurement of joint space width (JSW). Trials off at least 1 and probably 2 years duration will be required. Although there is much less evidence for hip JSW, it may provide greater responsiveness than knee JSW. For MRI cartilage morphometry in knee OA, there is some evidence for construct and predictive validity, with good evidence for reliability and responsiveness. The responsiveness of semi-quantitative MRI assessment of cartilage morphology, bone marrow lesions and synovitis was also good in knee OA.
Conclusions: Radiographic JSW is still a recommended option for trials of structure modification, with the understanding that the construct represents a number of pathologies and trial duration may be long. MRI is now recommended for clinical trials in terms of cartilage morphology assessment. It is important to study all the joint tissues of the OA joint and the literature is growing on MRI quantification (and its responsiveness) of non-cartilage features. The research recommendations provided will focus researchers on important issues such as determining how structural change within the relatively short duration of a trial reflects long-term change in patient-centred outcomes.

Introduction

Current status of guidance for assessing osteoarthritis (OA) disease modification

The development of disease-modifying osteoarthritis drugs (DMOADs) is faced with many challenges. There remains an inadequate understanding of the primary endpoint for demonstrating DMOAD efficacy. The actual result of clinical OA symptomatic progression, arthroplasty, is associated with multiple problems as

* Address correspondence and reprint requests to: P.G. Conaghan, Section of Musculoskeletal Disease, Chapel Allerton Hospital, Leeds LS7 4SA, UK. Tel: 44-113-392-4883; Fax: 44-113-392-4991.
E-mail address: p.conaghan@leeds.ac.uk (P.G. Conaghan).

1063-4584/$ – see front matter
doi:10.1016/j.joca.2011.02.018

an endpoint in clinical trials including the variability in rates of surgery, in part related to socioeconomic disparities, different healthcare environments and the relatively low incidence rate of arthroplasties compared with the total OA burden[1,2]. Alternative clinical endpoints for DMOAD clinical trials have therefore been considered and the Food and Drug Administration (FDA) previously provided regulatory draft guidelines for use in DMOAD development[3]. The FDA Clinical Development Programs for Drugs, Devices and Biological Products Intended for the Treatment of OA draft guidelines defined the current acceptable structural endpoint for DMOAD clinical trials as a slowing in the loss of knee or hip joint space narrowing (JSN) using conventional radiographs (CR); depending on the structural change this would need to be accompanied by symptom improvement. Similar recommendations were adopted by the European Medicines Agency (EMEA) in Europe[4] (also adopted by the Therapeutic Goods Administration (TGA) in Australia) and remain in their recently revised Guideline[5].

The current hierarchy of claims for structural outcome as defined by the FDA Clinical Development Programs for Drugs, Devices and Biological Products Intended for the Treatment of OA draft guidance is as follows:

1. Normalise the X-ray (reverse progression).
2. Improve the X-ray (halt progression).
3. Slow JSN by at least a pre-specified amount (slow the rate of progression).

CR have traditionally been the method of choice in clinical trials because of their relative feasibility. Until recently, it was widely accepted that alteration in progression of JSN implies preservation of hyaline cartilage and consequently clinical benefit; measurement of joint space width (JSW) by X-ray was determined as the most appropriate structural endpoint measure[6,7]. However it was recognized that the nature and magnitude of structural changes that are likely to be clinically relevant remain uncertain. Whether parallel clinical outcomes should be included in the claim depends on what JSW outcome is achieved, but collection of these data (including measurement of pain, a patient global assessment, a self-administered questionnaire and the time to the need for total joint replacement surgery) was strongly recommended regardless of the anticipated outcome since their assessment is critical for analysis of the overall risks and benefit of a product[3]. Since the concept of structural improvement connotes an element of durability, trials to demonstrate structure improvement were recommended to last at least 1 year[3].

As well, owing to the rapid growth of magnetic resonance imaging (MRI) studies in the last decade, there has been an increased awareness that symptomatic OA represents a process involving all the tissues in the OA joint, not just cartilage[8,9]. MRI has evolved substantially over the last decade and its strengths include its ability to visualise individual tissue pathologies, as well as the interrelationship between tissue pathologies.

Limitations of JSN as an outcome

Although a product showing a slowing of JSN would be expected to also affect symptoms, it is possible that certain products may affect structural progression without associated symptomatic evidence[10]. It is also possible that slowing of structural progression may occur at an earlier time-point with later reduction in symptoms (acknowledged in the recent EMEA Guideline[5]). A claim of structural improvement (i.e., slowing of JSN) might conceivably be dissociated from other claims when the mechanism of action of the product, and/or the size of the effect on slowing of JSN, are suggestive of future clinical benefits. If products are not anticipated to have different effects on these parameters or show only small improvements in JSN without demonstrated effects in symptoms they will not generally be considered for approval or for separate claims. In other words, as long as an observed delay in JSN progression is correlated to an improvement of clinical outcomes it is considered as an appropriate primary endpoint and as a surrogate endpoint for total joint replacement, the critical event characteristic of medical treatment failure for OA. It is assumed that a delay in JSN will consequently delay the need for total joint surgery, and can hence be interpreted as a treatment success for DMOADs.

The use of JSN measured by CR as a structural endpoint is associated with some concerns. Since disease progression is generally slow, minimal and variable within and between subjects[10,11], the use of JSW as an endpoint measurement requires long-term treatment periods (>1 year) and inclusions of large patient numbers. Moreover, the inability of radiographs to visualise cartilage leads to lack of sensitivity to detect early and small changes in this tissue[12]. There is difficulty in obtaining high quality reproducible images of OA joints, despite state-of-the-art standardisation of radiographic protocols to reduce the variability related to joint repositioning[13]. MRI studies have demonstrated that JSN represents a complex of hyaline cartilage loss, meniscal extrusion and meniscal degeneration[14]. Although structure is a critical component of OA assessment, the relationships between structure and pain and/or function and between structure and future outcomes (e.g., arthroplasty) are not well developed and the definition of a clinically relevant change in JSN has not been established.

The use of JSW alone may not be entirely relevant as an outcome measure for DMOAD efficacy since it fails to capture the multitissue nature of OA[9,15]. As such, potential early beneficial changes in other components of the joint are missed by the use of JSN alone as the structural endpoint. Moreover, the insensitivity of JSN to early changes in cartilage and meniscus means that even "moderate" OA knees (Kellgren–Lawrence ≥ 2) may already represent a stage of the disease too molecularly and biochemically advanced for alteration of disease course by pharmacological intervention. Previous attempts at OA disease modification using JSN as an endpoint have provided important lessons about the design and conduct of such trials, including issues on radiographic positioning, measurement methods, and study "enrichment" for progressors in order to ensure progression in the placebo group; this has been previously well reviewed[7,16]. Despite the limitations as a measure for DMOAD efficacy, delay in JSN has been reported for a small number of potential DMOADs to date[7,16]. However the lack of associated symptomatic benefit in these studies has prevented any of these agents from being successfully registered.

Methods

In the last decade since the FDA produced its draft guidance for industry, much evidence has been accumulated on the assessment of structural change in OA. The Osteoarthritis Research Society International (OARSI) FDA OA Assessment of Structural Change (ASC) Working Group comprised a wide range of expertise including clinical trialists, methodologists, academics, imaging experts and pharmaceutical company representatives with relevant trials experience; the Group was tasked with:

1. Examining a number of key issues about the performance metrics (including predictive validity for relevant clinical outcomes and responsiveness) of the commonest imaging tools used to assess structural change in OA, focussing predominantly on CR and MRI, while briefly examining the information on other modalities especially the growing field of ultrasound. This was performed by conducting systematic literature reviews for CR and MRI. The draft strategy for the literature review was written in December 2008, sent to all members of

the ASC Working Group, underwent iterative revision, and a final version of the protocol was approved in February 2009. Details of the protocol and the search terms are published separately[17–20]. The literature search was conducted using articles published up to the time of the search in April 2009. For examining the role of ultrasound, where it was acknowledged there was a much smaller literature base, we used a recently published systematic literature review on ultrasound in OA[21].

2. Producing state-of-the-art recommendations in assessing OA structural change for the purposes of optimising utilisation in OA clinical trials, based on the findings of the systematic literature reviews and *via* a consensus approach.
3. To also develop research recommendations again based on the literature reviews and through a consensus process of the ASC Working Group.

The following summary and derived recommendations attempts to overview the large amount of literature reviewed during this process. Much of the work summarised below is detailed in the more detailed accompanying systematic reviews of both radiography and MRI[17–20]; for this reason, individual references supporting each statement are not provided in this summary.

Summary and recommendations

The underlying assumption of these recommendations is that the manifestations of joint pain and disability currently associated with OA are strongly related to the pathophysiology of OA seen in joint structures. This postulate is to some extent supported by epidemiological evidence of the association between radiographic OA, joint pain and disability in the general population.

Much of the published evidence on imaging ASC relates to OA of the knee, with much less evidence (especially for modern imaging modalities) relating to OA of the hip and very limited information available for hand OA. This summary must therefore be seen as largely related to trials for OA of the knee and to a lesser extent, the hip.

Importantly most of the therapeutic studies on OA have included symptomatic and radiographically moderate to severe OA, so there is an absence of literature and definitions for "early" OA, especially studies entering people before the currently recognized clinical syndrome is apparent and when structural pathology is presumably minimal. So the literature on the performance of existing imaging modalities at this important stage of the OA process is sparse.

When mentioned, the term '*therapies*' refers to drugs, devices and biological products entered into the treatment of OA and regulated by the FDA; this could also include interventions such as weight loss.

Conventional radiography

- CR presents an image of the joint space of a diarthrodial joint, the width of which represents the thickness of articular cartilage. In some joints, notably the knee, JSW also reflects the presence, location and condition of other structures (e.g., meniscus), and JSW is a composite measure of the combined thickness of those structures. This should be considered when defining a relevant knee trial outcome.
- Much is now known of the performance metrics of CR JSW in the knee and to a lesser extent in the hip. There is some evidence for construct and predictive validity, with good evidence for reliability and responsiveness of metric measurement of JSW. In terms of correlations with concurrent symptoms, there is a weak association between progression in JSN and progression of symptoms. There is little information on how progression in JSN during the course of a study reflects post-study (long-term) change in symptoms. JSN progression is associated with increased rate of subsequent total joint replacement, but these may not be truly independent events as radiographic JSN is one of the features used to select people for joint replacement surgery.
- In the knee, the use of fluoroscopic positioning and semi-flexed views improve responsiveness, although it is acknowledged that access to fluoroscopic facilities is restricted. Studies will generally need to be at least 12 and more likely 24 months duration.
- It is possible and therefore advisable to 'enrich' a knee OA study population to increase the rate of joint space loss, for example, by including higher Kellgren Lawrence (KL) grade.
- Automated methods for assessing parameters of JSW offer promise of improved precision and therefore improved responsiveness.
- The natural history of hip OA appears different to that of knee OA and although the literature concerning the hip is much less extensive, there is some evidence for better responsiveness for JSW measurement at the hip. Hip JSW as a construct does not include a meniscus. There is little evidence on enriching cohorts for purposes of increasing rate of JSN progression.
- We support the continued use of CR JSW as one option for assessing structural OA change, taking all the previous points into account when deciding on study endpoint.

Research recommendations

To further understand the relationship between JSN and symptoms, future studies should focus on the following areas:

- Cross-sectional studies in which the patients are their own controls (such as one recently published[22]) to better evaluate the potential correlation.
- Longitudinal studies evaluating the relationship between changes in symptoms and changes in joint space.
- Predictive validity studies, i.e., does joint space predict subsequent pain and disability and subsequent joint replacement? For example, does JSN between month 0 and month 12 correlates with joint replacement by month 60?
- Construct validity studies, i.e., correlation between JSN and mean pain or function. For example, is JSW between month 0 and month 12 correlated with mean pain and function evaluated every 6 months between month 0 and month 12, or between month 12 and month 24?

For Knee OA:

- Studies of the relationship between symptoms and radiographic joint space evaluated on semi-flexed X-rays with fluoroscopy.
- Studies on predictors of joint loss evaluated on semi-flexed X-rays with fluoroscopy and optimal serial tibial plateau alignment.
- Studies on the effect of rate of joint space loss on treatment effect.

For Hand OA:

- Studies comparing the metrological properties of hand OA scoring systems.

MRI

- For assessing MRI cartilage morphometry in knee OA, there is some evidence for construct and predictive validity, with good evidence for reliability and responsiveness. Using MRI it is

possible to accurately and feasibly measure change in cartilage morphology over 12 months for knee OA and we recommend the use of MRI for assessing cartilage morphology in trials of OA structure modification.
- It is possible to 'enrich' a knee OA study population with MRI outcomes in order to increase the rate of cartilage loss, for example, by including higher KL grade.
- In terms of correlations with concurrent symptoms, there is a weak association between progression of cartilage loss and increasing symptoms. There is little information on how change in cartilage parameters during the course of a study reflects post-study change in symptoms. There is some predictive validity with progression of cartilage loss predicting subsequent total joint replacement.
- More information is required on the performance metrics of MRI semi-quantitative and compositional measures of cartilage morphology. There may be a role for semi-quantitative assessments for assessing focal cartilage defects.
- Structure modification should be considered in a broader context than that of cartilage alone. Since MRI has the capacity to image the other tissues, further work is needed on the quantification and predictive validity of non-cartilage MRI pathologies. The performance metrics of non-cartilage MRI features have not been extensively studied but there is a rapidly emerging literature in this field.

Research recommendations

As well as the areas recommended above, including research into 'whole organ' or multi-tissue assessment (currently *via* semi-quantitative scores) and improving the quantification of non-cartilage pathologies, the MRI OA field needs:

- Studies to define more responsive measures of structural change.
- Studies that measure change at an earlier stage of disease when it may be more suitable for DMOAD intervention.
- Studies to improve predictive validity of current structural measures for important clinical outcomes (e.g., total joint replacement (TJR), virtual TJR).
- Studies to improve assessment precision of structural measures more closely related to symptom change (e.g., bone marrow lesions, synovitis).

Other imaging modalities

The potential for non-CR or MRI modalities to assess relevant non-cartilage tissues should be considered. Ultrasound is currently the other recent imaging modality with most information available, and at this stage it appears it is most promising as a tool for evaluating OA synovitis. Ultrasound-detected pathologies have been associated with current OA symptoms. Further work is required to better understand the performance metrics of ultrasonographic quantification of pathology, with such work requiring improved pathology definitions.

Conclusions

The amount of publications assessing OA structural change has dramatically increased over recent years, related to the availability of large cohort studies. Though we now understand CR JSW represents a complex of pathologies, gaps still remain in our understanding of both construct and predictive validity. The growth of MRI as an OA imaging biomarker has evolved to a point where it can be recommended for clinical trials in terms of cartilage morphology assessment. Much still needs to be understood about compositional cartilage measures and just as importantly, quantification of non-cartilage features. The literature reviews performed by the ASC Working Group, together with attention to the research recommendations listed here, should ensure that the current gaps in our knowledge regarding the performance metrics and clinical importance of existing tools will be filled.

Author contributions

PC: conception, drafting, critical revision, final approval of the manuscript.

DH, JFM, WR, EL: critical revision, final approval.

Declaration of funding and role of funding source

The OARSI FDA OA Initiative received financial support from the following professional organization:

American College of Rheumatology

Additionally the OARSI FDA OA Initiative received financial support from the following companies:

Amgen
ArthroLab
AstraZeneca
Bayer Healthcare
Chondrometrics
CombinatoRx
Cypress BioScience
DePuy Mitek
Expanscience
4QImaging
Genevrier/IBSA
Genzyme
King (Alpharma)
Merck
Merck Serono
NicOx
Pfizer
Rottapharm
Smith & Nephew
Wyeth

While individuals from pharmaceutical, biotechnology and device companies actively participated in on-going working group discussions, due to the conflict of interest policy enacted by OARSI, these individuals were not allowed to vote on the final recommendations made by OARSI to the FDA.

Conflict of interest statement

PC, JFM, WR, EL: no conflict of interest to declare.

DH: receives research or institutional support from DonJoy, NIH, and Stryker.

Acknowledgements

We acknowledge the contributions of the Assessment for Structural Change working group to this work: Jeffrey Duryea, PhD, Assistant Professor, Harvard Medical School; Garry Gold, MD, Associate Professor Radiology & Bioengineering, Stanford University; Steven Mazzuca, PhD, Director – Education, Epidemiology & Health Services, Indiana University; Timothy Mosher, MD, Professor of Radiology & Orthopaedics, Penn State University; Hollis Potter, MD, Chief, Division of MRI, Hospital for Special Surgery; David Felson, MD, MPH, Professor of Medicine & Public Health, Boston University; Gayle Lester, PhD, Program Director, Osteoarthritis &

Diagnostic Imaging, NIAMS/NIH; Erika Schneider, PhD, Diagnostic Radiology, The Cleveland Clinic Foundation; Helen Keen, MBBS, Senior Lecturer, University of Western Australia; Ali Guermazi, MD, Associate Professor Radiology, Boston University; Felix Eckstein, MD, Director, Institute of Anatomy & Musculoskeletal Research, Paracelsus Medical Center; Jean Pierre Pelletier, MD, Prof of Medicine, Hopitalier de l'Universite de Montreal; Saara Totterman, MD, PhD, Chief Medical Officer, 4QImaging; Wayne Tsuji, MD, Director & Early Development Leader Medical Science Amgen; John Randle, PhD (Previous), Senior Director Commercial Development, CombinatoRx; Rose Maciewicz, PhD, Chief Scientist, Respiratory & Inflammation, AstraZeneca; Bernard Dardzinski, PhD, Senior Director – Merck & Co; Mona Wahba, MD, Executive Director, US Arthritis, Novartis; Elisabeth Morris, DVM (Previous), Senior Director Osteoarthritis Research, Wyeth; Jeffrey Kraines, MD, Senior Medical Director, Genzyme Corporation; Marie-Pierre Hellio Le Graverand-Gastineau, MD, PhD, DSc, Pfizer Global Research & Development.

References

1. Ackerman IN, Dieppe PA, March LM, Roos EM, Nilsdotter AK, Brown GC, *et al*. Variation in age and physical status prior to total knee and hip replacement surgery: a comparison of centers in Australia and Europe. Arthritis Rheum 2009;61:166–73.
2. Hawker GA. Who, when, and why total joint replacement surgery? The patient's perspective. Curr Opin Rheumatol 2006;18:526–30.
3. FDA. Clinical Development Programs for Drugs, Devices and Biological Products Intended for the Treatment of OA 1999.
4. CHMP. Points to Consider – Clinical Investigation of Medicinal Products Used in the Treatment of OA 1998.
5. CHMP. Guideline on Clinical Investigation of Medicinal Products Used in the Treatment of Osteoarthritis 2010.
6. Mazzuca SA, Brandt KD. Plain radiography as an outcome measure in clinical trials involving patients with knee osteoarthritis. Rheum Dis Clin North Am 1999;25:467–80.
7. Brandt KD, Mazzucca SA. Lessons learned from nine clinical trials of disease-modifying osteoarthritis drugs. Arthritis Rheum 2005;52:3349–59.
8. Guermazi A, Burstein D, Conaghan P, Eckstein F, Hellio Le Graverand-Gastineau MP, Keen H, *et al*. Imaging in osteoarthritis. Rheum Dis Clin North Am 2008;34:645–87.
9. Wenham CY, Conaghan PG. Imaging the painful osteoarthritic knee joint: what have we learned? Nat Clin Pract Rheumatol 2009;5:149–58.
10. Brandt KD, Mazzucca SA, Katz BP, Lane KA, Buckwalter KA, Yocum DE, *et al*. Effects of doxycycline on progression of osteoarthritis: results of a randomised, placebo-controlled, double-blind trial. Arthritis Rheum 2005;52:2015–25.
11. Bingham CO, Buckland-Wright JC, Garnero P, Cohen SB, Dougados M, Adami S, *et al*. Risedronate decreases biochemical markers of cartilage degradation but does not decrease symptoms or slow radiographic progression in patients with medial compartment osteoarthritis of the knee: results of the two-year multinational knee osteoarthritis structural arthritis study. Arthritis Rheum 2006;54:3494–507.
12. Amin S, LaValley MP, Guermazi A, Grigoryan M, Hunter DJ, Clancy M, *et al*. The relationship between cartilage loss on magnetic resonance imaging and radiographic progression in men and women with knee osteoarthritis. Arthritis Rheum 2005;52:3152–9.
13. Hunter DJ, Le Graverand MP, Eckstein F. Radiologic markers of osteoarthritis progression. Curr Opin Rheumatol 2009;21:110–7.
14. Hunter DJ, Zhang YQ, Tu X, LaValley M, Niu JB, Amin S, *et al*. Change in joint space width: hyaline articular cartilage loss or alteration in meniscus? Arthritis Rheum 2006;54:2488–95.
15. Brandt KD, Radin EL, Dieppe PA, van de Putte L. Yet more evidence that osteoarthritis is not a cartilage disease. Ann Rheum Dis 2006;65:1261–4.
16. Hellio Le Graverand-Gastineau MP. OA clinical trials: current targets and trials for OA. Choosing molecular targets: what have we learned and where we are headed? Osteoarthritis Cartilage 2009;17:1393–401.
17. Reichmann WM, Maillefert JF, Hunter DJ, Conaghan PG, Katz JN, Losina E. Responsiveness to change and reliability of measurement of radiographic joint space width in osteoarthritis of the knee: a systematic review. Osteoarthritis Cartilage 2011;19:550–6.
18. Chu Miow Lin D, Reichmann WM, Gossec L, Losina E, Conaghan PG, Maillefert JF. Validity and responsiveness of radiographic joint space width metric measurement in hip osteoarthritis: a systematic review. Osteoarthritis Cartilage 2011;19:543–9.
19. Hunter DJ, Zhang W, Conaghan PG, Hirko K, Menashe L, Reichmann WM, Losina E. Responsiveness and reliability of MRI in knee osteoarthritis: a meta-analysis of published evidence. Osteoarthritis Cartilage 2011;19:589–605.
20. Hunter D, Zhang W, Conaghan PG, Hirko K, Menashe L, Li L, Reichmann WM, Losina E. Systematic review of the concurrent and predictive validity of MRI biomarkers in OA. Osteoarthritis Cartilage 2011;19:557–88.
21. Keen HI, Wakefield RJ, Conaghan PG. A systematic review of ultrasonography in osteoarthritis. Ann Rheum Dis 2009;68:611–9.
22. Neogi T, Felson D, Niu J, Nevitt M, Lewis CE, Aliabadi P, *et al*. Association between radiographic features of knee osteoarthritis and pain: results from two cohort studies. BMJ 2009;339:b2844.