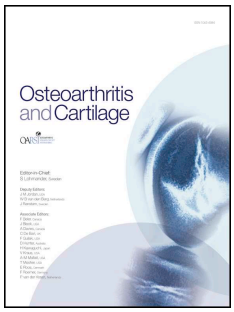
Article:
Leyland, KM, Gates, LS, Nevitt, M et al. (18 more authors) (2018) Harmonising measures of knee and hip osteoarthritis in population-based cohort studies: an international study. Osteoarthritis and Cartilage, 26 (7). pp. 872-879. ISSN 1063-4584

https://doi.org/10.1016/j.joca.2018.01.024

Accepted Manuscript



Harmonising measures of knee and hip osteoarthritis in population-based cohort studies: an international study

K.M. Leyland, L.S. Gates, M. Nevitt, D. Felson, S.M. Bierma-Zeinstra, P.G. Conaghan, L. Engebretsen, M. Hochberg, D.J. Hunter, G. Jones, J.M. Jordan, A. Judge, L.S. Lohmander, E.M. Roos, M.T. Sanchez-Santos, N. Yoshimura, J.B.J. van Meurs, M.E. Batt, J. Newton, C. Cooper, N.K. Arden

PII: S1063-4584(18)30090-6

DOI: 10.1016/j.joca.2018.01.024

Reference: YJOCA 4163

To appear in: *Osteoarthritis and Cartilage*

Received Date: 24 March 2017

Revised Date: 26 January 2018

Accepted Date: 30 January 2018

Please cite this article as: Leyland KM, Gates LS, Nevitt M, Felson D, Bierma-Zeinstra SM, Conaghan PG, Engebretsen L, Hochberg M, Hunter DJ, Jones G, Jordan JM, Judge A, Lohmander LS, Roos EM, Sanchez-Santos MT, Yoshimura N, van Meurs JBJ, Batt ME, Newton J, Cooper C, Arden NK, Harmonising measures of knee and hip osteoarthritis in population-based cohort studies: an international study, *Osteoarthritis and Cartilage* (2018), doi: 10.1016/j.joca.2018.01.024.

Harmonising measures of knee and hip osteoarthritis in population-based cohort studies: an international study

Authors: Leyland, K.M.[1,2], Gates, L.S.[1,3], Nevitt, M. [4], Felson, D. [5], Bierma-Zeinstra, S.M.[6, 7], Conaghan, P.G.[8], Engebretsen, L.[9], Hochberg, M.[10], Hunter, D.J.[11, 12], Jones, G.[13], Jordan, J.M.[14,15], Judge, A.[1] Lohmander, L.S.[16], Roos, E.M.[17], Sanchez-Santos, M.T.[1], Yoshimura, N.[18], van Meurs, J.B.J.[19], Batt, M.E.[20], Newton, J.[1], Cooper, C. [1,3], Arden, N.K [1,3]

[1]NIHR Musculoskeletal Biomedical Research Unit and Arthritis Research UK Centre for Sport, Exercise, and Osteoarthritis, University of Oxford, Oxford, UK

[2]MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

[3] MRC Lifecourse Epidemiology Unit, University of Southampton, Southampton, UK

[4]Department of Epidemiology and Biostatistics, University of California, San Francisco, CA, USA

[5]Clinical Epidemiology Research and Training Unit, Boston University School of Medicine, Boston, MA, USA

[6] Department of General Practice, Erasmus University Medical Centre, Rotterdam, the Netherlands

[7]Department of Orthopaedics, Erasmus University Medical Centre, Rotterdam, the Netherlands

[8] Leeds Institute of Rheumatic and Musculoskeletal Medicine, University of Leeds & NIHR Leeds Musculoskeletal Biomedical Research Unit, Leeds, UK

[9]Department of Orthopaedic Surgery, Oslo University Hospital and Oslo Sports Trauma Research Center, Norwegian School of Sports Sciences, Oslo, Norway

[10]University of Maryland School of Medicine, Baltimore, USA

[11]Institute of Bone and Joint Research, Kolling Institute, University of Sydney, Sydney, Australia

[12]Rheumatology Department, Royal North Shore Hospital, St Leonards, Sydney, Australia

[13]Menzies Research Institute Tasmania, University of Tasmania, Hobart, Australia

[14]Thurston Arthritis Research Center, University of North Carolina at Chapel Hill, Chapel Hill, NC USA

[15]Department of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC USA

[16]Lund University, Department of Clinical Sciences Lund, Orthopaedics, Lund, Sweden

[17] Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark

[18]Department of Joint Disease Research, 22nd Century Medical & Research Center, Faculty of Medicine, The University of Tokyo, Tokyo, Japan

[19]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands

[20]Centrefor Sports Medicine, Nottingham University Hospitals and Arthritis Research UK Centre for Sport, Exercise and Osteoarthritis, Nottingham, UK

Abstract

Objective: Population-based osteoarthritis (OA) cohorts provide vital data on risk factors and outcomes of OA, however the methods to define OA vary between cohorts. We aimed to provide recommendations for combining knee and hip OA data in extant and future population cohort studies, in order to facilitate informative individual participant level analyses. Method: International OA experts met to make recommendations on: 1) defining OA by x-ray and/or pain; 2) compare The National Health and Nutrition Examination Survey (NHANES)-type OA pain questions; 3) the comparability of the Western Ontario & McMaster Universities Osteoarthritis Index (WOMAC) scale to NHANES-type OA pain questions; 4) the best radiographic scoring method; 5) the usefulness of other OA outcome measures. Key issues were explored using new analyses in two population-based OA cohorts (Multicenter Osteoarthritis Study; MOST and Osteoarthritis Initiative OAI). Results: OA should be defined by both symptoms and radiographs, with symptoms alone as a secondary definition. Kellgren and Lawrence (K/L) grade ≥2 should be used to define radiographic OA. The variable wording of pain questions can result in varying prevalence between 41.0 and 75.4%, however questions where the time anchor is similar have high sensitivity and specificity (91.2% and 89.9% respectively). A threshold of 3 on a 0-20 scale (95% CI 2.1, 3.9) in the WOMAC pain subscale demonstrated equivalence with the preferred NHANES-type question. Conclusion: This research provides recommendations, based on expert agreement, for harmonising and combining OA data in existing and future population-based cohorts.

Keywords: Osteoarthritis; data; harmonisation; cohort; epidemiology

Introduction

OA is one of the most common causes of disability in the world (1). The prevention and management of OA is dependent on the understanding of modifiable risk factors for OA in the population at earlier stages of disease. To fully understand the risk factors for OA as well as its long-term effects, there is a need to combine data from population-based cohorts to provide sufficient statistical power. Traditional meta-analyses on OA rely on aggregate data obtained from study publications. These are vulnerable to outcome reporting and publication bias, and the quality and availability of data may vary across studies (2). An increasingly popular alternative to traditional meta-analysis is individual participant (IPD) meta-analysis, which utilises original raw data for the analysis. The key benefits of this type of analysis are the ability to better harmonise primary risk factors and outcomes between studies, the adjustment of identical confounders, the application of consistent inclusion and exclusion criteria, and the ability to include previously unpublished datasets into the analysis (3-5).

The critical limitation of traditional meta-analyses is the reliance upon the individual cohort definition of OA, some of which are over 50 years old. A diagnosis of OA is commonly established using radiographic features alone or in combination with joint pain, often defined using NHANES (National Health and Nutrition Examination Survey) type or WOMAC (Western Ontario and McMaster Universities Arthritis Index**)** questions (6). Many cohorts lack objective clinical assessment, which prevents the use of the American College of Rheumatology (ACR) criteria and the identification of pre-radiographic OA. More recently, self-reported pain, regardless of radiographic OA (ROA), has been used to measure disease burden. There are multiple ways to assess both radiographic OA and OA-related joint pain, and the comparability of these measurements is not yet completely understood. The choice of definition can substantially affect both OA prevalence and its association with risk factors. This has been demonstrated for ROA outcomes such as K/L grades and between the use of different individual feature atlases (7). Previous meetings have focused on defining early OA, however OA was outside the scope of their recommendations (8, 9).

The aim of this research was to generate recommendations for combining OA data within existing and future OA population cohort studies. A committee of international OA experts was convened to define OA for use in IPD meta-analyses using population-based cohorts. This paper presents the research and conclusions of the work performed by this committee.

Methods

Identification of key discussion points by the Steering Group

The steering group consisted of authors KML, LG, and NKA. Due to the variety of questionnaires and variables used to classify OA, the interest for this study were OA assessments used in previously collected longitudinal population-based cohort studies with concurrent OA-related pain and radiographic measures at multiple time points in the hip or knee. Cohorts were excluded if their non-OA subjects were recruited differently from their OA subjects, or did not have the same pain and ROA data available. Potential cohort studies were identified using two pathways: 1) literature review and 2) direct contact with principal investigators (PIs) of known osteoarthritis cohorts. The literature review sought to identify both cohorts matching the exact inclusion criteria, but also cohorts which appeared likely to have the data of interest (i.e. a published cross-sectional analysis of knee pain with indications that longitudinal and ROA data may exist) (appendix 1). Contact with PIs began with researchers with whom we had previous collaborative relationships, requesting their own unpublished variables and datasets along with any knowledge of additional cohorts matching the inclusion criteria. Additional PIs and datasets were identified through specialist OA meetings and conferences.

A comprehensive evaluation of OA variables available within the identified population-based and enhanced risk factor cohorts at baseline time-points, was undertaken by examining data dictionaries, liaising with cohort members or reviewing published cohort material. Cohorts were further excluded if their raw data and/or detailed data dictionaries were unavailable or inaccessible to the steering committee. Information was gathered to determine how each cohort utilised these OA variables in applied research and their methods of defining end-stage OA. Five key areas (outlined below) were identified as lacking sufficient published evidence to make decisions on combining OA data between data sources, and therefore opinions from international OA experts was sought.

Selection and endorsement of the Osteoarthritis Expert Committee

The definition and harmonisation of OA variables was determined within an expert group meeting. Participants contributed expert opinion on the key discussion points of the study (via video conference and email), recommended new statistical analyses, provided guidance on the post-hoc analyses, and contributed critical input on the manuscript. The panel consisted of multidisciplinary,

geographically diverse experts on OA and population-based cohort studies. Experts were selected based upon meeting one or more of the following criteria:

- Investigators with experience leading population cohorts who have an advanced knowledge of OA and thorough understanding of epidemiological cohort data collection
- Representatives with experience in producing guidelines for musculoskeletal disease definitions or investigative imaging techniques
- Members of the original IPD meta-analysis steering group to provide expertise and context for how the harmonised OA variable would be used for future research

Sixteen experts were invited to participate in the entire study. Nine of these attended the meeting by video link. All Sixteen contributed to the definition of new statistical analyses, the post hoc analysis and contributed to the manuscript.

The expert committee's work has been endorsed by Osteoarthritis Research Society International (OARSI), International Osteoporosis Foundation (IOF), European Society for Clinical and Economic Aspects of Osteoporosis and Osteoarthritis (ESCEO) and the British Association of Sport and Exercise Medicine (BASEM).

Meeting format

The process consisted of the following steps: 1) First steering committee meeting held in November 2014, where the decision was made to hold an expert meeting to address issues with existing OA data and produce recommendations for future research 2) Experts were contacted via email with aims and objectives of the meeting, points for discussion and all relevant background material identified by the steering committee including a summary of the type of variables each cohort appeared to contain from published literature and/or open access online data dictionaries; 3) A meeting was conducted in April 2015, using a structured discussion surrounding the five key points, led by NKA and KML; 4) Discussions on each point continued until agreement was reached using an iterative process, or it was determined that further action and/or information was required in order to reach agreement, which was provided by steering committee members; 5) A document containing the results from the April meeting along with the further recommended analysis was fed back to the group via email, with all experts indicating agreement, disagreement, or modification (November 2015); 6) To account for potential negative group dynamics, dissenting opinions could be voiced directly to the steering committee, where it was anonymously added to the feedback

document for discussion by all experts; 7) Final decisions were agreed via email by October 2015 8) First draft of manuscript produced in June 2016.

Five key discussion points

1. To determine the criteria to classify OA in population-based cohort studies
2. To determine the comparability of existing NHANES-type pain questions, which contain wording variations
3. To assess whether previously published thresholds used to determine pain using the WOMAC scale were appropriate for research, and determine comparability with the NHANES-type pain questions
4. To review the comparability of radiographic scoring methods and establish the 'best' measure to use based on available data
5. To assess the usability and comparability of alternate OA outcomes: self-reported OA, GP diagnosis, and joint replacement for OA

Results

1. To determine the criteria to classify OA in population-based cohort studies

Potential definitions of OA (radiographic, symptoms alone or symptomatic radiographic) were presented with supporting evidence to the expert committee for discussion.

Expert Discussion

The committee recognized that there has been a shift toward the importance of pain as a driving factor in the definition of OA, rather than structural factors alone. However, due to the risk of misclassification it was felt that the combination of symptoms and structural features would provide the most accurate definition. The committee also considered that symptoms alone, without radiographic data, could be an important aspect of the OA definition. Due to the lack of standardization and reliability of pain assessments available at multiple time-points, it was agreed that self-reported pain questions should not be used alone in the current state of knowledge.

Decision

Experts agreed to use symptomatic radiographic OA as the primary criteria to classify OA for the purpose of combining OA classifications across cohort studies. Pain alone was suggested as a secondary criterion. When defining pain, experts agreed that a binary, self-reported, joint-specific pain question would provide the best definition of OA-related symptoms in the majority of the population-based cohorts.

2. **To establish the comparability of existing NHANES-type pain questions which contain wording variations**

The committee was provided with details of the wording variation found in pain questions commonly used in population based studies to identify OA-related joint pain. NHANES in the 1970's used the question: "Have you ever had pain in or around a knee on most days for at least a month?" (10); a second question was added in the 1990's: "Have you had (any) pain in or around your knee for at least a month in the last year?". The ACR used a modified version of the question as part of criteria to diagnose OA: "Have you had (knee/hip) pain on most days in the last month?".

A wide range of these types of questions, with a variety of wording, was found among the international cohorts containing OA (appendix 2). The differences between these questions occurs in two places: first, the amount of time reported with pain (i.e. any, most days in the last month) and second, the period of recall (i.e. in the last month, last year, ever). In order to simplify a comparison between questions, they were grouped into five types by the steering group, where both the amount of time with pain and the period of recall were as similar as possible (figure 1).

Figure 1

Expert Discussion

Of the five variations of NHANES-type questions identified in the cohorts (figure 1), the two most commonly used were: A) most days in the last month and C) at least a month in the last year. The committee agreed that questions A-D appeared similar enough to be combined, however, question E (pain for at least a month ever) was deemed to be too different to be combined and that it should be analysed as part of a sensitivity analysis if necessary. Previous research by O'Reilly et al (11) compared three different variations of NHANES-type questions and found that knee pain prevalence varied between 19.3% and 28.3% depending on the questions. Two of these questions were comparable to our NHANES A and C variations, with

their reported prevalence differing by six percentage points (11). These results showed that although overall agreement was good, the estimates of knee pain are influenced by even minor changes in the wording of the question.

The committee ultimately decided that not enough was known to make an informed decision and suggested original research into the topic before making a final decision. In order to provide the necessary evidence, the steering group therefore undertook an analysis of these NHANES-type questions using an OA-related cohort (Action A), which was then reviewed by the full expert committee.

Action A

The experts suggested that the Multicenter Osteoarthritis Study (MOST) was the best cohort to examine the relationship of OA-pain assessments as it contains multiple NHANES questions at the same time point. The MOST study is a US-based observational study of subjects with or at high risk for knee OA recruited in 2003 with a greater number of subjects with high BMI, family history of OA and/or knee pain (12). Participants at baseline answered four binary NHANES-type questions: A) Knee pain on most days in the last month; B) Any knee pain in the last month; C) Knee pain lasting at least a month in the last year; D) Any knee pain in the last year. Sensitivity, specificity and area under the curve (AUC) from ROC curves were used to compare NHANES-type questions. NHANES A was selected as the reference question due to its similarity to the pain assessment used as part of the ACR OA diagnostic criteria, it was one of the more commonly used pain questions in the OA cohort studies, and it has been previously been used as part of a gold-standard definition of SROA to test the performance of ACR criteria in the general population (13).

Out of 3026 subjects, 2922 had all required data at baseline (basic demographics and pain questions) and were used for the cross-sectional analysis. NHANES A and C showed a similar prevalence of pain (41.0% and 43.4%), while NHANES B and D both produced a substantially higher prevalence (67.3 and 75.4%). NHANES C (pain lasting at least a month in the last year) showed the best sensitivity (91.2%) and specificity (89.9%) against the reference NHANES A, with both NHANES B and D having very low specificity (55.5% and 41.7% respectively) (table 1).

Table 1

Decision

The results of the analysis requested by the experts showed that the comparability of questions was influenced more by the duration of reported pain (i.e. pain lasting at least a month) than the period of pain recall (i.e. in the last year). NHANES A was felt to be the best wording based upon the frequency that it is found in OA cohorts, its use as part of the ACR clinical criteria and that the amount of time and period of recall used to identify pain occurs concurrent with the radiographic information. NHANES C had the best sensitivity and specificity for NHANES A, and was therefore identified as the most appropriate option in the instance of using existing data, where NHANES A is not available.

3. To assess whether previously published thresholds used to determine pain using the WOMAC scale are appropriate for research and determine comparability with the NHANES-type pain questions

The WOMAC is commonly used in addition to, or instead of, NHANES-type questions in OA-related population-based cohorts. It was felt important to investigate whether the WOMAC index could be used as an alternative pain measure. The WOMAC index is a standardized set of questions developed to evaluate knee or hip pain, function and disability (14). WOMAC pain scores are used as continuous measure (range 0-20).

Expert Discussion

Experts agreed that a threshold for WOMAC was needed so that all cohorts could be included into the IPD meta-analysis. Several issues were identified when using a threshold with a WOMAC scale to be comparable to NHANES-type questions, including that only the pain sub-scale, would be equivalent and that the period of recall for pain was not given in early versions of WOMAC (pre 3.0). It was thought that previous research where thresholds had been used (15-17) were not appropriate for current population cohorts due to their development primarily in, and for, clinical outcomes in patient populations. The committee believed that a threshold should be developed specifically for combining the data with the NHANES-type questions and suggested further work before an ultimate decision was made (Action B).

Action B

The MOST study (see Action A for cohort description) was used for this analysis. In addition to the NHANES-type questions assessed at baseline, participants completed the WOMAC pain sub-scale (range 0-20) asking for pain during daily activity in the past 30 days. A cut-point was established for the WOMAC pain sub-scale against the reference question (NHANES A), at the point at which sensitivity and specificity were closest together. 95% confidence intervals (CI) around the cut-points were estimated using bootstrap methods with 300 repeats. The Osteoarthritis Initiative cohort (OAI), which has similar inclusion criteria to MOST and is also an enhanced risk factor population-based cohort, was used to validate the WOMAC threshold against the gold-standard question using identical inclusion/inclusion criteria and statistical methods. OAI used the WOMAC pain sub-scale asking for pain during daily activity in the past 7 days.

The WOMAC pain sub-scale had a median of 2 (IQR 0, 6), and a cut point of 3 was found using both NHANES A (3 (95% CI 2.1, 3.9)) and C (3 (95%CI 2.8, 3.2)). When this cut-point was used to create a binary pain variable from the WOMAC pain sub-scale, the sensitivity and specificity of this new variable against the NHANES A question was 83.6% and 76.0%, respectively (table 2). In the OAI validation cohort (n=4,723), the WOMAC pain sub-scale had a median of 1 (IQR 0, 4) and also generated a cut-point of 3 (95% CI 2.3, 3.7).

Table 2

Decision

Action B analysis demonstrated that a cut-point of 3 in the WOMAC pain sub-scale had the best sensitivity and specificity against the gold standard NHANES question ‘pain on most days in the previous month’. The same cut-point of greater than or equal to 3 was found in the OAI validation cohort. Experts agreed that this threshold could be applied in cohorts where only WOMAC pain data was available to generate the symptomatic radiographic OA variable.

4. To assess the comparability of methods used to grade radiographic OA and determine the ‘best’ measure to use based on available data

There are a number of scoring methods to semi-quantitatively assess radiographic OA. Two of the most used in population-based cohorts are the K/L (a global grade) and the OARSI atlas of individual features which records features such as joint space narrowing and osteophyte size for each joint location (18, 19). Neogi et al found that in a within person matched case-control study

that K/L grade had a higher association with knee pain than either osteophytes or joint space narrowing alone (20). Most of the cohorts in our consortium used a K/L grade, however there is known variation between different versions of the grade. Kerkhof et al (7) found that the actual definition of K/L grade 2+ significantly varied across cohorts which substantially affected OA prevalence. Experts were presented with the x-ray views and scoring methods used in each cohort in order to inform decision making on the most appropriate scoring method and thresholds for determining radiographic OA in *existing* cohort studies.

Expert Discussion and Decision

The committee felt that the K/L grade should be used as it was available in the majority of the cohorts, and they did not feel a 'computed' grade (calculated using individual features of osteophytes and joint space narrowing) would add any benefit above and beyond K/L. All experts agreed that using the established cut-off for radiographic OA, K/L greater than or equal to 2 was appropriate for this current research to define more advanced stages of OA, rather than an alternate cut-off or individual features. However, there was interest in exploring the use of K/L as an ordinal measure in future research if the grading was found to be comparable between cohorts. The committee felt that the inclusion of the patellofemoral compartment was extremely important and were disappointed that it could not be included in this research due to the lack of data. For future research, the inclusion of the patellofemoral compartment was identified as a key area of improvement, in addition to the use of a high quality standardised atlas (such as the OARSI atlas) to grade at least osteophytes and joint space narrowing as individual radiographic features (19).

5. To assess the usability and comparability of alternate OA outcomes: self-reported OA, GP diagnosis, and joint replacement

Community-based cohort studies where OA and/or musculoskeletal conditions are not the primary interest often lack NHANES/WOMAC pain assessment and radiographic OA information, but may include questions relating to self-reported OA or to total joint replacement surgery (TJR). The addition of these types of cohorts increases the number of subjects and often provides more detailed risk factors. Two common variations of this type of question relate to self-perceived arthritis: "Do you have (knee/hip) osteoarthritis?" and self-reported physician diagnosed OA: "Have you ever been told that you have OA of your knee (hip) by a doctor?" Although evidence is limited, there is a known lack of comparability between these two question variations. Szoeke et al (21) demonstrated that within the same cohort of patients, 63.7% reported self-perceived arthritis versus 48.7% self-

reported physician diagnosed OA. More encouragingly, self-reported clinician diagnosed OA (hip and knee) has been found to have high positive predictive value (98% and 91%) when compared with clinical OA, as defined by ACR criteria (22).

Expert Discussion and Decision

The expert committee felt the 'self-perceived' measure would be more problematic for hip OA than knee OA, and suspected there would be little correlation between self-perceived OA and TJR. Joint replacement is also limited by variability in healthcare access across different countries and societies, and region and time-dependent variable contribution of indications other than OA for TJR, such as rheumatoid arthritis, fracture, and osteonecrosis. The experts agreed that further research, in cohorts with both variables reported to allow comparisons, was required before making a final decision.

Strengths and limitations

This study has several strengths; it is the first to create a standardised definition of knee and hip OA for use in combining data from cohort studies, which is becoming increasingly important to answer important questions in OA. We have demonstrated the importance of the exact wording of NHANES type questions and further more generate an equivalent WOMAC score for populations where NHANES questions are not recorded. The use of a comprehensive collection of existing cohort data and inclusion of the study PIs in addition to international experts facilitated the decision making process.

It also has several potential limitations. The cohorts included in this analysis are a subset which meet the inclusion criteria and may not contain the full range of OA assessments found in existing longitudinal population-based OA cohort studies.

Furthermore, the generation of "NHANES equivalent scores" using WOMAC, may allow the incorporation of other cohorts, however for the purpose of this study it was important to capture those with both symptomatic and radiographic knee and/or hip OA data and we do not feel that inclusion of additional cohorts would affect the results of this paper. The group of "experts", although covering most important stakeholders, may not have been complete, however we feel that due to the wide experience of the group in similar committees and processes mean that it is unlikely that the addition of other stakeholders would have changed our results.

Summary and Recommendations

This international study is the first to describe methods to define and harmonise OA data for population-based cohort studies. Combining OA data allows for the application of novel research techniques, such as IPD meta-analysis in existing studies as well as informing data collection recommendations for future OA cohorts.

This research has highlighted the disparity of OA data in existing cohort studies, making comparisons between cohorts and interpretation of previous research difficult. The effect of using different radiographic atlases, questionnaires and even the wording of OA related pain questions are important considerations when comparing OA data.

Recommendations for combining extant OA data

- Use a combination of symptoms and radiographic features to define OA as a primary outcome, or by symptoms alone when radiographic data is lacking
- Where possible, use NHANES-type questions where duration of pain is indicated as 'most days in a month' (NHANES A and NHANES C), due to wide variation in pain prevalence which was found depending on the question wording
- If a WOMAC pain subscale (0-20) is available, rather than NHANES question, a cut point of 3 or more can be used to reasonably equate to NHANES A or C questions
- For defining radiographic OA, experts recommended the use of a K/L grade 2 and above,
- Caution is recommended when trying to combine self-reported GP OA diagnoses or self-perceived OA, as the relationship between these is unknown. Experts believe these variables may be very different from symptomatic radiographic OA, and therefore require further research

Recommendations for collecting new OA data in cohort studies

- Use multiple pain assessments (i.e. NHANES pain question, WOMAC, clinical assessment, etc.) at multiple time-points to provide better comparability with existing cohorts and to use as outcome measures
- Include self-reported/GP-diagnosed OA and pain questions

- Use additional x-ray views (i.e. the patello-femoral compartment) to improve diagnosis of radiographic knee OA
- Record individual radiographic features (i.e. using OARSI atlas of individual features) in addition to K/L grades
- Wording of pain questions should be consistent for the duration of pain asked. 'Most days of the month' is the most commonly used wording in existing cohort studies.

Author contributions

KL, LG and NA were involved in the conception and design of the study. KL, LG and MN were involved in the acquisition and management of the data. KL, LG, MS, AJ, MN and NA were involved in the statistical analysis and interpretation of the data. KL, LG and NA drafted the manuscript. All authors reviewed the manuscript with critical revision of the article for important intellectual content and approved the final manuscript. KL, LG and NA took the responsibility for the integrity of the work as a whole, from inception to finished article.

Potential conflicts of interests:

JN, MEB, MTSS, NKA, LSG and KMLs institution received a grant from Arthritis Research UK Centre of Excellence Grant. EMR is deputy editor of Osteoarthritis and Cartilage, the developer of Knee injury and Osteoarthritis Outcome Score (KOOS) and several other freely available patient-reported outcome measures and founder of the Good Life with Osteoarthritis in Denmark (GLA:D), a not-for profit initiative to implement clinical guidelines in primary care. AJ has received consultancy, lecture fees and honoraria for unrelated work from Servier, UK Renal Registry, IDIAP Jordi GOI and Freshfields Bruckhaus Deringer and consortium research grants from Roche. CC has received consultancy fees for unrelated work from Alliance For Better Bone Health, Amgen, Eli Lilly, GSK, Medtronic, Merck, Novartis, Pfizer, Roche, Servier, Takeda and UCB. DF is supported by the National Institute of Health. GJ has received consultancy and lecture fees for unrelated work from multiple pharmacology companies. JMJ has received consultancy fees for unrelated work from Trinity partners, Samumed and Flexion, a grant from Johnson & Johnson and honorarium as Deputy Editor, Clinical Science for Society's Journal, Osteoarthritis and Cartilage. DJH has received consultancy fees for unrelated work from Merck Serono and Flexion. LEs institution received grants by Norwegian NIH, Helse Sør Øst, IOC, FIFA, Norwegian Lottery, Department of Culture, Norway, consultancy, lecture fees and/or royalties for unrelated work from Arthrex, Aspetar, Smith and Nephew and grants from Smith and Nephew, Biomet and Arthrex. MH has

received grants from National Institute of Health, consultancy fees for unrelated work from Bioiberica SA, IBSA SA, Novartis Pharma AG, Pfizer, Plexxikon, Proximagen, Theralogix LLC and EMD Serono and royalties from Elsevier. MNs institution has received grants from National Institute of Health. PGC is supported in part by the National Institute for Health Research (NIHR) Leeds Musculoskeletal Biomedical Research Unit. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health. NKA has received consultancy fees for unrelated work from Bioventus, Merck, Smith & Nephew, Flexion, Freshfields and Nicox and grants from Bioiberica. NYs institution has received Grants-in-Aid funding from the Ministry of Health, Labour and Welfare, Japan. SBZ has received board membership fees for Associate editorship from Osteoarthritis and Cartilage, consultancy fees for unrelated work from Regeneron, Infirst healthcare, has grants pending with the Dutch Arthritis Foundation, at the Netherlands Organisation for Health research and development, and EU Horizon 2020 and has received grants from the Dutch Arthritis Foundation, Netherlands organisation for Health research and development, Nuts-Ohra, and EU Fp7. SL has received consultancy fees for unrelated work from Galapagos NV, Flexion, Regeneron, Össur, Samumed and Johnson & Johnson. All other authors declare that they have no conflict of interest.

Funding:

The study was funded by the Arthritis Research UK Centre for Sport, Exercise and Osteoarthritis (Grant reference 20194) and was further supported by the Pre-Competitive Consortium for Osteoarthritis, Osteoarthritis Research Society International.

Acknowledgments:

The study was made possible by the contribution of many people, including the advisory board (A Judge and M Sanchez-Santos) and the expert committee (M Nevitt, D Felson, S Bierma-Zeinstra, P Conaghan, L Engebretsen, M Hochberg, D Hunter, G Jones, J Jordan, S Lohmander, E Roos, N Yoshimura, J van Meurs and C Cooper). We gratefully thank S Sheard for early contribution as study coordinator and K Ambrose for data advisory contributions. We would also like to express our gratitude to the participants of MOST and OAI.

References

1. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012;380(9859):2197-223. doi: 10.1016/S0140-6736(12)61689-4.

2. Tierney JF, Vale C, Riley R, Smith CT, Stewart L, Clarke M, et al. Individual Participant Data (IPD) Meta-analyses of Randomised Controlled Trials: Guidance on Their Use. PLoS Med. 2015;12(7):e1001855.
3. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340:c221.(doi):10.1136/bmj.c221.
4. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. JAMA. 2015;313(16):1657-65. doi: 10.001/jama.2015.3656.
5. Debray TP, Moons KG, van Valkenhoef G, Efthimiou O, Hummel N, Groenwold RH, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res Synth Methods. 2015;6(4):293-309. doi: 10.1002/jrsm.160. Epub 2015 Aug 19.
6. Felson DT, McAlindon TE, Anderson JJ, Naimark A, Weissman BW, Aliabadi P, et al. Defining radiographic osteoarthritis for the whole knee. Osteoarthritis Cartilage. 1997;5(4):241-50.
7. Kerkhof HJM, Meulenbelt I, Akune T, Arden NK, Aromaa A, Bierma-Zeinstra SMA, et al. Recommendations for standardization and phenotype definitions in genetic studies of osteoarthritis: the TREAT-OA consortium. Osteoarthritis and Cartilage. 2011;19(3):254-64.
8. Luyten FP, Denti M, Filardo G, Kon E, Engebretsen L. Definition and classification of early osteoarthritis of the knee. Knee Surg Sports Traumatol Arthrosc. 2012;20(3):401-6. doi: 10.1007/s00167-011-1743-2. Epub 2011 Nov 8.
9. Madry H, Kon E, Condello V, Peretti GM, Steinwachs M, Seil R, et al. Early osteoarthritis of the knee. Knee Surg Sports Traumatol Arthrosc. 2016;24(6):1753-62. doi: 10.007/s00167-016-4068-3. Epub 2016 Mar 21.
10. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey II; National Center for Health Statistics. Hyattsville MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention: http://cdc.gov/nchs/nhanes/nhanesii.htm
[
11. O'Reilly SC, Muir KR, M. D. Screening for pain in knee osteoarthritis: which question? Annals of the Rheumatic Diseases. 1996;55:931-3.
12. Segal NA, Nevitt MC, Gross KD, Hietpas J, Glass NA, Lewis CE, et al. The Multicenter Osteoarthritis Study: opportunities for rehabilitation research. PM & R : the journal of injury, function, and rehabilitation. 2013;5(8):647-54.
13. Peat G, Thomas E, Duncan R, Wood L, Hay E, Croft P. Clinical classification criteria for knee osteoarthritis: performance in the general population and primary care. Annals of the Rheumatic Diseases. 2006;65(10):1363-7.
14. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol. 1988;15(12):1833-40.
15. Tubach F, Ravaud P, Baron G, Falissard B, Logeart I, Bellamy N, et al. Evaluation of clinically relevant states in patient reported outcomes in knee and hip osteoarthritis: the patient acceptable symptom state. Annals of the Rheumatic Diseases. 2005;64(1):34-7.
16. Goggins J, Baker K, Felson D. What WOMAC pain score should make a patient eligible for a trial in knee osteoarthritis? The Journal of Rheumatology. 2005;32(3):540-2.
17. Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R, et al. Differences between Men and Women in the Rate of Use of Hip and Knee Arthroplasty. New England Journal of Medicine. 2000;342(14):1016-22.
18. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. Ann Rheum Dis. 1957;16(4):494-502.
19. Altman RD, Gold GE. Atlas of individual radiographic features in osteoarthritis, revised. Osteoarthritis Cartilage. 2007;15 Suppl A:A1-56.

20. 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.
21. Szoeke CEI, Dennerstein L, Wluka AE, Guthrie JR, Taffe J, Clark MS, et al. Physician diagnosed arthritis, reported arthritis and radiological non-axial osteoarthritis. Osteoarthritis and Cartilage. 2008;16(7):846-50.
22. Ratzlaff C, Koehoorn M, Cibere J, Kopec J. Clinical validation of an Internet-based questionnaire for ascertaining hip and knee osteoarthritis. Osteoarthritis and Cartilage. 2012;20(12):1568-73.

Appendix 1. Summary of the cohorts included within consensus study and potential OA variables identified within each

Cohort	**Self reported clinician diagnosed**	**Self perceived OA**	**TJR**	**Knee x-ray**	**NHANES- type questions**					**WOMAC**
					1	***2***	***3***	***4***	***5***	
OAI	✓		✓	✓		✓	✓			✓
MOST	✓		✓	✓	✓	✓	✓	✓		✓
SOF	✓		✓	✓					✓	✓
ROAD			✓	✓			✓			✓
Herts	✓		✓	✓	✓					
Johnston County	✓		✓	✓	✓					
TasOAC	✓		✓	✓						✓
Chingford	✓	✓ (hip only)	✓ (hip)	✓	✓	✓				
Framingham	✓			✓			✓			

Appendix 2. Wording variations of the binary NHANES-type pain questions found within the MILOS consortium cohorts

NHANES-Type Questions
"Pain, aching or stiffness in or around the knee most days" for at least 1 month of the past 12 months.
" [Any] Pain, aching, stiffness in (left/right)knee in past 12 months?"
"Pain, aching, stiffness in (right/left) knee on most days for more than 1 month in the last 12 months?"
"Pain, aching, stiffness on most days in the last month?"
NHANES I questionnaire "Have you ever had pain in or around your knee on most days for at least a month?"
"(Left/Right) Knee pain lasting at least a month during last 12 months"
"Knee pain lasting at least one month in the current or previous year"
"Number of months with knee pain for each year in the past 12 years since baseline visit"
"Have you had pain in or around your (left/right) knee on most days in the last month?"
"On most days do you have pain, aching or stiffness in your KNEES?"
"Have you had pain on most days of the last month?"
"Have you ever had pain in your knees for more than one month?"
"Have you had (any) knee pain within the last month?"
"Did you have [any] (knee/hip, R/L) pain in the last month?" "If yes, on how many days (0-5, 5-15, 15+)"
"Ever pain lasting at least one month (in previous 2 years)"

1

Table 1. Comparison of NHANES-type pain questions within the MOST cohort

	Prevalence (N)	Sensitivity	Specificity	AUC (95% CI)
NHANES A	41.0% (1198)	*Reference*	*Reference*	*Reference*
NHANES B	67.3% (1966)	100.0%	55.5%	0.78 (0.77, 0.79)
NHANES C	43.4% (1267)	91.2%	89.9%	0.91 (0.90, 0.92)
NHANES D	75.4% (2203)	100.0%	41.7%	0.71 (0.70, 0.72)

NHANES A “Knee pain on most days in the last month” NHANES B “Any knee pain in the last month”; NHANES C “Knee pain lasting at least a month in the last year”; D “Any knee pain in the last year”

Table 2. WOMAC thresholds (0-20 scale with 20 reflecting severe pain), and prevalence, sensitivity, and specificity after applying thresholds

	Cut point (Against NHANES A)	Applying a cut point of 3 (Tested against NHANES A)			
		Prevalence (N)	**Sensitivity**	**Specificity**	**AUC (95% CI)**
MOST	3 (95% CI 2.1, 3.9)	48.4% (1415/2922)	83.6%	76.0%	0.80 (0.78, 0.81)
OAI	3 (95% CI 2.3, 3.7)	35.9% (1695/4723)	70.7%	79.7%	0.75 (0.74, 0.77)

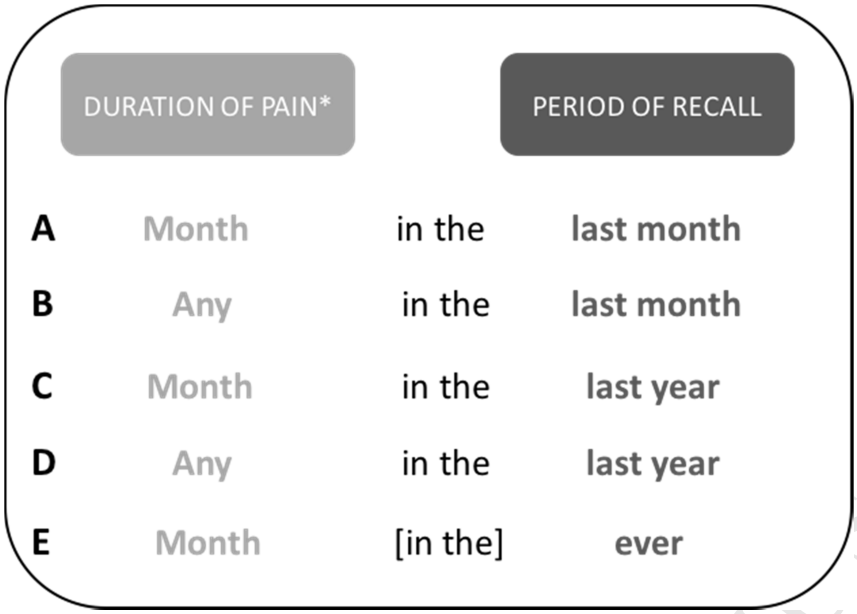


Figure 1. NHANES questions grouped into similar duration of pain and periods of recall
*****'Month' can represent the following: 'most days of a month', 'at least a month' or 'more than a month'