**Osteoarthritis
and Cartilage**

OARSI OSTEOARTHRITIS RESEARCH SOCIETY INTERNATIONAL

Review

Association between clinical findings and the presence of lumbar spine osteoarthritis imaging features: A systematic review

Mirna Chamoro [†] [*], Katie de Luca [‡], Omer Ozbulut [†], Edwin H.G. Oei [||],
Carmen L.A. Vleggeert-Lankamp [¶] [††], Bart W. Koes [†] [‡‡], Sita M.A. Bierma- Zeinstra [†] [§],
Alessandro Chiarotto [†]

† Department of General Practice, Erasmus MC, University Medical Center, Rotterdam, the Netherlands
‡ Discipline of Chiropractic, CQ University, Brisbane, Australia
§ Department of Orthopedics, Erasmus MC, University Medical Center, Rotterdam, the Netherlands
|| Department of Radiology & Nuclear Medicine, Erasmus MC, University Medical Center, Rotterdam, the Netherlands
¶ Department of Neurosurgery, Leiden University Medical Center, Leiden, the Netherlands
†† Department of Neurosurgery, Spaarne Hospital, Haarlem/Hoofddorp, the Netherlands
‡‡ Center for Muscle and Joint Health, University of Southern Denmark, Odense, Denmark

ARTICLE INFO

SUMMARY

Objective: Spinal osteoarthritis is difficult to study and diagnose, partly due to the lack of agreed diagnostic criteria. This systematic review aims to give an overview of the associations between clinical and imaging findings suggestive of spinal osteoarthritis in patients with low back pain to make a step towards agreed diagnostic criteria.
Design: We searched MEDLINE, Embase, Web of Science, and CINAHL from inception to April 29, 2021 to identify observational studies in adults that assessed the association between selected clinical and imaging findings suggestive of spinal osteoarthritis. Risk of bias was assessed using the Newcastle Ottawa Scale and the quality of evidence was graded using an adaptation of the GRADE approach.
Results: After screening 7902 studies, 30 met the inclusion criteria. High-quality evidence was found for the longitudinal association between low back pain (LBP) intensity, and both disc space narrowing and osteophytes, as well as for the association between LBP-related physical functioning and lumbar disc degeneration, the presence of spinal morning stiffness and disc space narrowing and for the lack of association between physical functioning and Schmorl's nodes.
Conclusions: There is high- and moderate-quality evidence of associations between clinical and imaging findings suggestive of spinal osteoarthritis. However, the majority of the studied outcomes had low or very low-quality of evidence. Furthermore, clinical and methodological heterogeneity was a serious limitation, adding to the need and importance of agreed criteria for spinal osteoarthritis, which should be the scope of future research.

Introduction

Low back pain (LBP) represents the greatest cause of disability globally and is a substantial health care and societal burden.[1,2] In general practice in the Netherlands, the large majority of patients with LBP are labeled as having 'nonspecific' LBP, and around 300,000 patients are given the diagnostic code reflective of spinal osteoarthritis.[3] Nonspecific LBP is diagnosed by exclusion of specific underlying causes, usually by means of history taking and physical examination.[4] However, within the group of patients with nonspecific LBP, there may be subgroups of patients, for example, those with symptomatic spinal osteoarthritis, who

* Corresponding author. Department of General Practice, Erasmus MC, University Medical Center, PO Box 2040, 3000 CA, Rotterdam, the Netherlands.
E-mail addresses: m.chamoro@erasmusmc.nl (M. Chamoro),
k.deluca@cqu.edu.au (K. de Luca), omerfarukozbulut@gmail.com (O. Ozbulut),
e.oei@erasmusmc.nl (E.H.G. Oei),
C.L.A.Vleggeert-Lankamp@lumc.nl (C.L.A. Vleggeert-Lankamp),
b.koes@erasmusmc.nl (B.W. Koes),
s.bierma-zeinstra@erasmusmc.nl (S.M.A. Bierma- Zeinstra),
a.chiarotto@erasmusmc.nl (A. Chiarotto).

are now unrecognized or misdiagnosed. In contrast with peripheral joints, such as the knee and hip, there are no diagnostic criteria for spinal osteoarthritis. This makes it unclear how and when LBP may be reflective of symptomatic spinal osteoarthritis.

To better understand and manage spinal osteoarthritis, the first, fundamental step is to establish agreed and empirically based criteria for structural and symptomatic spinal osteoarthritis to be used in research and clinical practice. In 2021, a Delphi study was conducted to reach an international consensus on statements regarding definitions of spinal osteoarthritis.[5] A high proportion of agreement was found for the statement "separate definitions for structural and symptomatic spinal osteoarthritis are needed" for research and clinical practice (94% and 88% respectively). In addition, a consensus was reached on the following clinical features as suggestive of symptomatic spinal osteoarthritis: spinal pain duration, spinal pain intensity, limitations in physical functioning, and self-reported spinal morning stiffness. Statements for which there was a lack of consensus included a limited or painful range of motion, and the statement that reached a consensus against inclusion in a definition of spinal osteoarthritis was palpable warmth. For research purposes, there was 95% agreement that an internationally recognized definition of symptomatic spinal osteoarthritis will aid the investigation of the prognosis and treatment of patients.

Three systematic reviews have investigated the association between clinical findings and spinal osteoarthritis features on imaging but only evaluated structural changes on radiographs,[6] focused on younger patients[7] or only involved Modic changes.[8] Up to date, no systematic review has comprehensively investigated the association of clinical features suggestive of symptomatic spinal osteoarthritis, with imaging features on lumbar radiographs, Magnetic Resonance Imaging (MRI), or Computer Tomography (CT)-scan that are potential indicators of structural spinal osteoarthritis. Hence, this systematic review aimed to determine the association present between clinical findings and imaging features that are potential indicators of lumbar spine osteoarthritis in patients with LBP. The clinical findings were selected among those for which there was consensus or uncertainty in the recent aforementioned Delphi study,[5] while the presence of imaging features possibly associated with lumbar spine osteoarthritis was investigated using different imaging modalities (i.e., radiography, MRI, CT-scan).

Methods

A research protocol was developed in advance and registered in the international register of systematic reviews, PROSPERO: https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42021253283. For reporting this systematic review we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses updated guidelines (see Appendix A).[9]

Search strategy

A systematic search was conducted in MEDLINE, Embase, Web of Science, and CINAHL from inception to April 29, 2021. Key terms referring to clinical features of symptomatic spinal osteoarthritis on which consensus has been reached (i.e., duration and intensity of LBP, LBP-related physical functioning limitations, spinal morning stiffness, and limited or painful range of motion) were combined with key terms referring to (possible) structural spinal osteoarthritis imaging features (i.e. disc degeneration, disc space narrowing, osteophytes, Modic changes, endplate lesions, Schmorl's nodes, annular tears, nucleus pulposus signal intensity, and facet joint abnormalities) on MRI, CT-scan, and radiographs. The search strategy

was developed in collaboration with a research librarian at Erasmus MC. A combination of MESH terms and free text terms were used in the search. No language restrictions were applied to the search. The full search strategy can be found in Appendix B. The studies included in previous systematic reviews by Raastad et al.,[6] Brinjikji et al.,[7] and Herlin et al.[8] were screened for eligibility, and forward citation tracking of eligible studies was performed in Google Scholar.

Selection process

Observational studies (i.e., prospective and retrospective cohort studies, case-control, and cross-sectional studies) in adults (≥ 18 years) were included if they evaluated the presence of structural imaging features possibly associated with spinal osteoarthritis and their association with one or more clinical features suggestive of spinal osteoarthritis. Clinical features included one or more of the following five: LBP duration, LBP intensity measured with patient-reported measurement instruments (e.g., Numeric Rating scale), physical functioning limitations measured with patient-reported measurement instruments (e.g., Oswestry Disability Index (ODI)), presence and/or duration of spinal morning stiffness and painful or limited range of motion. The included structural imaging features possibly associated with spinal osteoarthritis were: facet joint degeneration (joint space narrowing, subchondral bone sclerosis, osteophyte formation), vertebral endplate pathology (mineralization of cartilaginous endplate, edema, subchondral bone sclerosis, Modic changes), osteophyte formation at the margin of the intervertebral disc space, loss of intervertebral disc height or degeneration of the intervertebral disc (nucleus pulposus (black disc) and annulus fibrosis); with one or more imaging techniques (i.e., MRI, CT, conventional radiographs). Studies were excluded if they were reviews, case reports, case series, letters to the editor, or clinical trials. Studies involving adolescents (< 18 years) and/or focusing on LBP due to other (serious) pathologies, such as malignancy, axial spondyloarthritis, (spondylo)discitis, vascular causes of LBP including spinal cord infarction, recent trauma/fracture, radiotherapy in the lumbar region, spinal surgery and juvenile/idiopathic scoliosis, as well as lumbosacral radicular syndrome (sciatica), were also excluded.

Identification of eligible studies

Titles and abstracts from the literature search were screened for eligibility by two independent reviewers (M.C. and O.O.). Subsequently, full-text articles were screened by the same two independent reviewers, as well as the full-text articles which were retrieved through forward citation tracking. Studies were included if they were determined eligible by both reviewers. Any disagreement in the selection was first solved through consensus between the two reviewers. If consensus could not be reached, a third reviewer (A.C.) arbitrated the decision.

Data extraction and quality assessment

Data extraction was conducted using a standard form that was a-priori-designed (Appendix C). Study characteristics and demographics, investigated features, and reported measures of association, including their 95% confidence intervals and P-values, were extracted. In line with the Cochrane Handbook,[10] data on outcomes were extracted by two independent reviewers (M.C. and K.D.). The study characteristics were extracted by one reviewer (M.C.) and checked for accuracy by a second reviewer (K.D.). Disagreements in the data extraction process were resolved by consensus between the two reviewers.

The methodological quality of the studies was assessed by two reviewers (M.C. and K.D.) using the Newcastle-Ottawa scale, a validated tool for assessing the quality of non-randomized studies in systematic reviews and meta-analyses.[11] The Newcastle-Ottawa Scale uses a 'star system' to judge studies on 9 items concerning the selection and comparability of the study groups and ascertainment of either the exposure or outcome of interest, for case-control or cohort studies respectively. For the cross-sectional studies we used a modified version of the NOS consisting of 7 items concerning the same study aspects, as proposed by Patra et al.[12] In case of disagreement between the reviewers, a third reviewer (A.C.) arbitrated the decision. The quality of evidence was rated using an adaptation of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, which is a systematic approach to define the quality of a body of evidence in systematic reviews.[13,14] Since our aim was to identify associations and observational studies are appropriate for this purpose, we decided to start at a 'high' quality of evidence for all outcomes. The quality of evidence was downgraded, if necessary, for the following domains: study limitations, inconsistency, indirectness, and imprecision (see Box 1). Since we were not able to conduct a meta-analysis, publication bias was not quantitatively assessed. The presence of a dose-response gradient may increase the confidence in the findings of observational studies and thereby increase the quality of evidence.[13] Therefore, we included this as an upgrading domain.

Statistical analysis

First, clinical and methodological homogeneity was evaluated by comparing study characteristics, clinical features and measurement tools, imaging features, the manner in which they were graded, and the reported effect measures. We intended to perform a meta-analysis to produce pooled estimates, as stated in our protocol. However, since this was not possible due to the presence of substantial heterogeneity in the studied features and reported outcome measures, an approach following the Synthesis Without Meta-analysis guidelines was used.[15] We also investigated if the included studies assessed a dose-response relationship between the previously mentioned clinical features and imaging features possibly associated with spinal osteoarthritis.

Results

Literature search and selection of studies

Our search identified 13,807 studies, of which 7902 remained after removing duplicates. After the screening of titles and abstracts, 196 remained. Studies included in previous related systematic reviews were screened for eligibility. This resulted in an additional 26 studies, making a total of 222 full texts that were assessed for eligibility. From these, 30 studies were selected for inclusion in this systematic review. The flowchart (Fig. 1) presents the study selection process.

Study characteristics

Table 1 presents the characteristics of the included studies. Five studies were longitudinal cohort studies, two studies were case-control, and 23 studies were cross-sectional studies. Fifteen studies were population-based, two studies were from primary care, 10 studies were from secondary care, one study was from tertiary care and two studies were occupational studies. The mean age in the included studies ranged from 21.2 to 73.6 years. The included studies reported on the five clinical features of interest and on 10 structural imaging features possibly associated with spinal osteoarthritis (i.e., (intervertebral) disc degeneration, disc space narrowing, intervertebral disc height, osteophytes, Schmorl's nodes, annular tear, Modic changes, endplate abnormalities, nucleus pulposus signal intensity, and facet joint abnormalities). Twelve studies evaluated imaging features on lumbar radiographs, while the remaining 18 studies assessed the features on MRI.

All the following results, including the results on association between clinical features and imaging findings, and the GRADE score, are presented in Table 2. Our considerations for downgrading or upgrading the GRADE score are presented in Appendix D. Tables 3 and 4.

Cohort studies
Low back pain duration. There was moderate-quality evidence for the lack of association between LBP duration and Modic changes, in a baseline analysis of a longitudinal study.[16] This was based on one study.

Box 1
Upgrading and downgrading criteria for the quality of evidence (GRADE-framework).

Downgrading domains:	Criteria for downgrading:
Study limitations	– No downgrading if: NOS ≥ 7 – By 1 level if: NOS 4–6 for 1–2 studies (depending on the total of studies) – By 2 levels if: NOS ≤ 3 for 1–2 studies (depending on the total of studies)
Inconsistency	– No downgrading if: ≥75% of the studies show similar results – By 1 level if: 50–75% of the studies show similar results – By 2 levels if: < 50% of the studies show similar results
Indirectness	– By 1 level if: differences in a population (applicability) – By 2 levels if: differences in population and in outcome measures (surrogate outcomes)
Imprecision	– No downgrading if total sample size N ≥ 400 – By 1 level if total sample size N < 400 – By 2 levels if total sample size N < 400 and if the 95% CI overlaps no effect (i.e., CI includes OR of 1.0) or if: total sample size < 50
Upgrading domain: Dose-response gradient	Criteria for upgrading: -By 1 level if there is a dose-response gradient present

Box 1

Upgrading and downgrading criteria for the quality of evidence (GRADE-framework).

Osteoarthritis and Cartilage

Low back pain intensity. High-quality evidence was found for the longitudinal association between LBP intensity, and both disc space narrowing and osteophytes, based on one study by van den Berg et al.[17] Inconsistent results were found regarding the longitudinal relationship between LBP intensity and Modic changes, based on low-quality evidence. One study, with a high risk of bias, reported a significant, positive association,[18] whilst the other, with a low risk of bias, reported no clear association.[16] Luoma et al. (very low-quality evidence) reported that an increase of bony endplate lesions, Schmorl's nodes and the decrease of disc height are associated to an increase or persistence of pain and that degenerated discs with increased nucleus pulposus signals were more common among people who had persisting or increasing pain.[18]

LBP-related physical functioning limitation. High-quality evidence was found for the association between LBP-related physical
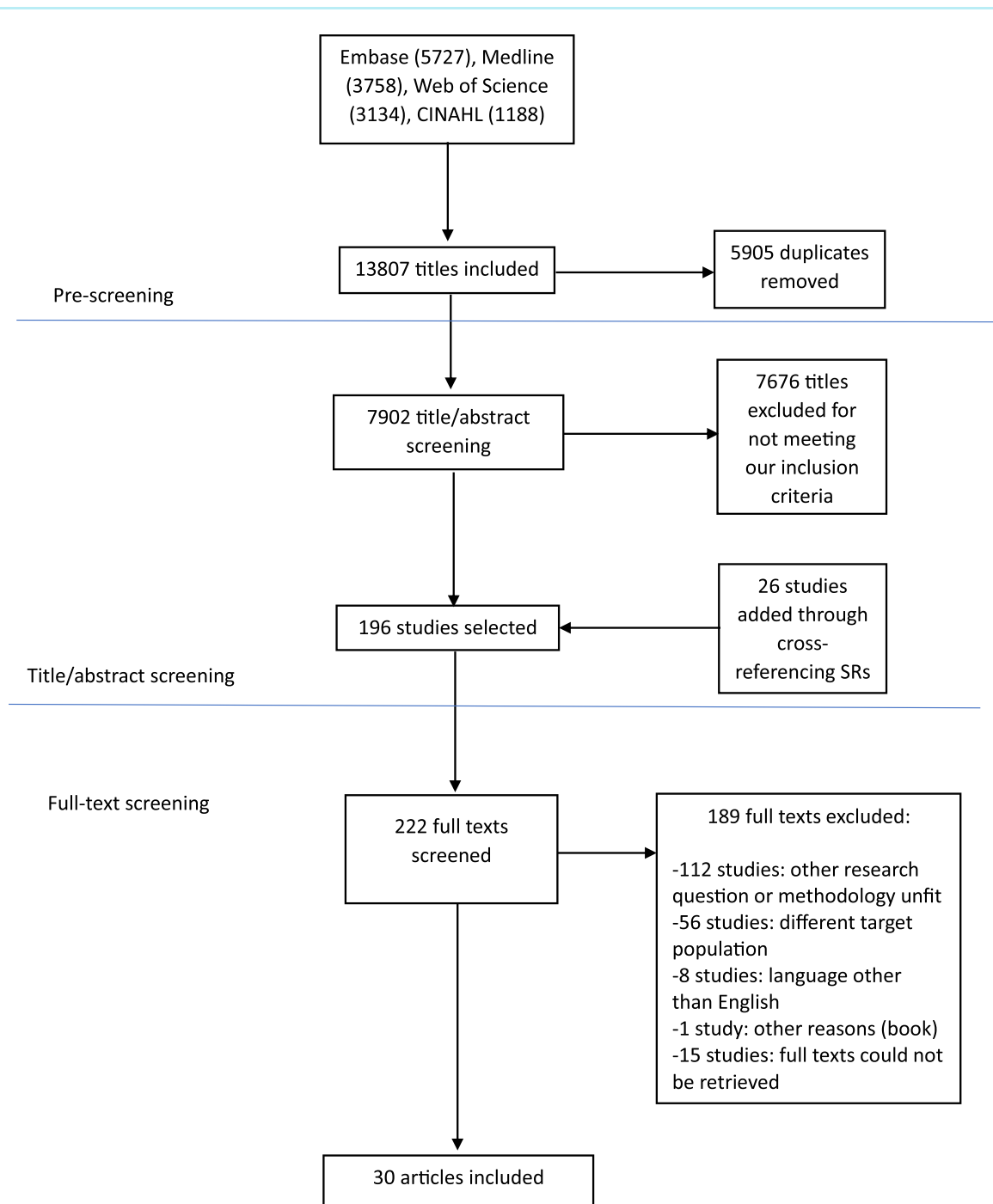

Fig. 1

Flowchart selection process.

Osteoarthritis and Cartilage

Study	Study population	LBP definition	Clinical features	Imaging modality	Imaging features
Cohort studies					
Keller et al.[20]	Rehabilitation center, Norway, FU 1 year, N = 269, N female = 153 (50.2%), mean age 49.7 years	Questionnaire: Chronic non-specific LBP	LBP-related physical function limitation was measured using the validated Norwegian ODI	Lumbar MRI	MC was reported as type I or type II
Luoma et al.[18]	Hospital population, Finland, FU 1 year, N = 49, N female = 42 (86%), mean age 43.7 years	Questionnaire: Chronic LBP lasting ≥3 months	LBP intensity was measured at baseline and follow-up using the Numeric Rating Scale (NRS) with a range from 0 to 10, where 0 = no pain and 10 = worst pain imaginable LBP-related physical function limitation was measured using the ODI	Lumbar MRI	All measurements were done at baseline and 1-year follow-up (changes: no (abnormal) finding, disappearance, decrease, persistence, increase (or enlargement), or appearance of a (new) finding) END were classified as focal subchondral hypointensity, small defect or larger Schmorl lesion-like bony defect, multifocal lesions, diffuse irregularity, and combined MC presence and size were measured NPS was classified on a 5-point scale IDH was measured
Määttä et al.[19]	General population (twins), UK, FU 10 years, N = 823 at baseline, N = 429 at FU, N female = 790 (96%), mean age at baseline was 54.0 years, 64.1 years at FU	Questionnaire: Severe and disabling LBP lasting more than 1 month at any point in total lifetime	LBP-related physical function limitation was determined as an inability to perform 1 or more following activities: walking around the house, standing for 15 min, getting up from a low chair, getting out of the bath, getting in and out of a car, going up and down the stairs, putting on socks /tights, and cutting toenails	Lumbar MRI	LDD (evaluated by loss of intervertebral disc space and NPS, presence of disc bulge and anterior OST)MC was reported as present or absent SN were reported as present or absent
Udby et al.[16]	Multidisciplinary non-surgical Back Center, Denmark, FU 13 years, N = 204 at baseline, N = 170 at FU, N female = 92 (54%), mean age 53.3 years	Questionnaire: A mean LBP score ≥4/10 for the last 14 days & pain for a minimum of four out of the past 12 months	LBP duration was measured, a method not reportedLBP intensity was measured at baseline and follow-up using the NRS (0–10) LBP-related physical function limitation was measured using the RMDQ and LBP Rating Scale for activity limitation	Lumbar MRI	MC was reported as present or absent
Van den Berg et al. (2022)	General population, Netherlands, FU 1 year, N = 543, N female = 320 (59%), mean age 67years	Questionnaire: Consulting a GP with a new episode of LBP Persisting LBP defined NRS≥1 after 1 year	LBP intensity was measured using the NRS (0–10)	Lumbar radiographs	LDD features OST and DSN were graded using the Lane atlas: a scale from 0 to 3, in which 0 = none, 1 = mild, 2 = moderate, 3 = severe
Case-control studies					
Hicks et al.[21]	General population (≥65years), USA, N = 321, N cases = 162, N controls=158, N female cases=80 (49%), N female controls = 66 (42%), mean age cases was 73.5 years, mean age controls was 73.6 years	Chronic LBP of at least moderate intensity occurring daily or almost every day for at least the previous 3 months	LBP intensity was scored using the McGill Pain Questionnaire-Short Form (MPQ-SF)	Lumbar radiograph	FJA was graded as: 0 = normal; 1 =DSN and/or mild eburnation; 2 = moderate narrowing + moderate eburnation or hypertrophy; 3 = severe OA with narrowing, eburnation, and OST LDD was graded as: 0 = no disease (normal Disc height, no spur formation, no ebur- nation, and no gas); 1 = mild disease (<25% DSN, small spur formation, minimal ebur- nation, and no gas); 2 = moderate disease (25–75% DSN, moderate spur formation, moderate eburnation, and no gas); 3 =advanced disease (> 75% DSN, large spur formation, marked eburnation, gas)
Quack et al.[22]	Population of office workers and nurses, Europe, N = 112, N female = 112 (100%), mean age was 53 years	Questionnaire: LBP for more than 7 days during the last year	Limited or painful range of motion was assessed by a trained physio-therapist using the modified Schober test, the fingertip-to-floor distance in the sagittal plane, and the lateral bending	Lumbar MRI	LDD, END, and FJA were all assessed using the Weishaupt grading system: 0 =normal; 1 =narrowing of the facet joint space < 2 mm, small OST and/or mild hypertrophy of the articular process; 2 = narrowing of the facet joint space, moderate OST, moderate hypertrophy of the articular process, and/ or mild subarticular bone erosion; and 3 = narrowing of the facet joint space, large OST, severe hypertrophy of the articular process, severe subarticular bone erosion, and/or subchondral cysts

6   *M. Chamoro et al. / Osteoarthritis and Cartilage xxx (xxxx) xxx–xxx*

Table 1 (*continued*)

Study	Study population	LBP definition	Clinical features	Imaging modality	Imaging features
Cross-sectional studies					
Ashraf et al.[42]	Rehabilitation center, Iran, N = 150, N female = 119 (79%), mean age 42.1 years	Interview by expert physiatrist: Presence of mechanical LBP > 3 months	LBP-related physical function limitation was measured with the Persian version of the Oswestry Disability Index, 0–20 = minimal disability, 21–40 = moderate disability, 41–60 = severe disability, 60–80 = crippled, 81–100 = bedridden	Lumbar radiograph	OA graded using the Kellgren and Lawrence scale: 0 = none; 1 = doubtful DSN and possible OST lipping; 2 = definite OST and possible DSN; 3 = moderate multiple OST, definite DSN, some sclerosis and possible deformity of bone contour; 4 = large OST, marked DSN, severe sclerosis and definite deformity of bone contour
Berg et al.[33]	Hospital population, Norway, N = 170, N female = 88 (52%), mean age 41 years	Clinical measure: Presence of chronic non-radicular LBP as a main symptom for > 1 year	LBP intensity was reported as the maximal current pain intensity in the last week on a visual analog scale (VAS), range from 0 to 100, where 0 = no pain and 100 = worst pain imaginable physical function limitation was measured using the validated Norwegian version of the ODI	Lumbar MRI	MC grades as: non, type I or type II HIZ graded as: present/not present NPS graded as: 1 = bright, 2 = gray, 3 = dark, 4 = black IDH decrease: <40% or >40% compared to nearest normal vertebra FJA graded as 0 = normal, 1 = mild (DSN or mild OST, 2 = moderate (sclerosis or moderate OST), 3 = severe (marked OST) Combined MRI-score: all features above combined, 0–5 points could be assigned at each level, MRI total score range 0 = 10
De Schepper et al. (2013)	General population, Netherlands; N = 2819, N female = 1615 (57%), mean age = 65.6 years	Interviews during home visits: LBP present if yes to: "Did you have complaints of the low back during the last month?"Chronic LBP if lasting > 1 year	LBP-related physical function limitation was graded using the Stanford Health Assessment Questionnaire (HAQ): A mean score 0.5 indicated moderate to severe disability	Lumbar radiographs	Both DSN and OST were graded using the Lane atlas as mentioned above (Van den Berg et al, 2022) DSN was defined as disc space narrowing (grade≥1) at 2 or more vertebral levels OST was graded as present (grade ≥2) at 2 or more vertebral levels
Frymoyer et al.[37]	Primary care, USA, N = 292, N female = 0 (0%), only men included mean age 32–32.6 years	Questionnaire: No LBP, moderate LBP ('mild', 'dis-comforting', 'distressing') or severe LBP ('horrible' or 'excruciating')	LBP intensity was measured using a Likert scale based on the McGill Pain Questionnaire	Lumbar radiograph	DSN was assessed by radiologists impression and the ratio of the anterior height of the space to its posterior height, and compared to the level above OST were graded as present or absent of anterior, posterior, and lateral OST at each lumbar intervertebral level and, when they were present, classified them as either traction or claw spurs SN was graded as present or absent
Hanmoğlu et al.[39]	Hospital outpatient population, Turkey, N = 49, N female = 37 (75.5%), mean age 40.4 years	Questionnaire at a routine visit	LBP-related physical function limitation was measured using the validated Turkish version of the ODI	Lumbar MRI	LDD - graded using the Pfirrmann classification as described above (Udby et al, 2019)IDH was measured from the middle of the superior border of the disc to the middle of its inferior border with the inclusion of both end plates SN were graded as present or not, and number of SN was graded MC: present or absent; type 1 involvement area was measured
Kuisma et al.[23]	Train engineers, Finland, N = 228, N female = 0(0%), mean age 46.9 years	Questionnaire: LBP lasting at least 14 days	LBP duration was assessed as: ever experiencing LBP and having LBP on the day of the assessment, also the number of previous episodes lasting >14 daysLBP intensity was measured using the NRS (0–10)	Lumbar MRI	MC were assessed as type I, type II or type III, and mixed types I/II or II/III The extent of MC was quantified as none, minimal (<25% of the vertebral height), or extensive (>25% of the vertebral height)
Lee et al.[41]	Community-based population, Korea, N = 1512, N female = 879 (58.1%), mean age 61.3 years	Questionnaire: LBP if yes to "Do you have LBP at the present time, that is, right now?"	LBP-related physical function limitation was measured using the validated Korean version of the ODI	Lumbar radiograph	OST presence and severity were graded as: 0 = none, 1 = barely visible, 2 = definite, 3 = large DSN graded as 0 = none, 1 = probable, 2 = definite, 3 = severe, bone to bone
Määttä et al.[24]	General population, China, N = 1142, N female = 717 (62.8%) mean age 52.9 years	Questionnaire: Prolonged severe LBP was defined as LBP lasting ≥ 30 days during the past year and a VAS pain intensity of ≥ 6	LBP duration in the past year was Measured as: 0 day, 1–7 days, 8–30 days, over 30 days, or daily LBPLBP intensity was measured on the VAS (0–10); it was divided into 3 categories: no/mild pain (VAS <3), moderate (VAS 3–5.9), and severe (VAS ≥6)LBP-related physical function limitation was measured using the validated Chinese version of the ODI; ODI scores ≥15% were defined to note back-related disability	Lumbar MRI	MC was assessed as type I, type II, type II/III, and type III
Mera et al.[35]			LBP intensity was measured using the VAS (0–100)	Lumbar MRI	MC was defined as type I, type II, and type III Modic change was defined more than 50% of the width of the vertebral body

Table 1 (continued)

Study	Study population	LBP definition	Clinical features	Imaging modality	Imaging features
Middendorp et al.[40]	General population, Japan, N = 814, N female = 568 (69.8%), mean age 63.6 years	Interview: LBP if yes to "Have you experienced LBP on most days during the past month?" Clinical measure: Prior history of LBP without a history of spinal surgery	LBP-related physical function limitation was measured using ODI	Lumbar MRI	LDD was graded using the modified Pfirrmann classification as described above (Udby et al., 2019)
Mok et al.[36]	Hospital outpatient population, Germany, N = 591, N female = 327 (55%), mean age 47.3 years	Questionnaire: Continuous localized pain for ≥2 weeks	LBP intensity was measured using the VAS (0–100)LBP-related physical function limitation was Measured using the validated Chinese version of the ODI and the Roland-Morris Disability Questionnaire (RMDQ)	Lumbar MRI	MC was graded as absent or present at each lumbar level
Perera et al.[29]	General population, China, N = 2449, mean age 40.4 years	Clinical measure: Presence of chronic LBP (defined as pain, muscle tension, or stiffness) on most days of the week for at least three months	LBP intensity was measured using the NRS (0–100)LBP-related physical function limitation was measured using the Modified version of the ODI (MODI)	Lumbar radiograph	LDD was graded as follows: 0 = normal (grade 0 DSN and grade 0 anterior OST), 1 = grade 1 DSN and/or grade 1 anterior OST, 2 = grade 2 DSN and/or grade 2 anterior OST DSN was graded as: 0 = none, 1 = definite mild) narrowing, 2 =moderate to severe narrowing OST were graded as: 0 = none, 1 = small OST, 2 = moderate to large OST
Peterson et al.[30]	Rheumatology clinic population, Sri Lanka, N = 439, N female = 323 (73.58%), mean age 48.99 years	Questionnaire: Presence of LBP	LBP intensity was measured using the NRS (0–10)LBP-related physical function limitation was measured using a revised ODI: 10 sections, each dealing with either pain or an area of daily activity likely to be affected by back pain	Lumbar radiograph	LDD was graded on a scale from 1 to 3: 1 = slight degeneration (OST were noted but with no visible DSN); 2 = moderate degeneration (OST > 2 mm and visible DSN), 3 = severe degeneration (complete loss of the disc height with or without a vacuum phenomenon) FJA was graded as: slight=increase in sclerosis compared with adjacent levels but with no hypertrophy, moderate=facet hypertrophy, no narrowing of the intervertebral foramen, severe=intervertebral foraminal stenosis
Rannou et al.[25]	Hospital outpatient population, UK, N = 172, N female = 12 (33.3%), mean age 52 years	Clinical evaluation: LBP persisted for 3 months with no response to 3-month conservative treatment and severe interference with lifestyle	LBP duration was measured in months LBP intensity was measured using the VAS (0–100)LBP-related physical function limitation was measured using the Quebec disability score (20 items, scored from 0 =no disability to 5 =impossible to do; the range of final score 0–100)Spinal morning stiffness: presence and duration (minutes) were measuredLimited or painful range of motion was assessed using the modified Schober test and finger-to-floor test. Exacerbation of pain in anteflexion, hyperextension, and lateral bending was also assessed	Lumbar MRI	MC were defined as MC1 (> 50% end plate marrow edema), MC2 (> 50% fatty deposits) or MC0 (no signal change)
Saukkonen et al.[26]	Hospital population, France, N = 36, N female = 80 (46.5%), mean age 51.5 years	Questionnaire: Prolonged disabling LBP over 30 days of LBP during the last12 months and perceived bothersomeness ≥ 6	LBP duration was reported as the answer to: "How often have you had aches or pains during the last 12 months?" with answer options: 1–7 days, 8–30 days, > 30 days, or dailyLBP intensity (total pain intensity and the total bothersomeness during sleep, work, and leisure time) was measured using the NRS (0–10)	Lumbar MRI	MC was graded by type (MC1, MC2, or MC3), horizontal and vertical height (1 = along the endplate, 2 = <25% of vertebral height, 3 = 25%–50% of vertebral height, 4 = > 50% of vertebral height) and horizontal location of MC
Scheele et al.[44]	General population, Finland, N = 1512, N female = 801 (53%), mean age 47 years	Interview during a home visit: Presence of LBP during the last month	Spinal morning stiffness was assessed using a questionnaire regarding the presence, duration (< half an hour, half an hour to 1 h, or > 1 h), and the location of morning stiffness	Lumbar radiograph	DSN and OST were graded using the Lane atlas as described above (Van den Berg et al., 2022)
Takatalo et al. (2012)	General population, Netherlands, N = 2819, N female = 1615 (57.3%), mean age 65.7 years	Questionnaire: Presence of LBP during the past six months,	LBP intensity was measured using the NRS (0–10)LBP-related physical function	Lumbar MRI	MC was graded as absent or present and by type (I or II). SN was defined as vertical intervertebral disc protrusion through the

Table 1 (*continued*)

Study	Study population	LBP definition	Clinical features	Imaging modality	Imaging features
	General population, Finland, N = 554, N female = 321 (57.9%), mean age 21.2 years	supported by a drawing of the low back region	limitation was measured using a questionnaire assessing the restriction of sports and daily activities		end plate AT was defined as a hyperintense linear area from the nucleus pulposus toward the outer annulus fibrosus HIZ was defined as a high-intensity signal located in the substance of the posterior annulus fibrosus
Takatalo et al. (2011)	General population, Finland, N = 554, N female = 321 (57.9%), mean age 21.2 years	Questionnaire: Presence of LBP during the past six months, supported by a drawing of the low back region	LBP intensity was measured using the NRS (0–10)LBP-related physical function limitation was measured using a questionnaire assessing the restriction of sports activities and daily activities	Lumbar MRI	LDD was measured using a modified Pfirrmann classification as described above (Udby et al., 2019)
Teraguchi et al.[34]	General population, China, N = 1214, N female= 761 (63%), mean age 48.1 years	Questionnaire: Prolonged severe LBP, defined as severe LBP (> VAS 6) lasting at least 30 days	LBP intensity was measured using the VAS (0–10)LBP-related physical function limitation was measured using a questionnaire	Lumbar MRI (only T2W)	HIZ: disc level, shape (round, fissure, vertical type, rim, and enlarged), and location within the disc (posterior or anterior)
Van den Berg et al. (2020)	General population, Netherlands, N = 457, N female = 373 (82%), mean age 64 years	Questionnaire: Presence of LBP in the last 4 weeks	Spinal morning stiffness severity and duration were measured: severity was rated on a 5-point scale with 0 = no spinal morning stiffness and 5 = very severe (present if ≥2). The duration was measured as no stiffness, <30 min, and > 30 min	Lumbar radiograph	LDD, divided in OST and DSN was graded using the Lane atlas as described above (Van den Berg et al., 2022)
Van den Berg et al. (2017)	General population, Netherlands, N = 699, N female = 557 (80%), mean age 64.3 years	Questionnaire: Presence of LBP	LBP duration was reported as 3 months, 3–12 months, and > 1 yearLBP intensity was measured using the NRS (0–10)	Lumbar radiograph	Both LDD definitions DSN and OST were graded using the Lane atlas as described above (Van den Berg et al., 2022)
Videman et al.[32]	General population (male twins), Finland, N = 230, N female = 0 (0%), mean age 49.4 years	Interview: Presence of LBP today and history of LBP	LBP intensity in the last year and in the worst episode was measured (method not reported)LBP-related physical function limitation was assessed using a disability scale based on the degree of interference with commonly performed daily tasks (lifting groceries, getting dressed, putting on socks, and being able to sleep at night)	Lumbar MRI	DSN, AT, and END were all separately rated using a scale from 0 to 3, with 0 = normal and 1–3 representing progressive degrees of abnormality
Weiner et al.[45]	Rehabilitation center, hospital, and the general population, USA, N = 35, N female = 17 (49%), age range 64–90 years	Clinical evaluation: Only people without LBP included	Limited or painful range of motion (lumbar flexion, extension, lateral flexion, axial rotation) was measured using the Back Range of Motion instrument (BROM)	Lumbar radiograph	FJA was graded as: 0 = normal; 1 = DSN and/or mild eburnation; 2 = moderate narrowing + moderate eburnation or hypertrophy; 3 = severe OA with narrowing, eburnation, and OST LDD was graded as: 0 = no disease (normal disc height, no spur formation, no eburnation, and no gas); 1 = mild disease (<25% DSN, small spur formation, minimal eburnation, and no gas); 2 = moderate disease (25–75% DSN, moderate spur formation, Moderate eburnation, and no gas); 3 = advanced disease (>75% DSN, large spur formation, marked eburnation, gas present)

FU, follow-up; LBP, low back pain
Imaging features: OA, Osteoarthritis; LDD, (intervertebral) disc degeneration; DSN, disc space narrowing; IDH, Intervertebral disc height; OST, osteophytes; SN, Schmor's nodes; AT, annular tear; MC, Modic changes; END, endplate abnormality; NPS, nucleus pulposus signal intensity; FJA, facet joint abnormalities.

Table 1

Study characteristics with the description of the assessment of the clinical and imaging features.

Osteoarthritis and Cartilage

limitation and lumbar disc degeneration, and for the lack of association between physical limitation and Schmorl's nodes in patients with LBP, based on one study by Määttä et al.[19] Inconsistent results (low-quality evidence) were found when addressing the longitudinal association between physical limitation and Modic changes, with two studies reporting no significant association,[18,20] but two other studies reporting a significant positive association.[16,19] Luoma et al. (very low-quality evidence) found that a decrease in the signal intensity of the nucleus pulposus was associated with a decrease in the ODI, and that degenerated discs with increased nucleus pulposus signals were more common among people who had persisting or increasing pain, but that there was no association between physical limitation and intervertebral disc height.[18]

Case-control studies
Low back pain intensity. Hicks et al. (moderate-quality evidence due to imprecision) found no difference in LBP intensity in individuals with and without disc or facet joint degeneration.[21] This study also reported no dose-response relationship between LBP intensity and the presence or absence of severe disc and facet joint degeneration at any spinal level.

Limited or painful range of motion. One study, with low-quality evidence, assessed the association between a range of motion and imaging features in LBP patients, displaying a weak negative association between lateral bending and endplate changes, disc and facet joint degeneration but only at the L5-S1 level.[2] The fingertip-to-floor distance was weakly associated with endplate changes at the L4-L5 level.

Cross-sectional studies
Low back pain duration. Moderate-quality evidence was found of a positive association between LBP duration and Modic changes[23–26] and for the lack of association between LBP duration and both disc space narrowing and osteophytes in patients with LBP (based on a single study).[27]

Low back pain intensity. There was moderate-quality evidence of an association between LBP intensity and lumbar disc degeneration[28–30] and annular tears,[31,32] and for the lack of association between LBP intensity and high-intensity zones[31,33,34] and disc space narrowing.[27,29,32] Eight studies assessed the association between the intensity of LBP and the presence of Modic changes. There was low-quality evidence for this association, because of inconsistency of the results and moderate to high risk of bias. Four studies reported a positive association, while four other studies reported no association.[23–26,31,33,35,36] LBP intensity was associated with Schmorl's nodes, but this association was no longer significant after adjusting for sex, sum score of disc degeneration, and socio-economic status (low-quality evidence).[31,37]

Low-quality evidence was found for the association between LBP intensity, and endplate lesions,[32] and for the lack of association between LBP intensity and nucleus pulposus signal intensity, intervertebral disc height, and MRI total degeneration score (an additive score of all abnormal osteoarthritis-related imaging).[33]

There was very low-quality evidence for the lack of association between pain intensity and osteophytes in patients with LBP.[27,29,37]

LBP-related physical functioning limitation. Overall, the included studies assessing the association between LBP-related physical limitation and imaging features provided low or very low quality of evidence, with the only exception being the association between physical limitation and high-intensity zones,[33,34] with moderate-quality evidence (Table 2). Low-quality evidence was found for the association between LBP-related physical limitation and both endplate

lesions and annular tears[32] and for the lack of association between LBP-related physical limitation and disc space narrowing,[38] facet joint degeneration,[30] signal intensity of the nucleus pulposus, and MRI total degeneration score.[33] Among four studies investigating the association between physical limitation and disc degeneration (low-quality evidence), three studies reported no significant association,[29,30,39] while one study found a significant, positive association, but only at the lower lumbar levels.[40] Inconsistent results were also seen when assessing the associations between physical limitation and osteophytes,[29,41] where there was one study reporting a significant association and one a negative association. Due to this inconsistency, as well as study limitations, this evidence is of low quality. There was very low-quality evidence for the lack of association between LBP-related physical functioning limitation and Modic changes,[24,25,33,36,39] intervertebral disc height,[33,39] Schmorl's nodes,[39] and lumbar spine osteoarthritis (assessed using the Kellgren & Lawrence scale).[42]

Spinal morning stiffness. High-quality evidence was found of an association between spinal morning stiffness and disc space narrowing.[43,44] The same two studies also assessed the relationship between spinal morning stiffness and osteophytes but found inconsistent results leading to moderate-quality evidence. There is very low-quality evidence of a positive association between the presence and duration of spinal morning stiffness and Modic changes.[25]

Limited or painful range of motion. Weiner et al.[45] reported a significant association between lumbar flexion, as well as lateral flexion, and both facet joint and disc degeneration. No association was found between lumbar extension and facet joint and disc degeneration (very low-quality evidence). There was also very low-quality evidence for the lack of an association between a limited range of motion and Modic changes.[25]

Dose-response relationship
Seven studies assessed the dose-response relationship between LBP intensity and imaging features (Appendix D). Peterson et al.[30] reported a weak dose-response relationship between LBP intensity and the amount and severity of degeneration as well as the number of levels of degeneration at the discs and at the facets. Kuisma et al.[23] found a dose-response relationship between the intensity of LBP and extensive Modic changes (25% of vertebral height), where extensive Modic changes were associated with higher LBP scores during the past week and the past 3 months. A dose-response relationship between MRI total degeneration score and pain intensity in patients with LBP was assessed by Berg et al., but they found no significant relationship.[33] Takatalo et al. reported that the most painful patient clusters had a higher sum score of disc degeneration than the two least painful clusters (P < 0.001), where moderately degenerated discs were more likely associated with the most severe low back symptoms than mildly degenerated discs.[28] One study, by Van den Berg et al., reported that the association between spinal morning stiffness and both disc space narrowing and osteophytes was stronger when the severity of spinal morning stiffness increased.[43] Määttä et al. found that subjects with larger Modic changes in the lumbar spine reported disabling LBP slightly more frequently than subjects with smaller Modic changes, but the difference was not significant (P = 0.057).[19]

Discussion

This is the first systematic review that has comprehensively investigated the association of clinical features suggestive of symptomatic spinal osteoarthritis, with imaging features on lumbar

Clinical feature	Imaging feature	Study	A measure of association [95% CI]	P-value	GRADE-score
Cohort studies					
LBP duration	MC	Udby et al.[16]	At baseline: No difference in duration of LBP between no MC and MC group, value NR	Not sign, value NR	⊕⊕⊕⊕
LBP intensity	DSN	Van den Berg et al. (2022)	Univariable analysis		⊕⊕⊕⊕
			Multilevel grade ≥1 DSN: beta = 0.65, 95% CI [0.13–1.17]	0.02	
			Adjusted model: beta = 0.45, 95% CI [−0.07–0.97]	0.09	
			Multivariable analysis		
	OST	Van den Berg et al. (2022)	Multilevel grade ≥2 osteophytes: beta = 0.59, 95% CI [0.08–1.10]	0.02	⊕⊕⊕⊕
			Univariable analysis		
			Multilevel grade ≥2 OST: beta = 0.85, 95% CI [0.37–1.33]	0.00	
			Adjusted model: beta= 0.75, 95% CI [0.26–1.24]	0.00	
			Multivariable analysis		
	MC	Luoma et al.[18]	Multilevel grade ≥1 disc space narrowing 0.18, 95% CI [−0.35–0.72]	0.50	⊕⊕○○
			Type 1 MC:		
			Decrease predicted decrease of pain intensity: decrease type 1 MC leads to a decrease of NRS by 3–8 points in 40%; a decrease of 1–2 points in NRS in 42% and an increase by 1–4 points in the NRS or no change was seen in 18% of the patients	0.011	
			Persistence predicted increase of pain intensity: no decrease in 3–8 NRS points group, decrease by 1–2 points: 37.5%, increase by 1–4 points on the NRS or no change: 62.5%	0.001	
			Enlargement was more common among those with persisting or increasing pain: decrease NRS 3–8: 26.2%, decrease NRS 1–2: 26.2%, increase NRS 1–4 points or no change: 47.6%	0.062	
			Type 2 MC:		
			Decrease of M2 size and new M2, value NR	0.000	
			Enlargement predicted decreasing in pain, value NR	0.025	
			Type 2 MC developing from type 1 MC predicted an increase in pain, value NR	0.050	
		Udby et al.[16]	Baseline: LBP intensity score in pt with MC: 6.0 vs without MC: 6.3	0.254	⊕○○○
			13-year FU: LBP intensity score in pt with MC: 4.2 vs without MC: 4.8	0.104	
	END, combined with SN	Luoma et al.[18]	An increase of bony endplate irregularities or SN predicted increase in pain: decrease NRS by 3–8 points in 23.9%; a decrease of 1–2 points on NRS in 30.4% and an increase by 1–4 points on the NRS or no change was seen in 45.7% of the patients	0.023; 0.041	⊕○○○
			Decrease of SN predicted decrease in pain: decrease NRS by 3–8 points in 100% of the patients	0.000	
	IDH	Luoma et al.[18]	The decrease in disc height predicted increasing or persisting pain, value NR	0.008	⊕⊕○○
	NPS	Luoma et al.[18]	Decrease of NPS predicted decrease of LBP intensity: decrease NRS by 3–8 points in 50%; a decrease of 1–2 points on NRS in 50% and an increase by 1–4 points on the NRS or no change was seen in 0% of the patients	0.001	⊕⊕○○
			Discs with normal signal intensity were more common among those with persisting or increasing LBP: a decrease in NRS by 3–8 points in 17.9%; a decrease of 1–2 points in NRS in 30.8% and an increase by 1–4 points on the NRS or no change was seen in 51.3% of the patients	0.048	
			Degenerated discs with increased NPS were more common among those with persisting or increasing pain; persisting: decrease NRS by 3–8 points in 33%; a decrease of 1–2 points on NRS in 32.2% and an increase by 1–4 points on the NRS or no change was seen in 34.4% of the patients; increased: decrease NRS by 3–8 points in 18.2%; a decrease of 1–2 points on NRS in 36.4% and an increase by 1–4 points on the NRS or no change was seen in 45.4% of the patients	0.099	
LBP-related physical function limitation/disability	MC	Keller et al.[20]	No significant associations between MC and ODI at baseline (mean difference ODI-MC1 0.6 (−4.1, 5.2) and ODI-MC2 0.5 (−4.1, 5.2)), data NR	Not sign, value NR	⊕⊕○○
			No significant association between MC1 or MC2 and the clinical course of ODI	Not sign, value NR	
		Määttä et al[19]	Multivariate Cox:ODI-MC1: beta 0.00, SE 0.02; HRR 1.00, 95% CI [0.98, 1.04]	Not sign, value NR	
			Episodes of severe and disabling LBP lasting at least 1 month at any point in life and MC at baseline were associated: without MC 89 (16.4%) vs with MC 90 (35%)	<0.001	
			Episodes of severe and disabling LBP lasting at least 1 month at any point in life and MC at follow-up were associated: without MC 44 (20%) vs with MC 73 (35.1%)	<0.001	
		Udby et al.[16]	Disabling LBP-MC: OR = 1.57, 95% CI [1.03–2.39]	0.037	
			Baseline RMDQ score MC+: 12.4 vs MC−: 12.6	0.752	
			At 13-year FU: RMDQ score MC+: 7.4 vs MC−: 9.6	0.024	
			Analysis of covariance: MC positively associated with 13-year RMDQ scores	0.031	
	LDD	Luoma et al.[18]	No significant association, value NR	Not sign, value NR	⊕⊕⊕○
		Määttä et al[19]	Disabling LBP-LDD: OR 1.080, 95% CI [1.047–1.115]	0.037	

Table 2 (continued)

Clinical feature	Imaging feature	Study	A measure of association [95% CI]	P-value	GRADE-score
	NPS	-Luoma et al.[18]	A decrease in NPS predicted a decrease in ODI, value NR	0.012	⊕○○○
	IDH	-Luoma et al.[18]	Overall association between change in disc SI and that in ODI was not significant, value NR	0.078	⊕○○○
Case-control studies	SN	-Määttä et al[19]	No significant association, value NR	Not sign, value NR	⊕⊕⊕○
			Disabling LBP-prevalence SN: OR 1.112, 95% CI [0.691–1.791]	0.662	
LBP intensity	FJA/LDD	-Hicks et al.[21]	No difference in pain intensity (MPQSF-score) between without and with LDD or FJA, values NR	>0.05	⊕⊕⊕○
Limited of painful range of motion	LDD	-Quack et al.[22]	Lateral bending (level L5-S1): rho = −0.23	<0.01	⊕⊕⊕○
			Modified Schober (level L4-L5 and L5-S1): rho = −0.02	<0.05	
	END	-Quack et al.[22]	Lateral bending (level L5-S1): rho = −0.23	<0.05	⊕⊕○○
			Finger-to floor-distance level L4-L5: rho = 0.36	<0.01	
			Reduction of observed extension	0.002	
	FJA	-Quack et al.[22]	Lateral bending (level of L5-S1): rho = −0.25	<0.01	⊕⊕○○
			Reduction of observed flexion at level L2–3	0.002	
Cross-sectional studies					
LBP duration	MC	-Kuisma et al.[23]	Increased number of LBP episodes	Sign, value NR	⊕⊕⊕○
			Any MC: OR 2.62, 95% CI [1.47–3.86]	NR	
			MC1: OR 2.53, 95% CI [1.24–3.90]	NR	
			MC2: OR 2.51, 95% CI [1.36–3.74]		
		-Määttä et al.[24]	LBP > 30 days: No MC 33% vs MC 43%	0.005	
		-Rannou et al.[25]	Duration of symptoms (months) no MC: 14, MC1: 52, MC2: 54	0.007	
		-Saukkonen et al.[26]	Duration > 30 days during the past 12 months- and MC: OR 1.57, 95% CI [1.22–2.01]	0.001	
	DSN	-Van den Berg et al. (2017)	OR 1.5, 95% CI [0.8–2.9]	Not sign, value NR	⊕⊕⊕○
	OST	-Van den Berg et al. (2017)	OR 1.6, 95% CI [0.8–3.0]	Not sign, value NR	⊕⊕⊕○
LBP intensity	LDD	-Perera et al.[29]	Grade 1 LDD-Intensity of pain: beta 3.642, 95% CI [−0.688 to 7.951]	0.098	⊕⊕⊕○
			Grade 2 LDD-Intensity of pain: beta 4.559, 95% CI [−1.227 to 10.345]	0.122	
		-Peterson et al.[30]	Without LDD 5.27 (SD 2.2) vs NRS with LDD 5.53 (SD 2.2)		
			Pain intensity-LDD severity: rho=0.20	<0.05	
		-Takatalo et al. (2011)	LDD grade 3, major vs minor symptoms: OR 1.88, 95% CI [1.02–3.48]	<0.001	
			LDD grade 4, major vs minor symptoms: 2.77, 95% CI [1.38–5.57]	<0.001	
		-Berg et al.[33]	L4/L5: beta = 0.699	0.59	⊕⊕○○
			L5/S1: beta = 0.881	0.49	
	MC	-Kuisma et al.[23]	Any MC: past week OR 1.47[1.13–1.87] and past 3 months OR 1.51, 95% CI [1.17–1.90]	NR	
			MC1: past week OR 1.69, [1.23–2.24] and past 3 months OR 1.67, 95% CI [1.23–2.20]	NR	
			MC2: past week OR 1.37, [1.00–1.80] and past 3 months OR 1.38, 95% CI [1.02–1.80]	NR	
			MC L5/S1: OR 1.36; [1.06–1.70]	*Sign, value NR*	
		-Määttä et al.[24]	No MC: VAS 49 (16–75) vs MC VAS 67 (29–87)	<0.001	
			Any MC was associated with prolonged severe LBP: OR 1.48, 95% CI [1.01–2.18]	*Sign, value NR*	
		-Mera et al.[35]	No MC VAS 9.9 ± 19.4 vs MC1 VAS 23.9 ± 26.3	<0.05	
			No MC VAS 9.9 ± 19.4 vs MC2 VAS 12.6 ± 21.3	0.9	
			No MC VAS 9.9 ± 19.4 vs MC 3 VAS 12.7 ± 23.6	0.7	
		-Mok et al.[36]	MC lower lumbar levels: mean VAS without MC 49.8 vs with MC 60.3; mean VAS difference = 10.5 mm; 95% CI [4.4–16.6]	0.001	⊕⊕○○
		-Rannou et al.[25]	VAS no MC: 48, MC1: 61 and MC2: 60	0.292	
		-Saukkonen et al.[26]	Any MC: OR 1.08 [1.02–1.15]	0.010	
		-Takatalo et al. (2012)	Major vs minor symptoms/pain: crude OR 9.13, 95% CI [0.94–88.59], aOR3.94, 95% CI [0.37–41.91]	Not sign, value NR	

Table 2 (continued)

Clinical feature	Imaging feature	Study	A measure of association [95% CI]	P-value	GRADE-score
	HIZ	-Berg et al.[33]	L4/L5: beta = 0.800	0.77	⊕⊕⊕○
			L5/S1: beta = 0.389	0.90	
		-Takatalo et al. (2012)	Major vs minor symptoms/pain: crude OR 1.09, 95% CI [0.34–3.47]; aOR 0.38, 95% CI [0.11–1.30]	NR	
		-Teraguchi et al.[34]	Single-level HIZ: OR 0.84, 95% CI [0.63–1.12]	Not sign, value NR	
			Multilevel HIZ (posterior): OR 2.18, 95% CI [1.42–3.37]	<0.05	
			Multilevel HIZ (anterior): OR 1.09, 95% CI [0.66–1.80]	Not sign, value NR	
			Higher risk of prolonged severe LBP for homogeneous HIZ: OR 1.57, 95% CI [1.10–2.23]	<0.05	
			Higher risk of prolonged severe LBP for heterogeneous HIZ: OR 1.53, 95% CI [1.02–2.31]	<0.05	
	NPS	-Berg et al.[33]	L4/L5: beta = −2.009	0.40	⊕⊕○○
			L5/S1: beta = 0.312	0.60	
	IDH (≥ 40% decrease)	-Berg et al.[33]	L4/L5: beta = −0.417	0.90	⊕⊕○○
			L5/S1: beta = 1.383	0.57	
	DSN	-Perera et al.[29]	Grade 1 DSN-Intensity of pain: beta = 3.896, 95% CI [−0.616 to 8.408]	0.091	⊕⊕⊕○
			Grade 2 DSN-Intensity of pain: beta = 2.00, 95% CI [−3.555 to 7.569]	0.479	
		-Van den Berg et al. (2017)	OR 1.8, 95% CI [1.3–2.6]	Not sign, value NR	
		-Videman et al.[32]	OR 1.8, 95% CI [1.0, 3.1]	NR	
	OST	-Frymoyer et al.[37]	Increased incidence of severe LBP if traction spurs (only L4-L5) present, value NR	Sign, value NR	⊕○○○
		-Perera et al.[29]	Grade 1 OST-Intensity of pain: beta = 4.836, 95% CI [0.764–8.908]	0.020	
			Grade 2 OST-intensity of pain: beta = 2.931, 95% CI [−5.235 to 11.097]	0.482	
		-Van den Berg et al. (2017)	OR 1.3, 95% CI [0.9–1.9]	Not sign, value NR	
	AT	-Takatalo et al. (2012)	Major vs minor symptoms/pain: crude OR 2.70, 95%CI [1.46–5.00]; aOR 1.52, 95% CI [0.76–3.02]	NR	⊕⊕⊕○
		-Videman et al.[32]	OR 2.2, 95% CI [1.3, 3.9]	NR	
	SN	-Frymoyer et al.[37]	No difference between different LBP intensity groups	Not sign, value NR	⊕⊕○○
		-Takatalo et al. (2012)	Major vs minor symptoms/pain: crude OR 1.82, 95% CI [1.09–3.06], aOR 1.08, 95% CI [0.60–1.96]	<0.001: Not sign after adjustment, value NR	
	MRI total score	-Berg et al.[33]	Beta = 0.591	0.41	⊕⊕○○
	END	-Videman et al.[32]	LBP intensity in worst episode: OR 1.4, 95% CI [1.1, 1.9]	Sign, value NR	⊕⊕○○
LBP-related physical function limitation/ disability	OA	-Ashraf et al.[42]	No association (Spearman's correlation, value NR)	0.169	⊕○○○
	LDD	-Hanmoğlu et al.[39]	Rho = 0.253	0.080	⊕⊕○○
		-Middendorp et al.[40]	ODI-LDD L4/5: rho < 0.113	<0.006	
			ODI-LDD L5/S1: rho < 0.104	<0.011	
		-Perera et al.[29]	Grade 1 LDD-severity of disability: beta = 0.976, 95% CI [−1.885 to 3.838]	0.252	
			Grade 2 LDD-severity of disability: beta = 2.243, 95% CI [−1.599 to 6.086]	0.504	
		-Peterson et al.[30]	ODI without LDD 34.82 (SD 15.6) vs with LDD 33.97 (SD 14.7), r = 0.013	Not sign, value NR	
	FJA	-Peterson et al.[30]	Rho = 0.064	>0.05	⊕⊕⊕○

Table 2 (continued)

Clinical feature	Imaging feature	Study	A measure of association [95% CI]	P-value	GRADE-score
	MC	-Berg et al.[33]	L4/L5: beta = 1.078	0.16	⊕◯◯◯
			L5/S1: beta = −0.129	0.87	
		-Hanmoğlu et al.[39]	MC1 involvement area: beta = 0.0138	<0.001	
			No MC ODI 4 (0–16) vs MC ODI 8 (0–20)	0.001	
		-Määttä et al.[24]	Unadjusted OR 1.68, 95% CI [1.22–2.32]; Adjusted: OR 1.47, 95% CI [1.04–2.10]	NR	
			Any MC: OR 1.47, 95% CI [1.04–2.10]	Sign, value NR	
			MC1: OR 1.23	Not sign, value NR	
			MC2: OR 1.56	Sign, value NR	
		-Mok et al.[36]	Subjects without MC had lower ODI scores than subjects with MC: median ODI 8 vs 4	<0.001	
		-Rannou et al.[25]	ODI and RMDQ are not significantly associated, values NR	>0.05	
			Quebec disability score no MC: 47, MC1: 48, and MC2: 51	0.698	
	HIZ	-Berg et al.[33]	L4/L5: beta = −0.007	1.00	⊕⊕⊕◯
			L5/S1: beta = −3.440	0.06	
		-Teraguchi et al.[34]	Presence of HIZ: OR 1.16, 95% CI [0.88–1.55]	Not sign, value NR	
			Homogeneous HIZ: OR 1.31, 95% CI [0.83–2.04]	Not sign, value NR	
			Heterogeneous HIZ: OR 1.21, 95% CI [0.81–1.77]	Not sign, value NR	
	NPS	-Berg et al.[33]	L4/L5: beta = 0.185	0.90	⊕⊕◯◯
			L5/S1: beta = 0.724	0.63	
	END	-Videman et al.[32]	Disability worst episode: OR 1.6, 95% CI [1.2, 2.0]	Sign, value NR	⊕⊕⊕◯
	IDH	-Berg et al.[33];	(≥ 40% decrease): L4/L5: beta = 1.564	0.44	⊕◯◯◯
			L5/S1: beta = 0.130	0.93	
		-Hanmoğlu et al.[39]	Rho = 0.110	0.451	
	DSN	-De Schepper et al. (2013)	OR = 1.9, 95% CI [1.4 −2.6]	NR	⊕⊕◯◯
		-Lee et al.[41]	Associated after adjusting for age and gender; value NR	<0.0001	
		-Perera et al.[29]	Grade 1 DSN-severity of disability: beta = −0.311, 95% CI [−3.301 to 2.679]	0.839	
			Grade 2 DSN-severity of disability: beta = 2.751, 95% CI [−0.934 to 6.437]	0.143	
		-Videman et al.[32]	OR 1.9, 95% CI [1.1, 3.3]	NR	
	SN	-Hanmoğlu et al.[39]	Rho = 0.111	0.449	⊕◯◯◯
	OST	-Lee et al.[41]	Associated after adjusting for age and gender; value NR	<0.0001	⊕⊕⊕◯
		-Perera et al.[29]	Grade 1 OST-severity of disability: beta = 2.136, 95% CI [−0.565 to 4.838]	0.121	
			Grade 2 OST-severity of disability: beta = −2.309, 95% CI [−7.727 to 3.108]	0.403	
	AT	-Videman et al.[32]	OR 1.9, 95% CI [1.1, 3.0]	NR	⊕⊕◯◯
	MRI total score				⊕⊕◯◯
	MC	-Berg et al.[33]	Beta = 0.257	0.54	
		-Rannou et al.[25]	Presence: no MC: 5, MC1: 11, MC2: 9	0.028	⊕◯◯◯
			Duration (minutes): no MC: 9, MC1: 49, MC2: 21	0.009	
Spinal morning stiffness	DSN	-Scheele et al.[44]	Duration < 0.5h: OR 1.7, 95% CI [1.3–2.1]	>0.01	⊕⊕⊕◯
			Duration ≥ 0.5 to ≤ 1h: OR 2.3, 95% CI [1.4–3.7],	>0.01	
			Duration > 1 h: OR 1.9, 95% CI [0.9–4.3]	Not sign, value NR	
			Only pt with LBP in analysis:	NR	
			- aOR 1.4, 95% CI [1.0–2.1]		
			Adjusted for LBP: aOR 1.8, 95% CI [1.4–2.2]	<0.01	
			Combined with LBP: aOR 2.5, 95% CI [1.9–3.4]	NR	
	DSN	-Van den Berg et al. (2020)	Presence of DSN: OR 1.4, 95% CI [0.9–2.1]	0.11	
			Grade 1 DSN: OR 1.6, 95% CI [1.0–2.5]	0.06	
			Grade 2 or 3 DSN: OR = 2.0, 95% CI [1.1–3.5]	0.03	
			Presence of DSN and severe spinal morning stiffness L1–S1: OR 1.8, 95% CI [1.1–3.1]	Sign, value NR	
			Duration ≥ 30 min and grade 1 DSN: OR 2.0, 95% CI [1.1–3.6]	0.02	

Table 2 (*continued*)

Clinical feature	Imaging feature	Study	A measure of association [95% CI]	P-value	GRADE-score
	OST	-Scheele et al.[44]	Duration < 0.5 h: OR 1.2, 95%CI [1.0–1.5]	>0.05	⊕⊕⊕◯
			Duration ≥ 0.5 to ≤1 h: OR 1.7, 95% CI [1.1–2.7]	>0.05	
			Duration > 1 h: OR 2.4, 95% CI [1.2–4.8]	>0.05	
			Only pt with LBP in analysis:		
			-aOR 1.2, 95% CI [0.8–1.8]	NR	
			Adjusted for LBP: aOR 1.3, 95% CI [1.0–1.5]	<0.05	
			Combined with LBP: aOR 1.5, 95% CI [1.1–2.0]	NR	
		-Van den Berg et al. (2020)	Presence of OST: OR 2.1, 95% CI [1.3–3.2]	0.00	
			Grade 1 OST: OR 1.2, 95% CI [0.8–2.0]	0.43	
			Grade 2–3 OST: OR 1.7, 95% CI [0.7–3.3]	0.13	
			Presence of OST and moderate spinal morning stiffness: OR 2.0, 95% CI [1.2–3.2]	Sign, value NR	
			Presence of OST and severe spinal morning stiffness L1–S1: OR 2.0, 95% CI [1.2–3.7]	Sign, value NR	
			Duration ≥ 30 min and grade 2 osteophytes: OR = 2.6, 95% CI [1.1–6.2]	0.03	
Limited of painful range of motion	MC	-Rannou et al.[25]	Modified Schober test: no MC: 20 cm, MC1: 21 cm, MC2: 20 cm	0.289	⊕◯◯◯
			Finger-to floor-distance: no MC: 20 cm, MC1: 14 cm, MC1: 15 cm, MC2: 14 cm	0.498	
			Exacerbation of pain in anteflexion: no MC: 42%, MC1: 58%, MC2: 50%	0.723	
			Exacerbation of pain in lateral bending: no MC: 33%, MC1: 67%, MC2: 42%	0.244	
			Exacerbation of pain in hyperextension: no MC: 25%, MC1: 83%, MC2: 67%	0.013	
	FJA	-Weiner et al.[45]	Flexion: rho = 0.29	0.09	⊕◯◯◯
			Extension: rho= 0.17	Not sign, value NR	
			Right Lateral Flexion: rho = 0.19	Not sign, value NR	
			Left Lateral Flexion: rho= 0.29	0.09	
	LDD	-Weiner et al.[45]	Flexion: rho = 0.34	0.04	⊕◯◯◯
			Extension: rho= 0.21	Not sign, value NR	
			Right Lateral Flexion: rho = 0.36	0.03	
			Left Lateral Flexion: rho = 0.35	0.04	

FU, follow-up; ODI, Oswestry Disability Index; RMDQ, Roland-Morris Disability Questionnaire

Imaging features: OA, Osteoarthritis; LDD, (intervertebral) disc degeneration; DSN, disc space narrowing; IDH, Intervertebral disc height; OST, osteophytes; SN, Schmorl's nodes; AT, annular tear; MC, Modic changes; END, endplate abnormality; NPS, nucleus pulposus signal intensity; FJA, facet joint abnormalities

GRADE quality of evidence: ⊕⊕⊕⊕ = high; ⊕⊕⊕◯ = moderate; ⊕⊕◯◯ = low; ⊕◯◯◯ = very low quality.

Osteoarthritis and Cartilage

Table 2

Association between clinical features and imaging features in patients with low back pain (LBP).

Table 3

Risk of bias assessment cohort and case-control studies (Newcastle Ottawa Scale).

Cohort studies	Selection				Comparability	Outcome			Total
	Representativeness of exposed cohort (★)	Selection of non-exposed cohort (★)	Ascertainment of exposure (★)	Demonstration outcome was not present at the start (★)	Comparability (★★)	Assessment of outcome (★)	Follow-up long enough for an outcome to occur (★)	Adequacy of follow-up of cohorts (★)	
Keller et al.[20]	★	★	★	×	★★	×	★	×	6★
Luoma et al.[18]	×	×	×	×	★	×	×	×	3★
Määttä et al[19]	★	×	★	×	★★	★	★	★	7★
Udby et al.[16]	★	★	★	★	★★	★	★	×	8★
Van den Berg et al. (2022)	★	★	★	★	★★	★	★	×	8★

Case-control studies	Selection				Comparability	Outcome			Total
	Case definition adequate (★)	Representativeness Cases (★)	Selection of controls (★)	Definition of controls (★)	Comparability (★★)	Ascertainment of exposure (★)	Same method ascertainment cases & controls (★)	Non-response rate (★)	
Hicks et al.[21]	★	★	★	★	★★	★	★	×	8★
Quack et al.[22]	★	×	★	★	×	×	★	×	4★

Osteoarthritis and Cartilage

radiographs, MRI, or CT-scan that are potential indicators of structural spinal osteoarthritis. Thirty studies were included, mainly cross-sectional designs, with the majority being population-based studies. In total, 12 evaluated structural spinal osteoarthritis features on lumbar radiographs and 18 assessed these on MRI, none on CT-scan. High-quality evidence was found for the longitudinal association between LBP intensity, and both disc space narrowing and osteophytes, as well as for the association between LBP-related physical limitation and lumbar disc degeneration, the presence of spinal morning stiffness and disc space narrowing and for the lack of association between physical limitation and Schmorl's nodes.

In line with our review, Herlin et al.[8] showed inconsistent results regarding the association of Modic changes with back-related activity limitation and the presence of LBP, due to the high risk of bias and the heterogeneity of study samples, clinical outcomes, and prevalence estimates of LBP and Modic changes. Other reviews mainly looked at the association between the presence of LBP and imaging features possibly associated with spinal osteoarthritis. An association between the presence of LBP and disc space narrowing was reported by Raastad et al.,[6] but they only looked at the presence of LBP, not in combination with the clinical features we investigated in this review. The systematic review by Brinjikji et al.[7] reported an association between the presence of LBP and Modic changes, whereas we found low-quality evidence of no association. Again, this review only investigated the association of Modic changes with solely the presence of LBP.

There was substantial clinical and methodological heterogeneity and heterogeneity in terminology among studies included in this review, making it unfeasible to perform a meta-analysis and add more certainty to our conclusions. This underlines the need for agreed criteria for symptomatic and structural spinal osteoarthritis, as highlighted by a recent international consensus study.[5] Additionally, guidelines on how to universally define and assess various clinical and imaging features related to spinal osteoarthritis would be useful to ease the research in this field and to possibly identify potential subgroups regarding etiology, prognosis, and treatment.

Strengths and limitations

The strengths of this systematic review include that the protocol was prospectively registered in PROSPERO, the search strategy was developed with an experienced medical librarian, and that the title/abstract screening, full-text screening, data extraction, and risk of bias assessment were performed by two independent reviewers. More importantly, this review is a continuation of the aforementioned Delphi study,[5] contributing to the development of established international diagnostic criteria for symptomatic and structural spinal osteoarthritis. The results from this review form a good base for future observational and diagnostic accuracy research, to further explore and confirm which of the investigated clinical features are more strongly related to spinal osteoarthritis structural features.

This systematic review has some limitations as well. Firstly, for this review, we chose to focus on the clinical features on which consensus reached by experts in the fields of back pain and osteoarthritis in the aforementioned Delphi study, in patients with LBP. We did not include other clinical features, such as solely the presence or recurrence of LBP, leg pain, cauda equina symptoms, and intermittent claudication. Associations between these features and imaging features possibly associated with spinal osteoarthritis are surely of interest, especially for patients in the general population, but since they can also have other etiologies, such as non-degenerative lumbar disc herniation or spinal stenosis, and are not evident potential indicators of symptomatic spinal osteoarthritis, these clinical features were not included in this review. Furthermore, we argue these are different research questions, and adding these

Cross-sectional studies	Selection			Comparability	Outcome		Total
	Representativeness of exposed cohort (★)	Non-respondents (★)	Ascertainment of exposure (★)	Comparability (★★)	Assessment of outcome (★)	Statistical test (★)	
Ashraf et al.[42]	★	x	★	x	x	★	3★
Berg et al.[33]	★	x	★	★★	★	★	6★
De Schepper et al. (2013)	★	x	★	★★	★	★	6★
Frymoyer et al.[37]	x	x	★	x	★	★	3★
Hanımoğlu et al.[39]	x	x	★	x	x	★	2★
Kuisma et al.[23]	x	x	★	★	★	x	3★
Lee et al.[41]	★	x	★	★	x	★	4★
Määttä et al.[24]	★	x	★	★★	★	★	6★
Mera et al.[35]	★	x	★	★	★	★	5★
Middendorp et al.[40]	★	x	★	x	★	x	3★
Mok et al.[36]	★	★	★	★★	★	★	7★
Perera et al.[29]	★	x	★	★	★	★	5★
Peterson et al.[30]	★	x	★	★	★	★	5★
Rannou et al.[25]	★	★	★	x	★	★	5★
Saukkonen et al.[26]	★	x	★	★★	★	★	6★
Scheele et al.[44]	★	x	x	★★	★	★	6★
Takatalo et al. (2012)	★	★	★	★	★	★	6★
Takatalo et al. (2011)	★	★	★	★	★	★	6★
Teraguchi et al.[34]	★	★	★	★★	★	★	7★
Van den Berg et al. (2020)	★	x	★	★★	★	★	6★
Van den Berg et al. (2017)	★	x	★	★★	★	★	6★
Videman et al.[32]	★	x	x	★	★	★	5★
Weiner et al.[45]	★	x	★	★	★	★	5★

Table 4

Osteoarthritis and Cartilage

Risk of bias assessment cross-sectional studies (Modified Newcastle Ottawa Scale).

clinical features would have resulted in more heterogeneity. Secondly, some of the clinical features included in this review, such as spinal morning stiffness and range of motion, are less investigated, making it difficult to reach solid conclusions regarding the role of these clinical features in the etiology of spinal osteoarthritis. Thirdly, to assess the quality of evidence we used an adaptation of the GRADE framework and started at a 'high-quality' level for all outcomes, since our aim was to identify associations and observational studies are appropriate for this purpose. This approach did lead, on some occasions, to a high-quality evidence grading of an outcome that was researched by just one study. However, we argue that if this one study has a substantial sample size and a low risk of bias, the certainty that the reported effect lies close to the true effect is significant and therefore should be graded as high-quality evidence. Lastly, given the low amount of studies per association, we combined the results of different imaging modalities (i.e., radiographs and MRI), which might lead to some bias in the results, since for some imaging features there is a preferred first choice imaging modality (e.g., MRIs for evaluating Modic changes).

In conclusion, we found high-quality evidence for the longitudinal association between LBP intensity, and both disc space narrowing and osteophytes, as well as for the association between LBP-related physical limitation and lumbar disc degeneration, the presence of spinal morning stiffness and disc space narrowing and for the lack of association between physical limitation and Schmorl's nodes. Clinical and methodological heterogeneity, as well as heterogeneity in terminology in included studies, is a limitation, of which agreed diagnostic criteria would greatly benefit the field of spinal osteoarthritis. Future observational and diagnostic accuracy research should further explore and confirm which of the studied clinical features are more strongly related to spinal osteoarthritis structural features.

CRediT authorship contribution statement

SBZ, BK, AC, and MC conceived and designed the study. MC and OO executed the title/abstract and full-text screening. MC and KD extracted the data and assessed the risk of bias. MC assessed the quality of evidence using an adaptation of the GRADE framework; AC, BK, and SBZ supervised (and when necessary corrected) the assessment. MC, KD, BK, SBZ, and AC discussed the results and interpretation. MC drafted the manuscript and carried out the first round of revisions. All authors have read and approved the final version of the manuscript and contributed to its critical revision. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Declaration of Competing Interest

There were no competing interests of any of the authors.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.joca.2023.04.014.

References

1. Disease GBD, Injury I, Prevalence C. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1789–858.
2. Lambeek LC, van Tulder MW, Swinkels IC, Koppes LL, Anema JR, van Mechelen W. The trend in total cost of back pain in The Netherlands in the period 2002 to 2007. Spine 2011;36:1050–8.
3. Volksgezondheid en zorg N.Iv.OvdG.Zel, Landelijke Informatievoorziening Paramedische Zorg (LIPZ), CBS Centraal Bureau voor de Statistiek Doodsoorzakenstatistie, Landelijke Medische Registratie (LBZ), Gezondheidsenquête, voorheen POLS, gezondheid en welzijn, Kosten van Ziektenstudie. Nek- en rugklachten. vol. 2022: VZinfo.nl 2019.
4. Chiarotto A, Koes BW. Nonspecific low back pain. N Engl J Med 2022;386:1732–40.
5. de Luca K., Chiarotto A., Cicuttini F., Creemers L., de Schepper E., Ferreira P.H., et al. Consensus for statements regarding a definition for spinal osteoarthritis for use in research and clinical practice: a Delphi study. Arthritis Care Res (Hoboken); 2021.
6. Raastad J, Reiman M, Coeytaux R, Ledbetter L, Goode AP. The association between lumbar spine radiographic features and low back pain: a systematic review and meta-analysis. Semin Arthritis Rheum 2015;44:571–85.
7. Brinjikji W, Diehn FE, Jarvik JG, Carr CM, Kallmes DF, Murad MH, et al. MRI findings of disc degeneration are more prevalent in adults with low back pain than in asymptomatic controls: a systematic review and meta-analysis. Am J Neuroradiol 2015;36:2394–9.
8. Herlin C, Kjaer P, Espeland A, Skouen JS, Leboeuf-Yde C, Karppinen J, et al. Modic changes—their associations with low back pain and activity limitation: a systematic literature review and meta-analysis. PLoS One 2018:13.
9. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71.
10. Higgins JPTTJ, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.1, vol. 2020. UK: The Cochrane Collaboration; 2020.
11. GA Wells BS, O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. The Newcastle-Ottawa Scale ((NOS)) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses Ottawa: The Ottawa Hospital; 2009.
12. Patra J, Bhatia M, Suraweera W, Morris SK, Patra C, Gupta PC, et al. Exposure to second-hand smoke and the risk of tuberculosis in children and adults: a systematic review and meta-analysis of 18 observational studies. PLoS Med 2015;12, e1001835. discussion e1001835.
13. Schünemann HBJ, Guyatt G, Oxman A, editors. GRADE Handbook for Grading Quality of Evidence and Strength of Recommendations. 2022. Canada: The GRADE Working Group; 2013.
14. Huguet A, Hayden JA, Stinson J, McGrath PJ, Chambers CT, Tougas ME, et al. Judging the quality of evidence in reviews of prognostic factor research: adapting the GRADE framework. Syst Rev 2013;2:71.
15. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. BMJ 2020;368:l6890.
16. Udby PM, Bendix T, Ohrt-Nissen S, Lassen MR, Sørensen JS, Brorson S, et al. Modic changes are not associated with long-term pain and disability: a cohort study with 13-year follow-up. Spine 2019;44:1186–92.
17. van den Berg R, Chiarotto A, Enthoven WT, de Schepper E, Oei EHG, Koes BW, et al. Clinical and radiographic features of spinal osteoarthritis predict long-term persistence and severity of back pain in older adults. Ann Phys Rehabil Med 2022;65(1), 101427.
18. Luoma K, Vehmas T, Kerttula L, Grönblad M, Rinne E. Chronic low back pain in relation to Modic changes, bony endplate lesions, and disc degeneration in a prospective MRI study. Eur Spine J 2016;25:2873–81.
19. Määttä JH, Karppinen JI, Luk KDK, Cheung KMC, Samartzis D. Phenotype profiling of Modic changes of the lumbar spine and its association with other MRI phenotypes: a large-scale population-based study. Spine J 2015;15:1933–42.
20. Keller A, Boyle E, Skog TA, Cassidy JD, Bautz-Holter E. Are Modic changes prognostic for recovery in a cohort of patients with non-specific low back pain? Eur Spine J 2012;21:418–24.
21. Hicks GE, Morone N, Weiner DK. Degenerative lumbar disc and facet disease in older adults: prevalence and clinical correlates. Spine 2009;34:1301–6.
22. Quack C, Schenk P, Laeubli T, Spillmann S, Hodler J, Michel BA, et al. Do MRI findings correlate with mobility tests? An explorative analysis of the test validity with regard to structure. Eur Spine J 2007;16:803–12.
23. Kuisma M, Karppinen J, Niinimäki J, Ojala R, Haapea M, Heliövaara M, et al. Modic changes in endplates of lumbar vertebral bodies: prevalence and association with low back and sciatic pain among middle-aged male workers. Spine 2007;32:1116–22.
24. Määttä JH, Karppinen J, Paananen M, Bow C, Luk KDK, Cheung KMC, et al. Refined phenotyping of modic changes. Medicine 2016;95(22):e3495.
25. Rannou F, Ouanes W, Boutron I, Lovisi B, Fayad F, Mace Y, et al. High-sensitivity C-reactive protein in chronic low back pain with vertebral end-plate modic signal changes. Arthritis Rheum-Arthritis Care Res 2007;57:1311–5.
26. Saukkonen J, Määttä J, Oura P, Kyllönen E, Tervonen O, Niinimäki J, et al. Association between Modic changes and low back pain in middle age: a northern Finland birth cohort study. Spine 2020;45:1360–7.
27. Van Den Berg R, Jongbloed LM, Kuchuk NO, Roorda LD, Oostveen JCM, Koes BW, et al. The association between self-reported low back pain and radiographic lumbar disc degeneration of the cohort hip and cohort knee (CHECK) study. Spine 2017;42:1464–71.
28. Takatalo J, Karppinen J, Niinimäki J, Taimela S, Näyhä S, Mutanen P, et al. Does lumbar disc degeneration on magnetic resonance imaging associate with low back symptom severity in young Finnish adults? Spine 2011;36:2180–9.
29. Perera RS, Dissanayake PH, Senarath U, Wijayaratne LS, Karunanayake AL, Dissanayake VHW. Associations between disc space narrowing, anterior osteophytes and disability in chronic mechanical low back pain: a cross sectional study. BMC Musculoskelet Disord 2017:18.

30. Peterson CK, Bolton JE, Wood AR. A cross-sectional study correlating lumbar spine degeneration with disability and pain. Spine 2000;25:218–23.

31. Takatalo J, Karppinen J, Niinimäki J, Taimela S, Mutanen P, Sequeiros RB, et al. Association of modic changes, Schmorl's nodes, Spondylolytic defects, high intensity zone lesions, disc herniations, and radial tears with low back symptom severity among young Finnish adults. Spine 2011.

32. Videman T, Battié MC, Gibbons LE, Maravilla K, Manninen H, Kaprio J. Associations between back pain history and lumbar MRI findings. Spine 2003;28:582–8.

33. Berg L, Hellum C, Gjertsen O, Neckelmann G, Johnsen LG, Storheim K, et al. Do more MRI findings imply worse disability or more intense low back pain? A cross-sectional study of candidates for lumbar disc prosthesis. Skelet Radiol 2013;42:1593–602.

34. Teraguchi M, Cheung JPY, Karppinen J, Bow C, Hashizume H, Luk KDK, et al. Lumbar high-intensity zones on MRI: imaging biomarkers for severe, prolonged low back pain and sciatica in a population-based cohort. Spine J 2020;20:1025–34.

35. Mera Y, Teraguchi M, Hashizume H, Oka H, Muraki S, Akune T, et al. Association between types of Modic changes in the lumbar region and low back pain in a large cohort: the Wakayama spine study. Eur Spine J 2020.

36. Mok FPS, Samartzis D, Karppinen J, Fong DYT, Luk KDK, Cheung KMC. Modic changes of the lumbar spine: prevalence, risk factors, and association with disc degeneration and low back pain in a large-scale population-based cohort. Spine J 2016;16:32–41.

37. Frymoyer JW, Newberg A, Pope MH. Spine radiographs in patients with low-back pain: an epidemiological study in men. J Bone Jt Surg Ser A 1984;66:1048–55.

38. De Schepper EIT, Damen J, Van Meurs JBJ, Ginai AZ, Popham M, Hofman A, et al. The association between lumbar disc degeneration and low back pain: the influence of age, gender, and individual radiographic features. Spine 2010;35:531–6.

39. Hanımoğlu H, Çevik S, Yılmaz H, Kaplan A, Çalış F, Katar S, et al. Effects of Modic type 1 changes in the vertebrae on low back pain. World Neurosurg 2019;121:e426–32.

40. Middendorp M, Vogl TJ, Kollias K, Kafchitsas K, Khan MF, Maataoui A. Association between intervertebral disc degeneration and the Oswestry Disability Index. J Back Musculoskelet Rehabil 2017;30:819–23.

41. Lee SY, Cho NH, Jung YO, Seo YI, Kim HA. Prevalence and risk factors for lumbar spondylosis and its association with low back pain among rural Korean residents. J Korean Neurosurg Soc 2017;60:67–74.

42. Ashraf A, Farahangiz S, Pakniat Jahromi B, Setayeshpour N, Naseri M. Correlation between degree of radiologic signs of osteoarthritis and functional status in patients with chronic mechanical low back pain. Malaysian J Med Sci 2014;21:28–33.

43. van den Berg R, Jongbloed EM, Kuchuk NO, Koes BW, Oei EHG, Bierma-Zeinstra SMA, et al. Association between self-reported spinal morning stiffness and radiographic evidence of lumbar disk degeneration in participants of the cohort hip and cohort knee (CHECK) study. Phys Ther 2020;100:255–67.

44. Scheele J, De Schepper EIT, Van Meurs JBJ, Hofman A, Koes BW, Luijsterburg PAJ, et al. Association between spinal morning stiffness and lumbar disc degeneration: the Rotterdam Study. Osteoarthritis Cartilage 2012;20:982–7.

45. Weiner DK, Distell B, Studenski S, Martinez S, Lomasney L, Bongiorni D. Does radiographic osteoarthritis correlate with flexibility of the lumbar spine. J Am Geriatr Soc 1994;42:257–63.